Lookups and mapping files – field names and definitions

Version 2 – October 2020

Introduction

Primary Care data in this UK Biobank interim data release contains codes from clinical coding classification systems, but not the text describing the code. Researchers using data provided by UK Biobank will need code lookup tables to obtain text descriptions for the codes used in the data and mapping tables to translate codes between coding systems. Spreadsheets have been provided to help researchers select the codes required to identify sets of patients with specific diseases, conditions, or other features of interest: see accompanying *'all_lkps_maps_v2'* file. This document provides tables for each, showing field names and descriptions. Information sources, copyright information and acknowledgements, variable names and descriptions have also been provided, where possible.

This release (Version 2) has been provided to correct corruptions present in the original file due to Excel reformatting some codes and descriptions as numbers rather than text. The fields in the *'all_lkps_maps_v2'* file have been corrected and manually set to character type.

Contents

BNF table

Information sources used for the BNF lookup table descriptions:

https://ebmdatalab.net/prescribing-data-bnf-codes/

https://digital.nhs.uk/data-and-information/areas-of-interest/prescribing/practice-level-prescribing-in-england-a-summary/practice-level-prescribing-glossary-of-terms

https://www.bnf.org/

BNF codes available for download here, from the NHS Business Services Authority: https://apps.nhsbsa.nhs.uk/infosystems/welcome

bnf_lkp | BNF version: 76

Field Name	Description
BNF_Presentation_Code	The full BNF presentation code for the medication, dressing or appliance. E.g., Yaltormin SR 500mg tablets, an antidiabetic drug with the chemical substance Metformin Hydrochloride, has a BNF code of 0601022B0BPAAAS.
BNF_Chapter	The associated BNF chapter, e.g. the endocrine system. Characters 1 and 2 of the BNF_Presentation_Code denote the chapter. Chapter 6 covers the endocrine system. Hence, this is shown as **06**01022B0BPAAAS.
BNF_Section	Relates to the section of the chapter in the BNF, e.g. drugs used in diabetes. Characters 3 and 4 of the BNF_Presentation_Code denote the chapter section. Section 1 of chapter 6 covers drugs used in diabetes. Hence, this is shown as 06**01**022B0BPAAAS.
BNF_Paragraph	Shows the paragraph of the section in the BNF, e.g. antidiabetic drugs. Characters 5 and 6 of the BNF_Presentation_Code denote the paragraph. Paragraph 2 of section 1 in chapter 6 covers antidiabetic drugs. Hence, this is shown as 0601**02**2B0BPAAAS.

Field Name	Description
BNF_Subparagraph	Relates to the subparagraph of the paragraph, e.g. biguanides. Character 7 of the BNF_Presentation_Code denote the subparagraph. Subparagraph 2 of paragraph 2 in section 1 of chapter 6 relates to biguanides. Hence, this is shown as 060102**2**B0BPAAAS.
BNF_Chemical_Substance	Provides the international non-proprietary name, which is the official, generic and non-proprietary name given to a pharmaceutical drug or its active ingredient. Characters 8 and 9 of the BNF_Presentation_Code denote the chemical substance. The chemical substance for Yaltormin SR 500mg tablets is Metformin Hydrochloride, shown as 0601022**B0**BPAAAS. Note: Chapters 20 to 23 contain no chemical name, as these are dressings and appliances.
BNF_Product	Provides product name which can be proprietary e.g. Yaltormin SR, or the generic variant of the drug. Characters 10 and 11 of the BNF_Presentation_Code denote the product name. Yaltormin SR 500mg tablets are shown as BP (0601022B0**BP**AAAS), as are all other dosages of Yaltormin SR. The generic variant of a drug is always coded as AA.
BNF_Presentation	Provides more information on the product, e.g. whether it is a capsule, tablet or liquid, what the strength of the drug is, its volume (if liquid), sizes for various dressings, etc. Characters 12 and 13 of the BNF_Presentation_Code denote this. Yaltormin SR 500mg tablets are shown as AA (0601022B0BP**AA**AS), Yaltormin SR 750mg are shown as AB, and Yaltormin SR 1000mg as AC. Letters denoting strength do not always refer to the same formulation in other drugs – so AA in this example does not always refer to 500mg tablets for other chemicals. Characters 14 and 15 of the BNF_Presentation_Code relate to equivalent products. To illustrate, characters 14 and 15 of Yaltormin SR 500mg are coded with AS (0601022B0BPAA**AS**). All other products with the chemical substance Metformin Hydrochloride, with a dosage of 500mg, e.g. Meijumet 500mg, will be coded with AS. This structure is useful. If we want to look at the prescribing of all antidiabetic drugs, we can search our database for any items prescribed that begin with 060102. If we want to look at all generics in a paragraph, we can search under BNF_Product for AA.

DM+D table

Information sources used for the DM+D lookup table descriptions:

TRUD files associated with DM+D are available for download from here: https://isd.digital.nhs.uk/trud3/user/guest/group/0/home

TRUD files used, from 'UK SNOMED CT Drug Extension, RF2: Full, Snapshot & Delta' file pack. [Date downloaded: 21/06/2019]

dmd_lkp | dm+d version: May 2019 release

Field Name	Description
concept_id	dm+d code
term	Description of the dm+d code

ICD-9 and ICD-10 tables

Information sources used for the ICD-9 and ICD-10 table descriptions:

TRUD files associated with ICD codes are available for download from here: https://isd.digital.nhs.uk/trud3/user/guest/group/0/home

TRUD files used from 'NHS ICD-10 5th Edition data files' file pack: ICD10_Edition5_CodesAndTitlesAndMetadataFileSpecification_GB_20160401, ICD10_Edition5_TablesOfCodingEquivalencesSpecification(analysis)_GB_20160401. [Date downloaded: 21/06/2019]

Additional files:

NHS Centre for Coding and Classification. Tables of Equivalence A Specification of the File Structure for Tables of Equivalence between ICD-9 and ICD-10 Version 1.1. Sep 1994.

icd9_lkp | ICD-9 version: April 1995 release

Field Name	Description
ICD9	ICD-9 code
DESCRIPTION_ICD9	Description of the ICD-9 code

icd10_lkp | ICD-10 version: 5th edition

Field Name	Description
CODE	ICD-10 Code
ALT_CODE	ICD-10 Code, alternate form: This form strips the decimal point from the code and appends the filler X where the 3 character category is undivided. For example: A00.1 = A001 A46 = A46X A15.0 = A150 I70.00 = I7000

Field Name	Description
USAGE	Dagger / Asterisk indication
USAGE_UK	Dagger / Asterisk indication
DESCRIPTION	Longest preferred rubric
MODIFIER_4	4th character modifier suffix
MODIFIER_5	5th character modifier suffix
QUALIFIERS	Dual classification (asterisk codes)
GENDER_MASK	Gender mask: Identifies single sex conditions
MIN_AGE	Minimum age that applies to this code
MAX_AGE	Maximum age that applies to this code
TREE_DESCRIPTION	This data field also contains descriptions that are functional in the context of the 3 character descriptions of their parent category, and are thus suitable for presentation where that context is available. For example: C00 Malignant neoplasm of lip C00.0 External upper lip C00.1 External lower lip C00.2 External lip, unspecified

icd9_icd10 | ICD-9 version: April 1995 release | ICD-10 version: April 1995 release

Field Name	Description
ICD9	ICD-9 code
DESCRIPTION_ICD9	Description of the ICD-9 code
ICD10	ICD-10 code
DESCRIPTION_ICD10	Description of the ICD-10 code

Read V2 and CTV3 tables

Information sources used for all Read V2 and CTV3 associated table descriptions:

TRUD files associated with Read codes are available for download from here: https://isd.digital.nhs.uk/trud3/user/guest/group/0/home

TRUD files used from 'NHS UK Read Codes Clinical Terms Version 3, Cross Maps' file pack: UK_READ_CTV3_Contents_20180401, UK_READ_Release_Overview_20180401, UK_READ_CTV3_Techupd_20180401, UK_READ_CTV3_Termup_20180401 [Date downloaded: 21/06/2019]

TRUD files used from 'NHS UK Read Codes Version 2' file pack: UKTC_READ_Release_Overview_20160401, UKTC_READ_V2_Techupd_20160401, UKTC_READ_V2_Termup_20160401 [Date downloaded: 22/01/2019]

TRUD files used from 'NHS Read Browser': UK_READ_Browser_Manual_20180401 [Date downloaded: 21/06/2019]

TRUD files used from 'The Read Codes Drug and Appliance Dictionary': UKTC_READ_DRUGS_BNF_20090102, UKTC_READ_DRUGS_Monthly_Release_Update_Report_20160401 [Date downloaded: 21/03/2019]

TRUD files used from the 'NHS Data Migration' file pack: ctv3rctmap_uk_documentation_20190601000001 [Date downloaded: 21/06/2019]

Additional files:

Brear H, Cogman G. The Read Codes Version 2 Definition Codes. NHS Data Standards & Products. 2012 Aug.

Computer Aided Medical Systems Limited. The Read Codes System Developers' Guide. 1993 Jan.

Scottish Clinical Information Management In Practice. Read Terms for General Practice A Quick Introduction. 2014 Oct.

UK Terminology Centre. The Clinical Terms Version 3 (The Read Codes) Cross Mapping File. International Health Terminology Standards Development Organisation UK Terminology Centre. 2008 Apr.

read_v2_lkp | Read V2 version: April 2016 release

Field Name	Description
read_code	Read V2 code
term_code	There is more than one way of describing the same clinical concept. For example, a 'Myocardial Infarction' may be referred to as a 'Heart Attack'. In Read V2, 'term codes' are typically synonyms that are different text descriptions of the same thing. Those marked with '0' denote the preferred term. Those marked with '11', '12' … are synonyms. Over time, some synonyms have been added that have a different meaning to the preferred term. Use of such synonyms where their meaning is different from the preferred term should be avoided.
term_description	Description of the Read V2 code

read_v2_drugs_lkp | Read V2 drugs version: April 2016 release

Field Name	Description
read_code	Read V2 code
term_description	Description of the Read V2 code
status_flag	No information could be sourced for the meaning of this variable

read_v2_icd9 | Read V2 drugs version: April 2016 release | ICD-9 version: April 1995 release

Field Name	Description
read_code	Read V2 code
icd9_code	Mapped ICD-9 code
icd9_code_def	Signifies, e.g., whether or not the Read V2 code matches to a single ICD-9 code, if the ICD-9 code is a parent code that links to several child codes, if the ICD-9 code is asterisk or dagger code, a dagger-asterisk combination, etc.

read_v2_icd10 | Read V2 version: April 2016 release | ICD-10 version: 5th edition

Field Name	Description
read_code	Read V2 code
icd10_code	Mapped ICD-10 code
icd10_code_def	Signifies, e.g., whether or not the Read V2 code matches to a single ICD-10 code, if the ICD-10 code is a parent code that links to several child codes, if the ICD-10 code is asterisk or dagger code, a dagger-asterisk combination, etc.

read_v2_opcs4 | Read V2 version: April 2016 release | OPCS-4 version: 4.7

Field Name	Description
read_code	Read V2 code
opcs_4.2_code	Mapped OPCS-4 code
opcs_code_def	Signifies, e.g., whether or not the Read V2 code matches to a single OPCS-4 code, if the OPCS-4 code is a parent code that links to several child codes, etc.

read_v2_drugs_bnf | Read V2 version: April 2016 release | BNF version: 68

Field Name	Description
read_code	Read V2 code
bnf_code	Associated BNF code. For this particular mapping file, the read_code - bnf_code combination is unique. The bnf_code field contains the code as a set of 8 (4 pairs) numeric characters, separated by ' . ' e.g. 02.05.01.00.

read_v2_read_ctv3 | Read V2 version: April 2016 release | Read CTV3 version: April 2018 release

Field Name	Description
CHAPTER	Read V2 chapter
READV2_CODE	Read V2 code
READV2_DESC	Description of the Read V2 code
TERMV2_DESC	There is more than one way of describing the same clinical concept. For example, a 'Myocardial Infarction' may be referred to as a 'Heart Attack'. In this instance, the term descriptions are typically synonyms that are different text descriptions of the same thing
TERMV2_ORDER	As above, descriptions come in two categories, preferred and synonymous. The preferred term is the accepted official terminology, and the synonymous terms provide alternative definitions to this terminology. In relation to TERMV2_DESC, those marked with '00' denote the preferred term. Those marked with '11', '12' … are synonyms
TERMV2_TYPE	P = preferred term, S = synonym
READV3_CODE	Read CTV3 code
TERMV3_CODE	The five character alphanumeric code for a CTV3 Term. Individual CTV3 concept codes, such as 'X20QM' in Table 2, can have multiple different terms associated with them. Each term is available in 30-, 60- and 198- character variants and each such triad has its own five-character term code (e.g. 'Y21Eu', 'Y21Ev', 'Y21Ew' and 'Y21Ex'). Of all the term codes associated with a given concept code, one is the 'preferred' term for that concept and the others 'synonyms'.
TERMV3_TYPE	P = preferred term, S = synonym
TERMV3_DESC	Description of the Read CTV3 code

read_ctv3_lkp | Read CTV3 version: April 2018 release

Field Name	Description
read_code	Read CTV3 code.
term_description	Description of the Read Code.
description_type	Descriptions come in two categories, preferred (P) and synonymous (S). The preferred term is the accepted official terminology, and the synonymous terms provide alternative definitions to this terminology. To illustrate, the concept of asthma can be described as "Asthma" (this is the default, or preferred term), but can also be described as "Bronchial Asthma" (a synonym).
status	Codes are marked as Current (C), Optional (O) or Extinct (E). No information could be found for values marked as R.

read_ctv3_icd9 | Read CTV3 version: April 2018 release | ICD-9 version: April 1995 release

Field Name	Description
read_code	Read CTV3 code.
icd9_code	Mapped ICD-9 code.
mapping_status	Denotes the nature of the mapping. Contains letters E, G, D, R, A and U. E = Exact one-to-one mapping. There is an exact match between host and target codes. There are no alternatives. G = Target concept more general. The mapping is correct, but Read coded concept is more detailed. There are no alternatives. D = Default mapping. Indicates either an alternative that is most acceptable in the absence of other information or a partial mapping in cases where one Read code maps onto a pair of target codes. There is not necessarily a default among a set of alternatives (other than in OPCS-4 and ICD-10). R = Requires checking. This alternative mapping must be checked against the default that has also been supplied in the table. (Different rubric from default). A = Alternative mapping. All alternatives that are not marked as D or R fall into this category. (Same rubric as default). U = Unspecified. Not used.
refine_flag	Denotes whether or not the target code is sufficiently detailed to be acceptable. 3-character ICD codes are usually not acceptable, for example. Covers addition of 4th and 5th digit extensions in ICD, 4th character in OPCS-4. C = Completely refined. M = Mandatory to refine further. P = Possible but not mandatory to refine further
add_code_flag	Denotes whether or not the target system specifies that extra codes might be added to the target code (e.g. aetiology codes in ICD and site codes in OPCS-4). C = Complete. No further codes need be added. M = Mandatory to add a further code. P = Possible but not mandatory to add a further code.

Field Name	Description
element_num	A Read code may need several target codes for a complete mapping, but each of these may have alternatives. Therefore, each set of alternatives is given a distinct element number. Element numbers start at 0, incrementing by 1. There are rarely more than two sets.
block_num	A block is a complete set of target codes for a mapping from any one Read code (including all alternatives to the suggested codes). There are a number of occasions where more than one block exists for a Read code. This occurs, for example, when a Read code maps either to a single target code or to a target code plus a second code (which may itself have alternatives). Blocks are numbered successively 0, 1, 2... Note that elements exist within blocks and that usually there is only one block.

read_ctv3_read_v2 | Read CTV3 version: April 2018 release | Read V2 version: April 2016 release

Field Name	Description
READV3_CODE	Read CTV3 code
TERMV3_CODE	Read CTV3 term code
TERMV3_TYPE	Descriptions come in two categories, preferred (P) and synonymous (S). The preferred term is the accepted official terminology, and the synonymous terms provide alternative definitions to this terminology. To illustrate, the concept of asthma can be described as "Asthma" (this is the default, or preferred term), but can also be described as "Bronchial Asthma" (a synonym)
TERMV3_DESC	Description of the Read CTV3 code
READV2_CODE	Read V2 code
READV2_DESC	Description of the Read V2 code
TERMV2_ORDER	Similar to the term_code variable in the read_v2_lkp file, those marked with '00' denote the preferred term. Those marked with '11', '12' … are synonyms
TERMV2_TYPE	See TERMV3_TYPE description
TERMV2_DESC	Description of the associated Read V2 preferred or synonymous term.
IS_ASSURED	The current clinical assurance status of each map assertion between a source CTV3 concept and term ID pair and a target READ2 concept and term ID pair, where 0 = Not assured and 1 = Assured. A semi-automatic quality assurance process was agreed with the NHS GP Systems of Choice (GPSoC) programme in October 2010 and subsequently with the Joint GP IT Committee. Under this process, maps are automatically assured if the 5-Byte READ target concept has a legitimate description that is exactly lexically equivalent to at least one of the three string-length variant strings encoded for by the original CTV3 TermID.

read_ctv3_icd10 | Read CTV3 version: April 2018 release | ICD-10 version: 5th edition

Field Name	Description
read_code	Read CTV3 code.
icd9_code	Mapped ICD-10 code.
mapping_status	Denotes the nature of the mapping. Contains letters E, G, D, R, A and U. E = Exact one-to-one mapping. There is an exact match between host and target codes. There are no alternatives. G = Target concept more general. The mapping is correct, but Read coded concept is more detailed. There are no alternatives. D = Default mapping. Indicates either an alternative that is most acceptable in the absence of other information or a partial mapping in cases where one Read Code maps onto a pair of target codes. There is not necessarily a default among a set of alternatives (other than in OPCS-4 and ICD-10). R = Requires checking. This alternative mapping must be checked against the default that has also been supplied in the table. (Different rubric from default). A = Alternative mapping. All alternatives that are not marked as D or R fall into this category. (Same rubric as default). U = Unspecified. Not used.
refine_flag	Denotes whether or not the target code is sufficiently detailed to be acceptable. 3-character ICD codes are usually not acceptable, for example. Covers addition of 4th and 5th digit extensions in ICD, 4th character in OPCS-4. C = Completely refined. M = Mandatory to refine further. P = Possible but not mandatory to refine further
add_code_flag	Denotes whether or not the target system specifies that extra codes might be added to the target code (e.g. aetiology codes in ICD and site codes in OPCS-4). C = Complete. No further codes need be added. M = Mandatory to add a further code. P = Possible but not mandatory to add a further code.
element_num	A Read Code may need several target codes for a complete mapping, but each of these may have alternatives. Therefore, each set of alternatives is given a distinct element number. Element numbers start at 0, incrementing by 1. There are rarely more than two sets.
block_num	A block is a complete set of target codes for a mapping from any one Read Code (including all alternatives to the suggested codes). There are a number of occasions where more than one block exists for a Read Code. This occurs, for example, when a Read Code maps either to a single target code or to a target code plus a second code (which may itself have alternatives). Blocks are numbered successively 0, 1, 2... Note that elements exist within blocks and that usually there is only one block.

read_ctv3_opcs4 | Read CTV3 version: April 2018 release | OPCS-4 version: 4.8

Field Name	Description
read_code	Read CTV3 code.
opcs4_code	Mapped OPCS-4 code.
mapping_status	Denotes the nature of the mapping. Contains letters E, G, D, R, A and U. E = Exact one-to-one mapping. There is an exact match between host and target codes. There are no alternatives. G = Target concept more general. The mapping is correct, but Read coded concept is more detailed. There are no alternatives. D = Default mapping. Indicates either an alternative that is most acceptable in the absence of other information or a partial mapping in cases where one Read Code maps onto a pair of target codes. There is not necessarily a default among a set of alternatives (other than in OPCS-4 and ICD-10). R = Requires checking. This alternative mapping must be checked against the default that has also been supplied in the table. (Different rubric from default). A = Alternative mapping. All alternatives that are not marked as D or R fall into this category. (Same rubric as default). U = Unspecified. Not used.
refine_flag	Denotes whether or not the target code is sufficiently detailed to be acceptable. 3-character ICD codes are usually not acceptable, for example. Covers addition of 4th and 5th digit extensions in ICD, 4th character in OPCS-4. C = Completely refined. M = Mandatory to refine further. P = Possible but not mandatory to refine further
add_code_flag	Denotes whether or not the target system specifies that extra codes might be added to the target code (e.g. aetiology codes in ICD and site codes in OPCS-4). C = Complete. No further codes need be added. M = Mandatory to add a further code. P = Possible but not mandatory to add a further code.
element_num	A Read Code may need several target codes for a complete mapping, but each of these may have alternatives. Therefore, each set of alternatives is given a distinct element number. Element numbers start at 0, incrementing by 1. There are rarely more than two sets.
block_num	A block is a complete set of target codes for a mapping from any one Read Code (including all alternatives to the suggested codes). There are a number of occasions where more than one block exists for a Read Code. This occurs, for example, when a Read Code maps either to a single target code or to a target code plus a second code (which may itself have alternatives). Blocks are numbered successively 0, 1, 2... Note that elements exist within blocks and that usually there is only one block.