Backdating of events in electronic primary health care data: should one censor at the date of last data collection.

Cormac Sammon[1] and Irene Petersen[1, 2]

1. Department of Primary care and Population Health, University College London, London, UK
2. Department of Clinical Epidemiology, Aarhus University, Denmark

Corresponding Author:

Cormac Sammon
Department of Primary care and Population Health,
University College London,
Royal Free Campus, Rowland Hill street
London, UK
Telephone: +44 (0)20 7794 0500 ext (31035)
Fax: +44 (0)20 7472 6871
c.sammon@ucl.ac.uk

- There is often a delay between the occurrence of medical events and the event being entered in a patient's primary care record.
- In studies using primary care databases, these delays will result in censoring of events towards the end of follow up.
- The problem is more apparent for events typically presenting/diagnosed outside primary care.
- We illustrate a graphical method to identify the problem and determine a point prior to the end of data collection to censor follow up at.

Word count: abstract: 247/250
body text: 2353/3000

Conflict of interest statement: CS has worked on a study funded by Novartis Vaccines and Diagnostics

This work has been presented at the 31st International Conference of Pharamcoepidemiology and Drug Safety, Boston, MA, August 2015

Abstract

Purpose: Studies using primary care databases often censor follow up at the date data is last collected from clinical computer systems (LCD). We explored whether this results in the selective exclusion of events entered in the electronic health records after their date of occurrence i.e. backdated events.

Methods: We used data from The Health Improvement Network (THIN). Using two versions of the database we identified events that were entered into a later (THIN14), but not an earlier version of the database (THIN13) and investigated how the number of entries changed as a function of time since LCD. Times between events and the dates they were recorded were plotted as a function of time since the LCD in an effort to determine appropriate points at which to censor follow up.

Results: There were 356 million eligible events in THIN14 and 355 million eligible events in THIN13. When comparing the two datasets, the proportion of missing events in THIN13 was highest in the month prior to the LCD (9.6%), decreasing to 5.2% at 6 months and 3.4% at 12 months. The proportion of missing events was largest for events typically diagnosed in secondary care such as neoplasms (28% in the month prior to LCD) and negligible for events typically diagnosed in primary care such as respiratory events (2% in the month prior to LCD).

Conclusions: Studies using primary care databases, particularly those investigating events typically diagnosed outside primary care, should censor follow up prior to the LCD to avoid underestimation of event rates.

Keywords: primary care, database, CPRD, THIN, censoring, epidemiology, pharmacoepidemiology

Introduction

Primary care databases such as the Clinical Practice Research Data Link (CPRD) and The Health Improvement Network (THIN) in the UK have become increasingly popular data sources for epidemiological and health care research.[1] However, as such databases are based on data that are entered for clinical management, rather than research, further work is needed to investigate best practice in the conduct and reporting of research carried out using them.[2]

As suggested by its name, primary care is commonly the initial point of contact between individuals and their healthcare system. As a result, a large amount of information about a patient is collected in primary care computer systems. Events which occur outside primary care, such as diagnoses made in secondary care may be reported to primary care after they have occurred e.g. via hospital discharge letters. Primary care computer systems typically allow such records to be entered into a patient's record alongside a backdated date of event occurrence.

When carrying out research using primary care databases, in the absence of a competing event (e.g. death, moving healthcare provider etc.), the date data is last uploaded from clinical computer systems to the database provider is typically used to censor follow up. As illustrated in Figure 1, censoring at the dates data are last collected from clinical computer systems, means that some events which have occurred may not yet have been reported to the primary care practice and entered into the clinical computer systems at the time of data collection. As a consequence the incidence and prevalence of specific diseases/conditions may be underestimated as follow up approaches the censoring date.

In this study we used a UK primary care database to examine the extent to which censoring at the date data is last collected results in the selective exclusion of backdated events and investigate a generally applicable method to identify and mitigate the problem in research using electronic health records.

Methods

Data source

We used data from The Health Improvement Network (THIN), which collects data from general practices (family practices) in the United Kingdom using Vision Software[3]. THIN is an electronic healthcare database containing the anonymised primary care medical records of more than 12 million individuals, more than 3.5 million of whom are currently actively providing data. Practice data available in the database includes the date a practice first computerised their records and the date a practice started to use Vision software. As data collection is ongoing, updated versions of the database are created on a monthly basis. Each time a new version is created information on when data was last collected from each practice (last collection date; LCD) is updated. Additional practice level metrics can be derived and used to define the date a practice begins recording acceptable levels of patient data.[4,5]

Patient data routinely available in the database include registration and demographic details, diagnoses and symptoms including those leading to hospital admissions, immunisations, pregnancies, laboratory tests, referrals to specialists, prescriptions issued by the General Practitioner, hospital discharge and clinic summaries and deaths. Clinical events and other patient characteristics in THIN are recorded as Read codes which are structured in a medical records table. There are currently over 100,000 Read codes each of which is associated with a short description of varying specificity. Read codes are arranged into mono-hierarchical chapters representing different clinical areas (Box 1).[6,7]

In general practices using Vision software, a system date is automatically entered on the computer when an event is recorded in a patient's electronic health record. However, the practice staff creating the record can enter a date on which the event they are recording is deemed to have occurred, called the event date. As outlined in the introduction this facilitates the backdating of events which occur outside general practice, for example, events that occurred before a patient registered with the general practice or events presenting in secondary or tertiary care such as cancer diagnoses.

In order to assess whether censoring at LCD results in the exclusion of events from the primary care record, this study took the novel approach of utilising two versions of the database, created approximately one year apart. In the first dataset (THIN 13) data lock occurred in January 2013 therefore the maximum possible LCD was in January 2013, while in the second dataset (THIN 14) data lock occurred in January 2014 therefore the maximum possible LCD was in January 2013. By excluding any events from both databases where the event date is greater than the LCDs recorded in THIN 13, we were able to carry out our analysis based on the premise that THIN 13 contains the data typically available to a research team conducting a study in a given year, and that the remaining THIN 14 data now simply represents an updated version of THIN 13 in which many events with an event date earlier than, but a system date greater than the THIN 13 LCD can now be observed. (Figure 2).

In both databases we focused on records which were made after 1 January 2000 and event dates which were at least 365 days after a patient had registered with the general practice. Further we only used information recorded after a practice had started using Vision software for patient management and after data recording in a practice reached acceptable levels.

Chapter	Contents
0	Occupations
1	History/symptoms
2	Examination/signs
3	Diagnostic procedures
4	Laboratory procedures
5	Radiology/medical physics
6	Preventative procedures
7	Operations and procedures
8	Other therapeutic procedures
9	Administration
A	Infectious/parasitic diseases
B	Neoplasms
C	Endocrine/metabolic
D	Blood diseases
E	Mental disorders
F	Nervous system/senses
G	Circulatory system
H	Respiratory system
J	Digestive system
K	Genito-urinary system
L	Pregnancy/childbirth/puerperium
M	Skin/subcutaneous tissue
N	Musculoskeletal
P	Congenital anomalies
Q	Perinatal conditions
R	Ill-defined conditions/working diagnoses
S	Injury/poisoning
T	Causes of injury/poisoning
U	External causes of morbidity and mortality
Z	Unspecified conditions

Box 1. Read codes are structured in a mono-hierarchical system within the following chapters.[6,7]

Data analysis

First, we calculated the total number of events recorded in the medical records table of THIN 13 and THIN 14 using a code from any of the Read code chapters (Box 1). As THIN 14 is considered the more complete dataset, we estimated the proportion of missing records in THIN 13 as the difference in the number of eligible events in THIN13 and THIN14 and expressed this as a percentage of the total number of eligible events in THIN 14. This percentage was described as a function of time between the event date and the THIN 13 LCD.

Subsequently, we calculated the same statistics for each individual Read code chapter and present the results for events from chapters B (neoplasms), G (circulatory system) M (skin/subcutaneous tissue) and H (respiratory system). As events in chapters B and G typically relate to cancer and cardiovascular diagnoses we suspect these are more likely to occur in secondary or tertiary care than events in chapters M and H, which are often diagnoses that can be made in the GP practice.

We then tested a graphical method that could potentially be used to identify and mitigate the problem in a normal study setting where two datasets are not typically available for comparison. This method is based on the assumption that at a certain point prior to the LCD (e.g. 5 years) the proportion of backdated events that are censored becomes negligible. The distribution of time between system and event dates for events occurring more than 5 years before the LCD can then be used to determine the probability that an event would be recorded at different points prior to the LCD. By choosing an 'acceptable' probability one can determine a point prior to the LCD at which to censor follow up. Cumulative incidence plots will be used to describe the time between system and event dates for events occurring greater than 5 years before the LCD for each of the four Read code chapters, B, G, H and M.

All data management and analysis was carried out in Stata 13.

Results

There were 355.96 million eligible events recorded in the THIN 13 dataset and 356.86 million events in the THIN 14 dataset during the same time period, a difference of 901,955 events. The proportion of events that were missing in THIN 13 was highest in the month prior to the LCD (9.6%) decreasing to 5.2% in the 6 months prior and 3.4% in the twelve months prior (Table 1). Expressed as a proportion of the overall number of events recorded in THIN 14 during the entire study period, 0.01%, 0.1% and 0.1% of codes were missing at 1, 6 and 12 months prior to the LCD.

The proportion of entries from specific Read code chapters differed. Thus, we observed that 27.7% of neoplasm events were missing in the first month prior to the LCD in THIN 13 (Table 1, figure 3). At 6 months and 12 months prior to the LCD the proportion of missing neoplasm records was 7.6% and 4.5% respectively. For circulatory system events the figures were 18.2%, 6.8% and 4.5%. In contrast, relatively few skin condition and respiratory events were missing in the same time periods (Table 1, figure 3).

Expressing these Read code chapter specific results as a proportion of the overall number of events in each of these chapters during the study period, the proportion of missing events at 1, 6 and 12 months prior to the LCD respectively was: 0.2%, 0.4% and 0.4% of neoplasm events and 0.1%, 0.2% and 0.3% of circulatory system events.

Figure 4 contains Read code chapter specific plots showing the distribution of time between event and system dates for first events occurring more than 5 years prior to the LCD. Among events occurring more than 5 years ago, 70% of circulatory system events, 67% of neoplasm events, 93% of skin events and 92% of respiratory events have the same event and system date, 81%, 84%, 96% and 97% of these events had <30 days difference between their event and system dates and 85%, 90%, 97% and 98% of events had <120 days difference between their event and system dates. Table S1 (supplementary data) shows the 50^{th}, 75^{th}, 90^{th} and 95^{th} percentile of time between the system and event dates for different Read codes (grouped at 2 digits); considerable variation within Read code chapters is apparent.

Discussion

This study illustrates that when using UK primary care databases, censoring at the last collection date may result in the selective exclusion of events that occur before, but are recorded after the last collection date. This appears to be a particular issue for events typically presenting/diagnosed in secondary care such as cancers and circulatory system events.

The main implication of these findings is that studies are likely to underestimate incidence and prevalence rates towards the end of data collection. The extent to which rates are underestimated will depend on what proportion of the numerator the missing events represent; the impact is likely to be largest for rates estimated for shorter periods near the end of data collection. Studies which compare incidence rates in two groups as a means of determining the safety or effectiveness of a treatment must consider whether censoring is informative (i.e. whether any underestimation is more likely to affect one group than another). For example, if individuals exposed to a drug 'X' of interest were under the care of a specialist and therefore more likely to present to them with an adverse

event, while unexposed individuals or those exposed to a drug 'Y' were not under the care of a specialist and therefore more likely to present directly to their GP with the same event, then it is more likely that there will be a difference in the system and event dates of adverse events in those exposed to drug X.

We suggest that the most suitable approach to deal with the issue is to determine an appropriate amount of time prior to the LCD to censor follow up at. The appropriate amount of time could be determined in a number of ways, most simply by carrying out analyses which test the sensitivity of results to censoring at a variety of time points (e.g. 1 month, 6 months and 12 months) prior to the LCD. We have illustrated an alternative, relatively simple graphical approach to determine an appropriate amount of time prior to the LCD to censor at using a single dataset. This method uses the distribution of times between event and system dates for events recorded in a period relatively unaffected by the censoring. This plot can be used to estimate the probability that an event will be observed at different points prior to the LCD, by choosing a cut-off at which the probability is deemed acceptable one can determine a period prior to the LCD at which to censor. For example, the probability of a respiratory or skin event not being recorded is less than 10% at any time prior to the LCD for respiratory and skin events but for neoplasm and circulatory events the probability an event is not recorded is ~30% on the day of the LCD, 15% and 18% respectively a month prior to the LCD and 10% and 14% respectively in the three months prior to the LCD. A limitation of this approach is that it assumes there are no secular trends in recording delays.

This study sought to provide a broad investigation of the issue across the entire database and therefore results are primarily presented for entire Read code chapters. As illustrated in Table S1, the relationships between system and event dates may vary for subsets of events within each of these chapters. For example, cancers which can be detected more rapidly or that are managed in general practice may have shorter durations between event and system dates than those which are more complicated. We would therefore suggest that while our results in Table S1 can be used as a guide, in specific studies the diagnostic plot described above should be created for the specific events under study.

While this study has focused on primary care databases, other electronic healthcare databases in which events can be backdated may also suffer from missing events in the period approaching the end of data collection.

Diagnostic plots such as those described by Lewis et al[8] have become a routine step in designing electronic healthcare database studies, our results suggest that diagnostic plots of time between event and system dates as a function of time since LCD should also be incorporated as a routine step in this phase of studies conducted in UK primary care databases, and potentially other databases which allow for the backdating of events. At a minimum, studies using these databases to investigate diagnoses typically made in secondary or tertiary care should test the sensitivity of their results to censoring at a series of time points earlier than the LCD.

References

1. Springate DA, Kontopantelis E, Ashcroft DM, *et al.* ClinicalCodes: An Online Clinical Codes Repository to Improve the Validity and Reproducibility of Research Using Electronic Medical Records. *PLoS ONE* 2014; **9**: e99825. doi:10.1371/journal.pone.0099825.

2. Sørensen H, Langan, Benchimol, *et al.* Setting the RECORD straight: developing a guideline for the REporting of studies Conducted using Observational Routinely collected Data. *Clin Epidemiol* 2013: 29. doi:10.2147/CLEP.S36885.

3. Vision. Available at: http://www.inps4.co.uk/vision. Accessed June 26, 2015.

4. Maguire A, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf* 2009; **18**: 76–83. doi:10.1002/pds.1688.

5. Horsfall L, Walters K, Petersen I. Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiol Drug Saf* 2013; **22**: 64–69. doi:10.1002/pds.3368.

6. Davé S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf* 2009; **18**: 704–707. doi:10.1002/pds.1770.

7. Chisholm J. The Read clinical classification. *BMJ* 1990; **300**: 1092.

8. Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2005; **14**: 443–451. doi:10.1002/pds.1115.

Table 1. Numbers of events in THIN14 and THIN13 in different periods before the last collection date.

	n events THIN2013	n events THIN2014	n(%) events missing THIN2013*	
All events				
1 month before LCD	3503886	3166627	337259	(9.63)
6 months before LCD	7184831	6808246	376585	(5.24)
12 months before LCD	11599088	11207025	392063	(3.38)
Ever before LCD	356859321	355957326	901995	(0.25)
Neoplasm events				
1 month before LCD	731100	725966	5134	(0.70)
6 months before LCD	4630	3348	1282	(27.69)
12 months before LCD	35639	32932	2707	(7.6)
Ever before LCD	71737	68490	3247	(4.53)
Circulatory system events				
1 month before LCD	13383	10944	2439	(18.22)
6 months before LCD	91386	85183	6203	(6.79)
12 months before LCD	189672	181094	8578	(4.52)
Ever before LCD	2756581	2733288	23293	(0.84)
Skin condition events				
1 month before LCD	46719	45427	1292	(2.77)
6 months before LCD	340150	337673	2477	(0.73)
12 months before LCD	693288	690411	2877	(0.41)
Ever before LCD	6935895	6930469	5426	(0.08)
Respiratory events				
1 month before LCD	84959	82866	2093	(2.46)
6 months before LCD	372311	367741	4570	(1.23)
12 months before LCD	776951	770814	6137	(0.79)
Ever before LCD	8800018	8784045	15973	(0.18)

*n events in THIN2014 in each period used as the denominator for % calculation

1

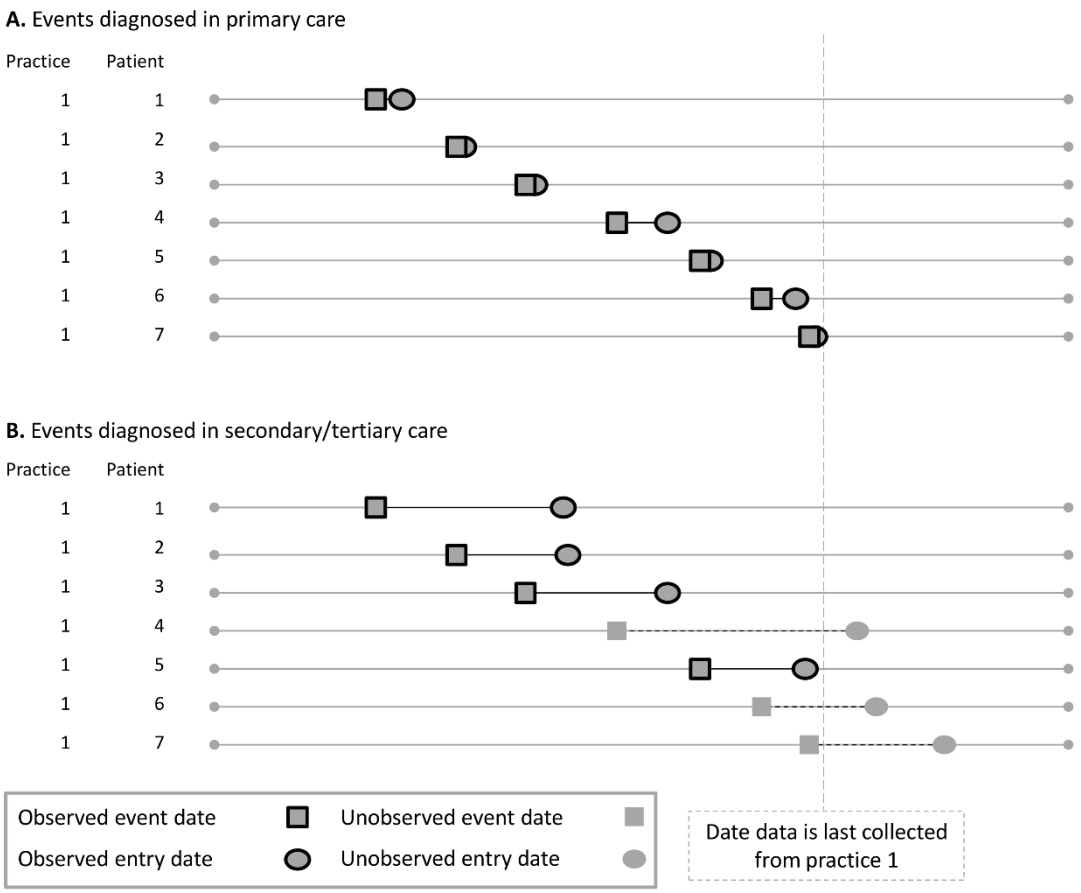

Figure 1: Schematic description of the censoring issue. In order to be recorded in the database, the entry date for an event must occur before the date data is last collected. As illustrated in Panel A, this should not be a problem for events diagnosed in primary care as the event date and entry dates will typically be the same (or very close together). However, Panel B illustrates that for events diagnosed in secondary or tertiary care, where the entry date and event date will typically be further apart, the entry date may occur after the date data are last collected from a clinical computer systems and therefore not be observed.

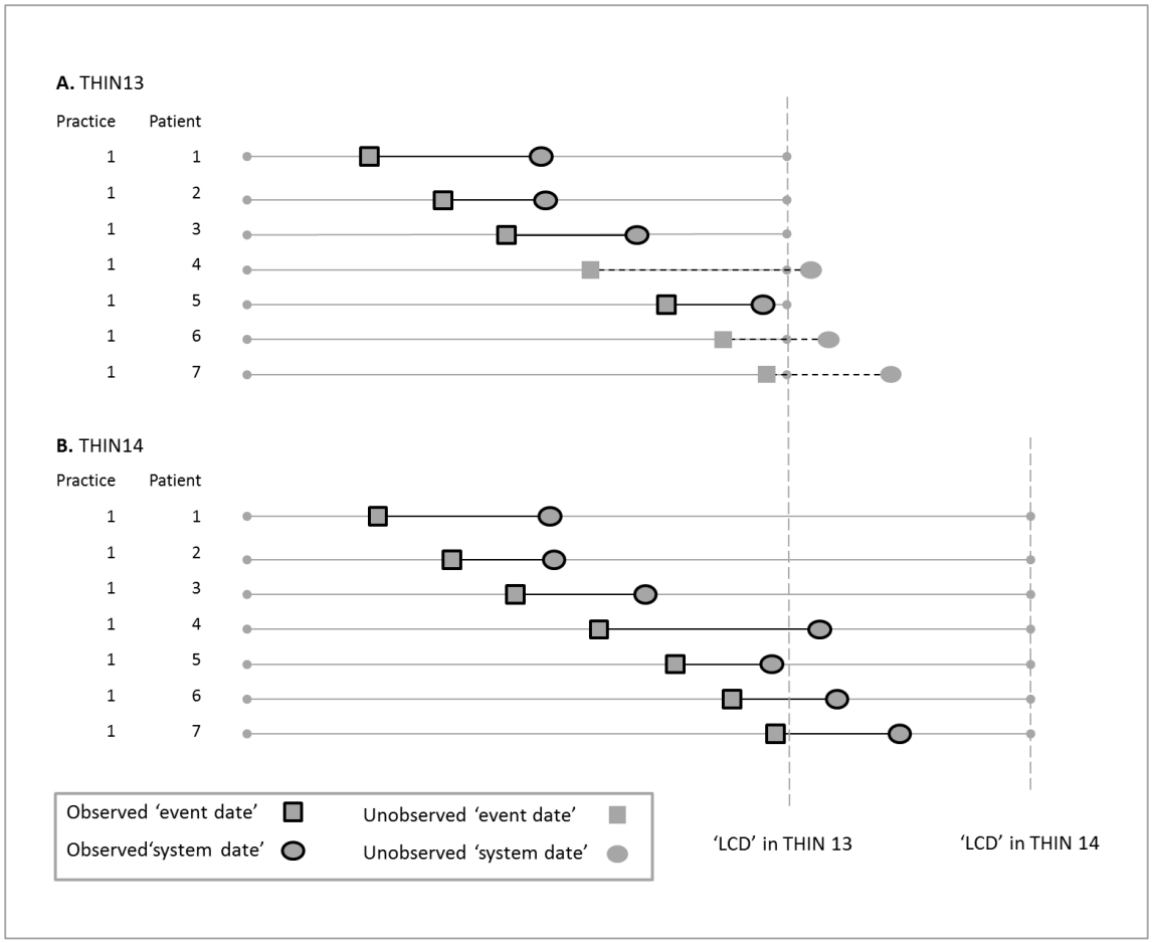

Figure 2: Reasoning for comparing two datasets in the analysis. Panel A shows a hypothetical sample of data available in THIN13. Events in patients 4, 6 and 7 are not recorded in THIN13 as their system date occurs after the practice LCD. Panel B shows data for the same patients in THIN14. System dates for events in patients 4, 6 and 7 occur before the THIN14 LCD therefore their event dates are recorded in this dataset even if follow up is censored at the THIN13 LCD.

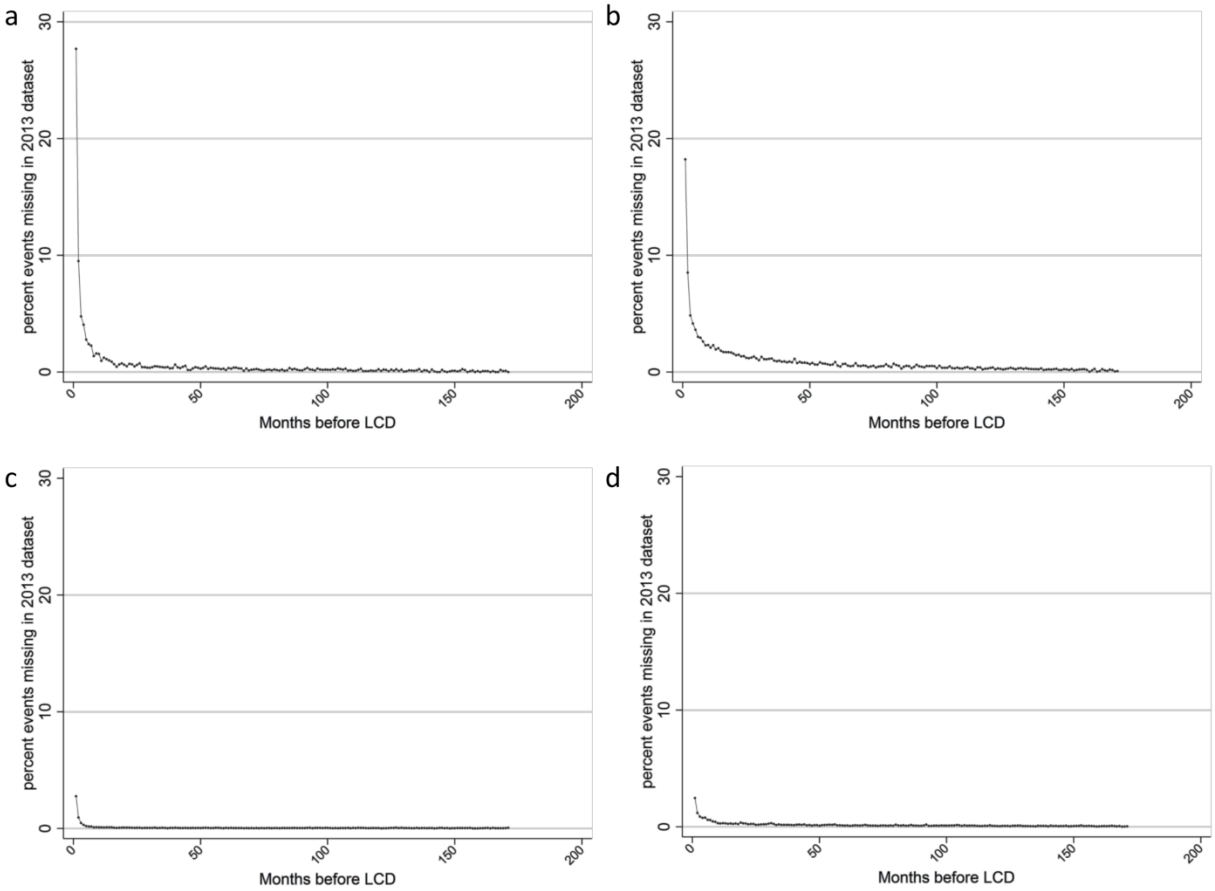

Figure 3 Line plots showing the difference in the number of events recorded in THIN2013 and THIN2014 in each month prior to the THIN2013 LCD as a proportion of the total number of events in the 2014 THIN dataset in that month, Plots shown for four different types of event: neoplasm (panel a) circulatory (panel b) skin (panel c) and respiratory conditions (panel d).

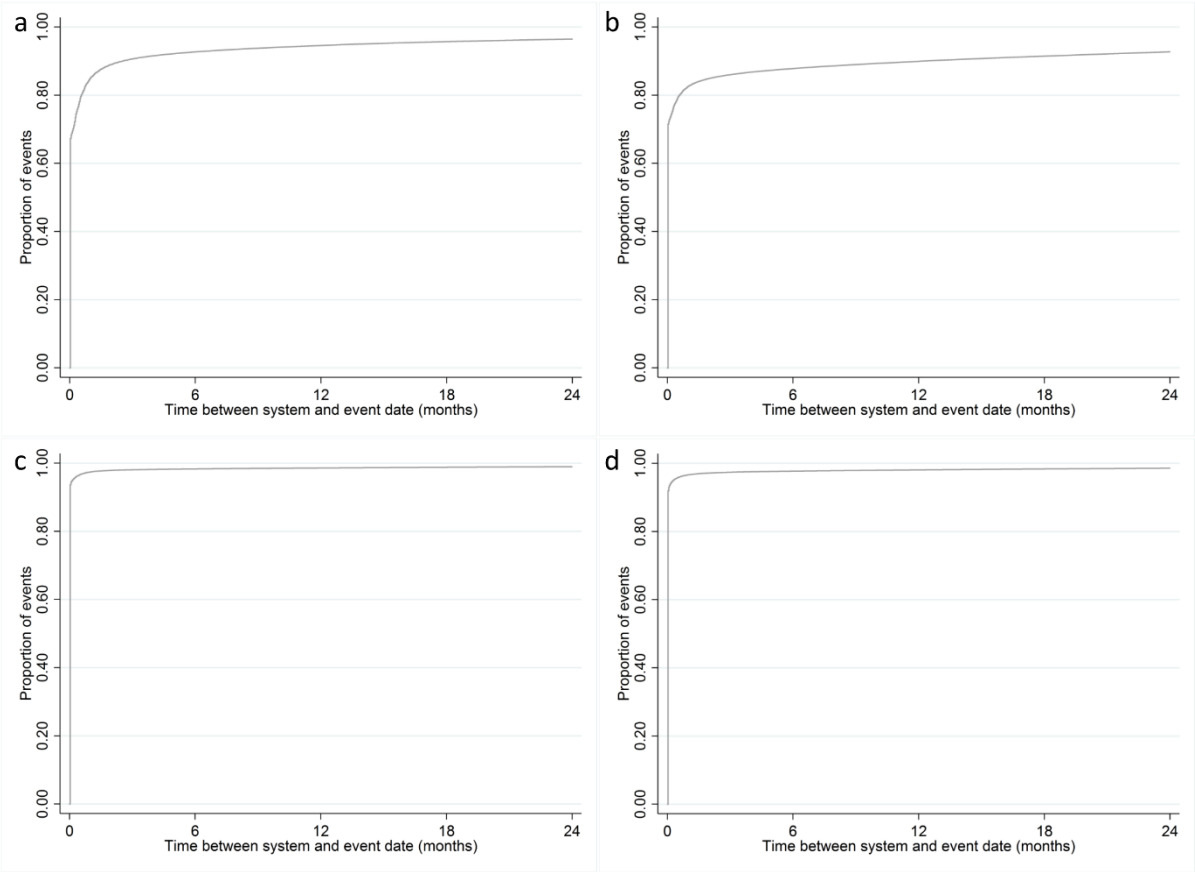

Figure 4 Cumulative incidence plots showing the distribution of time between event and system dates for first events occurring more than 5 years prior to the LCD for four different types of event: neoplasm (panel a) circulatory system (panel b) skin (panel c) and respiratory conditions (panel d). Plots truncated at 24 months.

Supplementary data 1

Supplementary to the main analysis which focused on four broad Read code chapters, the below table provides results for more specific Read code categories (categorised at 2 digit Read code level). That is, for events occurring more than 5 years prior to the LCD, we present the total number of events for each Read code category and the 50th, 75th, 90th and 95th percentiles of the distribution of time between system and event dates for each event type.

These data can be interpreted similarly to the data presented in Figure 4 of the manuscript in that for the 2 digit Read code category "B7", which contains codes related to benign neoplasms, there were 142,875 events occurring more than 5 years before the LCD, and the probability of a missing event was less than or equal to 10% 14 days before the LCD and as low as 5% (or less) 32 days before the LCD. For Read code category "B4" which contains code related to genitourinary neoplasms there were 16,655 events occurring more than 5 years before the LCD, and the probability of a missing event was 25% 13 days before the LCD, 10% 110 days before the LCD and 5% more than 1 year (391 days) before the LCD.

Note that categories with less than 5 events are marked "NA" as we cannot report results for categories containing so few events.

2 digit Read code		Total number of events	Days between system and event date			
			p50	p75	p90	p95
Infectious/parasitic diseases						
A0	Intestinal infectious diseases	21,039	0	0	2	10
A1	Tuberculosis	1,302	0	17	49	323
A2	Zoonotic bacterial diseases	13	0	0	0	18
A3	Other bacterial diseases	8,542	0	0	16	42
A4	Poliomyelitis & other non-arthropod-borne viral diseases-CNS	67	0	15	36	92
A5	Viral diseases with exanthem	125,480	0	0	0	7
A6	Arthropod-borne viral diseases	10	0	0	7	7
A7	Other viral and chlamydial diseases	317,178	0	0	0	3

A8	Rickettsioses and other arthropod-borne diseases	319	0	3	19	58
A9	Syphilis and other venereal diseases	834	0	0	14	48
AA	Other spirochaetal diseases	67	0	0	8	12
AB	Mycoses	460,321	0	0	0	0
AC	Helminthiases	9,269	0	0	0	0
AD	Other infectious and parasitic diseases	60,586	0	0	0	7
AE	Late effects of infectious and parasitic diseases	19	0	0	11	18
Ay	Other specified infectious or parasitic diseases	19	0	11	193	521
Az	Infectious and parasitic diseases NOS	NA	NA	NA	NA	NA
Neoplasms						
B0	Malignant neoplasm of lip, oral cavity and pharynx	657	0	14	93	407
B1	Malignant neoplasm of digestive organs and peritoneum	10,736	1	15	61	213
B2	Malig neop of respiratory tract and intrathoracic organs	6,990	1	11	36	106
B3	Malig neop of bone, connective tissue, skin and breast	53,366	6	21	102	523
B4	Malignant neoplasm of genitourinary organ	16,655	0	13	110	391
B5	Malignant neoplasm of other and unspecified sites	5,956	0	7	29	82
B6	Malignant neoplasm of lymphatic and haemopoietic tissue	10,900	7	20	146	473
B7	Benign neoplasms	142,875	0	0	14	32
B8	Carcinoma in situ	9,171	0	14	35	110
B9	Neoplasms of uncertain behaviour	4,354	5	14	43	289
BA	Unspecified nature neoplasm	59	0	14	28	54
BB	[M]Morphology of neoplasms	42,748	6	18	51	195
By	Neoplasms otherwise specified	80	7	18	67	249
Bz	Neoplasms NOS	3	16	37	37	37
Endocrine/metabolic						
C0	Disorders of thyroid gland	121,826	0	0	64	949
C1	Other endocrine gland diseases	308,420	0	0	49	797
C2	Nutritional deficiencies	16,004	0	0	0	20
C3	Other metabolic and immunity disorders	306,222	0	0	4	42
Cy	Endocrine, nutritional, metabolic or immunity disorders OS	NA	NA	NA	NA	NA

Cz	Endocrine, nutritional, metabolic or immunity disorder NOS	NA	NA	NA	NA	NA
Blood diseases						
D0	Deficiency anaemias	79,534	0	0	6	28
D1	Haemolytic anaemias	2,491	3	17	62	349
D2	Aplastic and other anaemias	23,961	0	0	14	41
D3	Clotting and bleeding disorders	5,429	0	12	39	228
D4	White blood cell and other blood disorders	6,357	0	6	23	95
Dy	Other specified disorders of blood or blood forming organ	248	0	0	0	23
Dz	Blood/blood forming organ NOS	NA	NA	NA	NA	NA
Mental disorders						
E0	Organic psychotic conditions	18,539	0	0	13	41
E1	Non-organic psychoses	95,386	0	0	9	118
E2	Neurotic, personality and other nonpsychotic disorders	1,141,400	0	0	4	184
E3	Mental retardation	185	0	14	1381	1381
Eu	[X]Mental and behavioural disorders	297,087	0	0	42	386
Ey	Other specified mental disorders	NA	NA	NA	NA	NA
Ez	Mental disorders NOS	12	0	0	0	7
Nervous system/senses						
F0	Inflammatory diseases of the central nervous system	1,042	0	2	26	66
F1	Hereditary and degenerative diseases of the CNS	53,302	0	0	15	45
F2	Other central nervous system disorders	223,687	0	0	11	47
F3	Peripheral nervous system disorders	84,596	0	0	20	55
F4	Disorders of eye and adnexa	1,017,805	0	1	22	61
F5	Diseases of the ear and mastoid process	1,540,049	0	0	0	6
Fy	Other specified diseases of nervous system or sense organ	9,382	0	0	14	35
Fz	Nervous system or sense organ disease NOS	NA	NA	NA	NA	NA
Circulatory system						
G0	Acute rheumatic fever	40	0	7	35	111
G1	Chronic rheumatic heart disease	1,505	9	26	90	472
G2	Hypertensive disease	502,416	0	0	1	304

G3	Ischaemic heart disease	246,952	0	5	54	660
G4	Pulmonary circulation diseases	4,105	1	18	106	621
G5	Other forms of heart disease	218,435	0	12	85	610
G6	Cerebrovascular disease	65,872	0	14	124	708
G7	Arterial, arteriole and capillary disease	67,205	0	7	38	278
G8	Vein, lymphatic and circulatory diseases NOS	308,995	0	0	8	27
G9	Vascular calcification	NA	NA	NA	NA	NA
Gy	Other specified diseases of circulatory system	NA	NA	NA	NA	NA
Gz	Circulatory system disease NOS	6	2076	2118	2129	2129
Respiratory system						
H0	Acute respiratory infections	3,873,071	0	0	0	1
H1	Other upper respiratory tract diseases	407,937	0	0	0	7
H2	Pneumonia and influenza	44,547	0	0	13	34
H3	Chronic obstructive pulmonary disease	572,282	0	0	35	630
H4	Lung disease due to external agents	1,093	0	17	75	750
H5	Other respiratory system diseases	20,758	0	6	26	66
Hy	Other specified diseases of respiratory system	28	0	0	0	0
Hz	Respiratory system diseases NOS	29	0	0	0	9
Digestive system						
J0	Oral cavity, salivary glands and jaw diseases	108,740	0	0	0	8
J1	Oesophageal, stomach and duodenal diseases	460,642	0	0	8	34
J2	Appendicitis and other disorders of the appendix	218	0	12	47	242
J3	Hernia of abdominal cavity	81,996	0	8	28	92
J4	Noninfective enteritis and colitis	74,149	0	1	21	64
J5	Other diseases of the intestines and peritoneum	385,628	0	0	14	42
J6	Liver, biliary, pancreas + gastrointestinal diseases NEC	64,977	0	14	50	258
Jy	Other specified diseases of digestive system	6	12	24	46	46
Jz	Digestive system diseases NOS	NA	NA	NA	NA	NA
Genito-urinary system						
K0	Nephritis, nephrosis and nephrotic syndrome	24,624	0	13	45	190

K1	Other urinary system diseases	737,010	0	0	2	14
K2	Male genital organ diseases	182,469	0	0	10	31
K3	Disorders of breast	124,394	0	0	11	29
K4	Female pelvic inflammatory diseases	42,799	0	0	0	11
K5	Other female genital tract disorders	697,672	0	0	1	17
Ky	Other specified diseases of genitourinary system	8	0	0	15	15
Kz	Genitourinary disease NOS	NA	NA	NA	NA	NA
Pregnancy/childbirth/puerperium						
L0	Pregnancy with abortive outcome	27,323	5	16	106	754
L1	Pregnancy complications	26,542	0	0	11	33
L2	Risk factors in pregnancy	5,743	9	21	84	684
L3	Complications occurring during labour and delivery	3,980	7	17	50	253
L4	Complications of the puerperium	1,109	0	0	0	6
L5	Maternal care for fetus	NA	NA	NA	NA	NA
Ly	Complications of pregnancy,childbirth or the puerperium OS	1,129	6	10	25	51
Lz	Complications of pregnancy,childbirth and the puerperium NOS	NA	NA	NA	NA	NA
Skin/subcutaneous tissue						
M0	Skin and subcutaneous tissue infections	676,370	0	0	0	4
M1	Other skin and subcutaneous tissue inflammatory conditions	1,228,714	0	0	0	7
M2	Other skin and subcutaneous tissue disorders	1,349,495	0	0	0	13
My	Other specified diseases of skin or subcutaneous tissue	123	0	0	1	12
Mz	Skin and subcutaneous tissue disease NOS	2,706	0	0	0	0
Musculoskeletal						
N0	Arthropathies and related disorders	1,218,997	0	0	6	26
N1	Vertebral column syndromes	1,189,506	0	0	1	12
N2	Rheumatism, excluding the back	1,984,830	0	0	0	9
N3	Osteopathy/chondropathy/acquired musculoskeletal deformity	83,750	0	0	28	228
Ny	Musculoskeletal or connective tissue diseases OS	6,631	0	0	24	120
Nz	Musculoskeletal and connective tissue diseases NOS	137	0	0	0	2
Congenital anomalies						

P0	Anencephalus and similar anomalies	NA	NA	NA	NA	NA
P1	Spina bifida	225	0	19	55	490
P2	Other nervous system congenital anomalies	143	12	30	62	151
P3	Congenital eye anomalies	460	9	22	52	100
P4	Ear, face and neck congenital anomalies	834	0	2	32	132
P5	Bulbus cordis and cardiac septal closure anomalies	1,623	14	33	183	918
P6	Other congenital heart anomalies	565	6	24	71	292
P7	Other congenital circulatory system anomalies	627	10	38	303	1106
P8	Respiratory system congenital anomalies	149	2	20	151	364
P9	Cleft palate and lip	142	8	34	90	267
PA	Other congenital upper alimentary tract anomalies	1,147	0	14	41	160
PB	Other congenital digestive system anomaly	252	3	24	50	120
PC	Congenital genital organ anomalies	2,485	0	14	41	126
PD	Urinary system congenital anomalies	1,609	6	19	53	481
PE	Certain congenital musculoskeletal deformities	1,592	0	12	35	87
PF	Other congenital limb anomalies	925	0	9	30	76
PG	Other congenital musculoskeletal anomalies	765	4	22	60	248
PH	Congenital integument anomalies	2,496	0	8	27	57
PJ	Chromosomal anomalies	1,012	0	19	50	193
PK	Other and unspecified congenital anomalies	1,164	4	20	60	253
Py	Other specified congenital anomaly	NA	NA	NA	NA	NA
Pz	Congenital anomaly NOS	NA	NA	NA	NA	NA
Perinatal conditions						
Q0	Fetus/neonate affected by maternal problem unrelated to preg	7	6	66	797	797
Q1	Disorders due to slow fetal growth, low and high birthweight	21	10	25	186	3640
Q2	Birth trauma, asphyxia and hypoxia	44	0	9	55	263
Q3	Fetus and newborn respiratory conditions	140	0	0	0	10
Q4	Other perinatal conditions	300	0	11	43	99
Qy	Other specified perinatal conditions	NA	NA	NA	NA	NA
Qz	Perinatal conditions NOS	NA	NA	NA	NA	NA

Ill-defined conditions/working diagnoses						
R0	[D]Symptoms	2,091,921	0	0	3	14
R1	[D]Nonspecific abnormal findings	110,299	0	0	0	21
R2	[D]Cause of morbidity and mortality unsure and ill-defined	45,432	0	0	3	20
R3	[D]Specific abnormal findings	24	0	0	10	20
Ry	[D]Other specified symptoms, signs or ill-defined conditions	1,265	0	0	12	153
Rz	[D]Symptoms, signs and ill-defined conditions NOS	NA	NA	NA	NA	NA
Injury/poisoning						
S0	Fracture of skull	1,620	3	14	40	139
S1	Fracture of neck and trunk	3,377	0	9	41	132
S2	Fracture of upper limb	41,641	8	20	58	252
S3	Fracture of lower limb	29,969	9	22	69	266
S4	Dislocations and subluxations	6,637	4	15	39	103
S5	Sprains and strains of joints and adjacent muscles	137,911	0	0	7	18
S6	Intracranial injury excluding those with skull fracture	20,619	3	9	23	48
S7	Internal injury of chest, abdomen and pelvis	47	0	15	62	115
S8	Open wound of head, neck and trunk	18,082	0	7	15	28
S9	Open wound of upper limb	4,195	3	9	21	43
SA	Open wound of lower limb	4,509	0	0	8	18
SB	Injury to blood vessels	21	0	0	1	7
SC	Late effects injury/poisoning/toxic effects/external causes	495	0	0	4	17
SD	Superficial injury	27,448	0	0	6	13
SE	Contusion (bruise) with intact skin	22,078	0	0	9	18
SF	Crushing injury	1,716	0	0	10	21
SG	Foreign body (FB) in orifice	2,106	3	9	21	34
SH	Burns	5,708	0	0	9	21
SJ	Nerve and spinal cord injuries	288	0	4	16	31
SK	Traumatic complications and unspecified injuries	140,201	0	6	15	28
SL	Poisoning	16,576	6	18	64	351
SM	Nonmedicinal agent causing toxic effects	74	0	6	15	55

SN	Other and unspecified external effect causes	20,989	0	0	15	56
SP	Surgical and medical care complications NEC	32,534	0	0	5	15
SQ	Open wounds involving multiple body regions	13	0	0	0	0
SR	Injury involving multiple body regions	281	0	0	0	0
Sy	[X] Injury and poisoning classification terms	145	0	0	13	26
Sz	Injury and poisoning NOS	1,229	0	0	0	2
Causes of injury/poisoning						
T0	Railway accidents	NA	NA	NA	NA	NA
T1	Motor vehicle traffic accidents (MVTA)	15,993	0	0	9	33
T2	Motor vehicle non-traffic accident, MVNTA	NA	NA	NA	NA	NA
T3	Other road vehicle accidents	2,168	0	0	9	20
T4	Water transport accidents	NA	NA	NA	NA	NA
T5	Air and space transport accidents	15	0	0	5	1746
T6	Vehicle accidents not elsewhere classifiable	NA	NA	NA	NA	NA
T7	Place of occurrence of accident or poisoning	1,223	0	0	0	0
T8	Accidental poisoning by drugs, medicines and biologicals	132	4	10	20	52
T9	Accidental poisoning by other non-drug substances	28	0	4	15	31
TA	Medical accidents to patients during surgical/medical care	26	0	0	35	56
TB	Medical + surgical procedures causing complications,no blame	63	0	0	4	21
TC	Accidental falls	94,954	0	0	6	15
TD	Accidents caused by fire and flames	NA	NA	NA	NA	NA
TE	Accidents due to natural and environmental factors	13,867	0	0	3	9
TF	Accidents caused by submersion, suffocation, foreign bodies	35	0	2	9	24
TG	Other accidents	2,791	0	0	16	61
TH	Late effects due to accidental injury	NA	NA	NA	NA	NA
TJ	Drugs and other substances-adverse effects in theraputic use	8,008	0	0	0	1
TK	Suicide and selfinflicted injury	1,764	6	17	71	565
TL	Homicide and injury purposely inflicted by other persons	618	0	0	9	29
TM	Legal intervention causing injury	NA	NA	NA	NA	NA
TN	Injury undetermined whether accidentally/purposely inflicted	29	0	0	14	50

TP	Injury resulting operations of war	14	0	0	7	1442
Tz	Causes of injury and poisoning NOS	NA	NA	NA	NA	NA
External causes of morbidity and mortality						
U0	[X]Transport accidents	91	0	0	0	8
U1	[X]Other external causes of accidental injury	422	0	0	7	18
U2	[X]Intentional self-harm	1,510	4	13	48	258
U3	[X]Assault	6,762	0	3	12	37
U4	[X]Event of undetermined intent	7	0	0	17	17
U5	[X]Legal intervention and operations of war	NA	NA	NA	NA	NA
U6	[X]Complications of medical and surgical care	913	0	0	0	0
U7	[X]Sequelae of external causes of morbidity and mortality	NA	NA	NA	NA	NA
U8	[X]Supplem factr relat to caus morbid+mortal classf elswhere	NA	NA	NA	NA	NA
Unspecified conditions						
Z1	Nursing care	189,608	0	0	0	0
Z2	Pregnancy, childbirth and puerperium observations	5,489	0	4	26	109
Z3	Child health procedures	1,326	4	20	78	150
Z4	Counselling	495,856	0	0	0	0
Z5	Psychotherapy	2,451	0	0	4	13
Z6	Physiotherapy	34,965	0	0	2	6
Z7	Occupational therapy	6,190	0	4	20	39
Z8	Ability to perform personal care activity	1,935	0	0	4	18
Z9	Indirect care procedures	90,027	0	0	11	29
ZA	Podiatry service	12,202	0	0	1	7
ZB	Orthoptic treatment	42	15	49	87	114
ZC	Feeding and dietary regimes	9,489	0	0	11	20
ZD	Speech and language therapy	380	7	21	49	78
ZE	Hearing observations	1,751	0	0	21	42
ZF	Audiological test observations	20	0	12	39	77
ZG	Advice	121,266	0	0	1	8
ZH	Surveillance	31,763	0	1	5	12

ZI	Physical and psychosocial approaches	64	1	7	10	15
ZJ	Complementary therapy	1,874	0	0	0	0
ZK	Social services procedures	31	0	1	9	13
ZL	Administrative statuses	1,018,247	2	11	24	41
ZM	Ability to perform domestic activities	1,142	0	0	10	29
ZN	Ability to perform community living activities	1,284	0	0	10	29
ZO	Ability to mobilise	100	0	0	7	20
ZP	Education and schooling	117	0	8	28	90
ZQ	Assessment regimes	98,673	0	0	0	3
ZR	Assessment scales	6,222	0	12	38	82
ZS	Speech and language disorder	1,328	0	14	40	77
ZT	Speech and language observations	154	0	0	13	20
ZU	Family details and household composition	552	0	0	8	236
ZV	[V]Health status and contact with health services factors	916,355	0	0	0	7
ZW	Care and support circumstances and networks	187	0	2	39	67
Zw	[Q] Temporary qualifying terms	NA	NA	NA	NA	NA
ZX	Self-harm	653	3	9	27	53