International Journal of Population Data Science

Journal Website: www.ijpds.org

Visualisation and optimisation of alcohol-related hospital admissions ICD-10 codes in Welsh e-cohort data

Laszlo Trefan[1], Ashley Akbari[2], Jennifer Siân Morgan[3], Daniel Mark Farewell[4], David Fone[5], Ronan A Lyons[6], Hywel Merfyn Jones[7,*], and Simon C Moore[8]

Submission History

Submitted:	15/06/2020
Accepted:	11/01/2021
Published:	24/03/2021

[1]Honorary Research Associate, Division of Population Medicine, School of Medicine, Cardiff University, Cardiff CF14 4YS
[2]Senior Research Manager and Data Scientist, Health Data Research UK Wales and Northern Ireland, Swansea University Medical School, Institute of Life Science 2, Sketty, Swansea SA2 8QA
[3]Performance Analysis Manager, NHS Wales Delivery Unit, Pencoed, Bridgend CF35 5LJ
[4]Senior Lecturer, Division of Population Medicine, School of Medicine, Cardiff University, Cardiff CF14 4YS
[5]Retired professor, Division of Population Medicine, School of Medicine, Cardiff University, Cardiff CF14 4YS
[6]Professor, Health Data Research UK Wales and Northern Ireland, Swansea University Medical School, Institute of Life Science 2, Sketty, Swansea SA2 8QA
[7]Research Associate / Statistician, Division of Population Medicine, Cardiff University, Cardiff CF14 4YS
[8]Professor of Public Health Research, Crime and Security Research Institute and School of Dentistry, Cardiff University, Cardiff CF14 4XY

Abstract

Introduction
The excessive consumption of alcohol is detrimental to long term health and increases the likelihood of hospital admission. However, definitions of alcohol-related hospital admission vary, giving rise to uncertainty in the effect of alcohol on alcohol-related health care utilization.

Objectives
To compare diagnostic codes on hospital admission and discharge and to determine the ideal combination of codes necessary for an accurate determination of alcohol-related hospital admission.

Methods
Routine population-linked e-cohort data were extracted from the Secure Anonymised Information Linkage (SAIL) Databank containing all alcohol-related hospital admissions (n,=92,553) from 2006 to 2011 in Wales, United Kingdom. The distributions of the diagnostic codes recorded at admission and discharge were compared. By calculating a misclassification rate (sensitivity-like measure) the appropriate number of coding fields to examine for alcohol-codes was established.

Results
There was agreement between admission and discharge codes. When more than ten coding fields were used the misclassification rate was less than 1%.

Conclusion
With the data at present and alcohol-related codes used, codes recorded at admission and discharge can be used equivalently to identify alcohol-related admissions. The appropriate number of coding fields to examine was established: fewer than ten is likely to lead to under-reporting of alcohol-related admissions. The methods developed here can be applied to other medical conditions that can be described using a certain set of diagnostic codes, each of which can be a known sole cause of the condition and recorded in multiple positions in e-cohort data.

Keywords
alcohol; hospital admission; e-cohort data; ICD-10 codes; optimisation

*Corresponding Author:
Email Address: jonesh75@cardiff.ac.uk (Hywel Merfyn Jones)

https://doi.org/10.23889/ijpds.v6i1.1373
March 24, 2021 © The Authors. Open Access under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.en)

Introduction

Excess alcohol consumption has adverse effects on health including liver cirrhosis [1], cancer [2], high blood pressure [3] and stroke [4]. There is also an increased risk of harm resulting from violence, including homicide [5], suicide [6], road traffic accidents [7], domestic violence [8], and assaults [9]. Globally, the net effect of alcohol consumption on health is detrimental, accounting for an estimated 3.8% of all mortality [10], at an estimated cost greater than 1% of the gross national product in high-income and middle-income countries [10] and additional costs associated with social harm.

Efforts to better understand the burden of alcohol on society has motivated the development of population wide statistics [11–15] that facilitate the study of factors that promote alcohol-related hospital admission [16–20]. However, there is a heterogeneity of datasets and varying definitions of alcohol-related admissions in use. Some studies used hospital episode statistics data [12–15, 18–20], which are routinely-collected administrative data that record any hospital activity and are very close to the original data source; others used processed, standardised data [11, 16, 17]. The former data typically use International Statistical Classification of Disease and Related Health Problems, Ninth Revision (ICD-9) [21] or Tenth Revision (ICD-10) [22] diagnostic codes that are recorded in up to 25 coding fields [13], the latter data use diagnostic codes that are recorded in up to 14 coding fields [11]. There are many different definitions in use, for both ICD-9 and ICD-10 codes, in defining alcohol-related admissions and for the number of coding fields in which to look for alcohol-related codes. Often these vary geographically between countries. In the United States of America (USA) ICD-9 alcohol-related diagnostic codes have been looked for in 15 coding fields [14], in Canada ICD-10 codes have been used in 25 fields [13]. In Australia both ICD-9 and ICD-10 codes were used [20] and in the United Kingdom (UK) ICD-10 codes were used in 20 diagnostic fields in one English study [18] and 14 fields in one Welsh study [11].

Generally, there are two sets of alcohol-related diagnostic codes used to define alcohol-related admission: a broad definition, [11, 18, 20], where alcohol-specific (e.g. 'alcoholic fatty liver') *and* alcohol-associated (e.g. 'oesophageal varices') diagnostic codes are used, and a narrow definition [13–17], where alcohol-specific diagnostic codes are used only. Furthermore, there is no consensus neither on which episode in the admission nor on which coding field of an episode defines an alcohol–related admission. (Episode means a continuous period of care under a single consultant doctor or medical team during an admission [18] in these data.) Some use the first episode [17, 18] others the discharge episode [13, 16] for identifying an alcohol-related admission. Often the first coding field of the first episode [18], in other cases the first three fields [17], or any fields [11] of this episode are used. In other cases any fields of the discharge episode [14, 16] are used for defining an alcohol-related admission.

There is a need to better understand the definition and patterns of alcohol-related diagnostic coding so that a consensus on the definition of an alcohol-related admission can be reached. Doing so will facilitate the derivation of reproducible and actionable epidemiological risk estimates.

The aim of this paper is to define methods to appropriately identify an alcohol-related admission in electronic hospital admission data. During a hospital admission, a patient may move through a number of specialities and receive a number of hospital procedures. Diagnostic codes that describe a patient's status therefore could be different at admission and at discharge. We compare the codes at admission and discharge, and assess and identify appropriate diagnostic (ICD-10) codes and number of diagnostic fields and derive a definition of an alcohol-related admission.

Methods

The data used in the current analyses are described in detail elsewhere [17, 23] and are summarised here.

Data sources

The Secure Anonymised Information Linkage (SAIL) Databank held within the Population Data Science department at Swansea University, contains health, social and education data on over three million residents of Wales, UK [24, 25]. The data used in this study can be accessed following an independent Information Governance Review Panel (IGRP) application approval. For all linked data within the SAIL Databank, each individual is assigned an Anonymous Linking Field (ALF), based on an encryption of the person's National Health Service (NHS) number either because the NHS number is present in the dataset, or because it is assigned based on the combination of unique identifiers including name, gender and date of birth [26].

The current study used the Patient Episode Database for Wales (PEDW) in SAIL. PEDW includes demographic and clinical data on all inpatient and day case admissions in NHS Wales hospitals and on all Welsh residents treated in hospitals in England. In PEDW records for an individual can be aggregated regardless of provider, where there is evidence that they are connected; these aggregated records are known as (person) spells. Spells involve at least one provider and one episode [27] and show continuous periods of inpatient care for a single patient which could take place under any number of different providers. Each record of an admission contains fields including, among others, date of admission; admission method (e.g. emergency or elective); spell number; episode number within the spell; provider unit code; patient classification (inpatient or day case), 14 ICD-10 diagnostic code fields [22]; six procedure code fields using the Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS) version 4.8 by date [28], which describe procedures applied; discharge destination (to identify inter hospital transfers); ICD-10 discharge codes; discharge method (to identify death in hospital) and date of discharge [29].

Characteristics of the e-cohort

The dataset used was extracted from PEDW data, on the basis of a set of alcohol-related ICD-10 codes (Supplementary Appendix 1), which was used in previous analysis [17] and of which had alcohol attributable fraction as one (Supplementary

Figure 1: Structure of the data used - extracted from Patient Episode Database for Wales (PEDW)

Person identifier	Arrival Date	Discharge date	Spell Identifier	Episode 1	Episode 2		Episode m
				Diagnostic coding field	Diagnostic coding field	...	Diagnostic coding field
ALF 1	01/01/2006	02/01/2006	Spell number 1	1|2|3|4|5|6|7|8|9|10|11|12|13|14		...	
ALF 1	09/02/2010	10/02/2010	Spell number 2	1|2|3|4|5|6|7|8|9|10|11|12|13|14	1|2|3|4|5|6|7|8|9|10|11|12|13|14	...	
ALF 2	06/06/2009	07/06/2009	Spell number 3	1|2|3|4|5|6|7|8|9|10|11|12|13|14		...	
...	...	...		...	...	...	
ALF n	07/09/2011	09/09/2011	Spell number n	1|2|3|4|5|6|7|8|9|10|11|12|13|14	1|2|3|4|5|6|7|8|9|10|11|12|13|14	..	1|2|3|4|5|6|7|8|9|10|11|12|13|14

*ALF: Anonymous Linking Field.

Appendix 1). The latter offered that each code used described a medical condition where alcohol was the known sole cause of the condition [20]. A request was made for the full record (including all diagnosis and all OPCS-4.6 procedure codes) for every episode with an admission date between 1 January 2006 and 31 December 2011 that contained any of these alcohol-related codes at any coding field at any point in the spell. Transfers between hospitals were taken into account if they had occurred within 24 hours. In these cases the relating episodes were collected into one admission under a new spell number and number of episodes were counted accordingly. A further criterion was that the person should have lived in Wales on 1 January 2006. Age was calculated as age at admission. After several cleaning steps the data were transposed into one row per single alcohol-related hospital admission [17] (Figure 1).

Definition and groups of alcohol-related admission and discharge codes

Our definition of an alcohol-related hospital admission was based on sets of ICD-10 diagnostic codes, as described earlier, and the details of these codes are presented in Supplementary Table 1 in Supplementary Appendix 1. The same sets of codes were used to identify both the alcohol-related codes at admission and at discharge. As admission, the first episode and as discharge, the last episode of an alcohol-related admission were used in this work (Figure 1). For chord diagram analysis the first alcohol code, which occurred in any of the 14 positions, recorded at admission and discharge was used. For chord diagram analysis admissions, which had more than one episode, were used (Figure 1). For further analyses both these admission and discharge codes were grouped into subcategories. The basis of these subcategories was the frequency of individual codes (subject to Information Governance disclosure rules such that all individual counts less than 5 were suppressed). Further details of these different subcategories can be found in Supplementary Table 1 in Supplementary Appendix 1, here just short examples are given: 'f104-f109' means alcohol-related ICD-10 diagnostic codes of from F10.4 to F10.9 *at admission*; 'K-NON LIVER' means alcohol-related ICD-10 diagnostic codes of K29.2,K85.2,K86.0 *at discharge* – alcohol-related non-liver diseases.

Statistical methods

All statistical analyses were conducted using the R-software version 3.2.1 [30].

Chord diagram

For the chord diagram, which represents the cross tabulations of first alcohol codes found in admission and in discharge, the chordDiagram tool of the circlize package [31] in R-software [30] was used. The chord diagram visualises the relationship between these codes, including the extent of agreement and discord between the different codes. In the cross tabulations both in the case of admission and discharge, these codes were grouped as we described earlier (see in Supplementary Appendix 1).

Misclassification rate calculation

Deciding how many coding fields to use to define an admission is an inexact science and clearly involves a trade-off between sensitivity (all 14 positions) and specificity (first position only) [17]. Sensitivity and specificity calculations require a standard, which describes the true (medical) condition [32]. In the absence of such a standard we defined the maximum

Table 1: Number of episodes in all alcohol-related admissions (n = 92,553)

Number of episode(s) within admission	Number of admissions	As percentage of all admissions
1	68,484	73.99%
2	17,854	19.29%
3	4,161	4.5%
4 or more	2054	2.22%
Total	**92,553**	**100.00%**

of alcohol-related admission as admissions where an alcohol related ICD10 code appeared in any coding position in any episode. We then determined the percentage of admissions that would be identified if only the first coding field was used, then second, third etc. position up to fourteenth position in the first episode. This process describes how many cases were "missed" potential, positive cases as a percentage by the number of coding fields used. This measure is called misclassification rate.

Results

There were 68,484 (74% of 92,553) admissions that had a single episode (Table 1) and therefore only one set of diagnostic codes exists. For these it is assumed that the admission and discharge codes are the same (Figure 1). There were of 24,069 (26%) admissions (Table 1) that had at least two episodes (Figure 1).

Figure 2 shows the number of alcohol codes in each of the 14 coding fields for the 68,484 alcohol-related admissions that had only one episode. It shows that alcohol-related codes are most likely to be found in the third coding followed by the first position. The second position is the least likely to contain alcohol-related code among the first four coding fields. From the fifth position onwards the frequency of alcohol-related codes declines.

In the case of the 24,069 admissions that had at least two episodes, the distribution of alcohol codes in both admission and discharge were very similar from the fifth coding field as previously described for admissions with a single episode (Figure 2 & Figure 3). However, the population of these codes was slightly different in the first four coding fields. In both admission and discharge, alcohol codes were least and second least populated in the fourth- and second coding field, respectively. The only difference between the population of alcohol codes in admission and discharge was that in the case of admission codes the most populated was the third coding field and then the first coding field while in the case of discharge codes it was the opposite (Figure 3).

The chord diagram analysis demonstrates good agreement between first found alcohol code recorded at admission (small letters) and the first found alcohol code recorded at discharge (capital letters). Most of the first found codes at admission (e.g. f100) or categories (e.g. f104-f109) were also found at discharge as well (Figure 4, Supplementary Table 2 in Supplementary Appendix 2). Discordant first found diagnosis codes were less than 1% of the (n = 24,069) admissions and no particular pattern was found in these cases. These results suggest agreement between alcohol-related diagnostic codes recorded at admission and discharge.

Since there was agreement between alcohol-related diagnostic codes recorded at admission and discharge, misclassification rates calculations were conducted for all

Figure 2: Distribution of alcohol-related codes in different coding fields in admissions with single episode (n = 68,484)

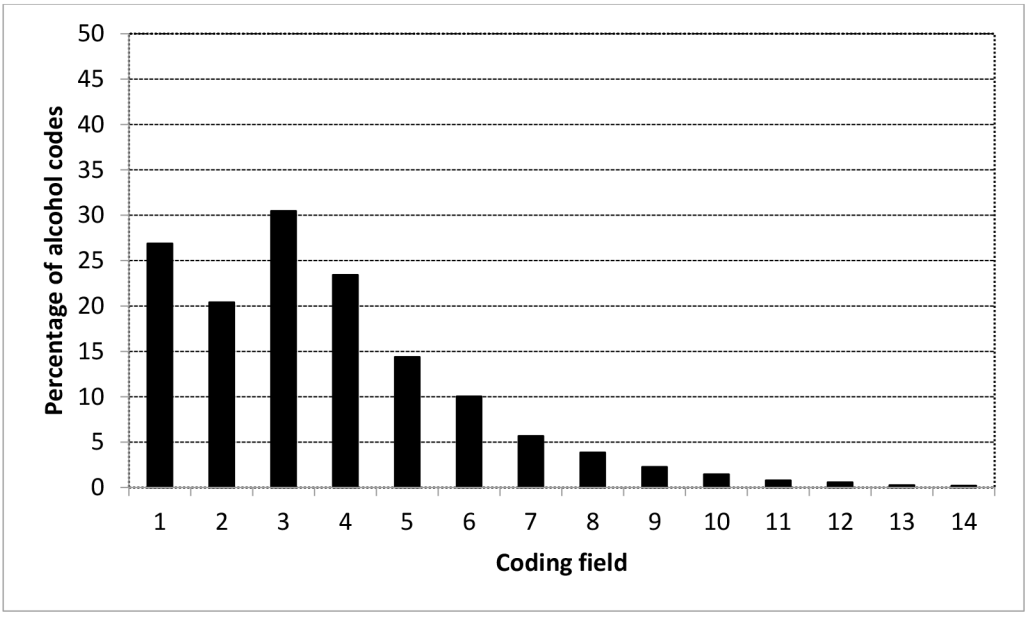


Figure 3: Distribution of alcohol-related codes in different coding fields in admission and discharge (n = 24,069)

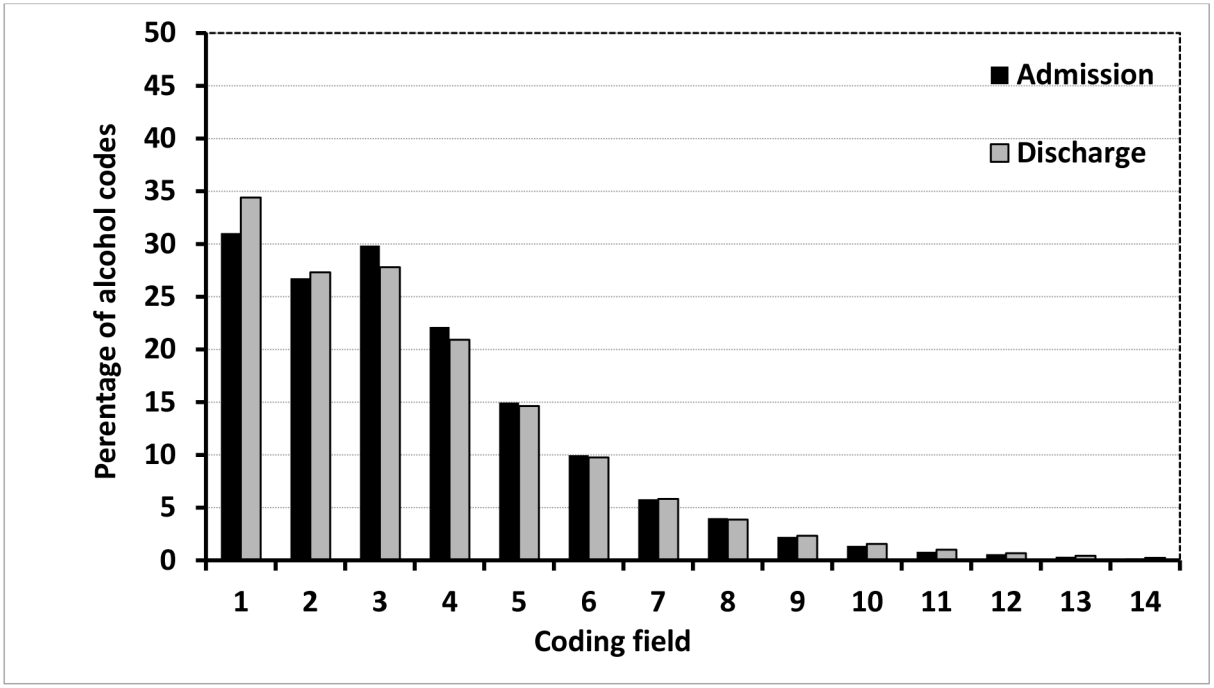
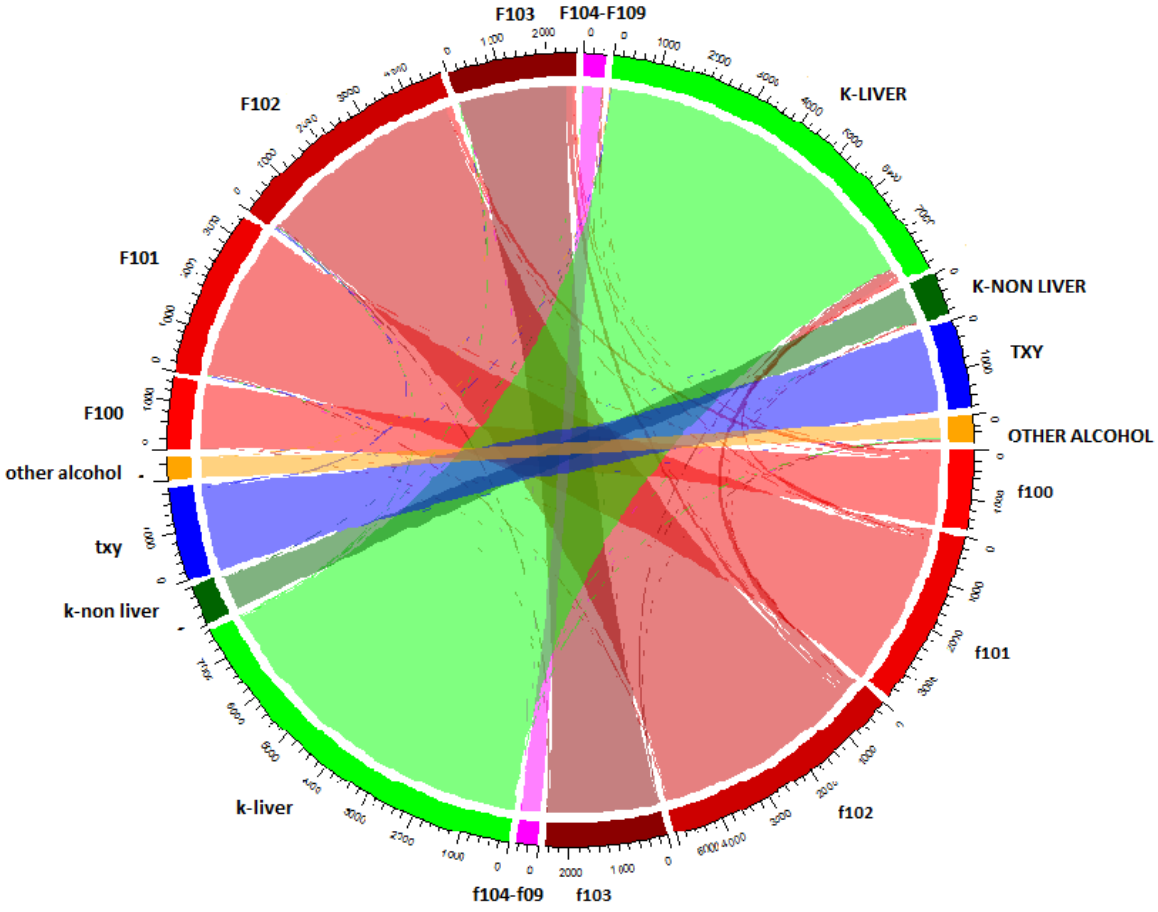


Figure 4: Chord diagram of first alcohol codes in alcohol-related admission and discharge* (n = 24,069)



*Small letters show first found alcohol-related diagnostic code(s) at admission; capital letters show these at discharge.

admissions (n = 92,553). Misclassification rates decrease as more and more coding fields were included (Table 2) and the misclassification rate falls less than one percent once more than ten coding fields are included.

Discussion

The diagnostic codes and the number of coding fields used to ascertain alcohol-related hospital admission varies across studies. This heterogeneity precludes opportunities to compare alcohol-related admission rates across jurisdictions when different methodologies are used. In addressing the need for an agreed universal approach, we developed robust methods to ascertain which code fields should be examined. The results of the analysis show that the distributions of alcohol-related codes in admission and discharge were similar. Furthermore, the first found alcohol-related codes in admission and discharge episodes demonstrated good agreement in the chord diagram comparison. These mean that both can be reasonably used. However, it was found that the chance of detecting an alcohol-related admission increases as the number of coding fields scrutinised increases, if more than the first ten alcohol-related diagnostic code fields were taken account of then the percent of alcohol-related admissions missed fell to less than 1% in the data used. According to our misclassification rate calculations, studies where only the first [18] or first three coding field(s) [17] are examined around 75% or 35% of all alcohol-related admissions, respectively, would have been missed. The results described here suggest that the optimal number of coding fields required to identify alcohol-related hospital admission is ten when the number of coding field is fourteen and each of these fields are populated. Optimality here refers to an appropriate grouping of codes that enable the identification of alcohol-related admission with the minimal set of data fields (number of codes). This work illustrates that using the first field only, which is very often used for primary diagnosis [18], can introduce misclassification of cases. The results can be extended to other countries and regions, especially when further (second, third etc.) coding fields are more populated than the first field.

The data used in this work organised all of a patient's episodes into one admission record. The methods described require only identification of a first and last episode which simplifies the data required and represents a minimum data set, potentially making the method applicable to other datasets.

The ICD-10 codes used in this work is similar to the list of 'wholly attributable conditions' [11, 18] or 'alcohol-attributable fractions' [33]. However, there are differences. The list used here does not include 'Methanol poisoning' (T51.1) and 'Toxic effect of alcohol, unspecified' (T51.9), which we regarded as not being specific to ethanol, and 'Foetal alcohol syndrome (dysmorphic)' (Q86.0), which is not relevant to adults, the target population of this study. By the same reasoning the list used here included 'Maternal care for (suspected) damage to foetus from alcohol' (O35.4). As such, the list here is slightly larger compared to other studies. For example, the Scottish Morbidity Record (SMR01) [16] is equivalent to PEDW data used here. Their definition of alcohol-related admission excluded E24.4, G72.1, K85.2, O35.4, and Y15-which were included here.

Based on several systematic reviews and meta-analyses [34–36], some studies extend the definition of alcohol-related hospital admissions to include partially alcohol attributable conditions [18, 18, 37], such as 'Ischaemic heart diseases' (I20-I25)), Diabetes mellitus (type II) (E11) or certain malignant neoplasm ('Malignant neoplasm of larynx' (C32)), conditions in which alcohol may be implicated in their aetiology [17]. The current analyses focused on wholly alcohol-related ICD-10 codes.

For the chord diagram only the first found alcohol-related diagnostic codes were used, which limits our findings related to this method. However, although we did not explore this, the same approach could be used for studying intermediate episodes e.g. comparison of diagnostic codes between any two episodes in the case of admissions where there are at least two episodes, therefore this method potentially offers

Table 2: Misclassification rates as percentage of all admissions (n = 92,553)

Coding fields taken account	Misclassification rate
1	70–75%*
1 to 2	55–60%*
1 to 3	30–35%*
1 to 4	20–25%*
1 to 5	10–15%*
1 to 6	5–10%*
1 to 7	<5%*
1 to 8	<3%*
1 to 9	<2%*
1 to 10	<1%*
1 to 11	<1%*
1 to 12	<1%*
1 to 13	<1%*
1 to 14	<1%*

*Information Governance Disclosure Control rules meant exact numbers could not be published.

the possibility of "in-depth" analysis of (first found) diagnostic codes.

Ideally sensitivity and specificity calculations would be used to identify optimal number of alcohol-related codes to use. The misclassification rate used in this work is a sensitivity like measure. For sensitivity and specificity calculations a "gold standard" is recommended [32]. In other work, codes of reviewed patients' charts data were used for this purpose [38]. We did not have these latter codings in our PEDW data, which is a limitation of our approach. In other cases [18, 20] when there were codes in multiple diagnostic positions within an episode, the classification of an episode eventually as an alcohol-related admission was done by given values of population attributable fractions to these codes. Population attributable fractions represent the proportion of cases at the population level that might be attributable to an exposure (i.e. alcohol), to different diagnostic codes [18]. In cases where alcohol was the known sole cause of a condition, the value of the population attributable fraction of the related diagnostic code is one [20]. All of the diagnostic codes used in this work had a value of one, which means they are wholly due to alcohol in every age and sex categories [33] therefore at population level. When partially attributable fraction(s) are used for identifying alcohol-related hospital admissions, the possibility of misclassification of non-alcohol-related admissions as alcohol-related admissions arises. In this case some form of external validation will be required, and sensitivity and specificity analysis should be conducted to establish the optimal number of coding fields.

Almost three quarters of our data had a single episode, in alcohol-related hospital episode statistics data higher than this, 86.7% were reported to have a single episode [18]. The difference might be due to the fact that the latter data are closer to the original data source than processed, standardised PEWD data, which was used in this work.

In this paper ICD-10 diagnostic codes were used. The defined and discussed methods could be implemented and used in the case of ICD-9 diagnostic codes as well.

Routinely collected electronic hospital records were used in this work, which were recorded in different hospitals and by different coders. Peng et al. (2018) [39] showed overall agreement (82.2%) and reliability (0.82) among 11 hospitals in emergency department ICD-10 (4-digits level) diagnostic codes. No influence of coder characteristics was found in almost half million ICD-10 hospital discharge records [40].

Conclusion

In the case of the data presented within this work alcohol-related ICD-10 diagnostic codes recorded at admission and discharge equivalently can be used for analysis. The appropriate number of these codes for analysis was established to be 10. Studies that consider only the first or first three coding fields are likely to miss 75% or 35%, respectively, of alcohol-related cases. The methods described and discussed could be applied to other medical conditions, which can be described with a certain set of diagnostic codes, each of which is a known sole cause of the condition and recorded in multiple positions in other routinely collected e-cohort data.

Research governance and ethics

The data used in this study are available in the SAIL Databank at Swansea University, Swansea, UK, but as privacy-preserving restrictions apply they are not publicly available. All proposals to use SAIL data are subject to review by an independent Information Governance Review Panel (IGRP). Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been granted, robust policies, structures, controls and special software are in place to protect privacy through a reliable matching, anonymisation and encryption process achieved in conjunction with NWIS, including presentation of data outside the SAIL Databank. SAIL has established an application process to be followed by anyone who would like to access data via SAIL at https://www.saildatabank.com/application-process.

Acknowledgements

This work was supported by funds from the Economic and Social Research Council, the Medical Research Council and Alcohol Research UK to the ELAStiC Project (ES/L015471/1).

This study used anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who enable SAIL to make anonymised data available for research.

This work was supported by Health Data Research UK, which receives its funding from HDR UK Ltd (NIWA1) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF) and the Wellcome Trust.

Conflict of interests

The authors declare that they do not have any conflict of interest.

Ethics approval and consent to participate

Approval for the use of anonymised data in this study, provisioned within the Secure Anonymised Information Linkage (SAIL) Databank was granted by an independent Information Governance Review Panel (IGRP) as project 0336. The IGRP has a membership comprised of senior representatives from the British Medical Association (BMA), the National Research Ethics Service (NRES), Public Health Wales and NHS Wales Informatics Service (NWIS). The use of anonymised data for research is outside the scope of the EU General Data Protection Regulations (GDPR) and the UK Data Protection Act.

Supplementary appendices

Appendix 1: Describes variable definitions for alcohol-related hospital admissions of PEDW data.

Appendix 2: Describes cross tabulation of first found diagnostic- and group of diagnostic codes at admission and discharge of alcohol-related hospital admissions.

References

1. Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. Lancet. 2006;367:52–6. https://doi.org/10.1016/s0140-6736(06)67924-5

2. Parkin DM. Cancers attributable to consumption of alcohol in the UK in 2010. Br J Cancer. 2011;105:S14–8.

3. Klatsky AL, Gunderson E. Alcohol and hypertension. J Am Soc Hypertens. 2008;2:307–17.

4. Reynolds K, Lewis LB, Nolen JDL, Kinney GL, Sathya B, He J. Alcohol consumption and risk of stroke: a meta-analysis. JAMA. 2003;289:579–88. https://doi.org/10.1001/jama.289.5.579

5. Parker RN. Alcohol, homicide, and cultural context: a cross national-analysis of gender-specific homicied victimization. Homicide Stud. 1998;2:6. https://doi.org/10.1177/1088767998002001002

6. Ramstedt M. Alcohol and suicide in 14 European countries. Addiction. 2001;96:59–75. https://doi.org/10.1046/j.1360-0443.96.1s1.6.x

7. del Rio MC, Gomez J, Sancho M, Alvarez F. Alcohol, illicit drugs and medicinal drugs in fatally injured drivers in Spain between 1991 and 2000. Forensic Sci Int. 2002;127:63–70. https://doi.org/10.1016/s0379-0738(02)00116-0

8. Abramsky T, Watts CH, Garcia-Moreno C, Devries K, Kiss L, Ellsberg M, et al. What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence. BMC Public Health. 2011;11:109. https://doi.org/10.1186/1471-2458-11-109

9. Sivarajasingam V, Morgan P, Matthews K, Shepherd JP, Walker R. Trends in violence in England and Wales 200-2004: an accident and emergency perspective. Injury. 2009;40:820–5. https://doi.org/10.1016/j.injury.2008.08.017

10. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. Lancet. 2009;373(9682):2223–33. https://doi.org/10.1016/s0140-6736(09)60746-7

11. Gartner A, Cosh H, Gibbon R, Lester N. A Profile of Alcohol and Health in Wales. Cardiff: National Public Health Service for Wales/Wales Centre for Health; 2009.

12. Public Health England [Public Health England: [Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/542889/Alcohol_consumption_and_harms_in_under_18s.pdf (accessed 26 Sept 2018).

13. Callaghan R, Macdonald SA. Changes in the rates of alcohol- and drug related hospital separations for Canadian provinces: 1996 to 2005. Can J Public Health. 2009;100(5):393–6. https://doi.org/10.1007/bf03405278

14. Yang AL, Vadhavkar S, Singh G, Omary B. Epidemiology of alcohol-related liver and pancreatic disease in the United States. Arch Intern Med. 2008;6:649–56.

15. Morleo M, Woolfall K, Dedman D, Mukherjee R, Bellis MA, Cook PA. Under-reporting of foetal alcohol spectrum disorders: an analysis of hospital episode statistics. BMC Pediatr. 2011;11:14–9. https://doi.org/10.1186/1471-2431-11-14

16. McDonald SA, Hutchinson SJ, Bird SM, Graham L, Robertson C, Mills PR. Association of self-reported alcohol use and hospitalization for an alcohol-related cause in Scotland: a record-linkage study of 23,183 individuals. Addiction. 2009;104:593–602. https://doi.org/10.1111/j.1360-0443.2009.02497.x

17. Fone D, Morgan J, Fry R, Rodgers S, Orford S, Farewell D, et al. Change in alcohol outlet density and alcohol-related harm to population health (CHALICE): a comprehensive record-linked database study in Wales. Public Health Res. 2016;4(3):1–222. https://doi.org/10.3310/phr04030

18. Green AM, Strong M, Conway L, Maheswaran R. Trends in alcohol-related admissions to hospital by age, sex and socioeconomic deprivation in England, 2002/03 to 2013/14. BMC Public Health. 2017;17:412. https://doi.org/10.1186/s12889-017-4265-0

19. Manhica H, Gauffin K, Almquist YB, Rostila M, Berg L, de Cortázar ARG, et al. Hospital admissions due to alcohol related disorders among young adult refugees who arrived in Sweden as teenagers – a national cohort study. BMC Public Health. 2017;17:644. https://doi.org/10.1186/s12889-017-4645-5

20. Pascal R, Liang W, Gilmore W, Chikritzhs T. Risks of alcohol-attributable hospitalisation and death in Australia over time: evidence of divergence by region, age and sex. Australas Med J. 2013;6(3):134–51. https://doi.org/10.4066/amj.2013.1618

21. Cartwright DJ. ICD-9-CM to ICD-10-CM Codes: What? Why? How? Adv Wound Care. 2013;2(10):588–92. https://doi.org/10.1089/wound.2013.0478

22. International Statistical Classification of Disease and Related Health Problems. 10th Edition ed. 20 Avenue Appia, 1211 Geneva 27,Switzerland: WHO Press, World Health Organization; 2010.

23. Fone D, Dunstan F, White J, Webster C, Rodgers S, Lee S, et al. Change in alcohol outlet density and alcohol-related harm to population health (CHALICE). BMC Public Health. 2012;12:428. https://doi.org/10.1186/1471-2458-12-428

24. Ford D, Jones K, Verplancke J-P. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Serv Res. 2009;9(3):24. https://doi.org/10.1186/1472-6963-9-157

25. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. BMC Med Inform Decis Mak. 2009;9:3. https://doi.org/10.1186/1472-6947-9-3

26. Demmler JC, Hill RA, Rahman MA, Bandyopadhyay A, Healy MA, Parajonthy S, et al. Educational Attainment at Age 10-11 Years Predicts Health Risk Behaviors and Injury Risk During Adolescence. The Journal of adolescent health : official publication of the Society for Adolescent Medicine. 2017;61:212–7. https://doi.org/10.1016/j.jadohealth.2017.02.003

27. Rees S, Akbari A, Collins H, Lee SC, Marchant A, Rees A, et al. Developing a standardised approach to the aggregation of inpatient episodes into person-based spells in all specialties and psychiatric specialties. BMC Med Inform Decis Mak. 2019;19:246. https://doi.org/10.1186/s12911-019-0953-2

28. Health and Social Care Information Centre [OPCS-4 Classification Version 4:[Available from: https://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp?shownav=1 (accessed 3 May 2018).

29. Wales NHS [Data Dictionary NHS Wales:[Available from: http://www.datadictionary.wales.nhs.uk/#!WordDocuments/nhswalesdatadictionary.htm (accessed 27 February 2018).

30. Team RDC. R: A language and environment for statitical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.

31. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811–2. https://doi.org/10.1093/bioinformatics/btu393

32. Altman DG. Practical Statistics for Medical Research. 1 st ed. 2–6 Boundary Row, London, SE1 8HN: Chapman & Hall; 1992.

33. Jones L, Bellis MA. Updating England-Specific Alcohol-Attributable Fractions. Liverpool: Center for Public Health, Faculty of Education Health & Community, Liverpool John Moores University; 2014.

34. Roerecke M, Rehm J. The cardioprotective association of average alcohol consumption and ischaemic heart disease: a systematic review and meta-analysis. Addiction. 2012;107:1246–60. https://doi.org/10.1111/j.1360-0443.2012.03780.x

35. Baliunas DO, Taylor BJ, Roerecke M, Jayadeep P, Mohapatka S, Rehm J. Alcohol as Risk Factor for Type 2 Diabetes: A systematic review and meta-analysis. Diabetes Care. 2009;132(11):2123–32. https://doi.org/10.2337/dc09-0227

36. Islami F, Tramacere I, Rota M, Bagnardi V, Fedirko V, Scotti L, et al. Alcohol drinking and laryngeal cancer: Overall and dose–risk relation – A systematic review and meta-analysis. Oral Oncol. 2010;46:802–10. https://doi.org/10.1016/j.oraloncology.2010.07.015

37. Stockwell T, Zhao J, Martin G, Macdonald S, Vallance K, Treno A, et al. Minimum alcohol prices and outlet densities in British Columbia, Canada: Estimated impacts on alcohol-attributable hospital admissions. American journal of public health. 2013(11):2020. https://doi.org/10.2105/ajph.2013.301289

38. So L, Evans D, Quan H. ICD-10 coding algortihms for defining comorbidities of acute myocardial infarction. BMC Health Serv Res. 2006;6:161. https://doi.org/10.1186/1472-6963-6-161

39. Peng M, Eastwood C, Boxill A, Jolley RJ, Rutherford L, Cralson K, et al. Coding reliabilty and agreement of International Classification of Disease, 10th revision (ICD-10) codes in emergency department data. Int J Popul Data Sci. 2018;3(4):1–6. https://doi.org/10.23889/ijpds.v3i1.445

40. Henessy DA, Quan H, Beck CA. Do coder characteristics influence validity of ICD-10 hospital discharge data? BMC Health Serv Res. 2010;10:99. https://doi.org/10.1186/1472-6963-10-99

Abbreviations

ICD-9:	International Statistical Classification of Disease and Related Health problems 9th edition
ICD-10:	International Statistical Classification of Disease and Related Health problems 10th edition
SAIL:	Secure Anonymised Information Linkage
USA:	United States of America
UK:	United Kingdom
IGRP:	Information Governance Review Panel
ALF:	Anonymous Linking Field
NHS:	National Health Service
PEDW:	Patient Episode Database for Wales
OPCS-4.6:	Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures version 4.6

Supplementary Table 1: ICD-10 codes, which define alcohol-related hospital admission adopted Fone et al (2016)*

ICD-10code	Description	Alcohol attributable fraction**	Alcohol code group in admission	Alcohol code group in discharge
F10.0	Acute intoxication	1	f100	F100
F10.1	Harmful use	1	f101	F101
F10.2	Dependence syndrome	1	f102	F102
F10.3	Withdrawal state	1	f103	F103
F10.4	Withdrawal state with delirium	1	f104-f109	F104-F109
F10.5	Psychotic disorder	1	f104-f109	F104-F109
F10.6	Amnesic syndrome	1	f104-f109	F104-F109
F10.7	Residual and late-onset psychotic disorder	1	f104-f109	F104-F109
F10.8	Other mental and behavioural disorders	1	f104-f109	F104-F109
F10.9	Unspecified mental and behavioural disorder	1	f104-f109	F104-F109
K70.0	Alcoholic fatty liver	1	k-liver	K-LIVER
K70.1	Alcoholic hepatitis	1	k-liver	K-LIVER
K70.2	Alcoholic fibrosis and sclerosis of liver	1	k-liver	K-LIVER
K70.3	Alcoholic cirrhosis of liver	1	k-liver	K-LIVER
K70.4	Alcoholic hepatic failure	1	k-liver	K-LIVER
K70.9	Alcoholic liver disease, unspecified	1	k-liver	K-LIVER
K29.2	Alcoholic gastritis	1	k-non liver	K-NON LIVER
K85.2	Alcohol-induced acute pancreatitis	1	k-non liver	K-NON LIVER
K86.0	Alcohol-induced chronic pancreatitis	1	k-non liver	K-NON LIVER
E24.4	Alcohol-induced pseudo-Cushing's syndrome	1	other alcohol	OTHER ALCOHOL
E51.2	Wernicke's encephalopathy	Not known		
G31.2	Degeneration of nervous system due to alcohol	1	other alcohol	OTHER ALCOHOL
G40.5	Special epileptic syndromes – if paired with other alcohol code	0.24–0.27 (males) 0.15–0.24 (females)	N/A***	N/A***
G62.1	Alcoholic polyneuropathy	1	other alcohol	OTHER ALCOHOL
G72.1	Alcoholic myopathy	1	other alcohol	OTHER ALCOHOL
I42.6	Alcoholic cardiomyopathy	1	other alcohol	OTHER ALCOHOL
O35.4	Maternal care for (suspected) damage to fetus from alcohol	1	other alcohol	OTHER ALCOHOL
R78.0	Finding of alcohol in blood	1	other alcohol	OTHER ALCOHOL
T51.0	Toxic effect: Ethanol Excl.: acute alcohol intoxication or "hangover" effects (F10.0), drunkenness (F10.0), pathological alcohol intoxication (F10.0)	1	txy	TXY
X45.0	Accidental poisoning by and exposure to alcohol: Occurrence at home	1	txy	TXY
X45.1	Accidental poisoning by and exposure to alcohol: Occurrence in residential institution	1	txy	TXY
X45.2	Accidental poisoning by and exposure to alcohol: Occurrence at school other institution/public admin area	1	txy	TXY
X45.4	Accidental poisoning by and exposure to alcohol: Occurrence on street/highway	1	txy	TXY
X45.5	Accidental poisoning by and exposure to alcohol: Occurrence at trade/service area	1	txy	TXY
X45.6	Accidental poisoning by and exposure to alcohol: Occurrence at industrial/construction area	1	txy	TXY
X45.8	Accidental poisoning by and exposure to alcohol: Occurrence at other specified place	1	txy	TXY
X45.9	Accidental poisoning by and exposure to alcohol: Occurrence at unspecified place	1	txy	TXY
X65.0	Intentional self-poisoning by and exposure to alcohol: Occurrence at home	1	txy	TXY

Continued

Supplementary Table 1: Continued

ICD-10code	Description	Alcohol attributable fraction**	Alcohol code group in admission	Alcohol code group in discharge
X65.1	Intentional self-poisoning by and exposure to alcohol: Occurrence in residential institution	1	txy	TXY
X65.2	Intentional self-poisoning by and exposure to alcohol: Occurrence at school other institution/public admin area	1	txy	TXY
X65.4	Intentional self-poisoning by and exposure to alcohol: Occurrence on street/highway	1	txy	TXY
X65.5	Intentional self-poisoning by and exposure to alcohol: Occurrence at trade/service area	1	txy	TXY
X65.6	Intentional self-poisoning by and exposure to alcohol: Occurrence at industrial/construction area	1	txy	TXY
X65.8	Intentional self-poisoning by and exposure to alcohol: Occurrence at other specified place	1	txy	TXY
X65.9	Intentional self-poisoning by and exposure to alcohol: Occurrence at unspecified place	1	txy	TXY
Y15.0	Poisoning by and exposure to alcohol, undetermined intent: Occurrence at home	1	txy	TXY
Y15.2	Poisoning by and exposure to alcohol, undetermined intent: Occurrence at school other institution/public admin area	1	txy	TXY
Y15.4	Poisoning by and exposure to alcohol, undetermined intent: Occurrence on street/highway	1	txy	TXY
Y15.8	Poisoning by and exposure to alcohol, undetermined intent: Occurrence at other specified place	1	txy	TXY
Y15.9	Poisoning by and exposure to alcohol, undetermined intent: Occurrence at unspecified place	1	txy	TXY
Y90.0	Blood alcohol level of less than 20 mg/100 ml	1	txy	TXY
Y90.1	Blood alcohol level of 20–39 mg/100 ml	1	txy	TXY
Y90.2	Blood alcohol level of 40–59 mg/100 ml	1	txy	TXY
Y90.3	Blood alcohol level of 60–79 mg/100 ml	1	txy	TXY
Y90.4	Blood alcohol level of 80–99 mg/100 ml	1	txy	TXY
Y90.5	Blood alcohol level of 100–119 mg/100 ml	1	txy	TXY
Y90.6	Blood alcohol level of 120–199 mg/100 ml	1	txy	TXY
Y90.7	Blood alcohol level of 200–239 mg/100 ml	1	txy	TXY
Y90.8	Blood alcohol level of 240 mg/100 ml or more	1	txy	TXY
Y90.9	Presence of alcohol in blood, level not specified	1	txy	TXY
Y91.0	Mild alcohol intoxication	1	txy	TXY
Y91.1	Moderate alcohol intoxication	1	txy	TXY
Y91.2	Severe alcohol intoxication	1	txy	TXY
Y91.3	Very severe alcohol intoxication	1	txy	TXY
Y91.9	Alcohol involvement, not otherwise specified	1	txy	TXY
Z50.2	Alcohol rehabilitation	Not known	N/A***	N/A***
Z71.4	Alcohol abuse counselling and surveillance	Not known	N/A***	N/A***
Z72.1	Alcohol use	Not Known	N/A***	N/A***

*Fone D, Morgan J, Fry R, et al. Change in alcohol outlet density and alcohol-related harm to population health (CHALICE): a comprehensive record-linked database study in Wales. Public Health Res 2016; 4(3): 1–222.

**Jones L, Bellis MA. Updating England-Specific Alcohol-Attributable Fractions. Liverpool: Center for Public Health, Faculty of Education Health & Community, Liverpool John Moores University; 2014.

***The ICD-10 codes, which either had alcohol attributable fraction <1 (G40.5) or were not known (E51.2, ,Z50.2,Z71.4,Z72.1) were found as part of the review of available, but considered out of scope of this study.

other alcohol, OTHER ALCOHOL: ICD-10 alcohol codes others.

k-liver, K-LIVER: ICD-10 alcohol codes, started by "K" related to liver disease.

k-non liver, K-NON LIVER: ICD10-10 alcohol codes related to non-liver disease.

txy, TXY: ICD-10 alcohol codes started by "T" or "X" or "Y".

Supplementary Table 2: Cross tabulation of first found diagnostic- and group of diagnostic codes at admission and discharge of alcohol-related hospital admissions. These hospital admissions had at least 2 episodes, happened between 2006 and 2011 in Wales. (<5 is shown where Information Governance Disclosure Control rules were not allowed to publish number)

n = 27,464									
Admission codes	**Discharge codes**								
	F100	**F101**	**F102**	**F103**	**F104-F109**	**K-LIVER**	**K-NON LIVER**	**TXY**	**OTHER ALCOHOL**
f100	1,452	30	107	38	<5	19	<5	<5	10
f101	8	3,360	32	48	12	102	16	<5	16
f102	10	35	4,544	114	20	212	38	<5	39
f103	6	7	35	2,367	11	38	8	<5	13
f104-f109	<5	<5	7	<5	423	<5	<5	<5	<5
k-liver	<5	14	21	27	<5	7,538	9	<5	26
k-non liver	<5	<5	6	<5	<5	10	784	<5	<5
txy	<5	20	36	<5	<5	7	<5	1,828	<5
other alcohol	<5	<5	9	<5	<5	14	<5	<5	486

f100/F100, f101/F101, f102/F102, f103/F103: F10.0, F10.1, F10.2, F10.3 alcohol-related ICD-10 diagnostic codes, respectively
f104-f109, F104-F109: F10.4, F10.5, F10.6, F10.7, F10.8, F10.9 alcohol-related ICD-10 diagnostic codes
k-liver, K-LIVER: K70.0, K70.1, K70.2, K70.3, K70.4, K70.9 alcohol-related, liver disease ICD-10 diagnostic codes
k-non liver, K-NON LIVER: K29.2, K85.2, K86.0 alcohol-related, non-liver disease ICD-10 diagnostic codes
other alcohol, OTHER ALCOHOL: E24.4, G31.2, G62.1, G72.1, I42.6, O35.4, R78.0 alcohol-related ICD-10 diagnostic codes
txy, TXY: T51.0, X45.0, X45.1, X45.2, X45.4, X45.5, X45.6, X45.8, X45.9, X65.0, X65.1, X65.2, X65.4, X65.5, X65.6, X65.8, X65.9, Y15.0, Y15.2, Y15.4, Y15.8, Y15.9, Y90.0, Y90.1, Y90.2, Y90.3, Y90.4, Y90.5, Y90.6, Y90.7, Y90.8, Y90.9, Y91.0, Y91.1, Y91.2, Y91.3, Y91.9 alcohol-related ICD-10 diagnostic codes.