

UKMED – 2021 V1 – May (interim) Data Dictionary

Authors: Daniel Smith, Elizabeth Twigg, Ganesan Gurusamy and Peter Tang
last updated: 30/4/21

Acknowledgements

Material has been taken from data dictionaries and documentation provided by data contributors: HESA, UCAS UKCAT, BMAT and the UKFPO.

Contents

UKMED Population definition	3
De-identification	4
Version control	5
Person and Event tables.....	7
ARCP_OUTCOMES	8
ARCP_REASON	11
ARCP_N_CODES	12
BMAT_SCORES_APP	13
EXAM_SECTION_MARKS	16
EXAM_TOTAL_MARKS.....	17
EXAM_DATA_FLAT.....	19
FP.....	20
VW_FP_PERSON.....	25
GAMSAT_APP	31
HESA_ACAD_YEARS.....	33
VW_HESA_ACAD_YEAR_PERSON.....	37
HESA_QUALIFICATION	38
IELTS_LIVE.....	40
MMI_UCAS.....	42
VW_NTS_ITEM_TRACKER	44
NTS_TRAINEE_IND_SCORES_LIVE	45
NTS_2020_SURVEY_DATA	46
ORIEL_INTERVIEW_SCORES.....	47
ORIEL_RECRUIT_OUTCOMES.....	48
PLAB1_DTL_LIVE.....	51
PLAB2_DTL_LIVE.....	53
PLAB_TOTAL_LIVE	56
RECRUIT_MSRA	57
ROD_01	60
STUDENT_FTP.....	61
UCAS_APPLICATIONS	63
APPLICANT_MEDICAL_SCHOOL_DISTANCES.....	67
UCAS_QUALS_VER.....	68
UCAS_QUALS_DEC	69
VW_UKCAT_RESULTS_APP.....	70
UKCAT_PROGRESSION	73
UKMED_GEOGRAPHY	75
VW_GP_ROUTE	80
VW_NTS_TRAINEE_HISTORY	81
VW_SPEC_ROUTE.....	84
VW_UKCAT_RESULTS_NONCOG.....	85
VW_UKMED_ALLEGATIONS.....	86
VW_UKMED_FTPCHARDECLARATIONS	88
VW_UKMED_FTP_SUMMARY	90
VW_UKMED_PERSON_APPLICANT	92
VW_UKMED_PERSON_FULL	101
VW_UKMED_DISABILITY_AUDIT	113

VW_UKMED_PRACTICEHISTORY	114
VW_UKMED_REGHISTORY	117
VW_UKMED_SPECIALTIES	118
Reference Tables	119
CR_UKPRN	120
CR_COURSE	121
CR_EXAM	123
CR_FP	124
CR_ONS_POSTCODE	125
CR_PSA_APPROACH	127
CR_SPECIALTIES	128
CR_UKPRN_NAME	129
ORGANISATION_LIVE	130

UKMED Population definition

UKMED – 2021 V1 – June (interim) of the UK Medical Education Database includes:

All students who started at a UK medical school from 2002 until 2019 as defined by data supplied to the GMC by the Higher Education Statistics Agency (HESA).

and

All doctors listed in the NTS_TRAINEE table at least once since 2012 through to 2020 (i.e. all those listed as belonging to a UK-based training scheme in one or more the annual censuses conducted to administer the GMC's National Trainee Survey [NTS]).

Note that the 2020 NTS_TRAINEE data was not validated by survey respondents as the usual NTS did not run due to coronavirus (COVID-19) pandemic

and

All Doctors completing the NTS survey between 2009 and 2011.

and

All those who applied via The Universities and Colleges Admissions Service (UCAS) to a degree that leads to a Primary Medical Qualification between 2007 and 2020. This means that test scores and entry qualifications are available for applicants to medicine from 2007 onwards, not just entrants.

HESA extract their data from their student records using the following query:

Population:

Population: REGBODY= 01 General Medical Council (GMC) or:

for The University of St Andrews those students coded A100 or A300 in SBJCA1-3 and include the Pathway to Medicine course title.

Include students who are recorded on an intercalated year in the INTERCALATE field If they were also recorded as REGBODY=01 in their previous year of study.

for The University of Glasgow include BSc (Med Sci) Clinical Medicine (This is a 1 year course) and Medical Science, CertHE (The length of course is unknown) course titles.

Inclusion in the UKMED population is defined by the following:

- A commencement date between 2002 and 2019 for the first year of the student's programme and the first year of their study.
- Any records for courses not associated with medicine that were included in the HESA extract are excluded.

UCAS base their query on a list of valid courses supplied to them by UKMED following clerical review of all likely courses.

De-identification

Fields in grey rows potentially allow re-identification and are not included in data extracts within the Safe Haven. They are included here to show the construction of the database. PERSON_UID is always replaced by a STUDY_ID that contains a person identifier and a unique study identifier. It is unique to the study for which the extract was generated.

Version control

Changes to this document since UKMED 2019 V2 – October

- The UKMED Population definition has updated dates to reflect the data we are currently loading – UCAS (2019 and 2020) and associated scores from test providers and HESA (2019/20)
- This is an interim document and we will republish shortly with updated tables. The main addition will be ARCP outcomes from 2020 that relate to COVID using the new COVID codes - https://www.copmed.org.uk/images/docs/Coding_for_ARCPs/Coding_for_ARCPs.pdf.
- The UCAS_QUALS_VER table will contain two sets of grades for 2020: the grades on results day and the final grades used for admission to medical school.

Changes to this document since UKMED 2019 V1 – March

- The UCAS data now include applicants' qualifications
- The test provider tables are now based on the applicant population and include Z-scores calculated from the applicant population statistics.
- The VW_FTP_SUMMARY now contains dates for when the sanctions were applied to the doctor.

Changes to this document since UKMED 2019 V2 – October

- ARCP_N_CODES table is included
- We have included person level views where appropriate: VW_FP_PERSON and VW_HESA_ACAD_YEAR_PERSON
- The population definition has been incremented to include cases listed in the 2019 NTS because 2019 postgraduate outcomes are now available.
- A table containing all applicant demographics - VW_UKMED_PERSON_APPLICANT
- Demographic data now include TUNDRA see <https://www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-the-data/>

Changes to this document since UKMED 2020 V1 – May

- The MMI table is now called MMI_UCAS as it is now matched to UCAS applications.
- VW_UKMED_DISABILITY_AUDIT which records when changes to PG_DISABILITY captured in Siebel occurred.
- VW_NTS_ITEM_TRACKER details the NTS items that can be requested for inclusion in extracts
- NTS_TRAINEE_IND_SCORES_LIVE and VW_NTS_TRAINEE_HISTORY have noted pertaining to the usual NTS not running in 2020 due to the coronavirus (COVID-19) pandemic.
- STUDENT_FTP table replaces SFTP_ALLYEARS the data come from the same collection, we have tidied and rename for clarity
- UCAS_APPLICATIONS now includes applicants' ages
- APPLICANT_MEDICAL_SCHOOL_DISTANCES gives distances between each applicant's home address and each medical school the applicant applied to.
- VW_UKMED_PERSON_FULL includes disability data for postgraduate trainees

- EXAM_DATA_FLAT contains postgraduate exam data re-arranged with one row per person per exam, with columns for each attempt
- The exam data now includes PSA data
- The usual NTS, normally administered from the end of March, did not run in 2020, instead a COVID specific, non-mandatory survey was run later in the year. This was not run on the usual platform and items are different. The relevant data are held in NTS_2020_SURVEY_DATA.
- CR tables which do not contain person or event data but can be useful for reference purposes have been placed at the back of the document.

Person and Event tables

These tables contain person and event level data for use in research extracts.

ARCP_OUTCOMES

Each year, Local Education Training Boards (LETBs) and deaneries provide the GMC with ARCP (Annual Review of Competence Progression) data that describes trainees' progression (or not) through their medical training.

A trainee may have more than one ARCP record in the annual submission, as an outcome is awarded for each specialty undertaken by the trainee. A trainee may also have more than one ARCP event per year (e.g. some core programmes rotate every eight months, so a trainee could potentially have 2 ARCPs in a given reporting year). Some trainees receive an outcome 5 (insufficient evidence); followed by another outcome upon presentation of the evidence in the same year.

The data collection notices that the GMC issues for this data collection are available here arranged by data collection year: <https://www.gmc-uk.org/education/reports-and-reviews/progression-reports/downloads-resources-and-briefing-notes>

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ARCP_EVENT_ID	VARCHAR2(50 CHAR)	Unique ID for each record in the ARCP data. The ID is generated by the GMC after the files are returned to the GMC and will be used to join to records in the ARCP_REASONS table.
ARCP_SUBMITTING_DEANERY	VARCHAR2(100 CHAR)	The deanery that returned the ARCP record to the GMC.
ARCP_DATA_YEAR	VARCHAR2(4 CHAR)	The year the data were returned to the GMC.
ARCP_REVIEW_DATE	DATE	Date of RITA or ARCP event.
ARCP_PERIOD_START_DATE	DATE	Start of the period that assessment refers to.
ARCP_PERIOD_END_DATE	DATE	End of the period that the assessment refers to.
ARCP_LEVEL_ASSESSED	VARCHAR2(50 CHAR)	Grade of trainee at the point of assessment, or at the point of the missed assessment.
ARCP_POST_TYPE	VARCHAR2(50 CHAR)	Trainee's Post type (e.g. Core, Specialty, Foundation).
ARCP_MILITARY_IND	VARCHAR2(3 CHAR)	Indicates if the trainee was a defence DPMD trainee (at the time of the ARCP/RITA).

COLUMN NAME	DATATYPE	COMMENTS
ARCP_ACADEMIC_IND	VARCHAR2(3 CHAR)	Indicates if the trainee was an academic trainee (at the time of the ARCP/RITA).
ARCP_REMAIN_ON_ACADEMIC_PROG	VARCHAR2(250 CHAR)	List of values that describes if (and why) the trainee will remain on their academic programme.
ARCP_SPECIALTY	VARCHAR2(250 CHAR)	The specialty to which the ARCP event referred.
ARCP_NOT_FULL_TIME_IND	VARCHAR2(50 CHAR)	Indicates if the trainee was less than full time during any part of the period covered by the ARCP.
ARCP_OUTCOME_CODE	VARCHAR2(250 CHAR)	ARCP / RITA Outcome. Possible values: 1; 2; 3; 4; 5; 6; 7.1; 7.2; 7.3; 7.4; 8; 9; C; D; E; F; G.
ARCP_OUTCOME_CLASSIF	VARCHAR2(50 CHAR)	Description for OUTCOME_CODE available in the progression report documentation here: https://www.gmc-uk.org/-/media/documents/dc11341-progreportutr-arcp-pg-jun2018_pdf-75128581.pdf
ARCP_OUTCOME_DESCRIPTION	VARCHAR2(250 CHAR)	Description for OUTCOME_CODE – available in the progression report documentation here: https://www.gmc-uk.org/-/media/documents/dc11341-progreportutr-arcp-pg-jun2018_pdf-75128581.pdf
ARCP_OUTCOME_APPL_TO_FOUND	VARCHAR2(50 CHAR)	Please see progression report documentation here: https://www.gmc-uk.org/-/media/documents/dc11341-progreportutr-arcp-pg-jun2018_pdf-75128581.pdf
ARCP_OUTCOME_TYPE	VARCHAR2(50 CHAR)	Describes if this is a clinical or academic outcome.
ARCP_BENCHMARK_GROUP	VARCHAR2(50 CHAR)	Derived variable for grouping specialties for reporting purposes.
ARCP_OUTCOME_ORDERED	NUMBER	Outcomes on an ordinal scale '1'=1 '2'=3 '3'=4 '4'=4 '5'=2 '6'=1 '7'=1 '7.1'=1 '7.2'=3 '7.3'=4 '7.4'=2 '8'=100 '9'=100 'C'=1 'D'=3 'E'=4 'F'=100 'G'=1 1 'Satisfactory progression' 2 'Insufficient evidence presented' 3 'Targeted training required (but training time not extended)' 4 'Extended training time required/left programme'

COLUMN NAME	DATATYPE	COMMENTS
		<p>100 'Out of programme'</p> <p>The outcomes awarded were coded into an ordinal scale following the method developed by Tiffin et al (2014). Tiffin, P. A., Illing, J., Kasim, A. S. & McLachlan, J. C. (2014). Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. BMJ 348: g2622. Available at: http://www.bmj.com/content/348/bmj.g2622</p>
ARCP_OUTCOME_ORDERED_DES	VARCHAR2(100 CHAR)	Label for ordered number.

ARCP_REASON

This table contains the U codes which describe the reasons for an unsatisfactory outcome. These are detailed on the Gold Guide ARCP form here:

[https://www.copmed.org.uk/images/docs/gold_guide_7th_edition/Appendix_3 -
_ARCP_Outcome_Form.pdf](https://www.copmed.org.uk/images/docs/gold_guide_7th_edition/Appendix_3_-_ARCP_Outcome_Form.pdf)

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ARCP_EVENT_ID	VARCHAR2(50 CHAR)	The ARCP event to which this reason relates. The event id should be used for linking to the ARCP_OUTCOMES table.
ARCP_YEAR	VARCHAR2(4 CHAR)	The ARCP year to which this reason relates.
REASON_CODE	VARCHAR2(5 CHAR)	Reason code as provided in the ARCP data. U1 to U10
REASON_DESCRIPTION	VARCHAR2(250 CHAR)	Description of the ARCP Reason Code (as held in REASON_CODE).

ARCP_N_CODES

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ARCP_EVENT_ID	VARCHAR2(50 CHAR)	The ARCP event to which this reason relates. The event id should be used for linking to the ARCP_OUTCOMES table.
ARCP_YEAR	VARCHAR2(4 CHAR)	The ARCP year to which this reason relates.
N_CODE	VARCHAR2(5 CHAR)	Reason code as provided in the ARCP data. N1 to N
N_DESCRIPTION	VARCHAR2(250 CHAR)	Description of the ARCP N Code (as held in N_CODE).

BMAT_SCORES_APP

UKMED holds BMAT scores for doctors who sat the test between 2003 and 2017.

Two matching processes were used depending on which privacy notice had been displayed to the BMAT candidate

BMAT test takers from 2014 onwards who saw a privacy notice that allowed sharing were matched to UCAS data, so for these cases UKMED contains data for applicants to medical school.

For those that sat BMAT prior to 2014 UKMED only contains data from candidates who entered medical school. who meet the following criteria:

- Graduated and registered with the GMC;
- GMC were able to contact the individual by email; and
- The doctor did not opt out including their data in UKMED. It was necessary to give people the option of opting out because the privacy notice they saw that stated the data would not be shared. Two rounds of opt out emails have been sent
 - May 2017: 16,062 doctors were emailed, 277 cases (1.7%) opted out of the transfer.
 - February 2019: 4,576 cases, of these 64 (1.4%) opted out of the transfer
- There are still BMAT test takers from prior to 2014 who have not appeared on the GMC's register, so it has not been possible to send them an opt-out email and load their data.

BMAT scores are valid for courses commencing the year following the test. For example, a 2006 score can be used for courses commencing in 2007. A very small number of candidates may have been given permission to defer their entry for one year by the medical school.

The schools and courses using BMAT vary over time. An applicant would have taken BMAT if they applied to at least one school requiring BMAT. UKMED holds test results for anyone who took BMAT. The student may have gained admission to a medical school that did not require it.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(64 BYTE)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. For cases that relate to unsuccessful medical school applications only, the ID will be the UCAS PERSON_ID with a 'U' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
UCAS_APPLICANT_ID	VARCHAR2(64 BYTE)	<i>Internal use only. Used for linking to HESA data.</i>
UCAS_RPD_ID_ALL	VARCHAR2(70 BYTE)	<i>Internal use only. Used for linking to HESA data.</i>
BMAT_CAND_SESSION	VARCHAR2(40 BYTE)	Session BMAT test sat – not all years had more than one

COLUMN NAME	DATATYPE	COMMENTS
		session
BMAT_CAND_YEAR	NUMBER(4)	Year BMAT test sat
BMAT_ID	VARCHAR2(20 BYTE)	<i>Internal use only.</i> Note this ID is only unique within BMAT data year.
BMAT_UCASPERID	VARCHAR2(10 BYTE)	<i>Internal use only.</i> Used for linking to HESA data.
BMAT_UCASAPPID	VARCHAR2(10 BYTE)	<i>Internal use only.</i> Used for linking to HESA data.
BMAT_SECTION_1	NUMBER	Aptitude and Skills score.
BMAT_SECTION_2	NUMBER	Scientific Knowledge and Applications.
BMAT_SECTION_3	VARCHAR2(15 BYTE)	Writing task score. Available for BMAT years 2004–2009.
BMAT_SECTION_3_CONTENT	VARCHAR2(15 BYTE)	Writing task – Quality of Content score Available for BMAT years 2010–2016.
BMAT_SECTION_3_ENGLISH	VARCHAR2(15 BYTE)	Writing task – Quality of English score Available for BMAT years 2010–2016.
BMAT_ABSENT_FLAG	VARCHAR2(1 BYTE)	Flag to indicate if the candidate attended their test session. Cases with this flag have no test score data
BMAT_MATCH_METHOD	VARCHAR2(50 BYTE)	Method used to assign PERSON_UID.
BMAT_SECTION_1_Z	NUMBER(2,3)	Aptitude and Skills score. Based on the Means and SD for BMAT candidature by session 2013-2017 supplied by BMAT and Means and SD released to UK med schools for 2003 to 2012.
BMAT_SECTION_2_Z	NUMBER(2,3)	Scientific Knowledge and Applications. Based on the Means and SD for BMAT candidature by session 2013-2017 supplied by BMAT and Means and SD released to UK med schools for 2003 to 2012.
BMAT_SECTION_3_Z	NUMBER(2,3)	Writing task score. Available for BMAT years 2004–2009. Based on the Means and SD for BMAT candidature by session 2013-2017 supplied by BMAT and Means and SD released to UK med schools for 2003 to 2012.
BMAT_SECTION_3_CONTENT_Z	NUMBER(2,3)	Writing task – Quality of Content score Available for BMAT years 2010–2016. Based on the Means and SD for BMAT candidature by session 2013-2017 supplied by BMAT and Means and SD released to UK med schools for 2003 to 2012.

COLUMN NAME	DATATYPE	COMMENTS
BMAT_SECTION_3_ENGLISH_Z	NUMBER(2,3)	Writing task – Quality of English score Available for BMAT years 2010–2016. Based on the Means and SD for BMAT candidature by session 2013-2017 supplied by BMAT and Means and SD released to UK med schools for 2003 to 2012.

EXAM_SECTION_MARKS

This table contains:

1. Exam data collected by the GMC from the medical royal colleges and faculties annually since 1 August 2013. The data collection notices arranged by year are here: <https://www.gmc-uk.org/education/reports-and-reviews/progression-reports/downloads-resources-and-briefing-notes>.
2. Data on The British Pharmacological Society (BPS) and MSC Assessment from 2014 until 2018 supplied to UKMED by the Medical Schools Council. CR_PSA_APPROACH records whether the PSA was formative or summative by medical school and year.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(7 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. In this table, this applies to same records for attempts at the Prescribing Safety Assessment. Replaced by extract specific STUDY_ID in de-identified research extracts.
EXAM_DATERESULT	DATE	The date the exam result was issued by the college.
EXAM_ABBREVNNAME	VARCHAR2(100 CHAR)	CR_EXAM indicates which exams have section makes available
EXAM_EXAMID	VARCHAR2(100 CHAR)	The unique ID for this table, a concatenation of PERSON_UID, date of result and abbreviated exam name. <i>Internal use only.</i>
EXAM_SECTIONNAME	VARCHAR2(100 CHAR)	The name of the exam section.
EXAM_SECTIONSCORE	NUMBER(10,2)	The score achieved for the section.

EXAM_TOTAL_MARKS

This table contains:

1. Exam data collected by the GMC from the medical royal colleges and faculties annually since 1 August 2013. The data collection notices arranged by year are here: <http://www.gmc-uk.org/education/29409.asp>; <https://www.gmc-uk.org/education/reports-and-reviews/progression-reports/downloads-resources-and-briefing-notes>. This covers all exams sat since 1 August 2013. Note that some left censoring of this table is possible as some cases will have sat exams prior to the 1 August 2013.
2. Data on The British Pharmacological Society (BPS) and MSC Assessment from 2014 until 2020 supplied to UKMED by the Medical Schools Council. CR_PSA_APPROACH records whether the PSA was formative or summative by medical school and year.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(255 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. In this table, this applies to same records for attempts at the Prescribing Safety Assessment. Replaced by extract specific STUDY_ID in de-identified research extracts.
EXAM_DATERESULT	DATE	The date the exam result was issued by the college.
EXAM_EXAMID	VARCHAR2(255 CHAR)	The unique ID for this table, a concatenation of PERSON_UID and date of result <i>Internal use only.</i>
EXAM_COLLEGE	VARCHAR2(255 CHAR)	All colleges and almost all faculties have signed a data sharing agreement the table CR_EXAM gives the names of colleges whose data are contained in UKMED.
EXAM_COLLEGEID	VARCHAR2(1020 BYTE)	The record ID received from the college in their original submission. Not all colleges provided IDs. It is included here to ensure traceability. <i>Internal use only.</i>
EXAM_ABBREVMNAME	VARCHAR2(255 CHAR)	The name of the exam contained in the college's submission to the GMC.
EXAM_TOTALMARK	NUMBER	The total mark achieved in the exam, where supplied by the college. For some college this is a score; for others it a percent. The format is given in CR_EXAM.
EXAM_PASSMARK	NUMBER	The pass mark for the exam where supplied by the college.
EXAM_PASSFAIL	NUMBER	1= Pass and 0 =Fail.
EXAM_ATTEMPTS	NUMBER	As shown in CR_EXAM not all colleges have recorded the number of attempts. In some cases, it is the number of attempts up to the attempt recorded in the record, in others it

COLUMN NAME	DATATYPE	COMMENTS
		is a total including the attempts in the record.

EXAM_DATA_FLAT

This table contains the data from EXAM_TOTAL_MARKS rearranged with one row per person per exam with columns D1 through to D15 for each attempt where D1 is the first attempts held within UKMED

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(255 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. In this table, this applies to same records for attempts at the Prescribing Safety Assessment. Replaced by extract specific STUDY_ID in de-identified research extracts.
COLLEGE_EXAM	VARCHAR2(511 CHAR)	A concatenation of the college and the exam
D1_EXAM_DATERESULT	DATE	The date the exam result was issued by the college for the first attempt at this exam.
D1_EXAM_TOTALMARK	NUMBER	The total mark achieved in the exam for the first attempt, where supplied by the college. For some college this is a score; for others it a percent. The format is given in CR_EXAM.
D1_EXAM_PASSMARK	NUMBER	The pass mark for the exam on the first attempt where supplied by the college.
D1_SCORE_RELATIVE_TO_PASS	NUMBER	D1_EXAM_TOTALMARK D1_EXAM_PASSMARK
D1_EXAM_PASSFAIL	NUMBER	1= Pass and 0 =Fail for the first attempt at this exam.
Continues for D2 through to D15 depending on the exam		

FP

The UK Foundation Programme Office (UKFPO) provided extracts of data from the Foundation Programme Application System (FPAS) for all applications received from UK Medical Students from 2012 to 2014; from 2015 onwards, the data also included applications from non-UK PMQ holders. Data from 2017 onwards come from the ORIEL system.

To understand the process, please see the Applicant handbook here:

<http://www.foundationprogramme.nhs.uk/applicant-guidance>.

The GMC hold the applicant guides for earlier years and these are available on request.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
FP_MATCH_SCRIPT	VARCHAR2(100 CHAR)	Method used to assign PERSON_UID.
FP_DEDUPMETHOD	VARCHAR2(100 CHAR)	Method applied for de-duplicating records where required.
FP_DATA_YEAR	NUMBER	Derived based on year the extract was received
FP_ID	VARCHAR2(20 CHAR)	To link back to source date if required. Needs to be set to String due to the IDs used in 2012. <i>Internal use only</i>
FP_SJT	NUMBER	The Situational Judgement Test (SJT) assumes that applicants have knowledge of and insight into the job role of an FY1 doctor and applicants will be asked to respond as they should as an FY1 doctor. The scenarios are set in a clinical setting but there is no requirement for specific clinical knowledge as the test is targeting professional attributes, such as: Commitment to Professionalism, Coping with Pressure, Effective Communication, Patient Focus and Working Effectively as Part of a Team See https://isfp.org.uk/sjt-development-and-piloting/ and http://www.foundationprogramme.nhs.uk/content/situational-judgement-test-sjt
FP_SJT_RAW	NUMBER(4,1)	Score prior to equating.
FP_SJT_EQUATED	NUMBER(8,4)	Equating is described in the technical reports. For example Patterson F, Murray, H, Baron, H, Aitkenhead A and Flaxman, C. (June 2014) <i>Analysis of the Situational Judgement Test for Selection to the Foundation Programme 2014 Technical Report</i> . Available at https://isfp.org.uk/fp-technical-reports/ .

COLUMN NAME	DATATYPE	COMMENTS
FP_QUESTIONPAPER	NUMBER	There are typically three papers a year.
FP_SJT_RAW_PART1	NUMBER	<p>In part one (two thirds of the questions) there are four different formats:</p> <ul style="list-style-type: none"> Rank in order the appropriateness of the following actions in response to this situation (1= Most appropriate; 5= Least appropriate). Rank the order in which the following tasks should be undertaken (1= Do first; 5= Do last). Rank in order the importance of the following considerations in the management of this situation (1= Most important; 5= Least important). Rank in order the extent to which you agree with the following statements in this situation (1= Most agree with; 5= Least agree with).
FP_SJT_RAW_PART2	NUMBER	<p>In part two there are two formats of lead-in:</p> <ul style="list-style-type: none"> Choose the three most appropriate actions to take in this situation. Choose the three most important considerations to consider in the management of this situation.
FP_EPM_QUARTILE	NUMBER	<p>FPAS 2012 only</p> <p>Medical school performance (34 - 43 points)</p> <p>Applicants' academic ranking is calculated by their medical school, which has divided your year group into four quartiles based on academic performance. Applicants in the first quartile (the top 25% of your year) will receive a score of 40; applicants in the second quartile 38; the third quartile 36 and the fourth quartile 34.</p>
FP_EPM_DECILE	NUMBER	<p>FPAS 2013 onwards only</p> <p>Medical school performance (34–43 points)</p> <p>Medical school performance score calculated by the applicant's medical school and divided into 10 equal groups (deciles) based on performance in a number of assessments. Each UK medical school has agreed with its students' which assessments will be included in this measure. This element of the EPM is known as the EPM decile score. The first decile (the top 10% of the applicant's year) will receive a score of 43; if you are in the second decile the applicant will receive a score of 42; the third decile 41 and so on. Students in the tenth decile will be awarded 34 points.</p> <p>For further details see: http://www.foundationprogramme.nhs.uk/content/epm-sjt</p>
FP_EPM_DEGREE_SCORE	NUMBER	<p>Additional degrees (up to 5 points):</p> <p>0 – no degree, with PhD 5 points</p> <p>Missing to be re-coded to 0.</p>
FP_EPM_PUB_SCORE	NUMBER	The UKFPO confirmed on 7 August 2014 that two source

COLUMN NAME	DATATYPE	COMMENTS
		<p>columns can be combined into one.</p> <p>They are not directly comparable over the years as the scoring mechanism varies as follows:</p> <p>2012 Other educational achievements (up to 5 points)</p> <p>2013 Other educational achievements (up to 2 points)</p> <p>2014 Other educational achievements (up to a maximum of 2 points) points could only be claimed for publications and presentations.</p> <p>2015 onwards points will only be awarded for publications with a maximum of 2 points</p>
FP_OTHER_QUAL_IND	VARCHAR2(3 CHAR)	
FP_OTHER_QUAL_TITLE	VARCHAR2(1000 CHAR)	
FP_DATE_OTHER_QUAL	DATE	Not present in 2012; 2013 onwards only.
FP_OTHER_QUALINSTITUTION	VARCHAR2(200 CHAR)	Not present in 2012; 2013 onwards only.
FP_WITHDRAW_DATE	DATE	Can be used to account for cases not linking to the register.
FP_WITHDRAW_REASON	VARCHAR2(100 CHAR)	Can be used to account for cases not linking to the register.
FP_PROG_REF	VARCHAR2(100 CHAR)	This identifier links to a description of the programme's providers and placement specialties and carries over to other databases such as the foundation e-portfolio.
FP_PROG_RANK	VARCHAR2(20 CHAR)	<p>Applicants can rank as many or as few programmes as they would like. If you do not secure one of the programmes you have ranked, you will be randomly allocated to one of the programmes you did not rank.</p> <p>This number gives the rank the applicant gave to the programme they have been allocated to. It only applies to those who did not accept an AFP offer.</p> <p>Higher scoring applicants get a higher preference.</p>
FP_PROG_TYPE	VARCHAR2(20 CHAR)	Identifies the type of foundation programme, FP = standard Foundation Programme, AFP = Academic Foundation Programme.
FP_PRIMARY	VARCHAR2(20 CHAR)	<p>The 'n' top scoring applicants are placed on the primary list at the time of the initial allocation to UoA, where 'n' equals the total number of vacancies available across the UK. Any remaining applicants are placed on the reserve list.</p> <p>Applicants who accepted an AFP offer are excluded from the FP allocation and are not included in either the primary or reserve list.</p>
FP_UOA	VARCHAR2(100 CHAR)	This shows the AUoA where AFP applicants have accepted an AFP offer, or the UoA where FP applicants have been

COLUMN NAME	DATATYPE	COMMENTS
		<p>allocated.</p> <p>It is possible that applicants may transfer to a different foundation school at a later date via an inter-foundation school transfer, but this is not captured on FPAS.</p> <p>A/UoAs are geographical groups of one or more foundation schools</p>
FP_UOA_PREF	VARCHAR2(20 CHAR)	<p>The rank the applicant gave to their FPAS_UOA. Where 1 is their first choice as so on. See the Applicants handbook for further details:</p> <p>http://www.foundationprogramme.nhs.uk/download.asp?file=FP2015_Applicants_Handbook_FINAL.pdf</p> <p>Only valid for FP applicants since AFP applicants do not rank the AUoAs they apply to.</p>
FP_APP_STAGE	VARCHAR2(100 CHAR)	Stage of the application at the point the extract was generated.
FP_GROUP_RANK	VARCHAR2(10 CHAR)	Where a UoA uses groups, this gives the preference the applicant was matched to. The programme rank would only be the preference with the matched group so isn't a true reflection of their preferences. For example, they could be matched to their first-choice programme, but this was in their 50th choice group so not really their first choice at all.
FP_NATIONALITY	VARCHAR2(255 BYTE)	Applicant's nationality. Available from 2017 onwards.
FP_MEDSCHOOLCOMDATE	DATE	Date the applicant started at medical school as recorded on their application to foundation. Available from 2017 only.
FP_MEDSCHOOLDATEQUAL	DATE	Date the applicant obtained their primary medical qualification as recorded on their application to foundation. Available from 2017 only.
FP_OTHER_QUALINSTITUTE_G PA	VARCHAR2(255 BYTE)	Available for applicants from overseas universities where they provide Grade Point Average (GPA) points. Available from 2017 onwards.
FP_OTHER_QUALTYPE	VARCHAR2(255 BYTE)	Type of other qualification achieved. Available from 2017.
FP_FIRST_RANKED_UOA	VARCHAR2(255 BYTE)	The unit of application that the applicant ranked as their first choice. Available from 2017 onwards.
FP_ALLOCATED_UOA_DATE	DATE	The date of programme allocation. Available from 2017 onwards.
FP_MATCHED_GROUP	VARCHAR2(255 BYTE)	The group to which the applicant was matched. Available from 2017 onwards.
FP_MATCHED_PROGRAMME	VARCHAR2(255 BYTE)	Programme to which the applicant was matched. Available from 2017 onwards.

COLUMN NAME	DATATYPE	COMMENTS
FP_APPLICANT_PROGRAMME_RANK	NUMBER	The rank given by the applicant to the Ooaa which they were matched to. Available from 2017 onwards.
FP_AFP_ALL_APP_STAGES	VARCHAR2(255 BYTE)	Details of applications to the academic foundation programme if an application was made. It is possible to apply to two UoA when applying for AFP.
FP_AFP_AFP_OFFACCEPTED	VARCHAR2(255 BYTE)	Indicates whether an AFP offer has been accepted. Available from 2017 onwards.
FP_AFP_OFFERPREF_ACCEPTED	VARCHAR2(255 BYTE)	Programme code of the AFP accepted.
FP_AFP_OFFERRANK_ACCEPTED	NUMBER	Rank given by the applicant to the accepted programme.
FP_AFP_FOUNDATION_ACCEPTED	VARCHAR2(255 BYTE)	Details of the AFP accepted.
FP_AFP_TOTAL_AFP_APPLICATIONS	NUMBER	Number of AFP applications

VW_FP_PERSON

This table contains the data from the FP table arranged so that each case only has one row of data with columns for each application year. There are additional columns containing derived scores as described below.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
FIRST_APP_YEAR	NUMBER	The first year an application was made.
FP_N_ENTRIES	NUMBER	The number of applications made
FP_ID_2012	VARCHAR2(20 CHAR)	Derived based on year the extract was received
FP_DATA_YEAR_2012	NUMBER	To link back to source date if required. Needs to be set to String due to the IDs used in 2012. <i>Internal use only</i>
FP_EPM_QUARTILE_2012	NUMBER	FPAS 2012 only Medical school performance (34 - 43 points) Applicants' academic ranking is calculated by their medical school, which has divided your year group into four quartiles based on academic performance. Applicants in the first quartile (the top 25% of your year) will receive a score of 40; applicants in the second quartile 38; the third quartile 36 and the fourth quartile 34.
FP_EPM_PUB_SCORE_2012	NUMBER	Additional degrees (up to 5 points): 0 – no degree, with PhD 5 points Missing to be re-coded to 0.
FP_EPM_DEGREE_SCORE_2012	NUMBER	The UKFPO confirmed on 7 August 2014 that two source columns can be combined into one. They are not directly comparable over the years as the scoring mechanism varies as follows: 2012 Other educational achievements (up to 5 points) 2013 Other educational achievements (up to 2 points) 2014 Other educational achievements (up to a maximum of 2 points) points could only be claimed for publications and presentations. 2015 onwards points will only be awarded for publications with a maximum of 2 points
FP_ID_2013	VARCHAR2(20 CHAR)	Derived based on year the extract was received
FP_DATA_YEAR_2013	NUMBER	To link back to source date if required. Needs to be set to

COLUMN NAME	DATATYPE	COMMENTS
		String due to the IDs used in 2012. <i>Internal use only</i>
FP_SJT_2013	NUMBER	<p>The Situational Judgement Test (SJT) assumes that applicants have knowledge of and insight into the job role of an FY1 doctor and applicants will be asked to respond as they should as an FY1 doctor. The scenarios are set in a clinical setting but there is no requirement for specific clinical knowledge as the test is targeting professional attributes, such as: Commitment to Professionalism, Coping with Pressure, Effective Communication, Patient Focus and Working Effectively as Part of a Team</p> <p>See https://isfp.org.uk/sjt-development-and-piloting/</p> <p>and</p> <p>http://www.foundationprogramme.nhs.uk/content/situational-judgement-test-sjt</p>
FP_SJT_RAW_2013	NUMBER(4,1)	Score prior to equating.
FP_SJT_EQUATED_2013	NUMBER (8,4)	Equating is described in the technical reports. For example, Patterson F, Murray, H, Baron, H, Aitkenhead A and Flaxman, C. (June 2014) <i>Analysis of the Situational Judgement Test for Selection to the Foundation Programme 2014 Technical Report</i> . Available at https://isfp.org.uk/fp-technical-reports/ .
FP_EPM_DECILE_2013	NUMBER	<p>FPAS 2013 onwards only</p> <p>Medical school performance (34–43 points)</p> <p>Medical school performance score calculated by the applicant's medical school and divided into 10 equal groups (deciles) based on performance in a number of assessments. Each UK medical school has agreed with its students which assessments will be included in this measure. This element of the EPM is known as the EPM decile score. The first decile (the top 10% of the applicant's year) will receive a score of 43; if you are in the second decile the applicant will receive a score of 42; the third decile 41 and so on. Students in the tenth decile will be awarded 34 points.</p> <p>For further details see: http://www.foundationprogramme.nhs.uk/content/epm-sjt</p>
FP_EPM_PUB_SCORE_2013	NUMBER	<p>Additional degrees (up to 5 points):</p> <p>0 – no degree, with PhD 5 points</p> <p>Missing to be re-coded to 0.</p>
FP_EPM_DEGREE_SCORE_2013	NUMBER	<p>The UKFPO confirmed on 7 August 2014 that two source columns can be combined into one.</p> <p>They are not directly comparable over the years as the scoring mechanism varies as follows:</p> <p>2012 Other educational achievements (up to 5 points)</p> <p>2013 Other educational achievements (up to 2 points)</p> <p>2014 Other educational achievements (up to a maximum of 2</p>

COLUMN NAME	DATATYPE	COMMENTS
		points) points could only be claimed for publications and presentations. 2015 onwards points will only be awarded for publications with a maximum of 2 points
FP_ID_2014	VARCHAR2(20 CHAR)	As per 2013
FP_DATA_YEAR_2014	NUMBER	As per 2013
FP_SJT_2014	NUMBER	As per 2013
FP_SJT_RAW_2014	NUMBER (4,1)	As per 2013
FP_SJT_EQUATED_2014	NUMBER (8,4)	As per 2013
FP_EPM_DECILE_2014	NUMBER	As per 2013
FP_EPM_PUB_SCORE_2014	NUMBER	As per 2013
FP_EPM_DEGREE_SCORE_2014	NUMBER	As per 2013
FP_ID_2015	VARCHAR2(20 CHAR)	As per 2013
FP_DATA_YEAR_2015	NUMBER	As per 2013
FP_SJT_2015	NUMBER	As per 2013
FP_SJT_RAW_2015	NUMBER (4,1)	As per 2013
FP_SJT_EQUATED_2015	NUMBER (8,4)	As per 2013
FP_EPM_DECILE_2015	NUMBER	As per 2013
FP_EPM_PUB_SCORE_2015	NUMBER	As per 2013
FP_EPM_DEGREE_SCORE_2015	NUMBER	As per 2013
FP_ID_2016	VARCHAR2(20 CHAR)	As per 2013
FP_DATA_YEAR_2016	NUMBER	As per 2013
FP_SJT_2016	NUMBER	As per 2013

COLUMN NAME	DATATYPE	COMMENTS
FP_SJT_RAW_2016	NUMBER (4,1)	As per 2013
FP_SJT_EQUATED_2016	NUMBER (8,4)	As per 2013
FP_EPM_DECILE_2016	NUMBER	As per 2013
FP_EPM_PUB_SCORE_2016	NUMBER	As per 2013
FP_EPM_DEGREE_SCORE_2016	NUMBER	As per 2013
FP_ID_2017	VARCHAR2(20 CHAR)	As per 2013
FP_DATA_YEAR_2017	NUMBER	As per 2013
FP_SJT_2017	NUMBER	As per 2013
FP_SJT_RAW_2017	NUMBER (4,1)	As per 2013
FP_SJT_EQUATED_2017	NUMBER (8,4)	As per 2013
FP_EPM_DECILE_2017	NUMBER	As per 2013
FP_EPM_PUB_SCORE_2017	NUMBER	As per 2013
FP_EPM_DEGREE_SCORE_2017	NUMBER	As per 2013
FP_ID_2018	VARCHAR2(20 CHAR)	As per 2013
FP_DATA_YEAR_2018	NUMBER	As per 2013
FP_SJT_2018	NUMBER	As per 2013
FP_SJT_RAW_2018	NUMBER (4,1)	As per 2013
FP_SJT_EQUATED_2018	NUMBER (8,4)	As per 2013
FP_EPM_DECILE_2018	NUMBER	As per 2013
FP_EPM_PUB_SCORE_2018	NUMBER	As per 2013
FP_EPM_DEGREE_SCORE_2018	NUMBER	As per 2013

COLUMN NAME	DATATYPE	COMMENTS
FP_ID_2019	VARCHAR2(20 CHAR)	As per 2013
FP_DATA_YEAR_2019	NUMBER	As per 2013
FP_SJT_2019	NUMBER	As per 2013
FP_SJT_RAW_2019	NUMBER (4,1)	As per 2013
FP_SJT_EQUATED_2019	NUMBER (8,4)	As per 2013
FP_EPM_DECILE_2019	NUMBER	As per 2013
FP_EPM_PUB_SCORE_2019	NUMBER	As per 2013
FP_EPM_DEGREE_SCORE_2019	NUMBER	As per 2013
EPM_DECILE_FIRST_YEAR	NUMBER	1 st year with an EPM decile score. E.g. if there are scores in 2013 and 2014 then this would be 2013.
EPM_DECILE_FIRST	NUMBER	EPM decile score from the 1 st year with a score
FP_EPM_YEAR	NUMBER	The year from which the 1 st EPM score (quartile or decile) was recorded
EPM_SCORE_FIRST	NUMBER	The first available EPM score, for 2012 this is a quartile; otherwise this is a decile.
EPM_HALF	VARCHAR2(10 BYTE)	Bottom 50% or top 50% based on 1 st quartile of decile
EPM_QUARTILE_INT	NUMBER	1 to 4
EPM_QUARTILE_STR	VARCHAR2(19 BYTE)	1 = 'Lowest quartile, 2 = '2nd lowest quartile' 3= '3rd quartile ' 4 = 'Top quartile'
FP_FIRST_DECILE_YEAR	NUMBER	The year with the first decile
FP_FIRST_DECILE	NUMBER	The decile score from the first year with a decile
FP_FIRST_DECILE__INT_RC	NUMBER	The decile score recoded - 34 = 1 through to 43 = 10
FP_FIRST_DECILE__STR_RC	VARCHAR2(24 BYTE)	34 = '10th decile - bottom 10%' 35 = '9th decile' through to 43 = '1st decile - Top 10%'

COLUMN NAME	DATATYPE	COMMENTS
FP_NORMAL_DEVIATE	NUMBER	<p>The normal deviate score to place deciles and quartiles on the same scale</p> <p>fp_epm_quartile_2012 = 34 then -1.0491 fp_epm_quartile_2012 = 36 then -0.2993 fp_epm_quartile_2012 = 38 then 0.2993 fp_epm_quartile_2012 = 40 then 0.2993 fp_epm_decile = 34 then -1.5466 fp_epm_decile = 35 then -1.005 fp_epm_decile = 36 then -0.6554 fp_epm_decile = 37 then -0.3755 fp_epm_decile = 38 then -0.1226 fp_epm_decile = 39 then 0.1226 fp_epm_decile = 40 then 0.3755 fp_epm_decile = 41 then 0.6554 fp_epm_decile = 42 then 1.0005 fp_epm_decile = 43 then 1.5466</p> <p>as per Garrud P, McManus IC. Impact of accelerated, graduate-entry medicine courses: a comparison of profile, success, and specialty destination between graduate entrants to accelerated or standard medicine courses in UK. BMC Med Educ. 2018;18(1):250.</p>
FP_DEVIATE_SRC	VARCHAR2(49 BYTE)	The year from which the quartile or decile score came to calculate FP_NORMAL_DEVIATE

GAMSAT_APP

This table contains test score for Graduate Medical School Admissions Test (GAMSAT). Please see: <https://gamsat.acer.org/> .

UKMED holds GAMSAT data for all applicants matched to UCAS applications from 2007 in addition to data matched to entrants in HESA for those taking GAMSAT for entry from 2005 onwards.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(64 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. For cases that relate to unsuccessful medical school applications only, the ID will be the UCAS PERSON_ID with a 'U' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
GAMSAT_ATTEMPTID	VARCHAR2(10 CHAR)	GAMSAT's own unique ID for a test attempt. <i>Internal use only</i>
GAMSAT_UCASPERSONID	VARCHAR2(20 CHAR)	UCAS Person ID, used for linking to HESA data. <i>Internal use only.</i>
GAMSAT_HIGHESTQUALIFICATION	VARCHAR2(50 CHAR)	Qualification information declared by GAMSAT candidate on the GAMSAT registration form.
GAMSAT_DEGREEYEARCOMPLETED	VARCHAR2(10 CHAR)	Qualification information declared by GAMSAT candidate on the GAMSAT registration form.
GAMSAT_MAJORSUBJECTAREA1	VARCHAR2(50 CHAR)	Qualification information declared by GAMSAT candidate on the GAMSAT registration form.
GAMSAT_MAJORSUBJECTAREA2	VARCHAR2(50 CHAR)	Qualification information declared by GAMSAT candidate on the GAMSAT registration form.
GAMSAT_CLASSACHIEVED	VARCHAR2(50 CHAR)	Qualification information declared by GAMSAT candidate on the GAMSAT registration form.
GAMSAT_SECTION1	VARCHAR2(3 CHAR)	Reasoning in Humanities and Social Sciences.
GAMSAT_SECTION2	VARCHAR2(3 CHAR)	Written Communication.
GAMSAT_SECTION3	VARCHAR2(3 CHAR)	Reasoning in Biological and Physical Sciences.

COLUMN NAME	DATATYPE	COMMENTS
GAMSAT_OVERALLSCORE	VARCHAR2(3 CHAR)	The overall score is calculated as a weighted average of the 3 section scores according to the following formula: $(S1+S2+S3 \times 3)/4$. As section scores are calculated to several decimal points and then rounded, it is not possible to calculate the overall score from the section scores reported to candidates.
GAMSAT_EURESIDENT	VARCHAR2(3 CHAR)	Candidate declared information on whether they are an EU resident.
GAMSAT_TEST_ID	VARCHAR2(15 CHAR)	GAMSAT'S ID for the test sitting includes month where applicable and year
GAMSAT_TESTCENTRE	VARCHAR2(50 CHAR)	Location test sat, not present for all TEST_IDS
GAMSAT_SPECIALTESTCONDITIONS	VARCHAR2(50 CHAR)	Whether special condition applied (Yes or No, no further details. Only available from 2009 onwards.
GAMSAT_MEDICALPROFESSIONAL	VARCHAR2(50 CHAR)	Medical profession of candidate – e.g. Pharmacist, Physio etc . Available from 2009 onwards
GAMSAT_REG_DATE	DATE	Date candidate registered for GAMSAT. Available from 2009 onwards.
GAMSAT_TEST_DATE	DATE	Date candidate took GAMSAT. Available from 2009 onwards.
GAMSAT_MATCH_METHOD	VARCHAR2(100 CHAR)	Method used to match the test result to HESA (and onwards to GMC).
GAMSAT_SECTION1_Z	NUMBER (2,3)	Reasoning in Humanities and Social Sciences. Z-score based on means and SDs for each sitting of GAMSAT UK, from 2006 – 2017 supplied by GAMSAT.
GAMSAT_SECTION2_Z	NUMBER (2,3)	Written Communication. Z-score based on means and SDs for each sitting of GAMSAT UK, from 2006 – 2017 supplied by GAMSAT.
GAMSAT_SECTION3_Z	NUMBER (2,3)	Reasoning in Biological and Physical Sciences. . Z-score based on means and SDs for each sitting of GAMSAT UK, from 2006 – 2017 supplied by GAMSAT.
GAMSAT_OVERALLSCORE_Z	NUMBER (2,3)	Overall score. Z-score based on means and SDs for each sitting of GAMSAT UK, from 2006 – 2017 supplied by GAMSAT.

HESA_ACAD_YEARS

This table contains one row per instance per subject (defined in SBJCA fields see table below) per year contained within the HESA data. Students have one or two rows of data per year depending on how their course is coded with respect to the subject. For UKMED it will be restricted to cases that commenced their medical degrees between 2002 and 2018 (based on HESA_COMDATE).

Note the following:

Records for students who are recorded on an intercalated year in the INTERCALATE field (from 2013/14) and COURSEAIM/QUALAIM field (2002/03 to 2002/13) if they were also recorded as REGBODY=01 in their previous year of study.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
HESA_HUSID	VARCHAR2(50 CHAR)	HUSID – HESA’s unique student identifier see: https://www.hesa.ac.uk/collection/c16051/a/husid/.html/.html/ <i>Internal use only.</i>
HESA_INSTANCEKEY	VARCHAR2(50 CHAR)	HESA Instance Identifier see: https://www.hesa.ac.uk/collection/c15051/a/numhus/.html/.html/ <i>Internal use only.</i>
HESA_NUMHUS	VARCHAR2(50 CHAR)	Student instance identifier. See https://www.hesa.ac.uk/collection/c15051/a/numhus/.html/.html/
HESA_ACYEAR	VARCHAR2(7 CHAR)	Academic year.
HESA_UKPRN	VARCHAR2(50 CHAR)	UK Provider Reference Number See: https://www.hesa.ac.uk/collection/c16041/a/ukprn/.html/.html/
HESA_UKPRN_NAME	VARCHAR2(100 CHAR)	As per the HESA provided label.
HESA_CAMPID	VARCHAR2(150 CHAR)	Campus identifier – The campus with which the student instance is associated. Only for campuses outside of locality of main campus. See https://www.hesa.ac.uk/collection/c16025/e/campid/.

COLUMN NAME	DATATYPE	COMMENTS
HESA_QUAL_AIM	VARCHAR2(200 CHAR)	This field describes the general qualification aim of the course and is intended to record the qualification that will be attained as a result of successful completion of studies. In use from 2002/2003 to 2006/2007.
HESA_COURSE_AIM	VARCHAR2(250 CHAR)	Course Aim This field describes the general qualification aim of the course and is intended to record the qualification that will be attained as a result of successful completion of studies. In use from 2007/08 onwards. See: https://www.hesa.ac.uk/collection/c16051/a/courseaim .
HESA_INTERCALATE	VARCHAR2(50 CHAR)	Indicates if the given year is an intercalated year (i.e. during this year they were studying a different degree, thus their corresponding QUALAIM/COURSAIM may differ) See: https://www.hesa.ac.uk/collection/c15051/a/intercalate Available from 2013/14 only – previously in COURSEAIM or QUALAIM fields). See for instance: https://www.bmh.manchester.ac.uk/medicine/study/undergraduate/intercalation/
HESA_QUAL_OBTAIN_POP	VARCHAR2(125 CHAR)	Derived value added by HESA in order to indicate students that were actively enrolled during the instance.
HESA_HIGH_ED_REG_POP	VARCHAR2(100 CHAR)	Indicates if the student is in the Standard Registration population.
HESA_SBJCA	VARCHAR2(75 CHAR)	This field describes the subject or subjects appropriate to the current course. See – https://www.hesa.ac.uk/support/definitions/students#subject-study-and-jacs-codes and https://www.hesa.ac.uk/collection/c15051/a/sbjca (4 digit for 2002/2003 – 2006/2007) merged with (4 digit for 2007/2008 – 2011/2012)
HESA_SBJCA_YEAR	VARCHAR2(50 CHAR)	The identifier for the HESA code set used in HESA_SBJCA Possible values are: SBJCA_1 SBJCA_2 SBJCA_3
HESA_XFPE01	VARCHAR2(50 CHAR)	Apportions the students to a nominal full person equivalent, according to the number of subjects submitted per course and the weighting allocated to a course's subjects. This may occur across for example where there are two rows for different subjects as part of the same medical degree in the same academic year (A300) Clinical medicine (A100) Pre-clinical medicine

COLUMN NAME	DATATYPE	COMMENTS
		See https://www.hesa.ac.uk/support/definitions/students#count-fpe-fte
HESA_COURSE_ID	VARCHAR2(30 CHAR)	The universities' own course identifiers. In use from 2007/2008 onwards. See https://www.hesa.ac.uk/collection/c16051/a/course_courseid
HESA_RSNEED	VARCHAR2(75 CHAR)	Reason for leaving – indicates the reason the student left medical school. See: https://www.hesa.ac.uk/collection/c16051/a/rsneed
HESA_YEAR_PRG	NUMBER	This field indicates the year number of the course that the student is currently studying. This could be different from the year of student if the student has changed course or re-taken a year. See: https://www.hesa.ac.uk/collection/c16051/a/yearprg .
HESA_YEAR_STUDY	NUMBER	This field indicates the year number that the student is in since enrolling for a course leading to the student's qualification aim (whether or not the intended subject or class has changed) i.e. number of years on this student instance. This could be different from the year of course if the student has changed course or retaken a year See: https://www.hesa.ac.uk/collection/c15051/a/yearstu
HESA_CTITLE	VARCHAR2(300 CHAR)	The medical school's title for the course The course title as studied in this academic year. See: https://www.hesa.ac.uk/collection/c16051/a/ctitle/
HESA_ENDDATE	DATE	The date the student left the student instance detailed in the given record. See: https://www.hesa.ac.uk/collection/c15051/a/enddate
HESA_HUSID_OLD	VARCHAR2(15 CHAR)	Where de-duplication has been required (to merge multiple HUSIDs belonging to one individual), the source HUSID from the original HESA Instance (relating to this academic year) is stored here.
HESA_DEDUP_SCRIPT	VARCHAR2(50 CHAR)	Where de-duplication has been required (to merge multiple HUSIDs belonging to one individual), the script used to identify this is a duplicate is stored here.
HESA_ZPROPFRAN	VARCHAR2(250 CHAR)	Franchise marker (ZPROPFRAN) – The field ZPROPFRAN gives the proportion not taught by the returning institution for each instance. The value ranges between 0 and 1. A value of 0 means that the student is entirely taught by the reporting institution (no franchising), a value of 1 means that the student is entirely taught at another institution. If a student doesn't have a record in the module table or has all modules with 0 FTE, it is assumed that the student is entirely taught by the reporting institution, so ZPROPFRAN is set to 0. See https://www.hesa.ac.uk/support/definitions/students#proporti

COLUMN NAME	DATATYPE	COMMENTS
		on-franchised
HESA_ZTINST1	VARCHAR2(250 CHAR)	Teaching Institution for franchised courses. See https://www.hesa.ac.uk/support/definitions/students#teaching-provider
HESA_ZTINST1_FTE	VARCHAR2(20 CHAR)	Teaching Institution – % time at franchised institution.
HESA_ZTINST2	VARCHAR2(250 CHAR)	Teaching Institution for franchised courses
HESA_ZTINST2_FTE	VARCHAR2(20 CHAR)	Teaching Institution – % time at franchised institution.
HESA_XMODE01	VARCHAR2(250 CHAR)	<p>Mode of study</p> <p>Possible values are Dormant Full-time Part-time Sabbatical Sandwich</p> <p>A student intercalating at another institution will be coded as dormant at the institution they came from.</p> <p>HESA RULE – “where the intercalation is to a different provider then the original provider is required to return a dormant record and the intercalation provider to return the intercalate as a separate instance on their submission with the HUSID maintained and the Instance. INTERCALATE field complete”</p> <p>See https://www.hesa.ac.uk/support/definitions/students#mode-study-qualification</p>
HESA_MEDICAL_SCHOOL	VARCHAR2(50 CHAR)	Medical school assigned from UKPRN table using CR_UKPRN_NAME. Some medical schools have cases assigned to more than one UKPRN as their funding is split across two universities, for example Hull York Medical School.
HESA_COURSE	VARCHAR2(100 CHAR)	Assigned from CR_COURSE USING values obtained from the Medical Schools Council publication on medical school course here: http://www.medschools.ac.uk/Publications/Pages/Entry-requirements-for-UK-medical-schools.aspx

VW_HESA_ACAD_YEAR_PERSON

This table contains one row per student with a set of columns for each year of data in HESA_ACAD_YEARS, where “_YR1” refers to the first ACYEAR for that case and so on. It contains a cleaned version of HESA_RSNEND for each year:

HESA_RSNEND_RC_YR1_CLEAN
HESA_RSNEND_RC_YR2_CLEAN
HESA_RSNEND_RC_YR3_CLEAN
HESA_RSNEND_RC_YR4_CLEAN
HESA_RSNEND_RC_YR5_CLEAN
HESA_RSNEND_RC_YR6_CLEAN
HESA_RSNEND_RC_YR7_CLEAN
HESA_RSNEND_RC_YR8_CLEAN
HESA_RSNEND_RC_YR9_CLEAN
HESA_RSNEND_RC_YR10_CLEAN
HESA_RSNEND_RC_YR11_CLEAN
HESA_RSNEND_RC_YR12_CLEAN
HESA_RSNEND_RC_YR13_CLEAN
HESA_RSNEND_RC_YR14_CLEAN

Where if the student returned the following year, RSNEND for the given year is recoded to “Returned following year RSNEND ignored “. These data anomalies occur because providers report to HESA before the conclusion of some appeals.

HESA_QUALIFICATION

HESA source data contains a list of the qualifications against every academic year at university (thereby capturing any update or change – over time – to the noted "qualifications upon entry"), we only import the set associated with the earliest instance/academic year, i.e. the qualifications stated to UCAS when they first applied.

While HESA joins qualifications onto an "instancekey", i.e. FK = HUSID + UKPRN + NUMHUS, we derive a 1:1 join between VW_UKMED_PERSON and HESA_QUALIFICATION after matching PERSON_UID on both (unique in both tables).

HESA note that institutions are only required to complete qualifications for those who applied via UCAS with Level 3 qualifications who attended a UK institution. Therefore, data are not present for graduate entry students.

Please note, qualifications on entry data was only required for those students entering HE from 2007/08 onwards. As such, earlier years in this item will have far less data than later years.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
HESA_INSTANCEKEY	VARCHAR2(13 CHAR)	Identifies the separate courses taken by an individual student See https://www.hesa.ac.uk/collection/c17051/a/numhus . <i>Internal use only.</i>
HESA_QUALGRADE	VARCHAR2(11 CHAR)	
HESA_QUALSIT	VARCHAR2(3 CHAR)	This field identifies which exam sitting the qualification was obtained. See: https://www.hesa.ac.uk/collection/c15051/a/qualsit
HESA_QUALSBJ	VARCHAR2(100 CHAR)	The subject in which the student obtained the qualification on entry to medical school See: https://www.hesa.ac.uk/collection/c15051/a/qualsbj
HESA_QUALYEAR	NUMBER (4)	The year the student gained the qualification. See: https://www.hesa.ac.uk/collection/c15051/a/qualyear
HESA_QUALTYPE	VARCHAR2(100 CHAR)	The type of qualification obtained on entry to medical school, for example GCE A Level or SQA Highers. See https://www.hesa.ac.uk/collection/c15051/a/qualtype
HESA_INC_TARIFF	VARCHAR2(1 CHAR)	Identifies if the qualification is included in the total HESA tariff calculation.

COLUMN NAME	DATATYPE	COMMENTS
HESA_POINTS	NUMBER	The tariff points associated with the student's qualification. See https://www.hesa.ac.uk/support/definitions/students#tariff

IELTS_LIVE

IELTS is the high-stakes English test for study, migration or work.

See <https://www.ielts.org/>

The GMC requires an IELTS certificate in the academic format of the test.

The GMC requires an IELTS certificate from the following groups:

- Those applying for the PLAB test
- International medical graduates applying for registration with a licence to practise
- European doctors applying for registration with a licence to practise
- Applying for your first licence if you have ever held registration or currently hold registration only

The required IELTS level for PLAB has varied over the years but is currently set at a score of at least 7.0 in each testing area and an overall score of 7.5. Candidates taking PLAB in earlier years may have had lower scores either overall or on subscales. Some PLAB candidates are exempted from the required IELTS level, primarily by demonstrating that their training was at a medical school where the great majority of teaching is in English.

The table also contains Occupational English Test (OET) scores.

See: <https://www.gmc-uk.org/registration-and-licensing/join-the-register/before-you-apply/evidence-of-your-knowledge-of-english/using-your-oet-certificate>

COLUMN NAME	DATATYPE	COMMENTS
AGORA_ID	VARCHAR2(12 CHAR)	GMC ID <i>For internal use only.</i>
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
READING_SCORE	NUMBER (8,2)	See https://www.ielts.org/about-the-test/test-format
WRITING_SCORE	NUMBER (8,2)	See https://www.ielts.org/about-the-test/test-format
LISTENING_SCORE	NUMBER (8,2)	See https://www.ielts.org/about-the-test/test-format
SPEAKING_SCORE	NUMBER (8,2)	See https://www.ielts.org/about-the-test/test-format
GRADE	NUMBER (8,2)	The Overall Band Score. This is the average of the four component scores, rounded to the nearest whole or half band. The component scores are weighted equally.
PASSED_DATE	DATE	

COLUMN NAME	DATATYPE	COMMENTS
STATUS	VARCHAR2(50 CHAR)	Gives the test type IELTS or OET and whether it met GMC requirements Accepted/Not accepted)

MMI_UCAS

These Multiple Mini Interview (MMI) data were submitted by the following schools. For some schools the MMI is only used for a particular course:

- Cambridge covering applications between 2015 and 2017
- Dundee covering applications between 2010 and 2017
- Edinburgh covering applications between 2014 and 2017
- Leicester covering applications between 2012 and 2016
- St Andrews covering applications between 2015 and 2016
- St George's covering applications between 2010 and 2017

For a brief description please see 'An admissions tutor's perspective on the multiple mini interview' here: <http://student.bmj.com/student/view-article.html?id=sbmj.i5502>.

Data are held at the level of station: 1 row per station attempt. So, for a given application there are multiple rows of data.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 BYTE)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. For cases that relate to unsuccessful medical school applications only, the ID will be the UCAS PERSON_ID with a 'U' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
MMI_DATA_YEAR_NUMBER	NUMBER	The year of entry the applicant attending the MMI was applying to as a number.
MMI_MATCH_METHOD_UCAS	VARCHAR2(255 BYTE)	Method used to match MMI data to UCAS records
MMI_STATION_SRC	VARCHAR2(255 BYTE)	Spreadsheet the data were uploaded from.
MMI_STATION_NAME	VARCHAR2(255 BYTE)	Name of the station. For some schools this is just a number.
MMI_STATION_TITLE	VARCHAR2(250 BYTE)	Station title.
MMI_STATION_DESCRIPTION	VARCHAR2(255 BYTE)	Provided by Dundee and Edinburgh.
MMI_STATION_SCORE	NUMBER (4,1)	Score on individual station.
MMI_TOTAL_SCORE	NUMBER (4,1)	The overall score achieved.
MMI_SCORE_PERCENTAGE	NUMBER (14,10)	Total score as a percentage. The total number of marks available varied by year for some schools so the total cannot be used.
MMI_MEDICAL_SCHOOL		

COLUMN NAME	DATATYPE	COMMENTS
	VARCHAR2(25 BYTE)	Medical school where the MMI was attended. This may not be the school the applicant went to when the application cycle was complete.
MMI_COURSE	VARCHAR2(25 BYTE)	The course the MMI candidate was applying to. Uses UCAS course codes.
MMI_DATA_YEAR	VARCHAR2(255 BYTE)	The year of entry the applicant attending the MMI was applying to as a number.
MMI_QUESTION_CYCLE	VARCHAR2(40 BYTE)	Question cycle A/B/C. Used by St George's for all years.
MMI_EXCELLENT	NUMBER (1)	Impression of the candidate's suitability. Used by St George's, all years.
MMI_ACCEPTABLE	NUMBER (1)	Impression of the candidate's suitability. Used by St George's, all years.
MMI_UNACCEPTABLE	NUMBER (1)	Impression of the candidate's suitability. Used by St George's, all years.
MMI_OUTCOME	VARCHAR2(40 BYTE)	The outcome of the MMI, i.e. successful or unsuccessful.
MMI_GAMSAT_ID	VARCHAR2(25 BYTE)	Used by St. George's for Graduate Entry applicants. <i>For internal use only.</i>
MMI_RED_CARD	NUMBER (1)	Severe concerns about a candidate's suitability. Used by St George's, all years.
MMI_UKCAT_ID	VARCHAR2(40 BYTE)	11-character alphanumeric UKCAT participant identifier. <i>Internal use only for matching purposes.</i>
MMI_SURNAME	VARCHAR2(40 BYTE)	Used for matching purposes. <i>Internal use only.</i>
MMI_NAME	VARCHAR2(40 BYTE)	Used for matching purposes. <i>Internal use only.</i>
MMI_ID	VARCHAR2(250 BYTE)	Station ID.
MMI_DOB	DATE	Date of birth used for linking purposes. <i>Internal use only</i>
MMI_GENDER	CHAR (1 BYTE)	
ID_ON_SUB	VARCHAR2(150 BYTE)	ID on submission.

VW_NTS_ITEM_TRACKER

It is possible to obtain individual item data from the NTS. The items have varied slightly over the year. They are listed by year here:

<https://www.gmc-uk.org/education/how-we-quality-assure/national-training-surveys/national-training-surveys---deaneries-and-hee-local-teams/resources-and-briefings-archive>.

Generic and demographic final questionnaire Annex A for each year.

This table contains a record of the items used in the NTS with one row per item per answer per year. Use this table to specify which items are required in extracts. It can be provided on request.

COLUMN NAME	DATATYPE	COMMENTS
NTS_YEAR	NUMBER	The year of the survey the item was used,
NTS_QUESTION_REF	VARCHAR2(100 CHAR)	The database ID. Use this to specify which items are required.
NTS_QUESTION	VARCHAR2(2000 CHAR)	The item question text.
NTS_ANSWER	VARCHAR2(255 CHAR)	The item answer text.

NTS_TRAINEE_IND_SCORES_LIVE

This table contains the indicator scores derived from trainees' responses to the NTS. Details of the indicators are here: https://www.gmc-uk.org/help/What_questions_are_in_the_surveys.htm.

This table does not contain data from 2020 as the standard survey did not run due to the coronavirus (COVID-19) pandemic

COLUMN NAME	DATATYPE	COMMENTS
AGORA_ID	VARCHAR2(15 CHAR)	
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
NTS_YEAR	VARCHAR2(4 CHAR)	Year to which the indicator score relates.
INDICATOR_TYPE	VARCHAR2(250 CHAR)	Indicator name.
INDICATOR_SCORE	NUMBER	Indicator score.

NTS_2020_SURVEY_DATA

One table contains the following: census data programme and post details and so forth usually recorded in VW_NTS_TRAINEE_HISTORY and item data from the survey. Indicators were not calculated for this survey.

ORIEL_INTERVIEW_SCORES

This table contains scores from specialty recruitment assessment centres. Data are only available from 2016 onwards. Details of the scoring from the assessment centres have been collated from the Recruitment Operational Group.

COLUMN NAME	DATATYPE	COMMENTS
ORIEL_APPID	VARCHAR2(100 CHAR)	HEE system generated application ID. <i>Internal use only.</i>
PERSON_UID	VARCHAR2(7 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ORIEL_DATA_YEAR	NUMBER	GMC generated to give the year the record was received.
ORIEL_INTERVIEW_SCORE_NAME	VARCHAR2(100 CHAR)	The name of the given section of the assessment centre. Documentation for each specialty is available.
ORIEL_INTERVIEW_SCORE	NUMBER	The score achieved.

ORIEL_RECRUIT_OUTCOMES

This table contains postgraduate recruitment data from Health Education England (HEE) covering recruitment to level 1 (CT1/ ST1) specialty training programmes UK-wide. For 2015 onwards, the data include recruitment data to level (3/4) higher specialty training programmes. Data from 2015 onwards come from the ORIEL system. Earlier data were collated from each of the lead recruiting offices' systems, see http://www.gmc-uk.org/User_guide_Recruitment_outcome_reports.pdf_60052036.pdf.

COLUMN NAME	DATATYPE	COMMENTS
ORIEL_APPID	VARCHAR2(100 CHAR)	HEE system generated application ID. <i>Internal use only.</i>
PERSON_UID	VARCHAR2(7 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ORIEL_DATA_YEAR	NUMBER	GMC generated to give the year the record was received.
ORIEL_PROG_ID	VARCHAR2(50 CHAR)	Entry in posts and applicant data return should exactly match. Round ID has been dropped from the Prog ID coding. New agreed format is: 3-digit UoA code* / 3-digit specialty code* / 1 digit for level For example: EMD_CST_1
ORIEL_ROUND_ID	VARCHAR2(10 CHAR)	Should match recruitment timetable.
ORIEL_UOAPP_PREF	VARCHAR2(100 CHAR)	For national recruitment this is the candidates' preferred region, for local recruitment this would be the region to which the application has been made
ORIEL_UOAPP_PROG	VARCHAR2(100 CHAR)	The region in which the application has progressed furthest. Region – Progressed. For locally recruited specialties, this is the region and takes the same value as Region_PREF. For nationally recruited specialties, this is the UoA to which the application has furthest progressed. (For example, in a national recruitment where an applicant's preference is for Yorkshire, if an offer is not made by Yorkshire but an offer is made by North Western, Region_PREF should take the value "Yorkshire" and Region PROGRESS should take the value "North Western".) Where applications have progressed to the same extent to more than one UoA, enter the preferred of these regions. The default value is Region_PREF.
ORIEL_SPECIALTY_APPLIED	VARCHAR2(100 CHAR)	Speciality to which the application has been made. Combination of specialty and level must have a corresponding entry in the posts data set.
ORIEL_LEVEL_APPLIED	NUMBER	Level to which the application has been made

COLUMN NAME	DATATYPE	COMMENTS
ORIEL_PART_TIME_APPLIED	VARCHAR2(20 CHAR)	Recoded to consistent values RECODE PART_TIME ('0'='Not requested') ('No'='Not requested') ('Not requested'='Not requested') ('Requested'='Requested') ('Yes'='Requested') INTO PartTimeAppliedTo.
ORIEL_APPOINTABLE	NUMBER	0 Not appointable 1 Appointable For 2012 only, the following transformation was applied RECODE APPOINTABLE_HEE ('1'=1) ('yes'=1) ('Yes'=1) ('No'=0) ('no'=0) ('0'=0) ('='=0) INTO Appointable. IF (Offered = 0) Appointable =0. IF (Offered = 1) Appointable =1.
ORIEL_OFFERED	NUMBER	Recoded to 0 not offered a post 1 offered a post any type.
ORIEL_POST_TYPE_OFFERED	VARCHAR2(80 CHAR)	Recoded to consistent values STRING PostTypeOffered (A20). RECODE OFFER_MADE ('ACF'='Academic') ('CT'='Full programme') ('FTSTA'='LAT/FTSTA') ('LAT'='FTSTA/LAT') ('ST'='Full programme') INTO PostTypeOffered.
ORIEL_ACCEPTED	NUMBER	Recoded to 0 did not accept a post 1 accepted a post of any type. If an acceptance is withdrawn this is set to a 0. Set to NULL if the same person is recorded as having accepted multiple places within the same year, as this is not possible and is indicative of a failure to record that an application has been withdrawn
ORIEL_POST_TYPE_ACCEPTED	VARCHAR2(80 CHAR)	Recoded to consistent values STRING PostTypeAccepted (A20). RECODE ACCEPTED_HEE ('ACF'='Academic') ('ACL'='Academic') ('CT'='Full programme') ('FTSTA'='FTSTA/LAT') ('FTTA'='FTSTA/LAT') ('LAT'='FTSTA/LAT') ('ST'='Full programme') ('SpR'='Full programme') INTO PostTypeAccepted.
ORIEL_SHORT_LISTING_SCORE	NUMBER	Short listing score. The rules for deriving these will be specific to the specialty applied for. Some documentation from the recruitment offices is available.
ORIEL_INTERVIEW_SCORE	NUMBER	Interview score. The rules for deriving these will be specific to the specialty applied for. Some documentation from the recruitment offices is available
ORIEL_MATCH_METHOD	VARCHAR2(100 CHAR)	Where possible, HEE doctor numbers have been validated. The match method used is stored here.

COLUMN NAME	DATATYPE	COMMENTS
ORIEL_NATIONALITY	VARCHAR2(100 CHAR)	Nationality as entered on the ORIEL system.
ORIEL_ACADEMIC	VARCHAR2(50 CHAR)	Application to a nationally recruited academic training programme [DN – more to follow].
ORIEL_INTERVIEW	NUMBER	
ORIEL_NAT_RECRUIT	NUMBER	

PLAB1_DTL_LIVE

PLAB Part 1 is an assessment of medical knowledge in four domains (context, diagnosis, investigation, management). It is a single best answer examination with 200 items of which a small number are removed because of problems in keying or scoring, a typical exam having 197 scored items. The pass mark is set by a variant of the Angoff method and is typically about 125 but has varied in the range 116 to 135.

(From I. C. McManus and R. E. Wakeford (2014) Data linkage comparison of PLAB and UK graduates' performance on MRCP(UK) and MRCGP examinations: Equivalent IMG career progress requires higher PLAB pass-marks. Brit.Med.J. 348 (17th April 2014):g2621. See <https://www.bmj.com/content/348/bmj.g2621>)

Since June 2018 the Part 1 exam has been made up of 180 multiple choice single best answer questions. The pass mark now also includes 1 Standard Error of Measurement (SEM) and is typically about 114-115. On occasions a small number of questions are removed from papers because of problems in keying or scoring.

The addition of a single SEM to the PLAB part 1 pass mark had been recommended by the PLAB Review Panel as "a reasonable compromise between the risk of false positives for patients and the consequences of failure for the slightly raised number of false positives"

The exam covers the common, important or acute conditions (those common in emergency departments) seen by trainees entering the second year of the Foundation Programme (F2), and the management of long-term conditions seen in primary care. It excludes the advanced duties of a general practitioner (GP).

See - <https://www.gmc-uk.org/registration-and-licensing/join-the-register/plab/plab-1-guide> for information on the current version of PLAB1.

Data from 30/05/2001 onwards are included.

COLUMN NAME	DATATYPE	COMMENTS
TEST_ID	NUMBER	ID for test
TEST_DATE	DATE	Date of test
PERSON_UID	VARCHAR2(10 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
QUESTION_ID	VARCHAR2(15 CHAR)	ID for question
QUESTION_TITLE	VARCHAR2(100 CHAR)	Title for question
SKILL	VARCHAR2(150 CHAR)	Question skill area used for sampling
		See the test blueprint for a list of topics. https://www.gmc-

COLUMN NAME	DATATYPE	COMMENTS
TOPIC	VARCHAR2(100 CHAR)	uk.org/-/media/documents/plab-blueprint_pdf-65021787.pdf . Question topic used for sampling
TOPIC2	VARCHAR2(100 CHAR)	See the test blueprint for a list of topics. https://www.gmc-uk.org/-/media/documents/plab-blueprint_pdf-65021787.pdf . Second question topic used for sampling if applicable.
DOMAIN	VARCHAR2(250 CHAR)	Mapping to 3 of the 4 domains of <i>Good Medical Practice</i> : https://www.gmc-uk.org/-/media/documents/good-medical-practice---english-1215_pdf-51527435.pdf Domain 1: Knowledge, skills and performance Domain 2: Safety and quality 10 Domain 3: Communication, partnership and teamwork 13 Only available from 04/09/2012 to 11/09/2014.
ANSWER	VARCHAR2(1 CHAR)	Candidate's answer
RESULT	NUMBER	0 = Incorrect 1 = Correct
ORDER_BY	NUMBER	Question order
SOURCE	VARCHAR2(30 CHAR)	Data have been collated from different source in the GMC depending on the system used at the time PLAB1 exam was originally taken.

PLAB2_DTL_LIVE

PLAB Part 2 is an objective structured clinical exam (OSCE). Before June 2016 candidates were assessed on 15 stations one of which is a non-scoring pilot station. From June 2016 this was increased to 18 stations, all are scored. Data are on row per station.

See <https://www.gmc-uk.org/registration-and-licensing/join-the-register/plab/plab-2-guide/what-is-the-plab-2-exam> for information on the current version of PLAB2.

Data from 13/06/2001 are available.

COLUMN NAME	DATATYPE	COMMENTS
TEST_ID	VARCHAR2(15 CHAR)	Test ID
TEST_DATE	DATE	Date of test
PERSON_UID	VARCHAR2(10 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
EXAMINER_UID	VARCHAR2(15 CHAR)	GMC doctor identifier. Known externally as GMCRefNo. Replaced by extract-specific ID in de-identified research extracts.
MAX_POSS_SCORE	NUMBER	From 07/09/2016 - maximum score across all stations.
STATION_ID	VARCHAR2(15 CHAR)	ID for station.
STATION_V_NUM	VARCHAR2(15 CHAR)	Station version number
STATION_NAME	VARCHAR2(100 CHAR)	Station name.
SKILL	VARCHAR2(150 CHAR)	Skill assessed.
STATION_ACT_SCORE	NUMBER	Candidate's score on the station. From 07/09/2016 this is the sum of the domain scores. Prior to September 2016 this is the total score for the objectives. For each objective, the examiner gave the candidate a grade from A to E (as now). This was then turned into a number (0-4) and multiplied by the objective weighting. Each objective score was added together and divided by 100 to give a score between 0.00 and 5.00.
STATION_REQ_SCORE	NUMBER	Score required to pass the station.
STATION_GRADE	NUMBER	STATION_ACT_SCORE banded for cases prior to September

COLUMN NAME	DATATYPE	COMMENTS
		2016. Bands are as follows 0 0.9 to 1.5 1 1.55 to 2.5 2 2.55 to 3.5 3 3.55 to 4.5 4 4.55 to 5
EXAMINER_OVERALL_JUDGEMENT	VARCHAR2(1 CHAR)	Does not count to individual result. Applies to the borderline standard set yearly. In use from 08/01/2007 to 14/07/2016.
STATION_RESULT	VARCHAR2(4 CHAR)	Pass or Fail the station
OBJECTIVE1	NUMBER	0 - 4. (Optional – stations do not have all 7 objective questions) Used from 13/06/2001 to 14/07/2016
OBJECTIVE2	NUMBER	0 - 4. (Optional – stations do not have all 7 objective questions) Used from 13/06/2001 to 14/07/2016
OBJECTIVE3	NUMBER	0 - 4. (Optional – stations do not have all 7 objective questions) Used from 13/06/2001 to 14/07/2016
OBJECTIVE4	NUMBER	0 - 4. (Optional – stations do not have all 7 objective questions) Used from 13/06/2001 to 14/07/2016
OBJECTIVE5	NUMBER	0 - 4. (Optional – stations do not have all 7 objective questions) Used from 13/06/2001 to 14/07/2016
OBJECTIVE6	NUMBER	0 - 4. (Optional – stations do not have all 7 objective questions) Used from 13/06/2001 to 14/07/2016
OBJECTIVE7	NUMBER	0 - 4. (Optional – stations do not have all 7 objective questions) Used from 13/06/2001 to 14/07/2016
DOMAIN_SCORE_1	NUMBER	In use from 07/09/2016. Data gathering, technical and assessment skills
DOMAIN_SCORE_2	NUMBER	In use from 07/09/2016. Clinical management skills
DOMAIN_SCORE_3	NUMBER	In use from 07/09/2016. Interpersonal skills.

COLUMN NAME	DATATYPE	COMMENTS
PILOT_STATION	VARCHAR2(15 CHAR)	Not always available
NUM_OF_STN_PASSES	NUMBER	Number of stations passed
ORDER_BY	NUMBER	Station order
SOURCE	VARCHAR2(30 CHAR)	Data have been collated from different source in the GMC depending on the system used at the time PLAB2 exam was originally taken.

PLAB_TOTAL_LIVE

This table contains the overall scores for all candidates' PLAB attempts at PLAB 1 (from 20/10/1997 onwards) and PLAB 2 (from 16/04/1998 onwards). These data come from the main GMC database (Siebel), more detailed data in PLAB1_DTL_LIVE and PLAB2_DTL_LIVE come from the examination specific databases which have changed over time.

COLUMN NAME	DATATYPE	COMMENTS
AGORA_ID	VARCHAR2(15 CHAR)	GMC database ID. <i>Internal use only.</i>
TEST_PART	VARCHAR2(100 CHAR)	Indicates whether the record relates to PLAB Part 1 or PLAB Part 2.
TEST_ID	VARCHAR2(30 CHAR)	.
TEST_DATE	DATE	Date of test
PERSON_UID	VARCHAR2(10 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ATTEMPT_NUM	NUMBER	The candidate's attempt number on that part of the test
TEST_SCORE	NUMBER	The candidate's overall test score for that attempt
TEST_PASS_MARK	NUMBER	The overall pass mark for that test
TEST_RESULT	VARCHAR2(4 CHAR)	Indicates whether a candidate passed or failed

RECRUIT_MSRA

This table contains applicants' scores on the Multi-Specialty Recruitment Assessment (MSRA) described in more detail here: <https://gprecruitment.hee.nhs.uk/Recruitment/Applicant-Guidance/MSRA>.

Scores are available for the following specialties and application years:

- General practice 2015 onwards
- Ophthalmology' 2015 onwards
- Core Psychiatry Training 2015 onwards
- Clinical radiology 2016 onwards
- Neurosurgery 2016 onwards
- Obstetrics and gynaecology 2016 onwards
- Community Sexual and Reproductive Health 2016 onwards

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(10 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ORIELPIN	NUMBER (10)	Internal use only Oriel system unique identifier used for linking to the oriel_appid in ORIEL_RECRUIT_OUTCOMES.
CANDIDATEID	NUMBER (11)	Pearson VUE Id from test sitting.
MSRA_YEAR	NUMBER (4)	Recruitment year MSRA scores for
MSRA_SRC	VARCHAR2(26 BYTE)	The source corresponds to the file name submitted to the GMC by HEE.
CLIENTCANDIDATEID	NUMBER (12)	Concatenation of Oriel Pin and DATE OF intake applied for in the format MMY. For 2015 these are the same as the ORIELPIN.
REGISTRATIONID	NUMBER (11)	Pearson VUE ID from test sitting.
GPNRO	VARCHAR2(26 BYTE)	Applied for specialty training in General Practice.
BBTNRO	VARCHAR2(26 BYTE)	Applied for specialty training in Broad Based Training.
PSYNRO	VARCHAR2(26 BYTE)	Applied for specialty training in Psychiatry
OPHTHAL	VARCHAR2(26 BYTE)	Applied for specialty training in Ophthalmology
CLINICALRADIOLOGY	VARCHAR2(26 BYTE)	Applied for specialty training in Clinical radiology

COLUMN NAME	DATATYPE	COMMENTS
NUCLEARMED	VARCHAR2(26 BYTE)	Applied for specialty training in Nuclear Medicine.
OPHTHALMOLOGY	VARCHAR2(26 BYTE)	Applied for specialty training in Ophthalmology
PSYCHIATRY	VARCHAR2(26 BYTE)	Applied for specialty training in Psychiatry
OANDG	VARCHAR2(26 BYTE)	Applied for specialty training in Obstetrics and Gynaecology
NEURO	VARCHAR2(26 BYTE)	Applied for specialty training in Neurosurgery
BBT	VARCHAR2(26 BYTE)	Applied for specialty training in Broad Based Training
BBTWALES	VARCHAR2(26 BYTE)	Applied for specialty training in Broad Based Training in Wales
CSRH	VARCHAR2(26 BYTE)	Applied for specialty training in Community Sexual and Reproductive Health.
BYPASS575	VARCHAR2(26 BYTE)	Same as direct pathway.
DIRECTPATHWAYNUMBER	NUMBER (3)	0 or 1 – depending on if the candidate was eligible for direct pathway (0 = no, 1 = yes). n/a values from DIRECTPATHWAY are classified here.
DIRECTPATHWAY	VARCHAR2(26 BYTE)	Yes/No did the candidate receive a direct pathway for their application
CPSRAWSCORE	NUMBER (4)	Raw score for the Clinical Problem-Solving test.
CPSRAWEQUATED	NUMBER (6,2)	Different versions of the test within a year are equated using common anchor items (up to a third of the questions may be the same in two versions to enable this equating).
CPS40	NUMBER (5)	Scores within a year are converted to scale with a mean of 250 and SD of 40, meaning that the scales are norm-referenced, and cannot be compared across years.
SJTRAWSCORE	NUMBER (7,2)	Raw score for the Professional Dilemmas test.
SJTRAWQUATED	NUMBER (5)	Different versions of the test within a year are equated using common anchor items (up to a third of the questions may be the same in two versions to enable this equating).
SJT40	NUMBER (3)	Scores within year are converted to scale with a mean of 250 and SD of 40, meaning that the scales are norm-referenced, and cannot be compared across years.
CPSBAND181	NUMBER (3)	1 = Very poor level of performance 2 = Below average

COLUMN NAME	DATATYPE	COMMENTS
		performance3 = Good level of performance 4 = Very good level of performance.
SJTBAND181	NUMBER (3)	1 = Very poor level of performance 2 = Below average performance 3 = Good level of performance 4 = Very good level of performance.
CPSBAND191	NUMBER (3)	1 = Very poor level of performance 2 = Below average performance 3 = Good level of performance 4 = Very good level of performance.
SJTBAND191	NUMBER (3)	1 = Very poor level of performance 2 = Below average performance 3 = Good level of performance 4 = Very good level of performance.
OUTCOME	NUMBER (5)	Overall score
OVERALL	NUMBER (5)	Overall score
OVERALL_SCALED	NUMBER (5,1)	
OVERALLSCALEDROUNDED	NUMBER (4)	
OUTCOME181	NUMBER (5)	
OUTCOME191	NUMBER (5)	
TGPT	VARCHAR2(3 BYTE)	
CAMHS	VARCHAR2(3 BYTE)	
NEUROSURGERYST1	VARCHAR2(3 BYTE)	Applied for specialty training in Neurosurgery ST1
NEUROSURGERYST3	VARCHAR2(3 BYTE)	Applied for specialty training in Neurosurgery ST3
CPSRESCALED_221	NUMBER (6,2)	
SJTRESCALED_221	NUMBER (6,2)	
STAGE2RESCALED	NUMBER (2)	

ROD_01

These data were supplied by Durham Medical School under the researcher own data process - https://www.ukmed.ac.uk/documents/UKMED_research_data.pdf . It will be available to other researchers once Durham Medical School have completed their project.

The table contains Conscientiousness Index data gathered on first and second year MBBS students at Durham University. The data comprise an Index score for year 1 and year 2 for students starting in Durham from 2006 to 2014 (N = 859).

McLachlan JC1, Finn G, Macnaughton J *Academic Medicine*. 2009 May;84(5):559-65. doi: 10.1097/ACM.0b013e31819fb7ff. The conscientiousness index: a novel tool to explore students' professionalism.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(10 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
R0D01_CI_R1Y1	NUMBER (25,20)	Conscientiousness Index Year 1.
R0D01_CI_R1Y2	NUMBER (25,20)	Conscientiousness Index Year 2.
R0D01_CI_AVG	NUMBER (25,20)	Conscientiousness Index average mean over the two years.
R0D01_COHORT	NUMBER (4)	Year the student started at Durham medical school.
R0D01_MATCH_TEST	VARCHAR2(100 CHAR)	Match test used to link data supplied by Durham Medical School to HESA data.

STUDENT_FTP

These data are collected in the GMC's Combined Return. The first combined return covered all students in medical school during the academic year 2017/2018. We now hold data for 2018/19 and 2019/20.

For students in their final year (i.e. applying to join the register for an August 2018 start as an F1 doctor) the medical school completed the final year return; for students in earlier years of their studies the medical school completed the non-final year worksheet.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(40 BYTE)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
SFTP_MATCHCHECK	VARCHAR2(100 BYTE)	
SFTP_SRC	VARCHAR2(26 BYTE)	Whether the case is final year or non-final year - the source worksheet of the combined return. Combined with the year of the return
SFTP_MEDICAL_SCHOOL	VARCHAR2(52 BYTE)	Medical school from which the submission was received.
SFTP_YEAR_OF_STUDY	VARCHAR2(26 BYTE)	Non-final year only – Year of study in which the proceedings occurred.
SFTP_HUSID_ORG	VARCHAR2(32 BYTE)	HUSID as per the school's original submission.
SFTP_STAGE_OF_PROCESS	VARCHAR2(160 BYTE)	Stage of the student fitness to practise process reached.
SFTP_ENTRY_METHOD	VARCHAR2(50 BYTE)	Non-final year only – Whether the student in question was a standard entry, graduate entry or unknown entry type.
SFTP_NATURE_OF_CONCERN	VARCHAR2(52 BYTE)	The nature of the student fitness to practise concern.
SFTP_ACADEMIC_YEAR	VARCHAR2(104 BYTE)	<i>Non-final year only</i> – The academic year in which the process occurred.
SFTP_DETAILS_OF_ISSUE	VARCHAR2(2050 BYTE)	<i>Final year only</i> – The details of the issue that warranted student fitness to practise proceedings

COLUMN NAME	DATATYPE	COMMENTS
SFTP_DATE_OCCURRED	VARCHAR2(850 BYTE)	<i>Final year only –</i> Date of the proceedings.
SFTP_FORMAL_PROCEEDINGS	VARCHAR2(64 BYTE)	Did the incident result in formal disciplinary proceedings? (Y/N)
SFTP_NATURE_OF_PROCEEDINGS	VARCHAR2(1100 BYTE)	The nature of the student fitness to practise proceedings (i.e. a type of committee, panel, meeting or board).
SFTP_OUTCOME_CATEGORY	VARCHAR2(104 BYTE)	Which category was the outcome/sanction of the proceedings under?
SFTP_CASE_ONGOING	VARCHAR2(26 BYTE)	Is the case still ongoing? (Y/N)
SFTP_APPEALED	VARCHAR2(52 BYTE)	Has the outcome been appealed to the university of OIA (office of the independent adjudicator)? (Y/N)
SFTP_COMPLIED_WITH	VARCHAR2(26 BYTE)	Was the CHECK complied with? (Y/N)
SFTP_ADDITIONAL_COMMENTS	VARCHAR2(1000 BYTE)	<i>Non-final year only –</i> Free text of any additional comments made.
SFTP_DETAILS_OF_OUTCOME	VARCHAR2(2050 BYTE)	Text provided by the university detailing the outcome of the SFTP event
SFTP_COMPLIED_WITH_FREETEXT	VARCHAR2(400 BYTE)	Text provided by the university detailing whether the student complied with the decision/conditions set by the SFTP panel
SFTP_DATE_OCCURRED_DF	DATE	Date occurred. Cleaned by the GMC to allow formatting as a date
SFTP_ID	VARCHAR2(20 BYTE)	GMC assigned ID.
SFTP_COMPLIED_WITH_SUMMARY	VARCHAR2(25 BYTE)	GMC recoded and cleaned from SFTP_COMPLIED_WITH_FREETEXT and SFTP_COMPLIED_WITH

UCAS_APPLICATIONS

This table contains one row per application to a medical school from any applicant who applied to a degree leading to a primary medical qualification from 2007 to 2018. In a given application cycle there are typically four applications to medical schools. Data are supplied by The Universities and Colleges Admissions Service (UCAS).

Age on application is calculated in different ways as per the table below to reflect the UCAS methodology.

“This statistical release uses country-specific age definitions which align with the cut off points for school/college cohorts within the different administrations of the UK. For England and Wales ages are defined on the 31 August, for Northern Ireland on the 1 July and for Scotland on the 28 February the following year. Defining ages in this way matches the assignment of children to school cohorts.”

For applicants outside of the UK the cohort cut off for England and Wales has been used.

<https://www.ucas.com/data-and-analysis/undergraduate-statistics-and-reports/ucas-undergraduate-releases/applicant-releases-2019-cycle/2019-cycle-applicant-figures-january-deadline>

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 BYTE)	<p>GMC unique doctor reference number for those that registered.</p> <p>For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix.</p> <p>For cases that relate to unsuccessful medical school applications only, the ID will be the UCAS PERSON_ID with a 'U' prefix.</p> <p>Replaced by extract specific STUDY_ID in de-identified research extracts.</p>
UCAS_PERSON_ID	VARCHAR2	UCAS person ID. <i>Internal use only</i>
UCAS_APPNO	VARCHAR2	Applicant identifier that is unique to each applicant in each cycle. <i>Internal use only</i>
UCAS_APP_IDENTIFIER	VARCHAR2	Applicant identifier that is unique to each applicant in each cycle. <i>Internal use only</i>
UCAS_UCAS_YEAR	VARCHAR2	The UCAS cycle in which the first application made by the applicant was processed.
UCAS_ENTRYYEAR	NUMBER	The academic year in which the course selected by the applicant starts. For courses starting between August and December the academic year is defined as starting in that year (I.e. for courses starting in August 2013 the entry year will be 2013.) For any courses starting earlier than September the academic year is defined as starting in the previous year. (I.e. for courses starting in January 2013 the entry year will be 2012 despite the course starting during the 2013 application cycle.) Please note: In each application cycle there are a small number of applications that are deferred for two years.
UCAS_JACS3_SUBGROUP	VARCHAR2	Single character code that classifies courses into a summary level of 26 subjects. Each course is assigned up to three valid JACS3 subject codes (e.g. G100 – Mathematics) and a course balance indicator by UCAS, which are available for review by the provider. The course is assigned a subject based on these JACS3 subject codes and balance indicator, it largely correlates to the first two characters of the subject codes.

COLUMN NAME	DATATYPE	COMMENTS
		Where there are more than one JACS3 subject codes for a given course, and the balance indicator is dual or triple, the first two characters of each subject code are reported in combination to a relevant category (e.g. Course with dual balance indicator with JACS3 subject codes B940 = Counselling and C843= Counselling psychology is assigned a subject 'Y Combined sciences'). Please note: Between 2007 and 2011 the subject codes assigned to each course were from the JACS2.0 classification, therefore JACS3 versions of the subject categories have been approximated for 2007 to 2011 to allow a consistent time series.
UCAS_JACS3_SUBLINE	VARCHAR2	Two-character code that classifies courses into a detailed level of 215 subjects. Each course is assigned up to three valid JACS3 subject codes (e.g. G100 – Mathematics) and a course balance indicator by UCAS, which are available for review by the provider. The course is assigned a subject based on these JACS3 subject codes and balance indicator, it largely correlates to the first two characters of the subject codes. Where there are more than one JACS3 subject codes for a given course, and the balance indicator is dual or triple, the first two characters of each subject code are reported in combination to a relevant category (e.g. Course with dual balance indicator with JACS3 subject codes L370 = Social Theory and M900 = Others in Law is assigned a subject 'Y Combs of soc studies/law'). Please note: Between 2007 and 2011 the subject codes assigned to each course were from the JACS2.0 classification, therefore JACS3 versions of the subject categories have been approximated for 2007 to 2011 to allow a consistent time series.
UCAS_SUB_GROUP_LABEL	VARCHAR2	Label for UCAS_JACS3_SUBLINE
UCAS_COURSE_ID	VARCHAR2	A combination of provider and course code. Please note: the course code is assigned to each course by the host provider and does not necessarily relate to JACS subject codes.
UCAS_INITIAL_COURSE_ID	VARCHAR2	A combination of provider and course code. Please note: the course code is assigned to each course by the host provider and does not necessarily relate to JACS subject codes. Some providers may offer a different course post application resulting in a different identifier.
UCAS_PROVIDER_CODE	VARCHAR2	Three-character code unique to each Higher Education Provider. This is a mapped view of the Higher Education Provider as-at 2018. In the interest of keeping a consistent time series the HEP displayed is mapped from a previous HEP if a merger has occurred or the HEP has been renamed. E.g. In 2013 HEPs K05 and H50 merged, so all instances of H50 prior to 2013 will be reported as K05.
UCAS_UCAS_HEE_PROVIDER_LABEL	VARCHAR2	Label for the provider code. 2018 labels have been used.
UCAS_COURSE_NAME	VARCHAR2	Provider's name for the course
UCAS_COURSE_TYPE	VARCHAR2	Type of course obtained by mapping course_ids to MSC categories. Possible values are: Graduate Entry Programme International Graduate Entry International Transfer Medicine Conversion Entry Programme Medicine with a Gateway Year Medicine with a Preliminary Year Standard Entry Medicine Standard Entry Medicine - direct entry to year 3 See https://www.medschools.ac.uk/studying-medicine/course-types
UCAS_PROVIDER_TARIFF_GRP	VARCHAR2	The grouping of providers based on the average levels of attainment of their accepted applicants (summarised through UCAS Tariff points) in a period of application cycles spanning from 2004 to 2011. Each group of providers accounted for around a third of all UK 18-year-old acceptances in these cycles. Split by the following values: 'Higher tariff provider', 'Medium tariff provider', 'Lower tariff provider'.
UCAS_PROVIDER_REGION	VARCHAR2	A mapped view of the UK region in which the provider is situated as-at 2018. Split by the following values: 'North East',

COLUMN NAME	DATATYPE	COMMENTS
		'Yorkshire and The Humber', 'North West', 'East Midlands', 'West Midlands', 'East of England', 'London', 'South East', 'South West', 'Wales', 'Northern Ireland', 'Scotland', 'Overseas'. In the interest of keeping a consistent time series the provider is mapped from a previous provider if a merger has occurred. E.g. In 2013 providers K05 and H50 merged, so the provider region of H50 will be reported as the provider region of K05.
UCAS_ROUTE_NAME	VARCHAR2	The application route with the following values attached to applications. Main scheme, Adjustment, Clearing, Extra, RPA.
UCAS_ACCEPT	NUMBER	Flag to indicate whether applicant was placed through application.
UCAS_PROVIDER_DEC_JUN30	VARCHAR2	The provider decision at the June 30 deadline with the following values attached to applications: Conditional offer, Invited for interview, Decision pending, Unconditional offer, Unconditional course change. Where no decision has been made for an application, the decision will be reported as missing.
UCAS_APP_REPLY_JUN30	VARCHAR2	The applicant response at the June 30 deadline with the following values attached to applications: Firm, Insurance, Declined by default, Declined, Cancelled. Where no reply has been made for an application, the applicant reply will be reported as missing.
UCAS_COND_UNCOND_OFFER	VARCHAR2	Flag to indicate whether an offer was a conditional unconditional offer (June 30 deadline).
UCAS_UCAS_UCOND_ANAL_FLAG	NUMBER	Flag to indicate whether this choice was included in UCAS' 2018 End of Cycle unconditional offer making analysis.
UCAS_JUNE_FLAG	NUMBER	Flag to indicate whether application was present at June 30 deadline
UCAS_EOC_FLAG	NUMBER	Flag to indicate whether application was present at End of Cycle
UCAS_APP_DATE	DATE	The date at which the applicant first applied in the application cycle.
UCAS_ACCEPTANCE_ROUTE	VARCHAR2	The acceptance route with the following values attached to placed applications. Firm Choice: where the applicant has been accepted to their first choice. Insurance choice: where the applicant has been accepted to their second choice. Main Scheme Clearing: where an applicant was unsuccessful in the main scheme (i.e. applied before 30 June) and subsequently found a place using Clearing. Direct Clearing: where the applicant has applied via Clearing without an initial application through the main scheme. Adjustment: where applicants who have met and exceeded the conditions of their firm choice choose to take up an alternative offer - introduced in 2009. Extra: where applicants who held no offers after using all of their main scheme choices, make additional choices. RPA: where an application is submitted to UCAS by an institution when an unconditional firm has already been offered and accepted by the applicant. Please note: 'Insurance choice' and 'Firm choice' values are based on the applicant's response to an offer as-at June deadline. There are 10,000 to 20,000 acceptances to a main scheme choice each year where the applicant has not responded, or is awaiting an offer, by June deadline. These acceptances are classified as 'Other Main Scheme Choice'.
UCAS_HESA_LINK_TEST	VARCHAR2(150 CHAR)	Test comparing UCAS to HESA values – school and entry year
UCAS_AGE_ON_START_NI	NUMBER	Age calculated on 1 July UCAS year (i.e. the year they applied to start – ignoring deferrals)
UCAS_AGE_ON_START_SCOT	NUMBER	Age calculated on 1 Feb the year after UCAS year.
UCAS_AGE_ON_START_EW	NUMBER	Age calculated on the 31 August UCAS year (i.e. the year they applied to start – ignoring deferrals)

COLUMN NAME	DATATYPE	COMMENTS
UCAS_AGE_ON_START_ALL_TO _EW	NUMBER	England and Wales date applied to all case regardless of domicile.
UCAS_MEDICAL_SCHOOL	VARCHAR2(50 BYTE)	UCAS provider coded to standard UKMED Medical school name.

APPLICANT_MEDICAL_SCHOOL_DISTANCES

This table contains the distances from medical school applicants' home addresses to each school they applied to, calculated by providing postcodes to the Google Driving distance API - <https://developers.google.com/maps/documentation/distance-matrix/overview>. This has only been calculated where the applicant's domicile was UK. A small number of postcodes did not return a distance.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 BYTE)	<p>GMC unique doctor reference number for those that registered.</p> <p>For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix.</p> <p>For cases that relate to unsuccessful medical school applications only, the ID will be the UCAS PERSON_ID with a 'U' prefix.</p> <p>Replaced by extract specific STUDY_ID in de-identified research extracts.</p>
UCAS_PERSON_ID	VARCHAR2	UCAS person ID. <i>Internal use only</i>
UCAS_APPNO	VARCHAR2	Applicant identifier that is unique to each applicant in each cycle. <i>Internal use only</i>
UCAS_APP_IDENTIFIER	VARCHAR2	Applicant identifier that is unique to each applicant in each cycle. <i>Internal use only</i>
UCAS_YEAR	VARCHAR2	The UCAS cycle in which the first application made by the applicant was processed.
UCAS_ENTRYYEAR	NUMBER	The academic year in which the course selected by the applicant starts. For courses starting between August and December the academic year is defined as starting in that year (I.e. for courses starting in August 2013 the entry year will be 2013.) For any courses starting earlier than September the academic year is defined as starting in the previous year. (I.e. for courses starting in January 2013 the entry year will be 2012 despite the course starting during the 2013 application cycle.) Please note: In each application cycle there are a small number of applications that are deferred for two years.
UCAS_PROVIDER_CODE	VARCHAR2	Three-character code unique to each Higher Education Provider. This is a mapped view of the Higher Education Provider as-at 2018. In the interest of keeping a consistent time series the HEP displayed is mapped from a previous HEP if a merger has occurred or the HEP has been renamed. E.g. In 2013 HEPs K05 and H50 merged, so all instances of H50 prior to 2013 will be reported as K05.
UCAS_HEE_PROVIDER_LABEL	VARCHAR2	Label for the provider code. 2018 labels have been used.
UCAS_COURSE_NAME	VARCHAR2	Provider's name for the course
UCAS_CORR_POSTCODE	VARCHAR2	Applicant's postcode as per UCAS. <i>Internal use only</i>
MEDICAL_SCHOOL_POSTCODE	VARCHAR2	Medical School postcode
GOOGLE_API_DISTANCE_METRES	NUMBER	Driving distance in metres
GOOGLE_API_DISTANCE_MILES	NUMBER	Driving distance in miles.

UCAS_QUALS_VER

A qualification level table containing all qualifications from applicants to the UCAS undergraduate scheme who are present in the UCAS extract. These qualifications were verified to UCAS by the Awarding Body Linkage.

COLUMN NAME	DATATYPE	COMMENTS
UCAS_APPNO	VARCHAR2	Applicant identifier that is unique to each applicant in each cycle. <i>Internal use only</i>
UCAS_YEAR	VARCHAR2	Cycle in which the applicant made their application.
UCAS_YEAR_OF_QUAL	VARCHAR2	Year in which the applicant took the qualification.
UCAS_QUAL_DESCRIP	VARCHAR2	Description of the type of qualification. For example 'GCE Advanced Level'
UCAS_SUBJECT_TITLE	VARCHAR2	Subject of the qualification.
UCAS_GRADE	VARCHAR2	Grade achieved by the applicant.
UCAS_SUBJECT_GROUP	VARCHAR2	Subject grouping as used in the UKCAT-12 study
UCAS_QUAL_TYPE	VARCHAR2	Mapped to HESA qualification types
UCAS_QUAL_GRADE	VARCHAR2	Mapped to HESA grades
UCAS_POINTS_UKCAT12	NUMBER	Points as per UKCAT-12 study with extension for A*
UCAS_TARIFF_POINTS	NUMBER	UCAS Tariff points as per
UCAS_EOC_FLAG	NUMBER	Flag to indicate whether application was present at End of Cycle
UCAS_APP_DATE	DATE	The date at which the applicant first applied in the application cycle.

UCAS_QUALS_DEC

A qualification level dataset containing all qualifications from applicants to the UCAS undergraduate scheme who are present in UCAS extract. These qualifications were declared by the applicant during their application. It includes predicted and achieved qualifications. This table includes all qualifications declared and so include GCSEs.

COLUMN NAME	DATATYPE	COMMENTS
UCAS_APPNO	VARCHAR2	Applicant identifier that is unique to each applicant in each cycle. <i>Internal use only</i>
UCAS_YEAR	VARCHAR2	Cycle in which the applicant made their application.
UCAS_YEAR_OF_QUAL	VARCHAR2	Year in which the applicant took the qualification.
UCAS_QUAL_STATUS	VARCHAR2	Pending or gained
UCAS_QUAL_DESCRIP	VARCHAR2	Description of the type of qualification. For example 'GCE Advanced Level'
UCAS_SUBJECT_TITLE	VARCHAR2	Subject of the qualification.
UCAS_GRADE	VARCHAR2	Grade achieved by the applicant.
UCAS_LEVEL	VARCHAR2	Applicant identifier that is unique to each applicant in each cycle.
UCAS_SUBJECT_GROUP	VARCHAR2	Subject grouping as used in the UKCAT-12 study
UCAS_QUAL_TYPE	VARCHAR2	Mapped to HESA qualification types
UCAS_QUAL_GRADE	VARCHAR2	Mapped to HESA grades
UCAS_POINTS_UKCAT12	NUMBER	Points as per UKCAT-12 study with extension for A*
UCAS_TARIFF_POINTS	NUMBER	Points as per https://www.hesa.ac.uk/collection/c17051/derived/xtpoints

VW_UKCAT_RESULTS_APP

UKMED hold UKCAT data for all cases for whom we could obtain a match to medical school applicants as listed in the UCAS data or a match to medical school entrants as listed in the HESA data.

Prior to 2019 the UCAT (University Clinical Aptitude Test) was known as the UKCAT. Whilst the name has changed, the test content remains the same. As these tables contain historic data we are still using the acronym UKCAT. Please see <https://www.ucat.ac.uk/ucat/>

The schools and courses using UKCAT vary overtime. An applicant would have taken UKCAT if they applied to at least one school requiring UKCAT. UKMED holds test results for anyone who took UKCAT. The student may have gained admission to a medical school that did not require it.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(64 BYTE)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. For cases that relate to unsuccessful medical school applications only, the ID will be the UCAS PERSON_ID with a 'U' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
UKCAT_TEST	VARCHAR2(20 BYTE)	Name of test variation registered for – indicates if the candidate sat the UKCATSEN - https://www.ucat.ac.uk/ucatsen-evidence
UKCAT_TESTREG_ID	VARCHAR2(50 BYTE)	<i>Internal use only</i> ukcatRegistration.id. Nine-digit test registration identifier. For linking to the UKCAT results table
UKCAT_DATETEST	DATE	Date test was taken.
UKCAT_YEARTEST	NUMBER(4)	Year test was taken.
UKCAT_ATTENDED	VARCHAR2(10 BYTE)	Boolean test attendance indicator
UKCAT_MATCH_SCRIPT	VARCHAR2(100 CHAR)	Match script used to join UKCAT record to HESA.
UKCAT_ATTEMPTNUMBER	NUMBER (4)	The number of attempts including the attempt sat on the date of this test.
UKCAT_SJT_BAND	NUMBER (2)	SJT scores are placed in 4 bands - https://www.ucat.ac.uk/ucat/results/scoring-and-marking/
UKCAT_YEAR_TESTING	NUMBER (4)	Year test was taken.

COLUMN NAME	DATATYPE	COMMENTS
UKCAT_PERCENTILE	NUMBER (4)	Available from 2016 onwards – overall score expressed as a percentile of all test takes in that year
UKCAT_ABSTRACT_REASONING	NUMBER (4)	https://www.ucat.ac.uk/ucat/test-format/abstract-reasoning/ . Score between 300 and 900 see https://www.ucat.ac.uk/ucat/results/scoring-and-marking/
UKCAT_DECISION_ANALYSIS	NUMBER (4)	https://www.ucat.ac.uk/ucat/test-format/decision-making/ Score between 300 and 900 see https://www.ucat.ac.uk/ucat/results/scoring-and-marking/
UKCAT_QUANTITATIVE_REASONING	NUMBER (4)	https://www.ucat.ac.uk/ucat/test-format/quantitative-reasoning/ Score between 300 and 900 see https://www.ucat.ac.uk/ucat/results/scoring-and-marking/
UKCAT_VERBAL_REASONING	NUMBER (4)	https://www.ucat.ac.uk/ucat/test-format/verbal-reasoning/ Score between 300 and 900 see https://www.ucat.ac.uk/ucat/results/scoring-and-marking/
UKCAT_SJT	NUMBER (4)	<p>The UKCAT Situational Judgement Test (SJT) is designed for a 'novice population'. This means that applicants do not require any experience in a healthcare context in order to score well. The scenarios are set in a healthcare or educational setting. However, there is no requirement for specific clinical knowledge or 'procedural' job knowledge. This is because the test is targeting professional attributes, namely Integrity, Perspective Taking (capacity to consider the perspective of others) and Team Involvement.</p> <p><i>Equating of test versions</i> For the purposes of fairness, test equating is used to align scores on the three test versions (formerly six) so that minor differences in their overall difficulty are controlled for statistically. Further information on test equating can be sought from the UKCAT office.</p> <p><i>Scaling and banding of equated scores</i> The 'raw' equated SJT scores are transformed onto a scale which preserves the original anticipated distribution and rescales it to a mean of 600 and a standard deviation of 70. This is a linear transformation, so it has no impact on applicants' scores relative to each other. The scaled scores are then banded into four bands, based on the anticipated distribution. Applicants and admissions departments receive the applicant's score band, i.e. the band into which the applicant's score falls, rather than the exact score. This is appropriate given that the UKCAT SJT is still relatively young, and feedback from medical/dental schools.</p>
UKCAT_COG_TOTAL	NUMBER (4)	A total scale score is generated by summing the individual scale scores of Verbal Reasoning, Decision Making, Quantitative Reasoning and Abstract Reasoning.
UKCAT_TOTAL	NUMBER (4)	A total scale score is generated by summing the individual scale scores of Verbal Reasoning, Decision Making, Quantitative Reasoning, Abstract Reasoning and SJT
UKCAT_SRC	VARCHAR2(50 BYTE)	Whether the cases came from a match to UCAS or HESA.
UKCAT_ABSTRACT_REASONING_Z	NUMBER (5,3)	Score converted into a Z-score using the population statistics from the UCAT technical report for the test year - https://www.ucat.ac.uk/research/technical-reports/
UKCAT_DECISION_ANALYSIS_Z	NUMBER (5,3)	Score converted into a Z-score using the population statistics from the UCAT technical report for the test year -

COLUMN NAME	DATATYPE	COMMENTS
		https://www.ucat.ac.uk/research/technical-reports/
UKCAT_QUANTITATIVE_REASONING_Z	NUMBER (5,3)	Score converted into a Z-score using the population statistics from the UCAT technical report for the test year - https://www.ucat.ac.uk/research/technical-reports/
UKCAT_VERBAL_REASONING_Z	NUMBER (5,3)	Score converted into a Z-score using the population statistics from the UCAT technical report for the test year - https://www.ucat.ac.uk/research/technical-reports/
UKCAT_COG_TOTAL_Z	NUMBER (5,3)	Score converted into a Z-score using the population statistics from the UCAT technical report for the test year - https://www.ucat.ac.uk/research/technical-reports/

UKCAT_PROGRESSION

This table contains progression data collected by the UKCAT consortium. Data for the following medical schools are included:

- Aberdeen
- Barts
- Brighton and Sussex
- Dundee
- Edinburgh
- Hull York
- Leicester
- Manchester
- Norwich
- Nottingham
- Peninsula
- Sheffield
- St Andrews
- St George's

The data are described in more detail in 'The UKCAT-12 study: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of twelve UK medical schools', *BMC Medicine* 2013, 11:244 IC McManus, Chris Dewberry, Sandra Nicholson and Jonathan S Dowell, <http://www.biomedcentral.com/1741-7015/11/244>.

COLUMN NAME	DATATYPE	COMMENTS
UKCAT_UKCATID	VARCHAR2(14 BYTE)	<i>Internal use only</i>
UKCAT_HUSID	VARCHAR2(14 BYTE)	HUSID – HESA unique student identifier. <i>Internal use only.</i>
PERSON_UID	VARCHAR2(7 BYTE)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
UKCAT_UNIVERSITY_CODE	VARCHAR2(3 BYTE)	
UKCAT_UNIVERSITY_NAME	VARCHAR2(400 BYTE)	Medical school name.
UKCAT_COURSE	VARCHAR2(5 BYTE)	UCAS course code
UKCAT_COURSE_NAME	VARCHAR2(400 BYTE)	Course name – not populated.
UKCAT_MATRICULATIONID	VARCHAR2(25 BYTE)	University student identifier.

COLUMN NAME	DATATYPE	COMMENTS
UKCAT_PROGRESSIONYEAR	NUMBER (1)	Year of study.
UKCAT_OUTCOME	VARCHAR2(50 BYTE)	Outcome classification of this year of study.
UKCAT_THEORYSCORE	NUMBER (5,2)	Student score in theory exam(s).
UKCAT_THEORYRANK	NUMBER (3)	Student rank in year group by theory exam score.
UKCAT_SKILLSCORE	NUMBER (5,2)	Student score in skills exam(s).
UKCAT_SKILLSRANK	NUMBER (3)	Student rank in year group by skills exam score.
UKCAT_OVERALLSCORE	NUMBER (4,2)	Student combined theory and skills score.
UKCAT_OVERALLRANK	NUMBER (3)	Student rank in year group by combined theory and skills score.
UKCAT_MTASRANK	NUMBER (3)	Student MTAS points in year group by quartile ranking
UKCAT_NOTES	VARCHAR2(400 BYTE)	Further details supplemental to outcome.
UKCAT_REPEATYEAR	NUMBER (1)	0 = first sitting 1 = repeat year

UKMED_GEOGRAPHY

This table contains attributes that are about a PERSON and are derived from reference tables that link to the person via their postcode on application to medical school. It contains the geographical data for all cases provided by HESA. Geographical data for applicants are in VW_UKMED_PERSON_APPLICANT.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
HESA_COMYEAR	NUMBER	Year the student started the course extracted from HESA_COMDATE.
PARENTAL_POSTCODE	VARCHAR2(8 CHAR)	Postcode of the student at the time of their application to UCAS (therefore presumably 'home' prior to university). <i>Internal use only.</i>
PARENTAL_POSTAL_OUTCODE	VARCHAR2(8 CHAR)	Outcode of the student at the time of their application to UCAS (therefore presumably "home" prior to university).
PARENTAL_POSTCODE_SRC	VARCHAR2(5 CHAR)	Outcode of the student at the time of their application to UCAS (therefore presumably "home" prior to university).
IMD_QUINTILE	NUMBER	Each small area within nation (England, Northern Ireland, Scotland and Wales) is ranked with a lower score indicating greater deprivation. These scores are put into quintiles 1 - Most deprived 2 3 4 5 - Least deprived See https://census.ukdataservice.ac.uk/get-data/related/deprivation for more information. The reference data covers the following years England 2004 https://webarchive.nationalarchives.gov.uk/20100407164233/http://www.communities.gov.uk/documents/communities/xls/dacidap04.xls 2007 https://webarchive.nationalarchives.gov.uk/20100411141238/http://www.communities.gov.uk/documents/communities/xls/576508.xls 2010 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6872/1871524.xls 2015 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467764/File_1_ID_2015_

COLUMN NAME	DATATYPE	COMMENTS
		<p>Index_of_Multiple_Deprivation.xlsx</p> <p>Scotland 2004 https://www2.gov.scot/Publications/2004/06/19421/38087 2006 https://www2.gov.scot/Resource/Doc/933/0041675.xls 2009 https://www2.gov.scot/Resource/Doc/933/0102096.xls 2012 https://www2.gov.scot/Resource/0041/00410767.xls 2016 https://www2.gov.scot/Resource/0053/00534450.xlsx</p> <p>Northern Ireland 2005 https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/NIMDM2005_SOA_level_0.xls 2010 https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/Copy%20of%20NIMDM_2010_Results_SOA_0.xls 2017 https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/NIMDM17_SOAreults.xls</p> <p>Wales 2004 https://gov.wales/docs/statistics/2011/111222wimd11scorese n.xls 2011 https://gov.wales/docs/statistics/2011/111222wimd11scorese n.xls 2014 http://gov.wales/docs/statistics/2015/150812-wimd-2014-overall-domain-ranks-each-Isa-revised-en.xlsx</p> <p>The reference data only hold a new value if it has changed. The value taken is the value from the closest proceeding year to the student commencing their studies as medical school. However, if the value has not changed over time the source year may be earlier.</p>
IMD_QUINTILE_LABEL	VARCHAR2(18 BYTE)	Label for quintile.
IMD_QUINTILE_NON_GRAD_ENTRY	NUMBER	IMD quintile for non-graduate only as postcode on entry more likely to be the parental postcode.
IMD_YEAR	NUMBER	Source year for IMD quintile. If the value has not changed it may be some years before the student's hesa comyear.
POLAR_QUINTILE	NUMBER	<p>Young participation quintile 1(low)-5(high) and (unclassified). The young participation classification (POLAR) is based the participation in high education of young people who reached 18 years of age between a given proceeding time period depending on the POLAR version.</p> <p>This field contains the POLAR value from the closest proceeding year to the student commencing their studies at medical school. The reference data only contains a new value if the value has changed.</p> <p>1999 - POLAR1 https://webarchive.nationalarchives.gov.uk/20120118175255/http://www.hefce.ac.uk/widen/polar/lookup/postcode_to_pola</p>

COLUMN NAME	DATATYPE	COMMENTS
		<p>r.zip</p> <p>2004 - POLAR2 https://webarchive.nationalarchives.gov.uk/20120118175306/http://www.hefce.ac.uk/widen/polar/polar2/postcode_to_area_groups_10_12_21b.zip</p> <p>2010 - POLAR3 https://www.officeforstudents.org.uk/media/9f7c40e6-4fa7-49c7-ad76-25168cd04302/postcode_lookup_file_sep18.xlsx</p> <p>2014 - POLAR4 https://www.officeforstudents.org.uk/media/9f7c40e6-4fa7-49c7-ad76-25168cd04302/postcode_lookup_file_sep18.xlsx</p>
POLAR_QUINTILE_LABEL	VARCHAR2(33 BYTE)	Label for POLAR quintile.
POLAR_QUINTILE_NON_GRAD_ENTRY	NUMBER	POLAR quintile for non-graduate only as postcode on entry more likely to be the parental postcode.
POLAR_YEAR	NUMBER	Source year for POLAR quintile. If the value has not changed it may be some years before the student's hesa comyear.
IDACI_QUINTILE	NUMBER	<p>Income Deprivation Affecting Children Index (IDACI) constructed by the Social Disadvantage Research Centre at the University of Oxford. Available for England and Wales only.</p> <p>England</p> <p>2004 https://webarchive.nationalarchives.gov.uk/20100407164233/http://www.communities.gov.uk/documents/communities/xls/idacidap04.xls</p> <p>2007 https://webarchive.nationalarchives.gov.uk/20100411141238/http://www.communities.gov.uk/documents/communities/xls/576508.xls</p> <p>2010 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6883/1871683.xls</p> <p>2015 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467766/File_3_ID_2015_Supplementary_Indices_-_Income_Deprivation_Affecting_Children_Index_and_Income_Deprivation_Affecting_Older_People_Index.xlsx</p> <p>Score are ranked and placed in quintiles where</p> <p>1 - Most deprived 2 3 4 5 - Least deprived</p>
IDACI_QUINTILE_LABEL	VARCHAR2(18 BYTE)	Label for IDACI quintile.
IDACI_QUINTILE_NON_GRAD_ENTRY	NUMBER	IDACI quintile for non-graduate only as postcode on entry more likely to be the parental postcode.
IDACI_YEAR	NUMBER	Source year for IDACI quintile. If the value has not changed it

COLUMN NAME	DATATYPE	COMMENTS
		may be some years before the student's hesa comyear.
IDAOPQI_QUINTILE	NUMBER	<p>Income Deprivation Affecting Older People Index constructed by the Social Disadvantage Research Centre at the University of Oxford. Available for England and Wales only.</p> <p>2004 https://webarchive.nationalarchives.gov.uk/20100407164233/http://www.communities.gov.uk/documents/communities/xls/idacdap04.xls</p> <p>2007 https://webarchive.nationalarchives.gov.uk/20100411141238/http://www.communities.gov.uk/documents/communities/xls/576508.xls</p> <p>2010 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6883/1871683.xls</p> <p>2015 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467766/File_3_ID_2015_Supplementary_Indices_-_Income_Deprivation_Affecting_Children_Index_and_Income_Deprivation_Affecting_Older_People_Index.xlsx</p> <p>Score are ranked and placed in quintiles where 1 - Most deprived 2 3 4 5 - Least deprived</p>
IDAOPQI_QUINTILE_LABEL	VARCHAR2(18 BYTE)	Label for IDAOPI quintile.
IDAOPQI_QUINTILE_NON_GRAD_ENTRY	NUMBER	IDAOPQI quintile for non-graduate only as postcode on entry more likely to be the parental postcode.
IDAOPQI_YEAR	NUMBER	Source year for IDAOPI quintile. If the value has not changed it may be some years before the student's hesa comyear.
ADULT_HE_QUINTILE	NUMBER	The proportion of adults in the area that hold a Higher Education level qualification (based on 2001 census data). Indicates which Adult Higher Education quintile the student's parental postcode is in within the country
ADULT_HE_LABEL	VARCHAR2	Label for quintile
AHE_QUINTILE_NON_GRAD_ENTRY	NUMBER	Adult Higher Education quintile for non-graduate only as postcode on entry more likely to be the parental postcode.
ADULT_HE_YEAR	NUMBER	Source year for ADULT_HE_QUINTILE. If the value has not changed it may be some years before the student's hesa comyear.
TUNDRA_QUINTILE	NUMBER	<p>TUNDRA (tracking underrepresentation by area) is an area-based measure that uses tracking of state-funded mainstream school pupils in England to calculate young participation. It is a new, experimental measure.</p> <p>TUNDRA classifies local areas across England into five equal groups – or quintiles - based on the proportion of 16-year-old state-funded mainstream school pupils who participate in higher education aged 18 or 19 years.</p> <p>Quintile one shows the lowest rate of participation. Quintile five shows the highest rate of participation.</p> <p>The measure focuses on state-funded mainstream students</p>

COLUMN NAME	DATATYPE	COMMENTS
		who are typically included in geographically based outreach. Its main objective is to help outreach programmes identify and target areas of low participation more effectively.
TUNDRA_LABEL	VARCHAR2	Label for TUNDRA quintile.
TUNDRA_QUINTILE_NON_GRAD_ENTRY	NUMBER	TUNDRA quintile for non-graduate only as postcode on entry more likely to be the parental postcode.
TUNDRA_YEAR	NUMBER	Source year for TUNDRA data

VW_GP_ROUTE

The route by which the Dr joined the GP register. See <https://www.gmc-uk.org/registration-and-licensing/the-medical-register/a-guide-to-the-medical-register/specialist-and-gp-application-types>

Applications have been stored in the GMC's database (Siebel) from 2007 onwards and specialist applications (CCT, CEGPR, CEGPR with Registration, CESR, CESR with Registration and Combined Programme) stored in Siebel from March 2011. Therefore, specialty route prior to March 2011 will not be complete.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(14 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ROUTE	VARCHAR2(15 CHAR)	
GRANT_TYPE	VARCHAR2(50 CHAR)	
GRANT_SUB_TYPE	VARCHAR2(30 CHAR)	This field allows CCTs to be distinguished from CEGPR and Combined Programme GP registrations.

VW_NTS_TRAINEE_HISTORY

These data are collected from LETBs and deaneries to administer the National Training Survey. The collection notices are organised by year here: <https://www.gmc-uk.org/education/how-we-quality-assure/national-training-surveys/national-training-surveys---deaneries-and-hee-local-teams/resources-and-briefings-archive>

For each year see Briefing Note 2: data collection and survey timeline.

In addition to data available from 2012 when the survey was administered using GMC systems, a reduced data set is available for years 2009 to 2011. 2009 to 2011 data also contain the information required to administer the survey, but not the entire population only those who responded to the survey (response rates were 85% [2009], 87.5% [2010] and 87% [2011]). More information on the historic data is available on request.

Data from the 2020 NTS census are included here. However, unlike 2012 through to 2019, these data were not validated by the trainees when responding the survey, as due to the coronavirus (COVID-19) pandemic the survey did not run using these data. A separate survey was run later in the year, as captured in

The table is also available in wide-format with one row per person and columns for each year. The wide version is called VW_NTS_TRAINEE_WIDE.

COLUMN NAME	DATATYPE	COMMENTS
AGORA_ID	VARCHAR2(15 CHAR)	GMC database ID. <i>Internal use only.</i>
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
SURVEY_YEAR	VARCHAR2(4 CHAR)	The year of the survey that this record relates to.
NTS_DATA_ID	VARCHAR2(25 CHAR)	Unique Siebel-generated reference ID for the NTS data record.
DEANERY_ID	VARCHAR2(15 CHAR)	Trainee's Management Deanery. FK onto Organisation table.
FOUNDATION_SCHOOL	VARCHAR2(100 CHAR)	The Trainee's Foundation School. Not taken from the Siebel NTS record, but derived based upon the PROGRAMME_APPROVAL_CODE (the Foundation School is stored on the PROGRAMME record. Stamped here for ease of reporting).
PROGRAMME_APPROVAL_CODE	VARCHAR2(12 CHAR)	Unique reference ID for the trainee's Programme.

COLUMN NAME	DATATYPE	COMMENTS
PROGRAMME_SPECIALTY	VARCHAR2(100 CHAR)	Curriculum reference ID for the trainee's Programme.
PROGRAMME_APPROVAL_CODE_2	VARCHAR2(12 CHAR)	Unique reference ID for the trainee's Programme (supplied when the trainee is in a dual training programme).
PROGRAMME_SPECIALTY_2	VARCHAR2(100 CHAR)	Curriculum reference ID for the trainee's second Programme (supplied if the trainee is in a dual training programme).
TRAINING_LEVEL	VARCHAR2(50 CHAR)	The trainee's grade/training level.
POST_START_DATE	DATE	Date trainee started their post.
POST_END_DATE	DATE	Date trainee ends their post (if relevant).
BOARD_TRUST_CODE	VARCHAR2(50 CHAR)	ODS code for the trainee's Board or Trust. (FK onto ORGANISATION table.)
BOARD_TRUST_NAME	VARCHAR2(250 CHAR)	Name of the trainee's Board or Trust. Also held on ORGANISATION table but included here for ease of reporting.
SITE_CODE	VARCHAR2(50 CHAR)	ODS Code for the trainee's Site.
SITE_NAME	VARCHAR2(250 CHAR)	Name of the trainee's site. Also held in ORGANISATIONS table but included here for ease of reporting.
POST_SPECIALTY	VARCHAR2(250 CHAR)	Name of the trainee's Post Specialty.
IN_TRAINING_IND	VARCHAR2(1 CHAR)	Indicates if the trainee is currently in a training post.
NIT_REASON	VARCHAR2(250 CHAR)	Reason trainee is not in training if IN_TRAINING_IND is false.
TRAINEE_TYPE	VARCHAR2(50 CHAR)	Type of trainee (e.g. LAT, FTSTA, CL, AFT, etc)
NTN_DRN	VARCHAR2(50 CHAR)	National Training Number / Deanery Reference Number.
NTN_PROGRAMME_CODE	VARCHAR2(50 CHAR)	Programme code value derived from the NTN string (this value is not recorded in Siebel. A script will be executed to write this value based on the rules utilised with the ARCP validation tool).
SURVEY_STATUS	VARCHAR2(50 CHAR)	NTS survey status. Possible values: 1. Not started 2. In progress 3. Excluded

COLUMN NAME	DATATYPE	COMMENTS
EXCLUSION_REASON	VARCHAR2(150 CHAR)	Describes why a trainee was excluded from the survey.
SURVEY_COMPLETION_CODE	VARCHAR2(50 CHAR)	Unique Siebel-generated code that is created once the trainee completes the survey.
GEOGRAPHIC_LETB	VARCHAR2(250 CHAR)	Geographic LETB name (value automatically set based on the Board or Trust associated with the NTS data record).
GEOGRAPHIC_DEANERY	VARCHAR2(250 CHAR)	Geographic Deanery name (value automatically set based on the Board or Trust associated with the NTS data record).
NTS_CURRICULUM	VARCHAR2(250 CHAR)	Referred to as "Programme Speciality Name", this is the Programme's related Curriculum Name from Siebel.
PROGRAMME_SPECIALTY_2_G ROUP	VARCHAR2(100 CHAR)	Programme specialty grouped by medical royal college.
PROGRAMME_SPECIALTY_GRO UP	VARCHAR2(100 CHAR)	Programme specialty grouped by medical royal college.
PRIMARY_PMQ_MEDSCHOOL	VARCHAR2(250 CHAR)	
LETB	VARCHAR2(100 CHAR)	

VW_SPEC_ROUTE

The route by which the Dr joined the specialist register. See <https://www.gmc-uk.org/registration-and-licensing/the-medical-register/a-guide-to-the-medical-register/specialist-and-gp-application-types>

Applications have been stored in the GMC's database (Siebel) from 2007 onwards and specialist applications (CCT, CEGPR, CEGPR with Registration, CESR, CESR with Registration and Combined Programme) stored in Siebel from March 2011. Therefore, specialty route prior to March 2011 will not be complete.

VW

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
SPECIALTY	VARCHAR2(100 CHAR)	
ROUTE	VARCHAR2(30 CHAR)	
GRANT_TYPE	VARCHAR2(50 CHAR)	
GRANT_SUB_TYPE	VARCHAR2(30 CHAR)	This field allows CCTs to be distinguished from CESR and Combined Programme specialty registrations.

VW_UKCAT_RESULTS_NONCOG

These tests were piloted in 2007 and are not presented in other years.

COLUMN NAME	DATATYPE	COMMENTS
UKCATREGID	VARCHAR2(10 CHAR)	<i>Internal use only</i> ukcatRegistration.id. Nine-digit test registration identifier.
TEST	VARCHAR2(50 CHAR)	Behavioural test name: ITQ100; IVQ33/ITQ50; IVQ49; MEARS; SA12.
SECTION	VARCHAR2(50 CHAR)	Behavioural sub test name: aloofness; conditional self-worth; confidence; control; dissimulation; emotional defensiveness; empathy; extreme control; faking; libertarian communitarian; narcissism; optimism; resilience; self-discipline; self-esteem.
SCORE	NUMBER	Score value of response > 0

VW_UKMED_ALLEGATIONS

VW_UKMED_ALLEGATIONS is a View of the existing Allegations table.

Allegation records are only included in the view where.

Case number is in VW_UKMED_FTP_SUMMARY.

AND

Investigation Outcome is either 'Warning' or 'Undertaking'.

OR

Adjudication Outcome is 'Proven'.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
CASE_NUMBER	VARCHAR2(25 CHAR)	The Case Number of the related FtP Case. A case may contain multiple allegations.
GMP_DOMAIN	VARCHAR2(50 CHAR)	The "Domain" (or high-level section) of Good Medical Practice that is referenced by this Allegation record. There are four domains: Communication, Partnership & Teamwork Knowledge, Skills & Performance Maintaining Trust Safety & Quality
ALLEG_TYPE	VARCHAR2(50 CHAR)	For more information see The state of medical education and practice in the UK report: 2014 – Chapter 2 Health: substance misuse, mental and behavioural issues, and other health issues affecting judgement. Criminality: violence, sexual issues, harassment, motoring offences, fraud, and other criminal activities. Acting honestly and fairly: failure to act with honesty and integrity, treating or prescribing for themselves or friends, and unfairness or discrimination. Professional performance: failure to follow guidance, codes or regulations, inadequate training and knowledge, inadequate leadership, poor record keeping, and inefficient use of resources. Clinical competence: bad judgement of own abilities, poor diagnosis and examination, prescribing problems, and other clinical issues. Communication and respect for patients: lack of appropriate communication, failure to coordinate care, and lack of respect for patients. Working with colleagues: not meeting teaching or training responsibilities, and not working well with colleagues. Safety and quality systems: inadequate use of safety and quality systems, inadequate response to risks, and delay or

COLUMN NAME	DATATYPE	COMMENTS
		failure to raise concerns.
ALLEG_SUB_TYPE	VARCHAR2(50 CHAR)	Allegation sub types

VW_UKMED_FTPCHARDECLARATIONS

VW_UKMED_FTPCHARDECLARATIONS is a view of the FTPCHARDECLARATIONS table. It only includes declarations relating to persons in the VW_UKMED_PERSON view.

This FTPCHARDECLARATIONS table contains character declarations as provided by doctors during applications.

Inclusion criteria

Only 'Positive' responses are included (i.e. where someone has declared 'yes' to one of the declarations). The data captured on application to the register is described here: http://www.gmc-uk.org/doctors/registration_applications/declaration_of ftp.asp.

In Siebel, the wording of the declaration questions has evolved over time, however within Siebel each question revision is a unique item. To ease reporting in Agora, an ETL mapping is used to group the declarations into categories, contained here as DECLARATION_TYPE.

COLUMN NAME	DATATYPE	COMMENTS
AGORA_ID	VARCHAR2(15 CHAR)	<i>Internal use only</i>
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
DECLARATION_TYPE	VARCHAR2(100 CHAR)	Describes the category of declaration question. This value is derived during extraction from Siebel (where, over time, a number of questions records have been used due to minor evolution in wording) Describes the category of declaration question. This value is derived during extraction from Siebel (where, over time, a number of questions records have been used due to minor evolution in wording) These include: <ol style="list-style-type: none"> 1. Cautions or convictions 2. Physical/mental health issues affecting a Dr's ftp 3. Formal disciplinary action/FTP procedures undertaken by medical school or university 4. Suspended from duty or complaint upheld 5. Refused registration or licence to practise 6. Fined or given a warning by a regulator 7. Conduct or capability that might raise ftp questions 8. Disciplinary action by employer 9. Current or future proceedings that might lead to ftp proceedings in another country 10. Settlement regarding malpractice/negligence 11. Fixed penalty notice 12. Penalty notice for disorder or harassment notice 13. Potentially refused CGS by MRA in countries worked

COLUMN NAME	DATATYPE	COMMENTS
DECLARATION_DATE	DATE	Date the declaration was made.
APPLICATION_TYPE	VARCHAR2(50 CHAR)	For the UKMED cohort the application type = 'Provisional'.
APPLICATION_STATUS	VARCHAR2(50 CHAR)	<p>The Status of the registration application that this declaration relates to. The possible values are:</p> <ul style="list-style-type: none"> • Application Withdrawn • Not Eligible • Registration Granted • Application Expired • Registration Refused
APPLICATION_STATUS_REASON	VARCHAR2(250 CHAR)	

VW_UKMED_FTP_SUMMARY

VW_FTP_SUMMARY is a view of an existing view that summarises the complex fitness to practise structure into a flat table, and only where the CASE_OUTCOME is either 'Sanctions applied at a Hearing' or 'Sanctions applied without a Hearing'.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
ENQUIRY_RECEIVED	DATE	The date the Enquiry (complaint or referral) was received.
CASE_OUTCOME	VARCHAR2(100 CHAR)	The Outcome of the FTP case.
CASE_OUTCOME_SANCTIONS	VARCHAR2(100 CHAR)	Sanction applied to the Dr, in decreasing order of severity: erased or suspended from the medical register, conditions imposed on their registration, agreeing to undertakings warning
CASE_NUMBER	VARCHAR2(50 CHAR)	The GMC FtP Case Number. This identifier links to the VW_UKMED_ALLEGATIONS
ENQUIRY_INCIDENT_DATE	DATE	Date of incident.
INCIDENT_ORGANISATION	VARCHAR2(15 CHAR)	The type of enquiry (complaint or referral).
INCIDENT_ORG_NAME	VARCHAR2(250 CHAR)	
ENQUIRY_TYPE	VARCHAR2(50 CHAR)	The type of enquiry (complaint or referral).
ENQUIRY_SOURCE_TYPE	VARCHAR2(50 CHAR)	The source of the Enquiry. Possible values: Public (Organisation) PAPC Public (Individual) Doctor Press Cuttings ISA PAPC - Performer List
FTP_FROM	DATE	Date FtP sanctions started
FTP_TO_DATE	DATE	End date for FtP sanctions

COLUMN NAME	DATATYPE	COMMENTS
FTP_TYPE	VARCHAR2	Type of FtP: Condition Erasure Suspension Undertaking Warning

VW_UKMED_PERSON_APPLICANT

This table combines the following:

- Demographic data from UCAS
- Demographic data from the UKCAT registration form
- Attributes derived from reference tables that link to the person via their postcode on their UCAS application form

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	<p>GMC unique doctor reference number for those that registered.</p> <p>For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix.</p> <p>For cases that relate to unsuccessful medical school applications only, the ID will be the UCAS PERSON_ID with a 'U' prefix.</p> <p>Replaced by extract specific STUDY_ID in de-identified research extracts.</p>
UCAS_PERSON_ID	VARCHAR2(10 CHAR)	UCAS person ID
UCAS_HESA_HUSID	VARCHAR2(15 CHAR)	HUSID that was matched to the UCAS data
UCAS_MATCH_METHOD_USED	VARCHAR2(25 CHAR)	Method for achieving the match between UCAS and HESA data. Only successful applicants are matched to HESA data.
UCAS_APPNO	VARCHAR2(30 CHAR)	Applicant identifier that is unique to each applicant in each cycle. For applicants who have more than one application in the dataset, this is their first application number (minimum number). <i>For internal use only.</i>
UCAS_YEAR	VARCHAR2(4 CHAR)	The UCAS cycle in which the first application made by the applicant was processed.
UCAS_GENDER	VARCHAR2(6 CHAR)	Sex as declared by the applicant. The value is taken from the 1st application present in the UCAS extract.
UCAS_YEAR_BIRTH	NUMBER	Year of birth
UCAS_CORR_POSTAL_OUTCODE	VARCHAR2(240 BYTE)	<i>Internal use only</i> for calculations such as distance from home address to each medical school.
UCAS_DOMICILE	VARCHAR2(25 CHAR)	Applicant's area of permanent residence summarised to global regions. Split by the following regions if Non-UK: 'Africa', 'Americas', 'Australasia', 'Europe', 'Asia'. Please note: The Channel Islands and the Isle of Man are assigned as 'Europe'.
UCAS_UK_REGION	VARCHAR2(50 CHAR)	Applicant's area of permanent residence summarised to regions. This variable is derived from domicile as declared by the applicant. Split by the following values if domiciled in the UK: 'North East', 'Yorkshire and The Humber', 'North West', 'East Midlands', 'West Midlands', 'East of England', 'London',

COLUMN NAME	DATATYPE	COMMENTS
		'South East', 'South West', 'Wales', 'Northern Ireland', 'Scotland'.
UCAS_ETHNIC_GROUP_SUM	VARCHAR2(50 CHAR)	High-level grouping of ethnic origin as declared by the applicant: 'White', 'Black', 'Asian', 'Mixed', 'Other', 'Unknown'. Please note: Ethnic origin is captured for UK-domiciled applicants only; therefore, all non-UK-domiciled applicants are assigned as 'Not Applicable' and is not available within the current live cycle.
UCAS_ETHNIC_GROUP_DETAIL	VARCHAR2(50 CHAR)	Low-level grouping of ethnic origin as declared by the applicant: 'White', 'Black – Caribbean', 'Black – African', 'Black - Other Black background', 'Asian – Indian', 'Asian – Pakistani', 'Asian – Bangladeshi', 'Asian – Chinese', 'Asian - Other Asian background', 'Mixed – White and Black Caribbean', 'Mixed – White and Black African', 'Mixed – White and Asian', 'Mixed – Other mixed background', 'Other', 'Unknown'. Please note: Ethnic origin is captured for UK domiciled applicants only, therefore all non-UK-domiciled applicants are assigned as 'Not Applicable' and is not available within the current live cycle.
UCAS_SOCIO_ECON_2000	VARCHAR2(100 CHAR)	The National Statistics Socio-economic Classification (NS-SEC) is an occupationally based system used to classify the adult population. Conditions such as pay, and seniority of position are used to determine class. This is declared by the applicant, however please note a change in question in 2008. 2008 question: 'If you are in full-time education, please state the occupation of the highest-earning family member of the household in which you live. If he or she is retired or unemployed, give their most recent occupation. If you are not in full-time education, please state just your own occupation'. Pre-2008 question: 'If you are under 21, please state the occupation of the highest-earning family member of the household in which you live. If he or she is retired or unemployed, give their most recent occupation. If you are 21 or over, please state just your own occupation'. The response is captured for UK domiciled applicants only, therefore all non-UK domiciled applicants are assigned as 'Not classified / unknown'.
UCAS_SOCIO_ECON_2010	VARCHAR2(100 CHAR)	The National Statistics Socio-economic Classification (NS-SEC) is an occupationally based system used to classify the adult population. The applicant is asked: 'If you are in full-time education, please state the occupation of the highest-earning family member of the household in which you live. If he or she is retired or unemployed, give their most recent occupation. If you are not in full-time education, please state just your own occupation'. The applicant may then choose from 28,000 ONS job descriptions. These job descriptions are then mapped to 8 Socio-Economic Group codes via a lower level set of around 380 '2010 SOC Codes'. The response is captured for UK domiciled applicants only, therefore all non-UK domiciled applicants are assigned as 'Not classified / unknown'. Please note that, although the same 8 Socio-Economic Group codes are displayed in the Socio-economic group 2000 variable available from 2004-2014, occupations are mapped via a different set of '2000 SOC Codes'. Therefore, some job descriptions are mapped to different Socio-Economic Group values.
UCAS_POLAR3	VARCHAR2(1 CHAR)	Developed by HEFCE, POLAR3 classifies small areas across the UK into five groups according to their level of young participation in Higher Education. Each of these groups represents around 20 per cent of young people and is ranked from Quintile 1 (areas with the lowest young participation rates, considered as the most disadvantaged) to Quintile 5 (highest young participation rates, considered most advantaged). POLAR3 is based on the participation rates of young people between 2005 and 2009, who entered HE

COLUMN NAME	DATATYPE	COMMENTS
		between 2005-06 and 2010-11 academic years. These groups are assigned using the postcode declared by the applicant. If a postcode is invalid, considered unsafe for measurement or there is no link to Census geography possible then the applicant is classified as missing. Please note: POLAR3 is only available for applicants domiciled in the UK, therefore any applicants domiciled outside of the UK are classified as missing. Also, although POLAR3 is available for applicants of all ages, it is recommended/most suitable for analysis of applicants aged 19 and under.
UCAS_POLAR4	VARCHAR2(1 CHAR)	Developed by HEFCE, POLAR4 classifies small areas across the UK into five groups according to their level of young participation in Higher Education. Each of these groups represents around 20 per cent of young people and is ranked from Quintile 1 (areas with the lowest young participation rates, considered as the most disadvantaged) to Quintile 5 (highest young participation rates, considered most advantaged). POLAR4 is based on the participation rates of young people between 2009 and 2014, who entered HE between 2009-10 and 2014-15 academic years. These groups are assigned using the postcode declared by the applicant. If a postcode is invalid, considered unsafe for measurement or there is no link to Census geography possible then the applicant is classified as missing. Please note: POLAR4 is only relevant for applicants domiciled in the UK, therefore any applicants outside of this cohort are classified as missing.
UCAS_ENGLISH_IMD	VARCHAR2(1 CHAR)	The Index of Multiple Deprivation (IMD) for 2015 identifies small area concentrations of multiple deprivation across all of England, providing a relative measure of deprivation amongst 32,844 small areas (data zones) based on 7 socio-economic domains. These small areas are classified into five groups ranked from Quintile 1 (considered the most deprived) to Quintile 5 (considered least deprived), with equal populations in each quintile. Quintiles are assigned using the postcode declared by the applicant, applicants declaring invalid postcodes are classified as missing. Please note: IMD is only relevant for applicants domiciled in England, therefore any applicants outside of this cohort are classified as missing.
UCAS_NORTHERNIRISH_MDM	VARCHAR2(1 CHAR)	The Northern Ireland Multiple Deprivation Measure (NIMDM) for 2017 identifies small area concentrations of multiple deprivation across all of England, providing a relative measure of deprivation amongst 890 small areas (data zones) based on 7 socio-economic domains. These small areas are classified into five groups ranked from Quintile 1 (considered the most deprived) to Quintile 5 (considered least deprived), with equal populations in each quintile. Quintiles are assigned using the postcode declared by the applicant, applicants declaring invalid postcodes are classified as missing. Please note: NIMDM is only relevant for applicants domiciled in Northern Ireland, therefore any applicants outside of this cohort are classified as missing.
UCAS_SCOTTISH_IMD	VARCHAR2(1 CHAR)	Scottish index of multiple deprivation (SIMD) for 2016 identifies small area concentrations of multiple deprivation across all of Scotland, providing a relative measure of deprivation amongst 6505 small areas (data zones) based on 7 socio-economic domains. These small areas are classified into five groups ranked from Quintile 1 (considered the most deprived) to Quintile 5 (considered least deprived), with equal populations in each quintile. Quintiles are assigned using the postcode declared by the applicant, applicants declaring invalid postcodes are classified as missing. Please note: SIMD is only relevant for applicants domiciled in Scotland, therefore any applicants outside of this cohort are classified as missing.
UCAS_WELSH_IMD	VARCHAR2(1 CHAR)	The Welsh Index of Multiple Deprivation (IMD) for 2014 identifies small area concentrations of multiple deprivation across all of England, providing a relative measure of deprivation amongst 1,909 small areas (data zones) based on 7 socio-economic domains. These small areas are classified into five groups ranked from Quintile 1 (considered the most deprived) to Quintile 5 (considered least deprived), with equal populations in each quintile. Quintiles are assigned using the postcode declared by the applicant, applicants declaring

COLUMN NAME	DATATYPE	COMMENTS
		invalid postcodes are classified as missing. Please note: WIMD is only relevant for applicants domiciled in Wales, therefore any applicants outside of this cohort are classified as missing.
UCAS_SCH_TYPE_MAP_2018	VARCHAR2(50 CHAR)	The type of the school or centre through which the application was submitted, giving an indication of the type of educational establishment attended by the applicant. The most recent school or centre type for each school code held by UCAS is displayed across the time series, regardless of the school or centre type at the time of the application. For example, academies were introduced 2012. Give school type, e.g. Comprehensive School, Further Education, Grammar School, Independent School and so on.
UCAS_NATIONAL_CENTRE_NUM	VARCHAR2(7 CHAR)	The National Centre Number of the school or centre through which the application was submitted. This is reported for schools for which UCAS holds a National Centre Number. The most recent National Centre Number for each school code held by UCAS is displayed across the time series, regardless of the National Centre Number at the time of application.
UCAS_UKPRN	VARCHAR2(8 CHAR)	The UK Provider Reference Number of the school or centre through which the application was submitted. This is reported for schools for which UCAS holds a UK Provider Reference Number. The most recent UK Provider Reference Number for each school code held by UCAS is displayed across the time series, regardless of the UK Provider Reference Number at the time of application.
UCAS_DFE_NUMBER	VARCHAR2(7 CHAR)	The DfE/LAESTAB number of the school or centre through which the application was submitted. This is reported for schools for which UCAS holds a DfE/LAESTAB number. The most recent DfE/LAESTAB number for each school code held by UCAS is displayed across the time series, regardless of the DfE/LAESTAB number at the time of application.
UKCAT_MUKCATID_MIN	VARCHAR2(15 BYTE)	<i>Internal use only.</i> 11-character alphanumeric ukcat participant identifier - lowest number identified as part of deduping UKCAT data.
UKCAT_MUKCATID	VARCHAR2(15 BYTE)	<i>Internal use only</i> 11-character alphanumeric ukcat participant identifier
UKCAT_GENDER	VARCHAR2(11 CHAR)	From the UKCAT registration form
UKCAT_ETHNICGROUP	VARCHAR2(10 BYTE)	From the UKCAT registration form
UKCAT_NATIONALIDENTITY	VARCHAR2(50 BYTE)	Delimited list of participant national identities from the UKCAT registration form.
UKCAT_REGIONALIDENTITY	VARCHAR2(30 BYTE)	Delimited list of participant British regional identities.
UKCAT_CORRESPONDENCECOUNTRY	VARCHAR2(10 BYTE)	UKCAT registration form - country
UKCAT_CORRESPONDENCESTATE	VARCHAR2(10 BYTE)	UKCAT registration form - state
UKCAT_PARENT1GENDER	VARCHAR2(20 BYTE)	Participant parent 1 gender description.
UKCAT_PARENT1OCCUPATION	VARCHAR2(50 BYTE)	Participant parent 1 socio-economic class occupation description.

COLUMN NAME	DATATYPE	COMMENTS
UKCAT_PARENT1EMPLOYMENT STATUS	VARCHAR2(40 BYTE)	Participant parent 1 socio-economic classification employment status description.
UKCAT_PARENT1EMPLOYERSIZE	VARCHAR2(20 BYTE)	Participant parent 1 socio-economic classification size of employer description.
UKCAT_PARENT1SUPERVISOR	VARCHAR2(20 BYTE)	Participant parent 1 socio-economic classification supervisory responsibility description.
UKCAT_PARENT2GENDER	VARCHAR2(50 BYTE)	Participant parent 2 socio-economic classification gender description.
UKCAT_PARENT2OCCUPATION	VARCHAR2(50 BYTE)	Participant parent 2 socio-economic classification occupation description.
UKCAT_PARENT2EMPLOYMENT STATUS	VARCHAR2(50 BYTE)	Participant parent 2 socio-economic classification employment status description.
UKCAT_PARENT2EMPLOYERSIZE	VARCHAR2(50 BYTE)	Participant parent 2 socio-economic classification size of employer description.
UKCAT_PARENT2SUPERVISOR	VARCHAR2(50 BYTE)	Participant parent 2 socio-economic classification supervisory responsibility description.
UKCAT_NSSEC	VARCHAR2(10 BYTE)	National Statistics socio-economic five-point scale classification 1 = managerial and professional occupations 2 = intermediate occupations 3 = small employers and own account workers 4 = lower supervisory and technical occupations 5 = semi-routine and routine occupations
UKCAT_HIGHESTQUALIFICATION	VARCHAR2(30 BYTE)	Highest qualification standard attained by participant prior to registration.
DATA_SOURCE	VARCHAR2(250 CHAR)	Data source from UKCAT load process
UKCAT_MATCH_SCRIPT	VARCHAR2(100 CHAR)	Match script used to join UKCAT record to HESA.
IMD_QUINTILE	NUMBER	Each small area within nation (England, Northern Ireland, Scotland and Wales) is ranked with a lower score indicating greater deprivation. These scores are put into quintiles 1 - Most deprived 2 3 4 5 - Least deprived See https://census.ukdataservice.ac.uk/get-data/related/deprivation for more information. The reference data covers the following years England 2004

COLUMN NAME	DATATYPE	COMMENTS
		<p> https://webarchive.nationalarchives.gov.uk/20100407164233/http://www.communities.gov.uk/documents/communities/xls/i dacidap04.xls 2007 https://webarchive.nationalarchives.gov.uk/20100411141238/http://www.communities.gov.uk/documents/communities/xls/ 576508.xls 2010 https://www.gov.uk/government/uploads/system/uploads/atta chment_data/file/6872/1871524.xls 2015 https://assets.publishing.service.gov.uk/government/uploads/s ystem/uploads/attachment_data/file/467764/File_1_ID_2015_ Index_of_Multiple_Deprivation.xlsx </p> <p> Scotland 2004 https://www2.gov.scot/Publications/2004/06/19421/38087 2006 https://www2.gov.scot/Resource/Doc/933/0041675.xls 2009 https://www2.gov.scot/Resource/Doc/933/0102096.xls 2012 https://www2.gov.scot/Resource/0041/00410767.xls 2016 https://www2.gov.scot/Resource/0053/00534450.xlsx </p> <p> Northern Ireland 2005 https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/ NIMDM2005_SOA_level_0.xls 2010 https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/ Copy%20of%20NIMDM_2010_Results_SOA_0.xls 2017 https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/ NIMDM17_SOAreults.xls </p> <p> Wales 2004 https://gov.wales/docs/statistics/2011/111222wimd11scorese n.xls 2011 https://gov.wales/docs/statistics/2011/111222wimd11scorese n.xls 2014 http://gov.wales/docs/statistics/2015/150812-wimd-2014- overall-domain-ranks-each-IsOA-revised-en.xlsx </p> <p> The reference data only hold a new value if it has changed. The value taken is the value from the closest proceeding year to the student commencing their studies as medical school. However, if the value has not changed over time the source year may be earlier. </p>
IMD_QUINTILE_LABEL	VARCHAR2(18 BYTE)	Label for quintile.
IMD_YEAR	NUMBER	Source year for IMD quintile. If the value has not changed it may be some years before the student's ucas year.
POLAR_QUINTILE	NUMBER	<p> Young participation quintile 1(low)-5(high) and (unclassified). The young participation classification (POLAR) is based the participation in high education of young people who reached 18 years of age between a given proceeding time period depending on the POLAR version. This field contains the POLAR value from the closest </p>

COLUMN NAME	DATATYPE	COMMENTS
		<p>proceeding year to the student commencing their studies at medical school. The reference data only contains a new value if the value has changed.</p> <p>1999 - POLAR1 https://webarchive.nationalarchives.gov.uk/20120118175255/http://www.hefce.ac.uk/widen/polar/lookup/postcode_to_polar.zip</p> <p>2004 - POLAR2 https://webarchive.nationalarchives.gov.uk/20120118175306/http://www.hefce.ac.uk/widen/polar/polar2/postcode_to_area_groups_10_12_21b.zip</p> <p>2010 - POLAR3 https://www.officeforstudents.org.uk/media/9f7c40e6-4fa7-49c7-ad76-25168cd04302/postcode_lookup_file_sep18.xlsb</p> <p>2014 - POLAR4 https://www.officeforstudents.org.uk/media/9f7c40e6-4fa7-49c7-ad76-25168cd04302/postcode_lookup_file_sep18.xlsb</p>
POLAR_QUINTILE_LABEL	VARCHAR2(33 BYTE)	Label for POLAR quintile.
POLAR_YEAR	NUMBER	Source year for POLAR quintile. If the value has not changed it may be some years before the student's ucas year.
IDACI_QUINTILE	NUMBER	<p>Income Deprivation Affecting Children Index (IDACI) constructed by the Social Disadvantage Research Centre at the University of Oxford. Available for England and Wales only.</p> <p>England</p> <p>2004 https://webarchive.nationalarchives.gov.uk/20100407164233/http://www.communities.gov.uk/documents/communities/xls/idacidap04.xls</p> <p>2007 https://webarchive.nationalarchives.gov.uk/20100411141238/http://www.communities.gov.uk/documents/communities/xls/576508.xls</p> <p>2010 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6883/1871683.xls</p> <p>2015 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467766/File_3_ID_2015_Supplementary_Indices_-_Income_Deprivation_Affecting_Children_Index_and_Income_Deprivation_Affecting_Older_People_Index.xlsx</p> <p>Score are ranked and placed in quintiles where</p> <p>1 - Most deprived 2 3 4 5 - Least deprived</p>
IDACI_QUINTILE_LABEL	VARCHAR2(18 BYTE)	Label for IDACI quintile.
IDACI_YEAR	NUMBER	Source year for IDACI quintile. If the value has not changed it may be some years before the student's hesa comyear.
IDAOPQI_QUINTILE	NUMBER	Income Deprivation Affecting Older People Index constructed

COLUMN NAME	DATATYPE	COMMENTS
		<p>by the Social Disadvantage Research Centre at the University of Oxford. Available for England and Wales only.</p> <p>2004 https://webarchive.nationalarchives.gov.uk/20100407164233/http://www.communities.gov.uk/documents/communities/xls/dacidap04.xls</p> <p>2007 https://webarchive.nationalarchives.gov.uk/20100411141238/http://www.communities.gov.uk/documents/communities/xls/576508.xls</p> <p>2010 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6883/1871683.xls</p> <p>2015 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467766/File_3_ID_2015_Supplementary_Indices_-_Income_Deprivation_Affecting_Children_Index_and_Income_Deprivation_Affecting_Older_People_Index.xlsx</p> <p>Score are ranked and placed in quintiles where 1 - Most deprived 2 3 4 5 - Least deprived</p>
IDAOPQI_QUINTILE_LABEL	VARCHAR2(18 BYTE)	Label for IDAOPI quintile.
IDAOPQI_YEAR	NUMBER	Source year for IDAOPI quintile. If the value has not changed it may be some years before the student's hesa comyear.
ADULT_HE_QUINTILE	NUMBER	The proportion of adults in the area that hold a Higher Education level qualification (based on 2001 census data). Indicates which Adult Higher Education quintile the student's parental postcode is in within the country
ADULT_HE_LABEL	VARCHAR2(33 BYTE)	Label for quintile
ADULT_HE_YEAR	NUMBER	Adult Higher Education quintile for non-graduate only as postcode on entry more likely to be the parental postcode.
TUNDRA_QUINTILE	NUMBER	<p>TUNDRA (tracking underrepresentation by area) is an area-based measure that uses tracking of state-funded mainstream school pupils in England to calculate young participation. It is a new, experimental measure.</p> <p>TUNDRA classifies local areas across England into five equal groups – or quintiles - based on the proportion of 16-year-old state-funded mainstream school pupils who participate in higher education aged 18 or 19 years.</p> <p>Quintile one shows the lowest rate of participation. Quintile five shows the highest rate of participation.</p> <p>The measure focuses on state-funded mainstream students who are typically included in geographically based outreach. Its main objective is to help outreach programmes identify and target areas of low participation more effectively.</p>
TUNDRA_LABEL	VARCHAR2(33 BYTE)	Label for TUNDRA quintile.
TUNDRA_YEAR	NUMBER	Source year for TUNDRA data

COLUMN NAME	DATATYPE	COMMENTS
APSFTE_ALEVA	NUMBER (6,2)	<p>Average point score per A level student (full-time equivalent) for the school attended. . From DfE file linked on UCAS_DFE_NUMBER</p> <p>Key Stage 5 Performance Tables – Published School Data for the year the student took A levels. Only available for England.</p>
TALLPPE_ALEVA	NUMBER (5,1)	<p>Average point score per A level entry from the school attended. From DfE file linked on UCAS_DFE_NUMBER</p> <p>Key Stage 5 Performance Tables – Published School Data for the year the student took A levels. Only available for England.</p>
TB3PTSE	NUMBER (6,2)	<p>Average point score in best 3 A level entries</p> <p>Key Stage 5 Performance Tables – Published School Data for the year the student took A levels. Only available for England. .From DfE file linked on UCAS_DFE_NUMBER</p>

VW_UKMED_PERSON_FULL

VW_UKMED_PERSON_FULL contains attributes that are considered to be about a PERSON, but particularly related to those for whom we hold student information from external data sources.

HESA and UKCAT data are used to derive student attributes.

In both HESA and UKCAT, many rows can occur with the same (or differing) attributes; therefore, logic has been applied in order to select the best value. As follows:

HESA

The following values are taken from the earliest available Instance:

- HESA_UKPRN_FIRST
- HESA_UKPRN_FIRST_NAME
- HESA_UCAS_APP_ID
- HESA_UCAS_PERSON_ID
- HESA_PREV_INST
- HESA_PREV_INST_NAME
- HESA_DOMICILE_COUNTRY
- HESA_DOMICILE_REGION
- HESA_PARENTAL_POSTCODE
- HESA_TARIFF
- HESA_QUALENT

The following values are taken from the latest available Instance:

- HESA_UKPRN_LAST
- HESA_UKPRN_LAST_NAME
- HESA_OWNSTU
- HESA_HIGHQUAL_OBTAIN

UKCAT

NOTE: For columns labelled UKCAT_* – where a person has done more than one UKCAT test registration, values are taken from the earliest available test registration.

Data that are derived from a person's postcode such as IMD quintiles have been moved to the UKMED_GEOGRAPHY table.

Protected characteristics are collected from HESA and the GMC's Siebel system. If a doctor states they do not wish to declare their given protected characteristic, such as ethnicity or religion to the GMC we do not include the given HESA value if the Doctor had declared it whilst a student.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. Replaced by extract specific STUDY_ID in de-identified

COLUMN NAME	DATATYPE	COMMENTS
		research extracts.
HESA_HUSID	VARCHAR2(50 CHAR)	HUSID – HESA’s unique student identifier see: https://www.hesa.ac.uk/collection/c16051/a/husid <i>Internal use only.</i>
BIRTH_DT	DATE	<i>Internal use only.</i>
BIRTH_MONTH	NUMBER	Birth month extracted from Date of Birth held on the GMC register. Date of birth is too identifiable for inclusion in extracts.
BIRTH_YEAR	NUMBER	Birth year extracted from Date of Birth held on the GMC register. Date of birth is too identifiable for inclusion in extracts.
GENDER	VARCHAR2(30 CHAR)	
GENDER_INT	NUMBER	Gender coded as integer for analysis: 1 = Female and 0 = Male.
PRIMARY_PMQ	VARCHAR2(250 CHAR)	Primary medical qualification as per the GMC’s register
PRIMARY_PMQ_YEAR	VARCHAR2(4 CHAR)	Primary PMQ year.
PRIMARY_PMQ_PLACE	VARCHAR2(150 CHAR)	Primary Medical Qualification awarding body.
PRIMARY_PMQ_SHORTCODE	VARCHAR2(50 CHAR)	Code used by the GMC to identify medical schools and their awarding bodies, due to changes in awarding bodies, the same medical school can have more than one shortcode.
PRIMARY_PMQ_WORLD_REGION	VARCHAR2(5 CHAR)	Primary Medical Qualification awarding body IMG UK EEA Based on the current classification of country of qualification not the classification at the time of registration.
PRIMARY_PMQ_WORLD_REGION_INT	NUMBER	World region codes as an integer.
CURR_REG_ADD_POSTCODE	VARCHAR2(30 CHAR)	<i>Internal use only</i>
CURR_REG_ADD_POSTCODE_OUTCODE	VARCHAR2(30 CHAR)	Outcode, derived from CURR_REG_ADD_POSTCODE – current postcode from the GMC register.
NATIONALITY	VARCHAR2(150 CHAR)	Collected at the point of checking a doctor’s ID – usually at the stage registration is granted, sometimes at the point where they are at PLAB ID checks. Doctors provide ID cards or passports and it is logged as part of checking them.

COLUMN NAME	DATATYPE	COMMENTS
NATIONALITY2	VARCHAR2(150 CHAR)	Collected at the point of checking a doctor's ID – usually at the stage we grant registration, sometimes at the point where they are at PLAB ID checks. Doctors provide ID cards or Passports and it's logged as part of checking them.
NATIONALITY_SRC		Nationality source. For cases on the GMC register this is the GMC, for cases not on registered it is HESA. However, nationality was only included in the GMC's HESA from the academic year 2015/16.
HESA_UKPRN_FIRST	VARCHAR2(8 CHAR)	The chronologically first UKPRN number that this student is found to have in HESA_ACAD_YEARS.
HESA_UKPRN_FIRST_NAME	VARCHAR2(100 CHAR)	The name of the university as found in HESA_UKPRN_FIRST. Included here for ease of reporting.
HESA_UKPRN_LAST	VARCHAR2(8 CHAR)	The chronologically last UKPRN number that this student is found to have in HESA_ACAD_YEARS.
HESA_UKPRN_LAST_NAME	VARCHAR2(100 CHAR)	The name of the university as found in HESA_UKPRN_LAST. Included here for ease of reporting.
MEDICAL_SCHOOL_FIRST	VARCHAR2(100 CHAR)	The first medical school derived from UKPRN using CR_UKPRN_NAME.
MEDICAL_SCHOOL_LAST	VARCHAR2(100 CHAR)	The last medical school derived from UKPRN using CR_UKPRN_NAME.
FIRST_UKPRN_AFTER_GAP		First UKPRN after any gap (non-continuous ACYEARS) in data from HESA. f
FIRST_MEDSCHOOL_AFTER_G AP	VARCHAR2(100 CHAR)	First medical school after any gap (non-continuous ACYEARS) in data from HESA. This variable is derived to capture first medical school of the course completed for students who leave and then start their studies at another school.
COURSE_TYPE	VARCHAR2(200 CHAR)	HESA CTITLE AND HESA COURSE_ID were manually mapped to course types in https://www.medschools.ac.uk/studying-medicine/course-types . Possible value include: <ul style="list-style-type: none"> • Standard Entry Medicine • Transferred from Oxbridge for year 3 onwards • Graduate Entry Programme • Medicine with a Gateway Year • Medicine with a Preliminary Year • Private • EU Medicine
COURSETYPE_AFTER_GAP	VARCHAR2(50)	First course type after any gap (non-continuous ACYEARS) in data from HESA. This variable is derived to capture course type of the course completed for students who leave and then start their studies at another school.
HESA_COMDATE	DATE	Commencement date of programme. The date of the student's initial commencement of studies.

COLUMN NAME	DATATYPE	COMMENTS
FIRST_COMDATE_AFTER_GAP		First commencement date of programme after a gap (non-continuous ACYEARS) in data from HESA.
HESA_COMYEAR	NUMBER	Year the student started the course extracted from HESA_COMDATE.
HESA_PREV_INST	VARCHAR2(10 CHAR)	Previous institution attended.
HESA_QUAL_YEAR	VARCHAR2(4 CHAR)	The earliest year the student obtained the qualifications used for entry to medical school.
HESA_PREV_INST_DCODE	VARCHAR2(8 CHAR)	<i>Internal use only.</i> The Department for Education (DfE) code for the students' previous institution was obtained to allow linkage to school A level results which are published by D code. Records from HESA contain codes in different formats as they migrate towards UKPRNs for all.
APSFTE_ALEVA	NUMBER	Average point score per A level student (full-time equivalent) for the school attended. From DfE file linked on HESA_PREV_INST_DCODE. Key Stage 5 Performance Tables – Published School Data for the year the student took A levels. Only available for England.
TALLPPE_ALEVA	NUMBER	Average point score per A level entry from the school attended. From DfE file linked on HESA_PREV_INST_DCODE. Key Stage 5 Performance Tables – Published School Data Key Stage 5 Performance Tables – Published School Data for the year the student took A levels. Only available for England.
HESA_DOMICILE_COUNTRY	VARCHAR2(100 CHAR)	Known in HESA as "Domicile - county/country". Contains the country of domicile if outside UK regions. Contains the domicile unitary authority level if within UK regions.
HESA_DOMICILE_REGION	VARCHAR2(50 CHAR)	Domicile – The students' domicile prior to commencement of the course. Contains the worldwide geographic region.
HESA_TARIFF	NUMBER	Tariff – Tariff points based on the qualifications on entry of the student. Tariff data were collected directly from UCAS prior to 2007/08 and were calculated by HESA for those entering from 2007/08 onwards. As such, tariff data will not be directly comparable across this time period.
HESA_QUALENT	VARCHAR2(200 CHAR)	Highest qualification on entry - The highest qualification that a student holds on entry. QUALENT3 (https://www.hesa.ac.uk/collection/c15051/a/qualent3) was compulsory for entrants from 2010/11, from that point only existing students could be given a QUALENT2 code.
HESA_HIGHQUAL_OBTAIN1	VARCHAR2(1000 BYTE)	Qualification obtained – Illustrates the qualification level achieved by the student. Students can obtain more than one qualification.

COLUMN NAME	DATATYPE	COMMENTS
HESA_HIGHQUAL_OBTAIN2	VARCHAR2(1000 BYTE)	Qualification obtained – Illustrates the qualification level achieved by the student. Students can obtain more than one qualification.
UKCAT_NATIONAL_IDENTITY	VARCHAR2(50 CHAR)	Delimited list of participant national identities from the UKCAT registration form.
UKCAT_NATIONAL_IDENTITY_INT	NUMBER	National identify coded as an integer.
UKCAT_REGIONAL_IDENTITY	VARCHAR2(50 CHAR)	Delimited list of participant British regional identities.
UKCAT_PARENT1GENDER	VARCHAR2(50 CHAR)	Participant parent 1 gender description.
UKCAT_PARENT1OCCUPATION	VARCHAR2(50 CHAR)	Participant parent 1 socio-economic class occupation description.
UKCAT_PARENT1EMPLOYMENT STATUS	VARCHAR2(50 CHAR)	Participant parent 1 socio-economic classification employment status description.
UKCAT_PARENT1EMPLOYERSIZE	VARCHAR2(50 CHAR)	Participant parent 1 socio-economic classification size of employer description.
UKCAT_PARENT1SUPERVISOR	VARCHAR2(50 CHAR)	Participant parent 1 socio-economic classification supervisory responsibility description.
UKCAT_PARENT2GENDER	VARCHAR2(50 CHAR)	Participant parent 2 socio-economic classification gender description.
UKCAT_PARENT2OCCUPATION	VARCHAR2(50 CHAR)	Participant parent 2 socio-economic classification occupation description.
UKCAT_PARENT2EMPLOYMENT STATUS	VARCHAR2(50 CHAR)	Participant parent 2 socio-economic classification employment status description.
UKCAT_PARENT2EMPLOYERSIZE	VARCHAR2(50 CHAR)	Participant parent 2 socio-economic classification size of employer description.
UKCAT_PARENT2SUPERVISOR	VARCHAR2(50 CHAR)	Participant parent 2 socio-economic classification supervisory responsibility description.
UKCAT_DOMICILE	VARCHAR2(50 CHAR)	Area of residence of applicant.
UKCAT_NATIONALITY1	VARCHAR2(50 CHAR)	First nationality of applicant.
UKCAT_NATIONALITY2	VARCHAR2(50 CHAR)	Second nationality of applicant.

COLUMN NAME	DATATYPE	COMMENTS
ARCP_MILITARY_IND	VARCHAR2(3 CHAR)	Trainee flagged as a military trainee on at least one ARCP return.
ARCP_ACADEMIC_IND	VARCHAR2(3 CHAR)	Trainee flagged as an academic trainee on at least one ARCP return.
ARCP_MILITARY_IND_INT	VARCHAR2(42 CHAR)	Trainee flagged as a military trainee on at least one ARCP return as an integer.
ARCP_ACADEMIC_IND_INT	VARCHAR2(42 CHAR)	Trainee flagged as an academic trainee on at least one ARCP return as an integer.
ETHNICITY_L1	VARCHAR2(50 CHAR)	Ethnicity Level 1 information.
ETHNICITY_L1_INT	NUMBER	Ethnicity Level 1 information coded as an integer.
ETHNICITY_L2	VARCHAR2(150 CHAR)	Ethnicity Level 2 information.
ETHNICITY_L2_INT	NUMBER	Ethnicity Level 2 information coded as an integer.
ETHNICITY_SRC	VARCHAR2(12 CHAR)	Source system of the ETHNICITY_L1 and ETHNICITY_L2 data stored for this record.
BME	VARCHAR2(7 BYTE)	Higher level ethnicity coding: BME or white.
BME_INT	VARCHAR2(2 BYTE)	Higher level ethnicity coded as an integer.
PG_DISABILITY	VARCHAR2(250 CHAR)	Disability as recorded in the GMC's Siebel system used for Doctors' registration from January 2016. Or prior to 2016 as captured in the NTS.
PG_DISABILITY_INT	NUMBER	Disability as recorded in the GMC's Siebel system used for Doctors' registration from January 2016. Or prior to 2016 as captured in the NTS. 0 = 'No', 1 = 'Yes', Null = 'Prefer not to say'
PG_DISABILITY_SRC	VARCHAR2(10 CHAR)	NTS year: 2012, 2013, 2014, 2015 or 2016. Or Siebel if 2017 onwards. Taking the most recent instance.
PG_BLIND_SIGHTLOSS	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
PG_COGNITIVE	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
PG_DEAF_HEARINGLOSS	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.

COLUMN NAME	DATATYPE	COMMENTS
PG_DISABLED_PREFERNOTTOSPEC	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
PG_LEARNING_DISABILITY	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
PG_MANUAL_DEXTERITY	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
PG_MENTAL_ILLNESS	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
PG_MOBILITY	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
PG_OTHER_IMPAIRMENT	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
PG_SPEECH	NUMBER	Populated with a 0 or 1 where PG_DISABILITY_SRC = 'SIEBEL' and PG_DISABILITY = 'Yes'.
HESA_DISABILITY_CODE_FIRST	VARCHAR2(5 CHAR)	First disability code in the HESA data for the student. See https://www.hesa.ac.uk/collection/c16051/a/disable
HESA_DISABILITY_LABEL_FIRST	VARCHAR2(250 CHAR)	First disability label in the HESA data for the student.
HESA_DISABILITY_CODE_LAST	VARCHAR2(5 CHAR)	Last disability code in the HESA data for the student.
HESA_DISABILITY_LABEL_LAST	VARCHAR2(250 CHAR)	Last disability label in the HESA data for the student.
LIMITED_ACTIVITIES	VARCHAR2(250 CHAR)	Describes whether the doctor's day-to-day activities are limited because of a health problem or disability which has lasted, or is expected to last, 12 months. Captured on the NTS.
LIMITED_ACTIVITIES_SRC	VARCHAR2(10 CHAR)	Year of NTS for column LIMITED_ACTIVITIES.
ADJUSTMENTS	VARCHAR2(250 CHAR)	Describes whether the doctor required any adjustment(s) to be made so they could carry out their work in their post. Captured on the NTS.
ADJUSTMENTS_SRC	VARCHAR2(10 CHAR)	Year of NTS for column ADJUSTMENTS
UK_EDUCATED	VARCHAR2(250 CHAR)	Describes whether the doctor completed secondary education AND their undergraduate medical degree in the UK.

COLUMN NAME	DATATYPE	COMMENTS
SCHOOL_TYPE	VARCHAR2(250 CHAR)	Describes which type of school the doctor mainly attended between the ages of 11 and 16. HESA State school marker or if no HESA data then as captured retrospectively on the NTS
SCHOOL_TYPE_SRC	VARCHAR2(10 CHAR)	Describes source of SCHOOL_TYPE. HESA or NTS year.
INCOME_SUPPORT	VARCHAR2(250 CHAR)	Describes whether the doctor's household received Income Support at any point during their school years. Captured retrospectively on the NTS
INCOME_SUPPORT_SRC	VARCHAR2(10 CHAR)	Year of NTS for column INCOME_SUPPORT.
FREE_SCHOOL_MEALS	VARCHAR2(250 CHAR)	Describes whether the doctor had free school meals as captured retrospectively on the NTS.
FREE_SCHOOL_MEALS_SRC	VARCHAR2(10 CHAR)	Year of NTS for column FREE_SCHOOL_MEALS.
PARENT_DEGREE	VARCHAR2(250 CHAR)	Describes whether the doctor's parent(s) or guardian(s) completed a university degree course of equivalent captured retrospectively on the NTS.
PARENT_DEGREE_SRC	VARCHAR2(10 CHAR)	Year of NTS for column PARENT_DEGREE
SEC	VARCHAR2(100 CHAR)	Socio-economic classification. Socio-economic classification of the student's parent (if under 21) or the student themselves when over 21. See https://www.hesa.ac.uk/collection/c16051/a/sec
SEC_INT	VARCHAR2(2 BYTE)	SEC coded as an integer.
SEC_SRC	VARCHAR2(10 CHAR)	Indicates HESA as the source of SEC.
NSSEC	VARCHAR2(100 CHAR)	National Statistics socio-economic five-point scale classification 1 = managerial and professional occupations 2 = intermediate occupations 3 = small employers and own account workers 4 = lower supervisory and technical occupations 5 = semi-routine and routine occupations
NSSEC_INT	VARCHAR2(1 CHAR)	NSSEC coded as an integer
NSSEC_SRC	VARCHAR2(10 CHAR)	Indicates that UKCAT registration form is the source of NSSEC.
SOC2000	VARCHAR2(100 CHAR)	Occupation code – SOC code of student if ages 21 or over at the start of course, or the parents SOC code if under 21. Occupational code of the student's parent (where the student is under 21), OR the occupation of the student themselves when they start training at over 21. See https://www.hesa.ac.uk/collection/c16051/e/soc2000

COLUMN NAME	DATATYPE	COMMENTS
SOC2000_SRC	VARCHAR2(10 CHAR)	Indicates HESA as the source of SOC2000.
SOC2010	VARCHAR2(50)	Occupation code – SOC code of student if ages 21 or over at the start of course, or the parents SOC code if under 21. Occupational code of the student's parent (where the student is under 21), OR the occupation of the student themselves when they start training at over 21. and https://www.hesa.ac.uk/collection/c16051/e/soc2010
SOC2010_SRC	VARCHAR2(50)	Indicates HESA as the source of SOC2010
PARED	VARCHAR2(50 CHAR)	Parental education records whether a student's parents had higher education qualifications. HESA provide further details of on how the data are collected here: https://www.hesa.ac.uk/collection/c16051/a/pared .
PARED_INT	VARCHAR2(2 BYTE)	PARED coded as an integer.
PARED_SRC	VARCHAR2(10 CHAR)	Indicates HESA as the source of PARED.
SEXUAL_ORIENTATION	VARCHAR2(30 CHAR)	Collected by the GMC as part of registration process from January 2016. NTS respondents are also asked to update their details. Possible values are: <ul style="list-style-type: none"> • Bisexual • Heterosexual/Straight • Lesbian/Gay • Other • Prefer not to say
SEXUAL_ORIENTATION_SRC	VARCHAR2(12 CHAR)	Indicates the GMC's database as the source of SEXUAL_ORIENTATION.
RELIGION	VARCHAR2(30 CHAR)	<ul style="list-style-type: none"> • Collected by the GMC as part of registration process from January 2016. NTS respondents are also asked to update their details. Buddhist • Christian - Baptist • Christian - Brethren • Christian - Catholic • Christian - Church of England • Christian - Church of Ireland • Christian - Church of Scotland • Christian - Free Presbyterian • Christian - Methodist • Christian - Other • Christian - Presbyterian • Christian - Protestant • Hindu • Jewish • Muslim • No religion • Other

COLUMN NAME	DATATYPE	COMMENTS
		<ul style="list-style-type: none"> • Prefer not to say • Sikh
RELIGION_SRC	VARCHAR2(12 CHAR)	Indicates the GMC's database as the source of RELIGION. Or HESA
SCHOOL_POSTCODE	VARCHAR2(50 BYTE)	Postcode of school in UKCAT data.
COUNTRY	VARCHAR2(50 BYTE)	Country where school is located.
DFEID	VARCHAR2(50 BYTE)	Schools code in UKCAT data. <i>Internal use only.</i>
CATEGORYID	VARCHAR2(50 BYTE)	UCAS school category identifier 1 = UK School 2 = Under Sixteen School 3 = Further Education College 4 = Southern Irish School 5 = EU School 6 = Non-EU School 7 = Overseas School 8 = unknown
SUBCATEGORYID	VARCHAR2(50 BYTE)	UCAS subcategory school classification 1 = agricultural and horticultural college 2 = art, design and performing arts 3 = comprehensive school 4 = further education 5 = grammar school 6 = grant maintained (special school) 7 = grant maintained secondary school (state) 8 = higher education 9 = independent school 10 = language school 11 = sixth form centre 12 = sixth form college 13 = special school 14 = technical college 15 = tertiary college 16 = other secondary school 17 = other school 18 = not applicable 19 = unknown
SCHOOL_STATUS	VARCHAR2(50 BYTE)	Open/closed status of school.
CATEGORYDESCRIPTION	VARCHAR2(50 BYTE)	Label for CATEGORYID.
SUBCATEGORYDESCRIPTION	VARCHAR2(50 CHAR)	Label for SUBCATEGORYID.
GRADUATE_ON_ENTRY	VARCHAR2(21 BYTE)	Based on HESA_QUALENT, the following values are defined as graduate entry

COLUMN NAME	DATATYPE	COMMENTS
		'First degree of UK institution', 'First degree with honours leading to Qualified Teacher Status (QTS)/registration with a General Teaching Council (GTC)', 'Higher degree of UK institution', 'Non-UK doctorate degree', 'Non-UK first degree', 'Non-UK masters degree', 'PGCE with QTS/GTC registration', 'PGCE without QTS/GTC registration', 'Postgraduate diploma or certificate, excluding PGCE', 'UK doctorate degree', 'UK first degree with honours', 'UK masters degree', 'UK ordinary (non-honours) first degree', 'Postgraduate Certificate in Education or Professional Graduate Diploma in Education', 'Graduate of other overseas institution', 'Professional Graduate Certificate in Education', 'Integrated undergraduate/postgraduate taught masters degree on the enhanced/extended pattern', 'Graduate of EU institution'
SEC_COMBINED	NUMBER	Uses the value from HESA if present; if not present uses the value from the UKCAT registration form.
SEC_COMBINED_LABEL	VARCHAR2(100 CHAR)	Label for SEC_COMBINED field.
BURSARY	CHAR(1 BYTE)	Flag to indicate presence in the UKCAT Bursary table. More details on the bursary scheme are given here: https://www.ucat.ac.uk/ucat/bursary-scheme/
NTS_TRAINEE	CHAR(1 BYTE)	Flag to indicate if the case is in the NTS data
CARELEAVER	VARCHAR2(250 CHAR)	HESA value that records whether a student is a care leaver. See https://www.hesa.ac.uk/collection/c16051/a/careleaver
XELSP01_FIRST	VARCHAR2(250 CHAR)	First value for expected length of programme. See https://www.hesa.ac.uk/support/definitions/students#expected-length-programme
XELSP01_LAST	VARCHAR2(250 CHAR)	Last value for expected length of programme. See https://www.hesa.ac.uk/support/definitions/students#expected-length-programme
SSN	VARCHAR2(15 CHAR)	HESA held Student Support Number See https://www.hesa.ac.uk/collection/c16051/a/ssn
WP_INDEX	NUMBER(8,2)	From UKMEDP41. Lambe P, Roberts M, Gale T, and Bristow D. Title: Evaluating the utility of contextual admissions data for widening participation in UK medical education.

COLUMN NAME	DATATYPE	COMMENTS
		<p>Calculated as follows POLAR quintile 1 = 1, quintile 2 = 0.9, quintile 3 = 0.8, quintile 4 = 0.7 and quintile 5= 0.6 IMD quintile 1 = 1, quintile 2 = 0.9, quintile 3 = 0.8, quintile 4 = 0.7 and quintile 5 = 0.6</p> <p>SCHOOL TYPE, PARED and BURSARY were weighted 1 = state funded school, 1= parent no HE qualifications, and 1= in receipt of a bursary, and zero otherwise.</p> <p>Only calculated for UK domiciled, aged under 21 years, entrants to Standard Entry Programmes, Medicine with a Gateway Year and Medicine with a Preliminary Year</p>
SAT_PLAB	VARCHAR2	Flag to indicate whether the doctor sat PLAB
PMQ_WORLD_REGION_AT_REGN	VARCHAR2	The PMQ world region of the country the doctor obtained their PMQ at the point of their registration
HESA_BIO_MED	VARCHAR2	Flag to indicate if the student transferred on to a PMQ course from a Biomedical science course at the same provider.
UKCAT_LANGUAGE_FLUENCY	VARCHAR2(190 BYTE)	Languages spoken as declared on the UKCAT registration form
UKCAT_MOTHERTONGUE	VARCHAR2(90 BYTE)	1 st language spoken as declared by the applicant on the UKCAT registration form.

VW_UKMED_DISABILITY_AUDIT

This table records changes to the disability values held in Siebel. It records changes from 9th July 2020 onwards.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. For those who did not complete their studies/have not completed to date this will be HESA_HUSID with an 'H' prefix. Replaced by extract specific STUDY_ID in de-identified research extracts.
OPERATION_DT	DATE	Date value changes
DESC_TEXT	VARCHAR2 (225 Char)	The disability value as per PG_DISABILITY where SRC = Siebel'
OLD_VALUE	VARCHAR2 (1 Char)	The original value – null, 'Y' or 'N'
NEW_VALUE	VARCHAR2 (1 Char)	The new value 'Y' or 'N'

VW_UKMED_PRACTICEHISTORY

Contains data from

PRACTICEHISTORY

This table contain data from:

- ESR – Electronic Staffing Records - updated weekly
- PCIS – Primary Care Information System, has been replaced by data from Primary Care Support England (PCSE) - updated weekly
- SWISS – Scottish Workforce Information Standard System - updated monthly
- Scottish primary care data – one off load

Northern Ireland Business Services Organisation. Note that very little data are available from Northern Ireland. Data has not been refreshed recently.

There is one row per instance of practice history – i.e. employment episode.

Collection of these data started in 2012 and therefore even if assignment start dates are prior to 2012, the data do not provide a complete picture as assignments with similar start dates that ended before 2012 are not present in the table.

One record should be provided for every 'assignment' that the doctor has with a given employer. An assignment should be uniquely defined as the combination of the following data items:

- GMC Reference Number (Doctor UID)
- Site code
- Employing (or Parent) Organisation Code
- Employment start date

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
AREA_OF_WORK	VARCHAR2(100 CHAR)	Source table: PRACTICEHISTORY. This is the area, function or specialty where the work activity takes place. See National Workforce Data Set (NWD) – http://content.digital.nhs.uk/datasets/nwd
ASSIGNMENT_CATEGORY	VARCHAR2(50 CHAR)	Source table: PRACTICEHISTORY Data are only available for ESR contains information on whether the person is active in the role
DATA_SOURCE	VARCHAR2(50 CHAR)	Source table: PRACTICEHISTORY. Gives the data source of the record, one of the following: ESR – Electronic Staffing Records PCIS – Primary Care Information System SWISS – Scottish Workforce Information Standard System Northern Ireland Business Services Organisation
		Source table: PRACTICEHISTORY. ESR only, gives the

COLUMN NAME	DATATYPE	COMMENTS
SPECIALITY_AREA	VARCHAR2(150 CHAR)	doctor's specialty area of work, more detailed than area of work. Some values are CCT specialty values.
START_DATE	DATE	Source table: PRACTICEHISTORY
END_DATE	DATE	Source table: PRACTICEHISTORY
WORK_PATTERN	VARCHAR2(50 CHAR)	Source table: PRACTICEHISTORY Full-time or part-time This is not available for Northern Ireland cases.
GRADE	VARCHAR2(25 CHAR)	Source table: PRACTICEHISTORY. ESR only, this contains codes for NHS pay scales and will allow identification of those working at consultant level versus those working as Associate Specialists. Staff Grade or locums etc. Further details are available here: http://www.nhsemployers.org/~media/Employers/Documents/Pay%20and%20reward/Pay%20and%20Conditions%20Circular%20MD%2012015.pdf
JOB_ROLE	VARCHAR2(100 CHAR)	Source table: PRACTICEHISTORY. Identifies GP Locum GP Registrar General Practitioner
JOB_ROLE_DTL	VARCHAR2(100 CHAR)	Source table: PRACTICEHISTORY. Include information on the types of GP from PCIS.
PRACTICE_TYPE	VARCHAR2(50 CHAR)	Source table: PRACTICEHISTORY Identifies locum and permanent contract types.
DB_IND	VARCHAR2(1 CHAR)	
ODS_CODE_SITE	VARCHAR2(50 CHAR)	ORGANISATION. NHS side code
ORG_NAME_SITE	VARCHAR2(250 CHAR)	
ODS_CODE_ORG	VARCHAR2(50 CHAR)	ORGANISATION. NHS organisation code
ORG_NAME_ORG	VARCHAR2(250 CHAR)	
ODS_PAR_NAME_ORG	VARCHAR2(250 CHAR)	
GOR_ORG_POSTCODE	VARCHAR2(50 CHAR)	Government office region from Group from National Statistics Postcode Lookup (NSPL) in CR_ONS_POSTCODE see National Statistics Postcode Lookup (February 2017) User Guide. Available from http://geoportal.statistics.gov.uk/ Office for National Statistics (Edition: February 2017) National Statistics Postcode Lookup User Guide Available from https://data.gov.uk/dataset/national-statistics-postcode-lookup-may-2017-user-guide .

COLUMN NAME	DATATYPE	COMMENTS
		The region code for each postcode. Pseudo codes are included for Wales, Scotland, Northern Ireland, Channel Island and Isle of Man.
CCG_ORGP_POSTCODE	VARCHAR2(100 CHAR)	Clinical Commissioning Group from National Statistics Postcode Lookup (NSPL) in CR_ONS_POSTCODE see National Statistics Postcode Lookup (February 2017) User Guide. Available from http://geoportal.statistics.gov.uk/ Office for National Statistics (Edition: February 2017) National Statistics Postcode Lookup User Guide Available from: https://data.gov.uk/dataset/national-statistics-postcode-lookup-may-2017-user-guide

VW_UKMED_REGHISTORY

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(12 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
PRIMARY_IND	VARCHAR2(1 CHAR)	Indicates whether the registration record is the doctor's current registration record (Y = current record)
START_DATE	DATE	Start date of the registration
END_DATE	DATE	End date of the registration
EXT_REG_TYPE	VARCHAR2(100 CHAR)	The registration type as per what was/is shown on the LRMP during this period.
LICENCE_IND	VARCHAR2(1 CHAR)	Indicates whether the doctor had a licence to practise during this period. NB: Licensing was introduced on 16/11/2009

VW_UKMED_SPECIALTIES

Contains data from

SPECIALITIES - entries to the specialist. This contains data that is found on the publicly available list of medical practitioner here - <http://www.gmc-uk.org/doctors/register/LRMP.asp>. To guard against re-identification of cases, dates are set to year only.

PERSON which contains information on GP register entries.

It contains one row per specialist register entry, so more than one row per Doctor is possible.

COLUMN NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15 CHAR)	GMC unique doctor reference number for those that registered. Replaced by extract specific STUDY_ID in de-identified research extracts.
CURR_INCOME_DISCOUNT_IN D	VARCHAR2(1 CHAR)	Whether the doctor currently receives income discount on their registration. Possible values are Y(es) or NULL See https://www.gmc-uk.org/registration-and-licensing/managing-your-registration/fees-and-funding/income-discount
CURR_LIC_STATUS_IND	VARCHAR2(1 CHAR)	Whether doctor currently holds a license to practice. Possible values are Y(es) or N(o) or or null (null means the doctor is not registered at all)
CURR_REG	VARCHAR2(100 CHAR)	Current registration status.
GP_IND	VARCHAR2(1 CHAR)	Dr is on the GP register. Possible values are Y(es) Y or null. Y indicates that the doctor has been on the GP register at some point (they may be currently erased). It is possible to be on the GP and specialist registers.
YEAR_GP_REGISTER	NUMBER	The year the Dr first joined the GP register.
SPECIALIST_IND	VARCHAR2(1 CHAR)	Dr is on the specialist register. Y means the doctor is currently on the specialist register. If the doctor has previously been on the specialist register but is not currently registered it will be N.
YEAR_SPECIALIST_REGISTER	NUMBER	The year of first specialty registration.
ROYAL_COLLEGE	VARCHAR2(15 CHAR)	The medical royal college which is responsible for the specialty.
ROYAL_COLLEGE_NAME	VARCHAR2(100 CHAR)	Royal college name in full.
SPECIALTY	VARCHAR2(100 CHAR)	Specialty. Approved curricula for CCTs are listed here https://www.gmc-uk.org/education/standards-guidance-and-curricula/curricula Doctors may have a specialty registration that is not from a currently approved curriculum. See https://www.gmc-uk.org/registration-and-licensing/the-medical-register/a-guide-to-the-medical-register/specialist-registration
SPECIALTY_GROUP	VARCHAR2(100 CHAR)	Specialty grouping.
SUB_SPECIALTY	VARCHAR2(100 CHAR)	Sub-specialty it the Dr has registered one. See https://www.gmc-uk.org/education/standards-guidance-and-curricula/curricula

Reference Tables

These tables are reference tables required for interpretation they do not contain identifiable data.

CR_UKPRN

Data from <https://www.ukrlp.co.uk/> giving the names and postcodes of UK schools. Data were supplied on 20 July 2020. We will request annual refreshed post UCAS load.

COLUMN NAME	DATATYPE	COMMENTS
UKPRN	INTEGER	UK Provider reference number see UK Register of Learning Providers. https://www.ukrlp.co.uk/
LEGAL_NAME	VARCHAR2(255 CHAR)	
STATUS	VARCHAR2(20 CHAR)	ACTIVE or DEACTIVATED
ADDRESS_1	VARCHAR2(255 CHAR)	
ADDRESS_2	VARCHAR2(255 CHAR)	
ADDRESS_3	VARCHAR2(255 CHAR)	
TOWN	VARCHAR2(255 CHAR)	
COUNTY	VARCHAR2(255 CHAR)	
POSTCODE	VARCHAR2(20 CHAR)	
WEBSITE	VARCHAR2(255 CHAR)	Not available for all providers

CR_COURSE

This table maps the HESA values (UKPRN, QUALAIM, COURSEAIM, COURSEID, and CTITLE) to the following:

Course types described in:

- the Medical Schools Council document <https://www.medschools.ac.uk/media/2032/msc-entry-requirements-for-uk-medical-schools.pdf>;
- UCAS course codes; and
- the test required for entry to the course.

The mappings were confirmed by colleagues in the medical schools.

This table is used to assign course type in VW_UKMED_PERSON_FULL by first populating cr_vwukmedper_course_type which gives the course type for each person. Some values in cr_vwukmedper_course_type are amended using UCAS values which are based on ucas_course_name. UCAS data were only available after CR_COURSE was first developed and do not cover the year prior to 2007.

COLUMN NAME	DATATYPE	COMMENTS
UKPRN	VARCHAR2(200 CHAR)	UK Provider reference number see UK Register of Learning Providers. https://www.ukrlp.co.uk/
QUALAIM	VARCHAR2(200 CHAR)	This field describes the general qualification aim of the course and is intended to record the qualification that will be attained as a result of successful completion of studies. In use from 2002/2003 to 2006/2007.
COURSEAIM	VARCHAR2(250 CHAR)	This field describes the general qualification aim of the course and is intended to record the qualification that will be attained as a result of successful completion of studies. From 2007/08 onwards. See: https://www.hesa.ac.uk/collection/c16051/a/courseaim
COURSEID	VARCHAR2(200 CHAR)	University's own course identifier (2007/2008 onwards).
CTITLE	VARCHAR2(200 CHAR)	The medical school's title for the course. The course title as studied in this academic year. See: https://www.hesa.ac.uk/collection/c16051/a/ctitle/
COURSETYPE	VARCHAR2(200 CHAR)	The type of medicine course offered by the medical school. The courses are grouped into four types: Standard Entry Medicine Graduate Entry Medicine Medicine with a Preliminary Year Medicine with a Gateway Year. Entry requirements for each of the course types differ. https://www.medschools.ac.uk/media/2357/msc-entry-requirements-for-uk-medical-schools.pdf

COLUMN NAME	DATATYPE	COMMENTS
UCASCOURSECODE	VARCHAR2(10 CHAR)	UCAS Course Code.
TEST	VARCHAR2(25 CHAR)	The admissions test used by the medicine course as part of the selection process. The majority of medicine courses will use one of the three admissions tests: BMAT - http://www.admissionstesting.org/for-institutions/about-our-tests/biomedical-admissions-test GAMSAT - https://gamsat.acer.org UKCAT - https://www.ukcat.ac.uk
COURSELENGTH	VARCHAR2(10 CHAR)	Length of the course.
COURSENAME	VARCHAR2(100 CHAR)	The name of the medicine course as used by the medical school.
COMMENTS	VARCHAR2(500 CHAR)	Comments on the mapping from UKMED staff and/or medical school staff.
MEDICAL_SCHOOL	VARCHAR2(100 CHAR)	As per CR_UKPRN_NAME.

CR_EXAM

This table contains details of each Royal College and faculty exam held in UKMED. It is updated annually as part the GMC's postgraduate exam data collection.

COLUMN NAME	DATATYPE	COMMENTS
COLLEGE_NAME	VARCHAR2(25 CHAR)	Abbreviated name of college.
COLLEGE_NAME_LABEL	VARCHAR2(100 CHAR)	Full name of college.
ABBREVIATED_EXAM_NAME	VARCHAR2(100 CHAR)	Abbreviated exam contained in the college's submission to the GMC.
EXAM_NAME	VARCHAR2(250 CHAR)	Full name of the exam as per the college web site.
DATA_YEAR INTRODUCED	VARCHAR2(4 CHAR)	The first GMC collection year containing this exam.
EXAM_TYPE	VARCHAR2(250 CHAR)	Example values (not a full list): Single best answer (SBA) questions Multiple Choice Questions Station Based
LEVEL_TAKEN	VARCHAR2(500 CHAR)	The stage of training at which the exam is typically taken.
ELIGIBILITY	VARCHAR2(2000 CHAR)	Who is eligible to take the exam, e.g. UK only, in training only and so forth.
REQUIRED_FOR_PROGRESSIO N_TO	VARCHAR2(500 CHAR)	The point in training at which the exam is required to progress – see the path for each specialty here: http://www.gmc-uk.org/education/approved_curricula_systems.asp
SOURCE_URL	VARCHAR2(250 CHAR)	The source of the information about the exam.
SECTION_SCORE_AVAILABLE	VARCHAR2(1 CHAR)	Whether section scores are available for the exam.
TOTAL_SCORE_AVAILABLE	VARCHAR2(1 CHAR)	Whether a total score is available for the exam.
TYPE_OF_TOTAL_SCORE	VARCHAR2(50 CHAR)	The format of the total score – percent or integer.

Contains details of the SJT used in the foundation recruitment as this can vary by year.

<http://www.foundationprogramme.nhs.uk/content/situational-judgement-test-sjt>

COLUMN NAME	DATATYPE	COMMENTS
DATA_YEAR	VARCHAR2(4 CHAR)	Derived based on year the extract was received
QUESTION_PAPER	NUMBER	There are three variants of the SJT test paper used within a year. Approximately 6,000 applicants take Paper 1 (December – first sit only); 2,000 take paper 2 (January – first sit plus contingency); <10 take Paper 3 (January – contingency date only).
ITEMS_SCORED	NUMBER	0-60 (in reality, 58, 59 or 60) The SJT test paper comprises 70 items; of these, 10 are trial items, and the SJT final score is based on performance of 60 live items with known psychometric properties. If one of the live items does not perform well psychometrically, it is removed before scoring and test-equating. The placement of trial items changes each year, and these items are not identified.
SJT_SCORED_ITEMS_PART1	NUMBER	Number of items in part 1.
SJT_SCORED_ITEMS_PART2	NUMBER	Number of items in part 2.
SJT_RAW_MAX	NUMBER	Overall maximum possible score.
SJT_RAW_PART1_MAX	NUMBER	Maximum possible score in part 1. Part 1 – ranking (5 options). Marks for near misses. Scores out of 20 though some items are scored out of 19.
SJT_RAW_PART2_MAX	NUMBER	Maximum possible score in part 2. Multiple choice (select 3 from 8). Questions marked out of 12.

CR_ONS_POSTCODE

Table imported from ONS to allow postcodes to be linked to various geographical data – please see <https://data.gov.uk/dataset/e7308379-35af-46e6-a570-8825fec1e008/national-statistics-postcode-lookup-may-2018-user-guide> for details of all fields.

COLUMN NAME	DATATYPE	COMMENTS
PCD	VARCHAR2(7 CHAR)	
PCD2	VARCHAR2(8 CHAR)	
PCDS	VARCHAR2(8 CHAR)	
DOINTR	VARCHAR2(6 CHAR)	
DOTERM	VARCHAR2(6 CHAR)	
USERTYPE	VARCHAR2(1 CHAR)	
OSEAST1M	VARCHAR2(6 CHAR)	
OSNRTH1M	VARCHAR2(7 CHAR)	
OSGRDIND	VARCHAR2(1 CHAR)	
OA11	VARCHAR2(9 CHAR)	
CTY	VARCHAR2(9 CHAR)	
LAUA	VARCHAR2(9 CHAR)	
WARD	VARCHAR2(9 CHAR)	
HLTHAU	VARCHAR2(9 CHAR)	
HRO	VARCHAR2(9 CHAR)	
CTRY	VARCHAR2(9 CHAR)	
GOR	VARCHAR2(9 CHAR)	
PCON	VARCHAR2(9 CHAR)	

COLUMN NAME	DATATYPE	COMMENTS
EER	VARCHAR2(9 CHAR)	
TECLEC	VARCHAR2(9 CHAR)	
TTWA	VARCHAR2(9 CHAR)	
PCT	VARCHAR2(9 CHAR)	
NUTS	VARCHAR2(10 CHAR)	
PARK	VARCHAR2(9 CHAR)	
LSOA11	VARCHAR2(9 CHAR)	
MSOA11	VARCHAR2(9 CHAR)	
WZ11	VARCHAR2(9 CHAR)	
CCG	VARCHAR2(9 CHAR)	
BUA11	VARCHAR2(9 CHAR)	
BUASD11	VARCHAR2(9 CHAR)	
RU11IND	VARCHAR2(2 CHAR)	
OAC11	VARCHAR2(3 CHAR)	
LAT	VARCHAR2(10 CHAR)	
LONGI	VARCHAR2(10 CHAR)	
LEP1	VARCHAR2(9 CHAR)	
LEP2	VARCHAR2(9 CHAR)	
PFA	VARCHAR2(9 CHAR)	
IMD	VARCHAR2(5 CHAR)	

CR_PSA_APPROACH

COLUMN NAME	DATATYPE	COMMENTS
MEDICAL_SCHOOL	VARCHAR2(8 CHAR)	Medical school as per CR_UKPRN_NAME
PSA_YEAR	NUMBER	Year the PSA was set
PSA_APPROACH	VARCHAR2(100 CHAR)	Whether the medical school regarded the PSA as formative or summative.

CR_SPECIALTIES

COLUMN NAME	DATATYPE	COMMENTS
SPECIALTY_NAME	VARCHAR2(400 CHAR)	GMC full specialty name
GMC_ABBR_NAME	VARCHAR2(400CHAR)	GMC abbreviated name where different
SPECIALTY_CODE_ISD	VARCHAR2(10 CHAR)	Specialty code as per https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions-by-Groups/Codes/Specialty-Codes.asp
SPECIALTY_CODE_ENGLAND	VARCHAR2(10 CHAR)	Specialty code as per https://digital.nhs.uk/data-and-information/areas-of-interest/workforce/nhs-occupation-codes
SPECIALTY_SYNONYM	VARCHAR2(400CHAR)	Specialty other name where applicable
SPECIALTY_TYPE	VARCHAR2(400CHAR)	Possible values: Core Dental Dual from ORIEL Foundation Special Interest Specialty - Higher Specialty - Higher/Specialty - Run through Specialty - Run through Sub-specialty
SPECIALTY_COLLEGE	VARCHAR2(400CHAR)	The medical royal college with responsibility for getting the specialty curriculum approved by the GMC: https://www.gmc-uk.org/education/standards-guidance-and-curricula/curricula
HSIC_SPECIALTY_GROUP	VARCHAR2(400CHAR)	Specialty groups used as per NHS Digital workforce reports. https://digital.nhs.uk/data-and-information/publications/statistical/nhs-workforce-statistics
SPECIALTY_GROUP	VARCHAR2(50 CHAR)	GMC specialty group based on Royal College/Faculty

CR_UKPRN_NAME

Table that maps HESA UKPRN values to the following:
GMC short codes
GMC awarding bodies.
Agreed abbreviated names for medical schools.

COLUMN NAME	DATATYPE	COMMENTS
UKPRN	VARCHAR2(8 CHAR)	UK Provider reference number see UK Register of Learning Providers https://www.ukrlp.co.uk/
NAME	VARCHAR2(100 CHAR)	Name for the UKPRN.
WEB_NAME	VARCHAR2(100 CHAR)	Name of medical school on their website.
MEDICAL_SCHOOL	VARCHAR2(50 CHAR)	Abbreviated medical school name.
SHORT_CODE	VARCHAR2(3 CHAR)	GMC medical school code.
AWARD_BODY	VARCHAR2(100 CHAR)	Awarding body – see https://www.gmc-uk.org/education/undergraduate/awarding_bodies.asp .
START_YEAR	VARCHAR2(4 CHAR)	The year the institution started.
HESA_INST1	VARCHAR2(100 CHAR)	A teaching institution that is separate from the main university.
UCAS_INSTIT_CODE	VARCHAR2(20 CHAR)	
MEDICAL_SCHOOL_POSTCODE	VARCHAR2(20 CHAR)	

ORGANISATION_LIVE

Table of organisations. It includes historic data. Data sources include:

- NHS Digital - <https://digital.nhs.uk/services/organisation-data-service/data-downloads>
- Information Services Division – <http://www.isdscotland.org/>
- Ad hoc updates from GMC staff

COLUMN NAME	DATATYPE	COMMENTS
AGORA_ID	VARCHAR2(15 CHAR)	
S_ORG_EXT_ROW_ID	VARCHAR2(15 CHAR)	
LEP_ADMIN_ROWID	VARCHAR2(15 CHAR)	
ORG_NAME	VARCHAR2(250 CHAR)	
SHORT_CODE	VARCHAR2(50 CHAR)	
ORG_CODE	VARCHAR2(50 CHAR)	
ODS_CODE	VARCHAR2(50 CHAR)	
STATUS	VARCHAR2(50 CHAR)	
ORG_TYPE	VARCHAR2(50 CHAR)	
STAKEHOLDER_TYPE	VARCHAR2(50 CHAR)	
DESIGNATED_BODY_NAME	VARCHAR2(250 CHAR)	
DB_TYPE	VARCHAR2(50 CHAR)	
DB_GROUP	VARCHAR2(20 CHAR)	
DB_IND	VARCHAR2(1 CHAR)	
RESPONSIBLE_OFFICER_UID	VARCHAR2(12 CHAR)	
EMPLOYER_LIAISON_ADVISER	VARCHAR2(150 CHAR)	
CITY	VARCHAR2(50 CHAR)	
POSTCODE	VARCHAR2(50 CHAR)	

COLUMN NAME	DATATYPE	COMMENTS
POSTCODE_OUTCODE	VARCHAR2(30 BYTE)	
POSTCODE_TOWN	VARCHAR2(50 CHAR)	
POSTCODE_COUNTY	VARCHAR2(30 CHAR)	
POSTCODE_COUNTRY	VARCHAR2(30 CHAR)	
COUNTY	VARCHAR2(250 CHAR)	
COUNTRY	VARCHAR2(250 CHAR)	
ODS_NAME	VARCHAR2(250 CHAR)	
ODS_PAR_CODE	VARCHAR2(50 CHAR)	
ODS_PAR_NAME	VARCHAR2(250 CHAR)	
ODS_CITY	VARCHAR2(50 CHAR)	
ODS_POSTCODE	VARCHAR2(50 CHAR)	
ODS_POSTCODE_OUTCODE	VARCHAR2(30 BYTE)	
ODS_POSTCODE_TOWN	VARCHAR2(50 CHAR)	
ODS_POSTCODE_COUNTY	VARCHAR2(30 CHAR)	
ODS_POSTCODE_COUNTRY	VARCHAR2(30 CHAR)	
ODS_COUNTY	VARCHAR2(250 CHAR)	
ODS_COUNTRY	VARCHAR2(250 CHAR)	
SOURCE	VARCHAR2(50 CHAR)	
RO_START_DATE	DATE	
LAT_ORG_CODE	VARCHAR2(50 CHAR)	
POSTCODE_WARD	VARCHAR2(100 CHAR)	

COLUMN NAME	DATATYPE	COMMENTS
POSTCODE_DISTRICT	VARCHAR2(50 CHAR)	
POSTCODE_REGION	VARCHAR2(50 CHAR)	
ODS_POSTCODE_WARD	VARCHAR2(100 CHAR)	
ODS_POSTCODE_DISTRICT	VARCHAR2(50 CHAR)	
ODS_POSTCODE_REGION	VARCHAR2(50 CHAR)	
SUPPRESS_LRMP_REVAL_FLAG	VARCHAR2(1 CHAR)	