User guide for generating bespoke tabulations in Showcase

Introduction

The data Showcase displays a univariate distribution of each data-field currently held within the central database. In some circumstances, researchers have requested cross-tabulations of data-fields to inform them of the numbers of participants in certain categories to guide them either during submission of a main application to use the resource. For example, cross-tabulations would be able to inform you of how many cases of breast cancer or self-reported diabetes there are by age category and by sex.

UK Biobank has developed a cross-tab facility so registered researchers can generate their own cross-tabulations between 2 or 3 data-fields in UK Biobank Data Showcase.

Who can generate cross-tabulations?

Only researchers who are registered with UK Biobank may use the cross-tabulation facility. For more information on how to get registered, please see the 'Register & apply' section on the main UK Biobank website. Currently, this facility is only available at the Main Application stage when you choose the data-fields you want via showcase for your research project.

Which data-fields can be cross-tabulated?

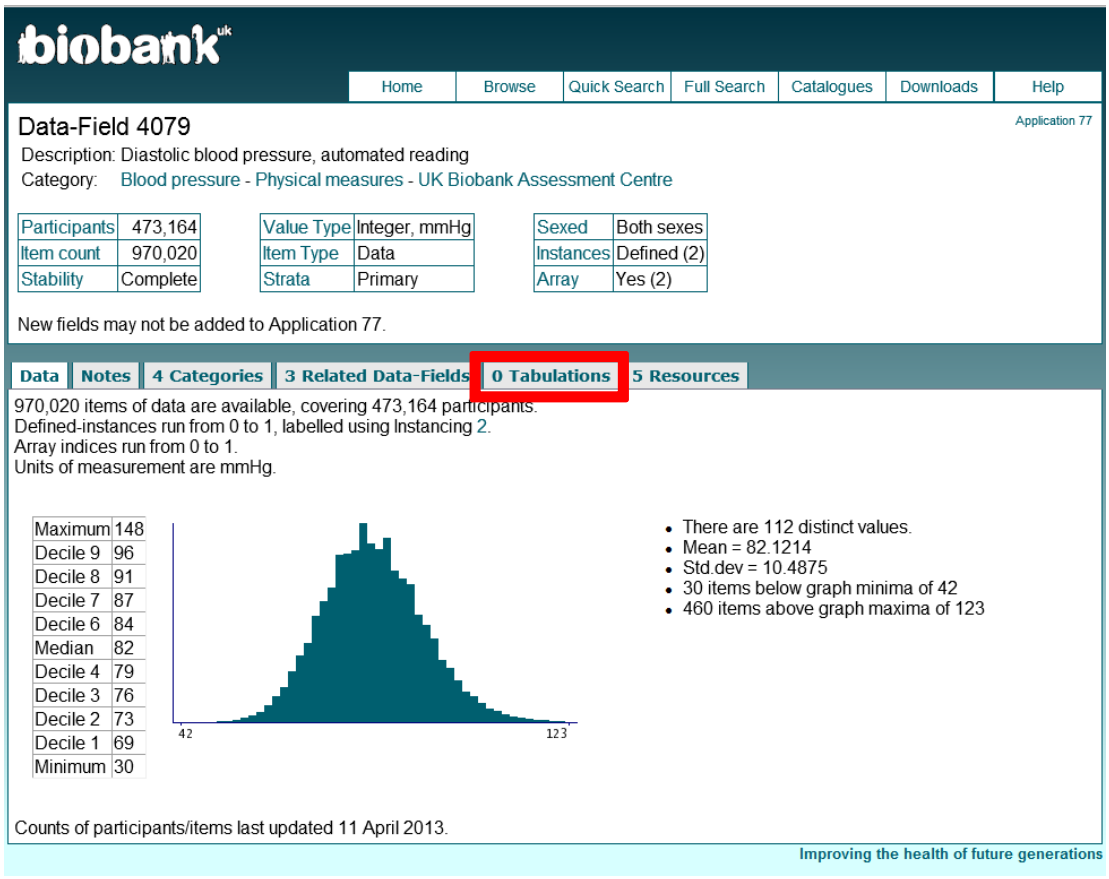
Some data-fields contain a large number of coding values (such as Treatment/medication code (20003), Non-cancer illness code, self-reported (20002), data-fields containing ICD or OPCS codes) and might not be accepted for tabulation or the result tables may be truncated. See Appendix A for a full list of data-fields with a high number of coding values.

If you require a cross-tabulation of a selected disease, procedure or treatment (for example), then please contact the Access Team requesting bespoke frequency tables.

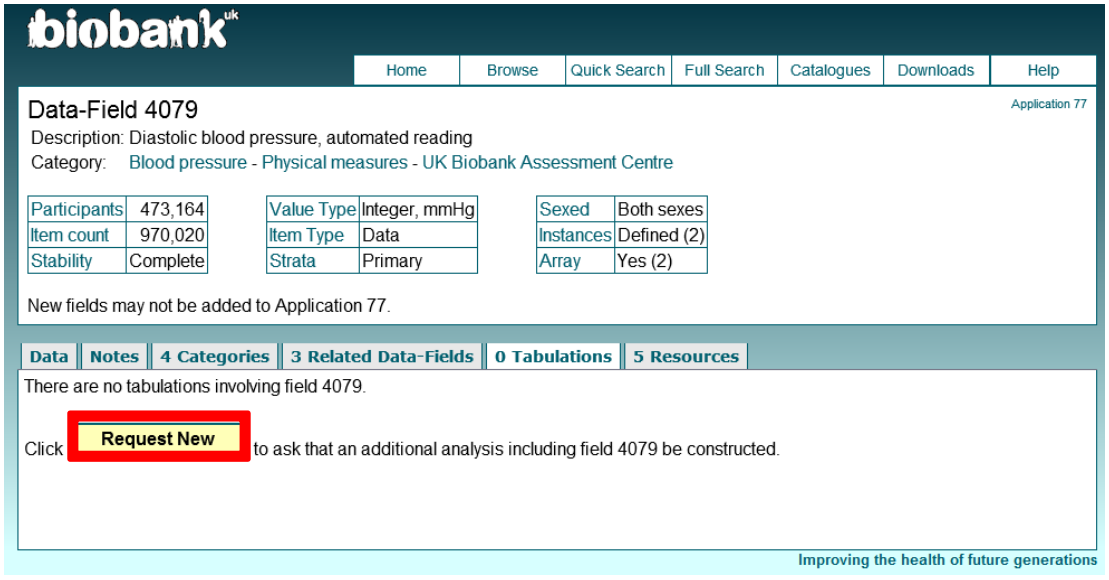
Step-by-step guide to requesting cross-tabulations

(Example: Diastolic blood pressure (4079) by sex (31))

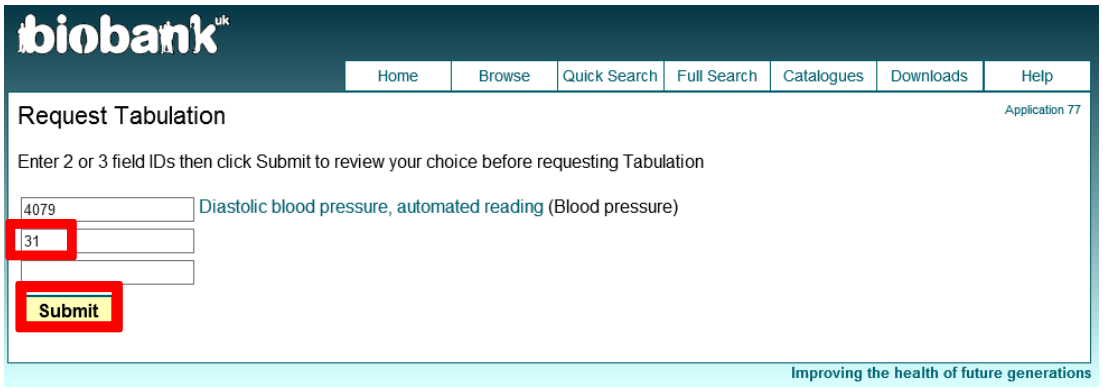
1. Log into the Access Management System (AMS).
2. Open '**Data Showcase**' and navigate to the data-field of interest (e.g. Diastolic blood pressure, automated reading (4079)). [note, currently you can only do this at a Main Application stage].



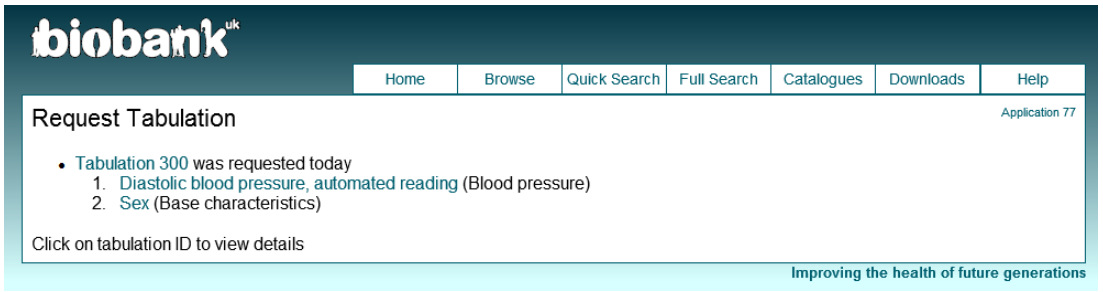
3. Go to the 'Tabulations' tab and select '**Request New**'.



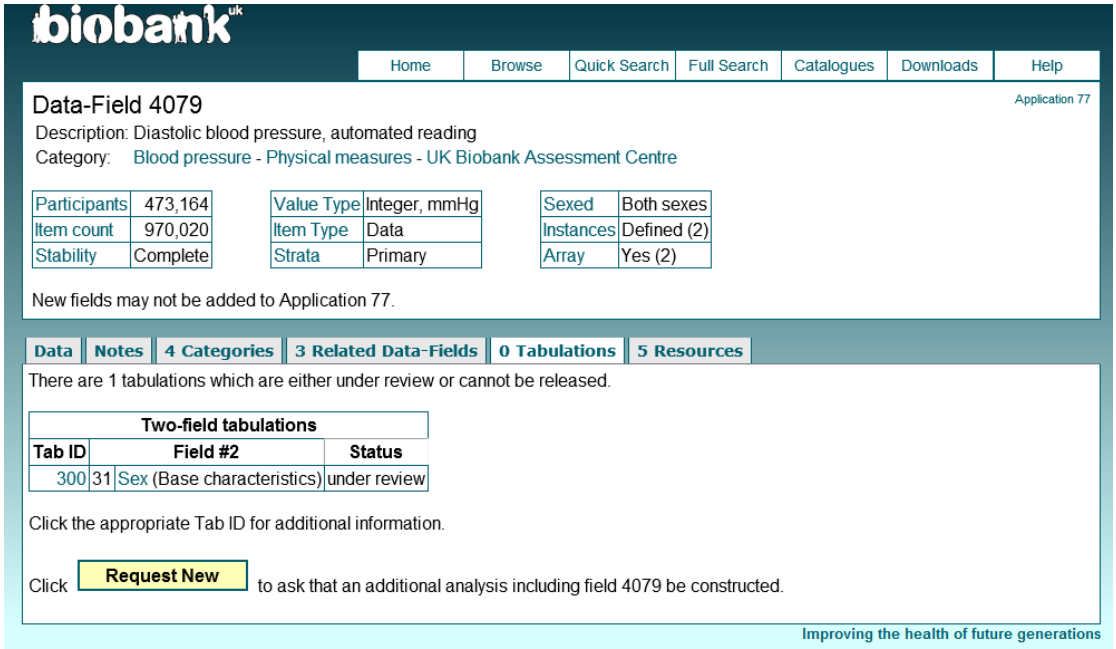
4. Enter the ID of second (and third) data-field to cross-tabulate (e.g. sex (31)) and press '**Submit**'.



5. 'Request Tabulation' screen will appear to acknowledge new tabulation request.



6. Tabulations usually take 1 to 2 working days to be generated and they will appear in the '**Tabulations**' tab in the data-field it was requested from.



Tabulations for numerical data-fields

All numerical data-fields will be converted to categorical for the purpose of tabulation and will be assigned arbitrary display bins by the Data Analysts team. For example 'Age when attended assessment centre' (21003) can be categorised into 5-year bins: <40, 40-45, 45-50, 50-55, 55-60, 60-65, over 65.

If you have a specific requirement of how the categories should be defined, please contact the Access Team quoting your Tabulation ID.

Example of 3x3 contingency tables

Below is an example of cross-tabulation of 3 data-fields:

- Age when attended assessment centre (21003)
- UK Biobank assessment centre (54)
- Sex (31)

For 3x3 tables, the tabulations are spilt into as many tables as many distinct values are there in the data-field with least coding values. In the example below, sex has 2 coding values (Males / Females) so the tabulation is displayed as two tables. The data-field with second least coding values will be displayed in table in a row across, while the data-field with most coding values will be displayed vertically.

The tabulations are showed separately for baseline and repeat assessment.

Tabulation 100

Application 684

This tabulation involves 3 fields

1. Age when attended assessment centre (Base characteristics)
2. UK Biobank assessment centre (Reception)
3. Sex (Base characteristics)

Analysis released today.

522650 data points Show all rows

init.init.all | rep1.rep1.all

- 54, UK Biobank assessment centre : Initial assessment visit (2006-2010) at which participants were recruited and consent given
- 21003, Age when attended assessment centre (years) : Initial assessment visit (2006-2010) at which participants were recruited and consent given
- 31, Sex

273469 data points

Sex: Female						
UK Biobank assessment centre	Age when attended assessment centre (years)					
	<45	45 to <50	50 to <55	55 to <60	60 to <65	65+
Stockport (pilot)	269	321	375	417	387	302
Manchester	1113	1015	1189	1332	1582	1106
Oxford	805	1103	1350	1600	1794	1423
Cardiff	1037	1379	1645	1981	2246	1416
Glasgow	1088	1497	1657	1955	2237	1857
Edinburgh	933	1379	1644	1865	2116	1723
Stoke	934	1225	1510	1922	2324	1656
Reading	1579	2118	2655	3022	3989	2732
Bury	1193	1874	2206	2920	3935	2797
Newcastle	1896	2575	3225	3746	5024	3750
Leeds	2290	3406	3852	4444	5976	4251
Bristol	2978	3370	3752	4344	5597	3988
Barts	941	1146	1217	1196	1418	978
Nottingham	1702	2320	2764	3523	4756	3414
Sheffield	1404	1998	2488	2945	4277	3195
Liverpool	1530	2267	2616	3245	4562	3494
Middlesborough	1113	1436	1713	2115	2961	2132
Hounslow	1860	2205	2483	2822	3734	2788
Croydon	1535	2179	2546	2779	3661	2765
Birmingham	1519	1836	2141	2339	3143	2487
Swansea	91	154	202	185	319	286
Wrexham	32	50	51	71	82	65

229182 data points

Sex: Male						
UK Biobank assessment centre	Age when attended assessment centre (years)					
	<45	45 to <50	50 to <55	55 to <60	60 to <65	65+
Stockport (pilot)	225	220	273	338	335	336
Manchester	971	915	1024	1214	1334	1146
Oxford	638	773	919	1087	1332	1239
Cardiff	872	1078	1251	1664	1882	1433
Glasgow	908	1188	1293	1475	1787	1711
Edinburgh	652	1024	1225	1412	1715	1514
Stoke	977	1216	1447	1838	2431	1960
Reading	1381	1741	1963	2404	3189	2649
Bury	1212	1683	1874	2375	3450	2803
Newcastle	1611	2088	2459	2980	4208	3447
Leeds	2055	2632	2919	3612	4701	4078
Bristol	2386	2446	2747	3270	4391	3748
Barts	852	959	934	941	1171	831
Nottingham	1412	1798	2040	2670	4025	3459
Sheffield	1212	1679	1877	2305	3729	3288
Liverpool	1306	1841	2052	2488	3868	3556
Middlesborough	975	1133	1330	1626	2603	2152
Hounslow	1610	1635	1837	2204	3041	2661
Croydon	1193	1550	1737	1915	3086	2442
Birmingham	1395	1503	1702	2030	2762	2649
Swansea	102	114	136	178	268	248
Wrexham	30	30	35	52	79	72

Appendix A: Data-fields with high number of coding values

Field ID	Description
132	Job code
20024	Job code - deduced
40004	Last known occupation
20003	Treatment/medication code
20004	Operation code
20002	Non-cancer illness code, self-reported
40001	Underlying (primary) cause of death: ICD10
40002	Contributory (secondary) causes of death: ICD10
40006	Type of cancer: ICD10
41201	External causes
41202	Diagnoses - main ICD10
41204	Diagnoses - secondary ICD10
40011	Histology of cancer tumour
40013	Type of cancer: ICD9
41203	Diagnoses - main ICD9
41205	Diagnoses - secondary ICD9
20115	Country of Birth (non-UK origin)
41229	PCT responsible for patient data
41230	PCT where patients GP was registered
41200	Operative procedures - main OPCS
41210	Operative procedures - secondary OPCS
41246	Treatment speciality of consultant (recoded)