

4 Using metadata to discover relevant variables

Researchers need better access to information about data so as to be able to discover variables relevant to their research and to find variables that, although collected through different sources, are nonetheless comparable.

Whilst the standards used and the information available has improved significantly over the years, the information provided to data repositories and archives has not (in general). It continues to require large amounts of effort on the part of researchers to turn it into high quality research resources. One reason is because information about the data collection is very often detached from the datasets made available for research, and high quality data requires them to be better connected.

4.1 Discovering relevant variables

The [CLOSER Discovery](#) search engine makes it easier for researchers to explore some of the most used longitudinal studies in the UK, by tying together rich layers of information about the data collected.

It is a user-friendly resource for locating the variables that best suit your research interests.

A search on CLOSER Discovery will show you:

- relevant survey questions and variables, which you can filter by study, topic and life stage
- the original question the variables are based on
- where questions appeared in the survey, including complete routing
- the sources of derived variables (but not the derivation methodology)
- an overview of each variable, including valid and invalid cases; maximum, minimum and mean values; and in many cases, a frequency count
- how to access the data from each study.

You can save information for questions and variables of interest to a List, and download it for your reference.

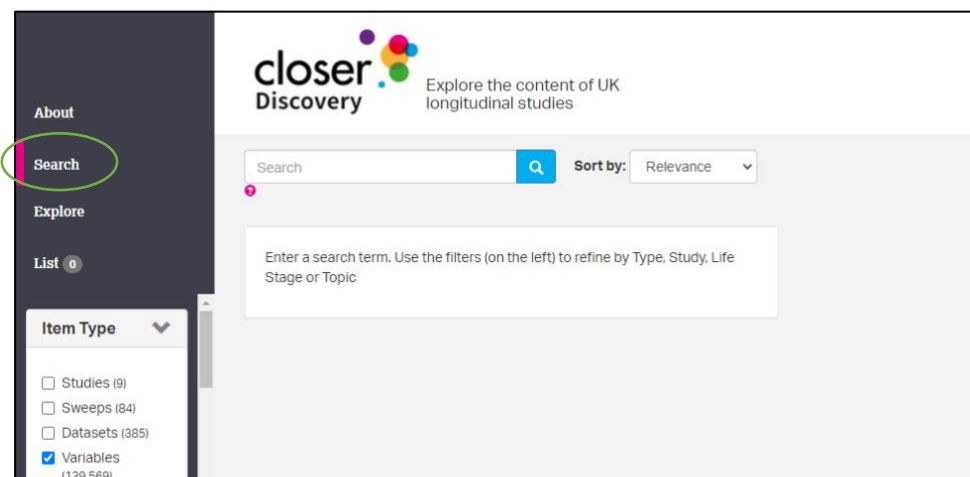


4.2 Using 'search' and 'explore' to find relevant variables

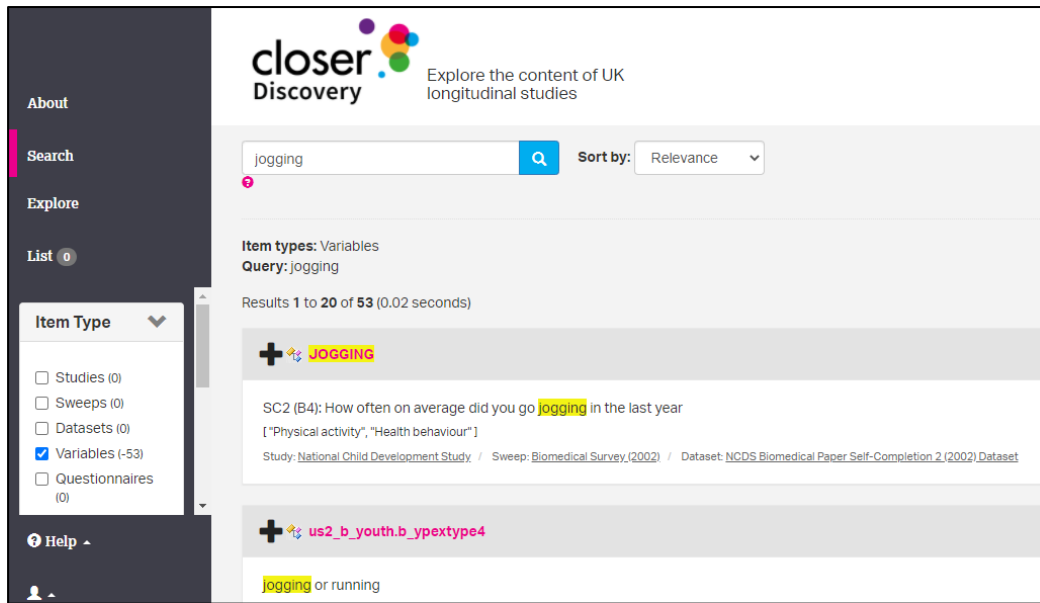
CLOSER Discovery enables you to search for variables relevant for your research. The search engine allows you to search not only for variable names and labels, but you can also search entire questionnaires to look for words or phrases contained in question texts.

There are two main ways of doing this, briefly outlined below:

4.2.1 Finding a variable via search



Clicking on the search menu takes you to a page where you can enter a text-based query to look for variables containing the word(s) in the search term.



In this example, we are looking for variables containing the word 'jogging'

The search returns 53 results

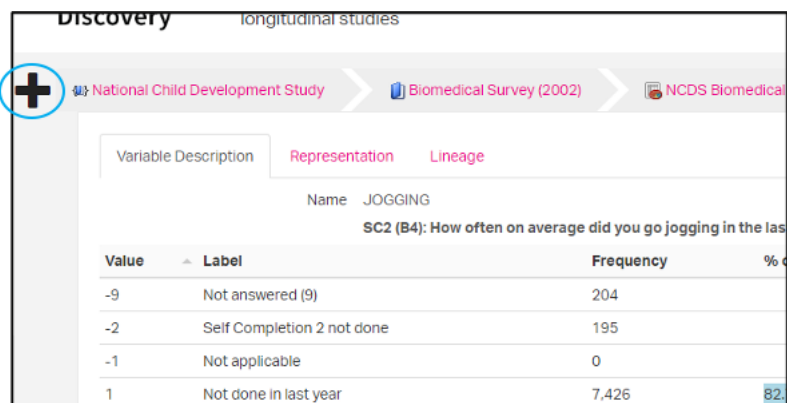
The screenshot shows the details for the variable 'JOGGING'. The variable description is 'SC2 (B4): How often on average did you go jogging in the last year'. The table below shows the frequency and percentage of responses for each value.

Value	Label	Frequency	% of valid	% of all
-9	Not answered (9)	204		0.00%
-2	Self Completion 2 not done	195		0.00%
-1	Not applicable	0		0.00%
1	Not done in last year	7,426	82.71%	0.00%
2	Less than once a month	509	5.67%	0.00%
3	Once a month	196	2.18%	0.00%
4	2 to 3 times a month	175	1.95%	0.00%
5	Once a week	226	2.52%	0.00%
6	2 to 3 times a week	343	3.82%	0.00%
7	4 to 5 times a week	79	0.88%	0.00%
8	Every day	24	0.27%	0.00%

Clicking on the first of the results you can see which study and dataset the variable comes from (1).

It also provides a description of the variable, including the wording of the question (2), missing value codes (3) and frequency counts (4) and percentages (5) of responses

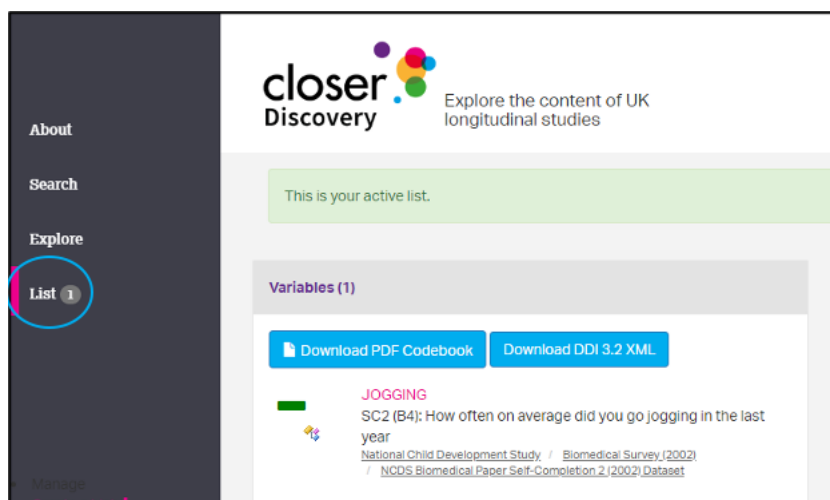
Authors: Neil Kaye, Hayley Mills and Jon Johnson



The screenshot shows the 'Discovery' interface for longitudinal studies. At the top, there are navigation arrows and a '+' icon circled in blue. Below this, there are tabs for 'Variable Description', 'Representation', and 'Lineage'. The 'Variable Description' tab is active, showing the variable name 'JOGGING' and its description: 'SC2 (B4): How often on average did you go jogging in the last year'. Below this is a table with columns for 'Value', 'Label', 'Frequency', and '% of total'.

Value	Label	Frequency	% of total
-9	Not answered (9)	204	
-2	Self Completion 2 not done	195	
-1	Not applicable	0	
1	Not done in last year	7,426	82.1

You can add this variable to your saved List by clicking the '+' symbol in the top-left...

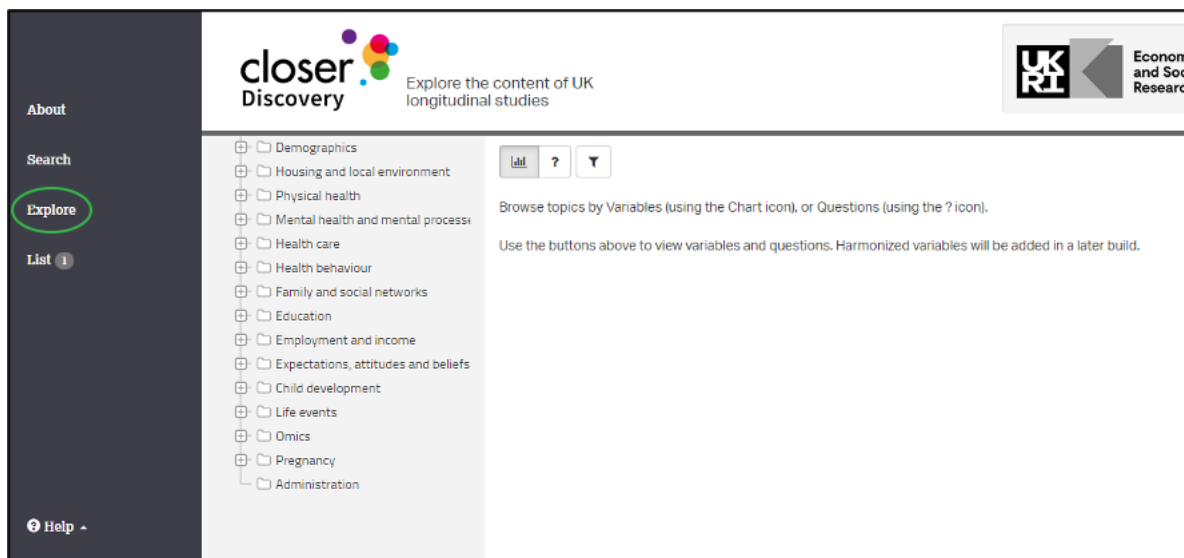


The screenshot shows the CLOSER Discovery interface. On the left is a dark sidebar menu with options: 'About', 'Search', 'Explore', 'List' (circled in blue), and 'Manage'. The main content area features the CLOSER Discovery logo and the tagline 'Explore the content of UK longitudinal studies'. Below this is a green box stating 'This is your active list.' and a section titled 'Variables (1)'. This section contains two buttons: 'Download PDF Codebook' and 'Download DDI 3.2 XML'. Below the buttons is a variable entry for 'JOGGING' with its description and source information.

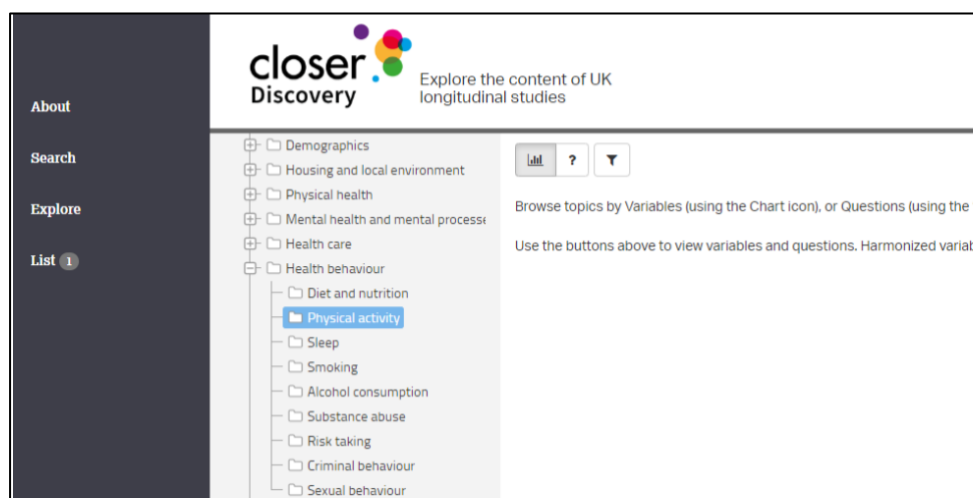
... and you can access your List from the menu on the sidebar.

Suggested citation: Kaye, N., Mills, H. & Johnson, J. (2020). *Understanding metadata*. CLOSER Learning Hub, London, UK: CLOSER

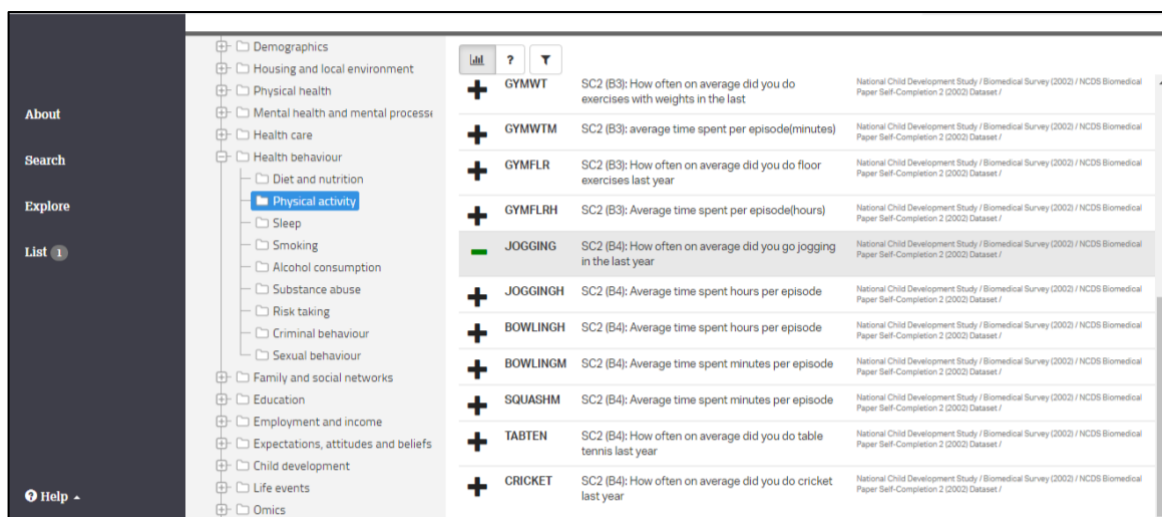
4.2.1 Finding relevant variables via explore



Another way of looking for variables using CLOSER Discovery is by exploring the datasets by theme, via the explore menu



Using the 'tree' menu we can explore topics by category and sub-category to home in on our particular focus – in this case 'Physical activity'



By scrolling through variables relevant to the topic, we can pick out those we are interested in, click into them to view a detailed description, and add them to our saved list

4.3 How-to guides

You can find out more about how to use CLOSER Discovery through the [FAQs](#) and how-to guides. The guides provide step-by-step instructions to help users find their way around the search engine.

[How do I find variables from similar life stage across different studies?](#)

This guide explains how researchers can use CLOSER Discovery to search for cross-study cross-sectional data, such as asthma in early childhood across all studies.

[How do I find information about a variable from a particular study?](#)

This guide shows how to use CLOSER Discovery to explore within-study cross-sectional data, for example discover how to find information on the sex of participants in the Millennium Cohort Study.

[How do I find how a variable was asked in the original questionnaire?](#)

Authors: Neil Kaye, Hayley Mills and Jon Johnson

This set of instructions demonstrate how CLOSER Discovery can be used to find information about a source question and the wider questionnaire, for instance search the questions used to measure alcohol intake in the CLOSER studies.

[How do I compile and save a list of variables?](#)

These step-by-step instructions explain how users can create and save lists in CLOSER Discovery for example how to compile and export a list of the diabetes-related variables collected by the Southampton Women's Survey.