Data sources, approaches to data collection and data collection timeline																		
Data source	Notes on data sources	Researcher(s) involved in data collection	Approaches to data collection	Notes on approaches to data	Timeline													
					Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18
Drug company websites	1	ER and PO independently	Manual scraping of online information, deductive and inductive coding	4,5	■							■						
Patient organisation websites	1			6,7		■	■	■					■	■	■			
Charity regulator websites	2			8,9					■	■			■	■	■	■		■
WHO ICD-10 website	3	PO	Manual scraping of online information	10,11				■	■						■	■		

Notes

1   Two of the main data sources, drug company and patient organisation websites, are not static and therefore may be open to change as organisations update or improve their website over time. The repeated independent web searches carried out by ER and PO were meant to ensure that the data collected was as accurate and up to date as possible. However, it can only be guaranteed that all extracted information was correct at the time of data collection.

2   The charity records held by the three charity regulators (the Charity Commission for England and Wales, the Scottish Charity Regulator, and the Charity Commission for Northern Ireland) are subject to change as some organisations have their accounts added or removed from the charity regulators' databases. Similarly, the details of their registration may change as a result of, for example, alterations in the organisation's governing documents (e.g. following a merger with another organisation).

3   ICD-10 Version:2010 website was used to collect data on the disease areas of patient organisations identified through drug company payment disclosure reports.

4   ER carried out an initial online search of drug company websites in June 2017 using the terms "[Drug company name]" AND "patient organisation/patient organization/patient group". In the same period, ER also carried out supplementary searches on the websites of companies for which no results were found using the Google search engine to ensure that no relevant payment disclosure report was missed. The same search strategy was subsequently repeated in January 2018 because the number of disclosure reports covering 2016 identified through the initial search was low, particularly when compared with the number of reports covering 2015. Carrying out the subsequent round of searches allowed for capturing a greater number of payment disclosure reports from companies that published them later than other companies.

5   Upon identifying relevant disclosure reports on drug company websites ER manually copied and pasted all relevant data to an Excel database, including, in particular, drug company name, recipient names and payment descriptions. Once the data entry was completed ER checked all entries in the spreadsheet twice against the original documents. The searches, data extraction and checking took place on a daily basis during the data collection periods. Once the data entry was completed PO cleaned the data (e.g. standardising patient organisation names, removing unnecessary spaces from payment descriptions). PO then coded the data deductively (e.g. VAT included in payment value or not) or inductively (e.g. payment goal - codes derived via iterative reading of payment descriptions). The validity of the coding was ensured via two rounds of checks, carried out at different points in time, involving the reading of payment descriptions linked with the same codes to ensure that they are sufficiently sematically similar; and comparing payment descriptions associated with different codes to ensure that they are sufficiently semantically different.

6   ER and PO carried out independent online searches of patient organisation websites in July-September 2017 using patient organisation names taken from drug company payment disclosure reports. ER and PO then repeated the searches in February-April 2018 so as to check the accuracy of the data entered during the first round of searches, address any missing data, and add information on patient organisations identified in payment disclosure reports found through the second round of searches on drug company websites.

7   In carrying out the searches of patient organisation websites, ER identified relevant data on the websites, manually copied and pasted key passages into an Excel spreadsheet, and coded the data deductively (e.g. membership - present or absent) or inductively (e.g. organisational goal - codes derived via iterative reading of the copied passages from organisational websites). PO then repeated this process independently. Any differences in data entry and coding were resolved through discussion. ER and PO conducted the searches and data entry on a daily basis throughout the search periods. PO checked the validity of coding in the same way as in relation to the coding of data from drug company payment disclosure reports.

8   ER and PO carried out independent online searches of the websites of the three charity regulators (the Charity Commission for England and Wales, the Scottish Charity Regulator and the Charity Commission for Northern Ireland) in October-November 2017 using the names of patient organisations as stated on their websites as well as charity regulator numbers if they were provided on the websites. ER and PO then repeated the searches in February-May 2018. Finally, ER carried out a small number of follow-up searches in July 2018 to establish whether the registration status of some of the organisations had changed.

9   In carrying out the searches of the charity regulator websites, ER identified relevant data on the websites, copied and pasted key passages into an Excel spreadsheet, and coded the data deductively (e.g. organisation included or not in the regulator's database; organisation form) based on iterative reading of the copied passages. PO then repeated this process independently. Any differences in data entry and coding were resolved through discussion. ER and PO conducted the searches and data entry on a daily basis throughout the search periods.

10   PO carried out searches on the ICD-10 Version:2010 website using the areas of interest provided by patient organisations on their websites in September-October 2017 and then April-May 2018.

11   For searches returning results on the ICD-10 Version:2010 website, PO copied into an Excel database the corresponding highest-level condition areas (ICD-10 chapters, indicated by the Roman numerals, e.g. Chapter II Neoplasms) and groups of conditions (e.g. C64-C68). PO conducted the searches and data entry on a daily basis throughout the search periods.