Clinical Epidemiology

Dovepress
Taylor & Francis Group

Open Access Full Text Article

REVIEW

Systematic Review and Development of Recommended Code Lists to Identify Smoking and Vaping Status in Electronic Health Records (EHR)

Rong Ding[1], Sarah Cook[2], Philip W Stone[3], Dharun Srirathan[4], Yashwin Shyam[4], Ruhan Anand[4], Palaniappa Sudharshan[4], Jennifer K Quint[1]

[1]School of Public Health, Imperial College London, London, UK; [2]Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK; [3]Research Department of Primary Care and Population Health, University College London (UCL), London, UK; [4]Faculty of Medicine, Imperial College London, London, UK

Correspondence: Rong Ding, School of Public Health, Imperial College London, 86 Wood Lane, London, W12 0BZ, UK, Email r.ding23@imperial.ac.uk

Introduction: Vaping and smoking are important health behaviours associated with many diseases. Evaluating the association of smoking and vaping with diseases using electronic health record (EHR) data requires accurate codelists to determine smoking and vaping status. However, codelists used in studies are not always published or consistent between studies. It is important to develop standard codelists for use in future studies, and transparency is required to ensure consistency and standardization.

Purpose: To provide an overview of the codes used in both peer-reviewed scientific literature and codelist repositories to identify smoking and vaping status in EHRs and derive a recommended codelist for use in EHRs to identify smoking and vaping status.

Methods: Publications (MEDLINE, Embase, and Scopus) and codelist repositories (LSHTM Data Compass, OpenCodelists, and the HDR UK Phenotype Library) were searched from January 2010 to April 2024. All publications or codelist repositories with codes referring to smoking/vaping status were included in this review (search terms are further addressed in Supplementary Table 1). All codes were extracted to review the frequency and consistency between studies.

Results: There were 100 codelists across different coding systems: 55 codelists from publications and 45 codelists from codelist repository entries. For vaping status, there were 23 codelists identified, 7 from publications, and 16 from codelist repositories. Only 10% of publications included codelists. A limited number of ICD codes were used, and more were reported using the Read or SNOMED CT codes. The codelists we subsequently developed were based on those found in the review.

Conclusion: Very few studies have reported the use of codelists despite smoking status being a widely used variable in many publications, and vaping status is increasing. Using the information from the review, we derived codelists for smoking and vaping using a transparent methodology that can be used in future studies.

Keywords: EHR, clinical codes, SNOMED CT, smoking status, vaping status, ICD

Introduction

Smoking is a leading cause of preventable morbidity and mortality in the UK.[1–4] Smoking led to 506,100 hospital admissions in the UK during 2019/2020, contributing to 4% of total hospital admissions.[5] In the UK, smoking also caused 74,600 (15%) deaths in 2019, and 32% of deaths are related to conditions caused by smoking.[6] Electronic cigarettes (or vaping) are battery powered devices that provide nicotine and other chemicals by heating up a liquid. The use of e-cigarettes or vaping was introduced to help smoking cessation,[7] however the use of e-cigarettes has been increasing among former smokers and young people.[8,9] Vaping is seen as less harmful than smoking, but it is still a risk factor for cardiovascular and respiratory diseases, and the effect of vaping is still unknown.[10,11]

Received: 10 April 2025
Accepted: 26 August 2025
Published: 9 September 2025

Electronic Health Records (EHRs) provide longitudinal data that are routinely collected from patients during healthcare use.[12] The main purpose of which is to record individual patients' healthcare information in clinical practice. Although these records also provide a valuable source of data for health research, they are not designed primarily for use in clinical research,[13] and their use in research can pose methodological challenges. Therefore, clinical codes are developed to select useful information from EHRs for clinical research.[13] There are different coding systems designed and used in EHRs. International Classification of Diseases and Related Health Problems (ICD) is used internationally in secondary healthcare records and reporting causes of death.[14] Read Codes were used by the National Health Service (NHS) in the UK to record primary care data from 1985 till 2016.[15,16] Built upon CTV3 (the third version of Read codes), Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) adds many more codes/terms and defines complex relationships between them, allowing much more patient information to be stored in EHRs in a coded manner. SNOMED CT has also been adapted by the NHS after the retirement of Read Codes.[17,18] As various codes can refer to a specific health condition/behavior, list of multiple codes, forming a "codelist" to refer to/define one condition/behavior. The use of codelists is important in clinical research to select samples from EHRs with certain types of health information. Differences in codelists may lead to differences in the conditions/behaviors identified and can cause bias in research, and poor-quality code lists can cause poor reporting quality.[19,20] Therefore, it is important to develop a set of accurate code lists for specific conditions/behavior that can be applied to all relevant studies to standardize the code lists for specific conditions/behaviors.

The transparency and consistency of code lists are very important for observational studies using EHRs.[21] Publishing the code lists alongside the paper can add reproducibility to the work, allowing the same research methods to be adapted to different EHRs, and also reducing the effort to create similar code lists in future studies[19] More publications include code lists used for the study, but these are mainly code lists for main exposures or outcomes. Despite calls for transparency and consistency in the code lists, there are still a limited number of publications reporting the full code lists used in the study.[13,19–21]

As smoking is a commonly included confounding variable in studies, there is a need to review and evaluate the codelists used to identify smoking status. While vaping is a new phenomenon, there has not yet been a code list that is widely recognized and used in studies. This study aimed to provide an overview of the codes used in both the literature and code-list repositories to identify vaping and smoking status in EHR and to estimate the frequency of the codes reported in previous studies. This review also aimed to develop a recommended SNOMED CT codelist based on the existing best practices.

Materials and Methods

Information Source

This review used literature databases and code-list repositories as sources of information to identify codes relevant to smoking and vaping. This approach has been used previously in different studies to review the codes used in EHR.[13,22–24]

This systematic review searched for published papers in three literature databases: MEDLINE, Embase, and Scopus.

The code list repositories included in this systematic review were the HDR UK Phenotype Library (https://phenotypes.healthdatagateway.org/), the London School of Hygiene & Tropical Medicine (LSHTM) Data Compass (https://datacompass.lshtm.ac.uk/), and Open Codelists (https://www.opencodelists.org/).

Search Strategy

The medical codes for smoking and vaping in healthcare-related studies were searched for separately. The search terms for codes to identify smoking were categorized into three groups according to their content: smoking, code, and database (see Supplementary Table 1). After searching for terms in the three categories, the results were combined using AND.

To review the codes used to identify vaping status, the search terms were categorized into two categories based on their content: vaping-specific and code/database-specific, and the two categories were combined using AND (see Supplementary Table 1).

The code list repositories were searched separately for smoking and vaping. The repositories were searched using search terms relevant to smoking or vaping. For a full list of search terms, see Supplementary Table 1.

Eligibility Criteria

There were no restrictions on the research topic or methods in the literature in this systematic review. Code lists relevant to identifying the smoking or vaping status were included in this study. All sources of information were written in English and were published after 2010. All codes or code lists included in this review must be publicly available.

Data Extraction and Management

Literature Database

Papers searched for smoking and vaping status from literature databases (Embase, MEDLINE, and Scopus) were exported separately into Covidence for further review. Duplicate papers were automatically removed using the Covidence software.

The inclusion criteria of this review is publication or codelist repositories that include codes from ICD-9, ICD-10, SNOMED CT, Read code, Medcode, Medcode ID coding systems to identify smoking/vaping status. The study method of the publications and other criteria will not be considered in this review. This systematic review skipped title and abstract screening because smoking and vaping status are often used as covariates instead of main exposures or outcomes; therefore, it is less likely for papers to mention defining smoking and vaping status in titles and abstracts.

Papers were manually screened for full-text availability (Figures 1 and 2). Studies that were conference abstracts or had no available full text were excluded from the review. The first reviewer (RD) and second reviewers (DS, YS, RA, and SP) independently reviewed all papers for full-text review. The full-text review for both smoking and vaping included any paper with codes relevant to identifying smoking or vaping status. Conflicts in the full-text review of the first and second reviewers were resolved by a third reviewer (SC) (Fourteen conflicts identified for smoking and 5 for vaping codelists). After any conflicts were resolved, the included articles and codelists were entered into a spreadsheet.

There were a total of 1027 papers extracted from MEDLINE, Scopus and Embase to Covidence for smoking codes review, 276 were deleted due to duplication and 751 were processed to screening for full text. There were in total 328 publications with full text available during the review for smoking codes, 423 were excluded due to either full text not available or conference abstract. The rest 328 publications were processed for full text review, and 299 were excluded either due to no clinical codes listed in publication or codes not relevant to smoking. Twenty-nine publications were finally extracted from the smoking codes review (Figure 1).

There were a total of 512 papers extracted from MEDLINE, Scopus and Embase to Covidence for vaping codes review, 53 were deleted due to duplication and 459 were processed to screening for full text. There were 375 publications with full text available during the review for vaping codes, 84 were excluded due to either full text not available or conference abstract. The rest 375 publications were processed for full text review, and 356 were excluded either due to no clinical codes listed in publication or codes not relevant to smoking. Nineteen publications were finally extracted from the smoking codes review (Figure 2).

The codes extracted into the spreadsheet were sorted into different categories (smoking/vaping status) and coding systems (ICD-9, ICD-10, SNOMED CT, Read code, CPRD GOLD medcode, and CPRD Aurum medcodeid). Then, the frequency of each code found in the literature databases was calculated to identify the most popular codes used to identify smoking and vaping status.

Code List Repositories

The search in the code list repositories was conducted after searching literature databases. The code list repositories were searched separately for smoking and vaping. The code lists extracted from the repositories were entered into StataMP 18 to review the frequency of each code.

Results

Code List Availability

A total of 48 publications reported code lists relevant to smoking and vaping. Of these, 29 publications were found in the review for smoking, and 19 publications were found in the review for vaping. Of the 19 papers with codes presented in

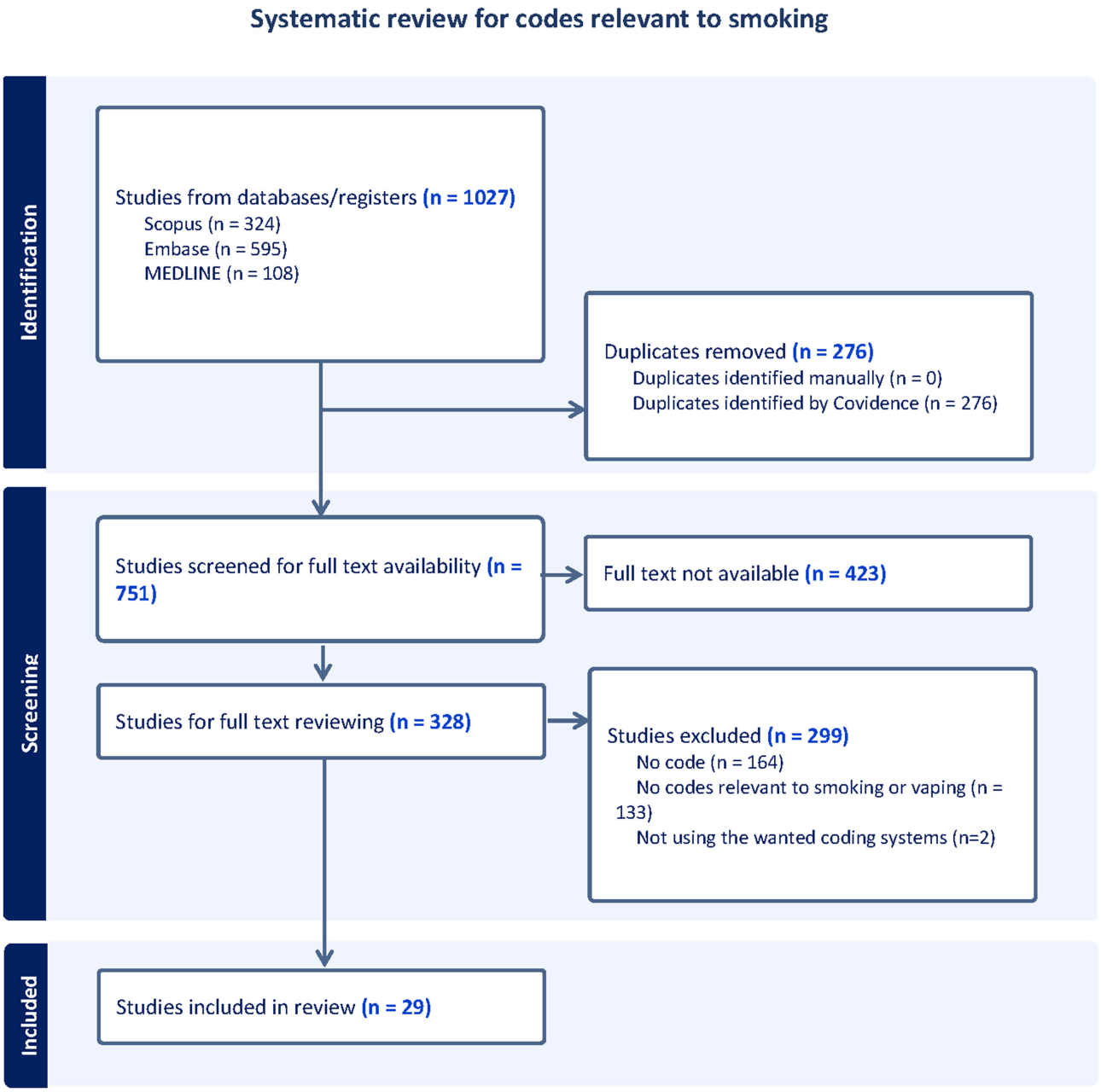

Figure 1 PRISMA flow chart for smoking (n = number of papers).

the review for vaping, 14 contained codes relevant to smoking status and were further categorized into the review for smoking (Figure 3). Finally, 43 publications contained 55 codelists relevant to smoking, and 5 publications contained 7 codelists relevant to vaping. Some publications contained multiple code lists for different coding systems.

A review of code list repositories found 45 codelists in 39 codelist repository submissions relevant to identifying smoking status and 16 codelists in 13 codelist repository submissions relevant to identifying vaping status. All three codelist repositories had codelists relevant to identifying smoking status; however, only the LSHTM Data Compass had search results relevant to vaping status. The code lists relevant to vaping status were mixed with the general code lists for smoking; however, only the codes relevant to vaping from these code lists were included in the review.

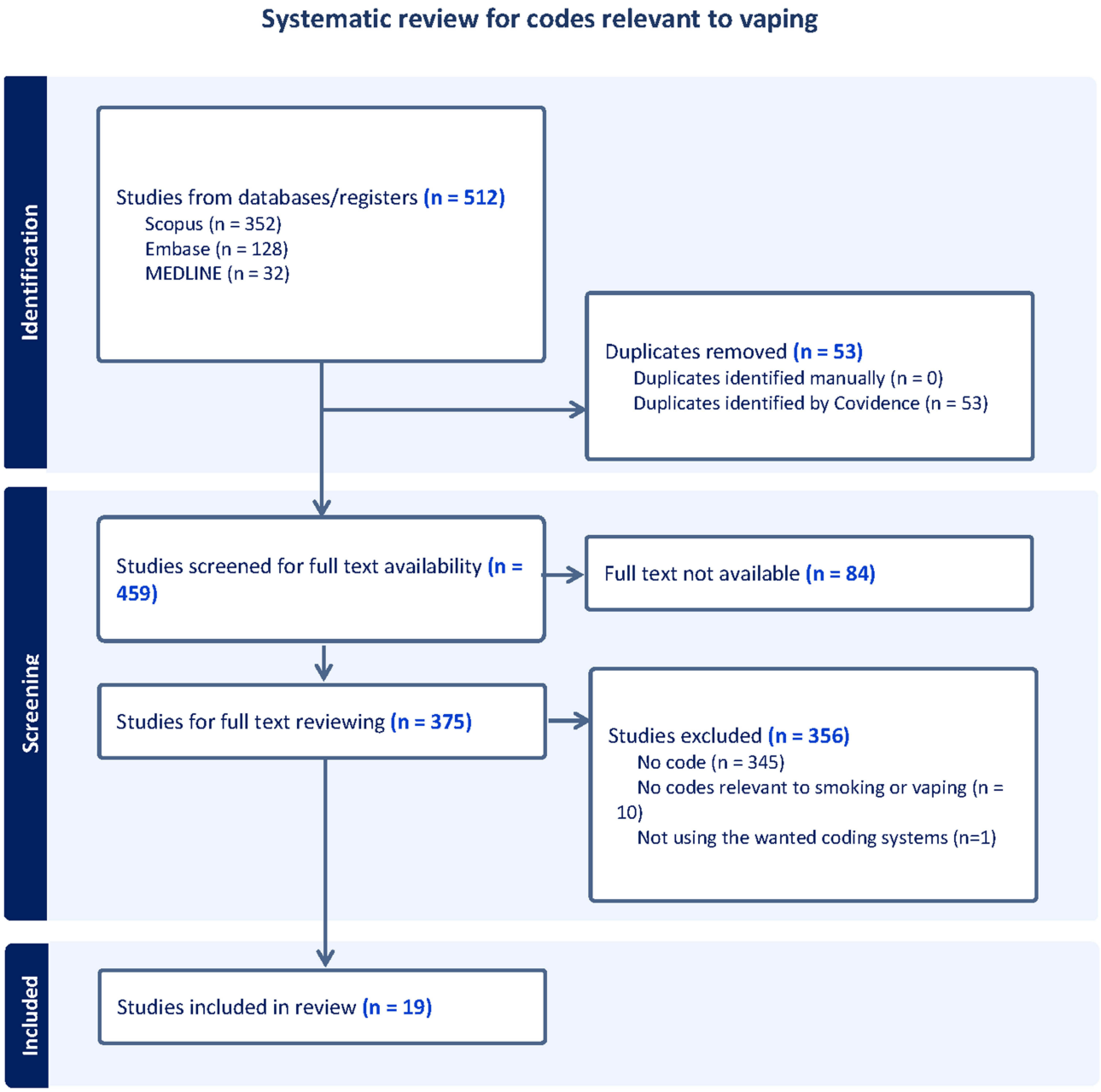

Figure 2 PRISMA flow chart for vaping (n = number of papers).

Smoking

In total, there were 100 codelists found relevant to smoking status: 55 from publications and 45 from codelist repositories (Figure 4). Read codes and Medcodes were the most commonly used coding systems in codelist repositories, whereas ICD codes were more common in publications. The reason for the large difference in the popularity of different coding systems in publications and codelist repositories might be the preference at the level of detail. Publications included in this review tended to include simpler codelists for example a shorter list of ICD-10 codes, whereas codelist repositories included longer and more complicated codelists for example using Readcodes and Medcodes. Readcodes and Medcodes are coding systems that have more detailed and specific codes for smoking as a health behavior. Another potential reason is that the codelist repositories were all British codelist repositories that contained only UK studies; code and code coding systems are used for primary care health recording in the UK, while ICD coding systems are used for secondary care health records used internationally.

https://doi.org/10.2147/CLEP.S529563

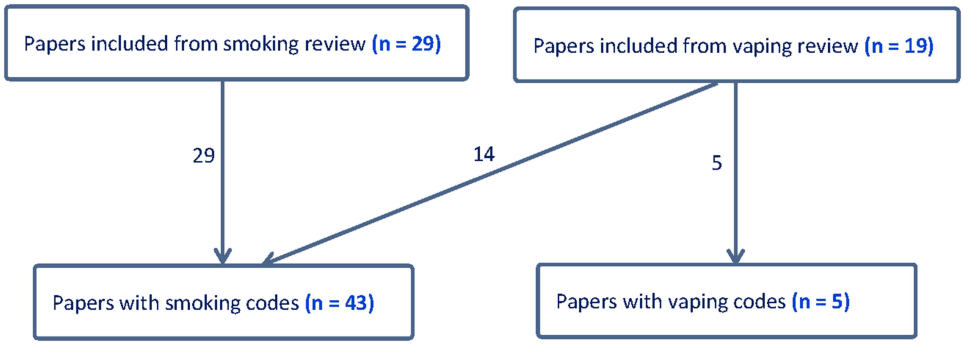

Figure 3 Number of papers extracted for vaping and smoking status (n = number of papers).

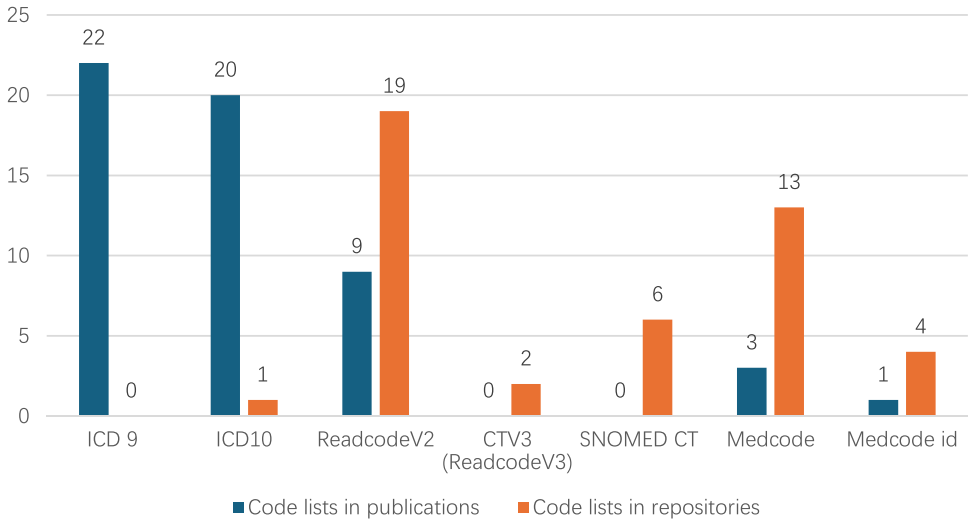

Figure 4 Smoking code lists summary.

ICD-9

There were in total 9 ICD-9 codes found in the review (Table 1). Most popular code was 305.1 "nondependent tobacco use disorder". This code only represented tobacco use disorder, but was not accurate in recording smoking status. Codes 649.0–649.04 refers to "tobacco use disorder complicating pregnancy", which did not clearly refer to smoking status. Code V15.82 "history of tobacco use" was more relevant to recording smoking status in all ICD-9 codes presented. (More details are provided in Supplementary Table 2).

Table 1 Summary of ICD-9 Codes Relevant to Smoking

ICD-9 Codes	Term	Frequency
305.1	Tobacco use disorder	20
V15.82	History of tobacco use	13
649.0	Tobacco use disorder complicating pregnancy, childbirth, or the puerperium, unspecified as to episode of care or not applicable	2
649.01	Tobacco use disorder complicating pregnancy, childbirth, or the puerperium, delivered, with or without	2

(*Continued*)

Table 1 (Continued).

ICD-9 Codes	Term	Frequency
649.02	Tobacco use disorder complicating pregnancy, childbirth, or the puerperium, delivered, with mention of postpartum complication	2
649.03	Tobacco use disorder complicating pregnancy, childbirth, or the puerperium, antepartum condition or complication	2
649.04	Tobacco use disorder complicating pregnancy, childbirth, or the puerperium, postpartum condition or complication	2
989.84	Toxic effect of tobacco	1
V69.8	Other problems related to lifestyle	1

Table 2 Summary of ICD-10 Codes Relevant to Smoking

ICD-10 Codes	Term	Frequency
F17.2-F17.299	Mental and behavioural disorders due to use of tobacco: dependence syndrome	15
Z72.0	Tobacco use	7
Z87.891	Personal history of nicotine dependence	7
F17	Nicotine dependence	3
O99.33	Tobacco use disorder complicating pregnancy, childbirth, and the puerperium	2
Z71.6	Tobacco abuse counselling	2
P04.2	Fetus and newborn affected by maternal use of tobacco	1
P96.81	Exposure to (parental) (environmental) tobacco smoke in the perinatal period	1
T65.2	Toxic effect: Tobacco and nicotine	1
Z57.31	Occupational exposure to environmental tobacco smoke	1
Z72	Problems related to lifestyle	1

ICD-10

There were in total 21 codelists found using ICD-10 coding system for smoking status (Table 2). Similar to the ICD-9 codes, the ICD-10 codes for "tobacco use disorder", "tobacco use", and tobacco exposure. There were also codes such as Z72.0 "tobacco use", just recording smoking status. (More details are provided in Supplementary Table 3).

Read Code

There were 28 codelists in total, and 288 codes relevant to smoking were identified using the read code V2 coding system (Supplementary Table 4). The codes were in more detail and could be categorized into different categories, including smoking status, smoking cessation, cigarette smoking amount, passive smoking, family smoking history, and smoking-relevant referrals. However, the main issue with codelists in Readcode is low consistency. Only 100 (34.72%) codes were used in 10 or more of the 28 codelists.

There were only two code lists using ReadcodeV3; therefore, we could not conclude or comment on the popularity of the codes.

Medcode

There were 16 code lists and 187 codes in total using the Medcode coding system. As medcodes are just the CPRD's version of read codes, they were matched with description terms and Readcodes from the CPRD medical dictionary

browser (Supplementary Table 5). Similar to the Readcode codelists, the Medcode codelists had more details about smoking, relaying different aspects relevant to smoking behavior. The consistency of the Medcode codelists used in studies was relatively high. Out of a total of 16 codelists, 148 (78.31%) codes were used in 10 or more codelists, and the codelists from the codelist repositories were mostly consistent.

SNOMED CT

SNOMED CT is a coding system that is used internationally. There were in total 6 code lists, 284 codes found using SNOMED CT coding system relevant to smoking (Supplementary Table 6). The SNOMED CT codelists used in different studies have been inconsistent. Only 16 (5.63%) codes were included in all 6 code lists and only 99 (34.86%) codes were used in two or more code lists. There were codes identifying the smoking status of patients and codes recording smoking cessation and environmental smoking; therefore, not all codes in the codelists were used to identify smoking status.

Medcodeid

Medcodeid is used in the Clinical Practice Research Datalink Aurum, which contains data from GP practices in England, using the EMIS software. A total of 420 codes were identified from the 5 Medcodeid codelists (Supplementary Table 7). There are 85 (20.24%) codes were used by all 5 codelists, and 350 (83.33%) codes were found in two or more codelists. The consistency of the codes used was relatively high among the five codelists; however, the number of codelists found was very limited.

Vaping

Limited codes were found for vaping in the codelist repositories (Figure 5). The codes for vaping in codelist repositories are normally listed in smoking codelists, and we only summarized the vaping codes for this section of the review.

Table 3 contains all codes that were found to be relevant to vaping (more details see Supplementary Table 8). Combining the results from publications and repositories, there are some common codes used in both publications and repositories for Medcodeid and Medcode, respectively. More Medcodeid codes were used to identify the vaping status than the other coding systems. The vaping status identified from the codes listed (Table 3) were all codes to identify current vaping and ex-vaping; no codes were to identify "non-vaping" in the review.

Codelist Development

Using information from this systematic review, we applied a standardized methodology (https://github.com/NHLI-Respiratory-Epi/SNOMED-CT-codelists) to develop a smoking codelist. The clinical terms (ie, descriptions) associated with the SNOMED CT codes found in this systematic review were searched using the NHS TRUD SNOMED CT UK

Figure 5 Vaping code lists summary.

Table 3 Summary of Codes Relevant to Vaping

	Code	Definition	Frequency
ICD 10	U07.0	Vaping-related disorder	4
Readcode	1PD.00	Ex user of electronic cigarette	5
	1PC.00	User of electronic cigarette	1
Medcode	108503	Ex user of electronic cigarette	6
	107292	User of electronic cigarette	3
Medcode id	187943100000 6110	User of electronic cigarette	5
	2265711000000117	Electronic cigarette user	5
	2336091000000117	Ex user of electronic cigarette	5
	1365349100000 6110	Nicotine-filled electronic cigarette vaper	3
	1365348100000 6112	Nicotine-filled electronic cigarette user	2
	3513803011	Electronic cigarette liquid	2
	7832521000006115	Electronic cigarette	2
	783254100000 6110	E-cigarette	1

Monolith Edition to check for the accuracy of the terms.[25] Then the codes from the systematic review were compared with an existing codelist (https://github.com/NHLI-Respiratory-Epi/Uncontrolled_COPD/blob/main/codelists/medical/CSV/smoking_status.csv). All codes present in existing codelists were found in this systematic review. The additional codes found in the systematic review that did not match the existing codelists were reviewed based on their clinical terms to determine whether the codes were used to identify the smoking status of the patients. These codes were found to focus on other aspects of smoking rather than on identifying patients' smoking status. Therefore, the codes present in the original codelist can be recommended to identify the smoking status.

The Vaping code list was developed following a publicly available algorithm for creating SNOMED CT codelists (https://github.com/NHLI-Respiratory-Epi/SNOMED-CT-codelists). As shown in Figure 6, the first step was to identify search terms for vaping status and apply them to the TRUD SNOMED CT code dictionary.[25] The search terms were the terms used for vaping specifically in the systematic review. After the initial search, we excluded some broad or undesired terms to limit search relevance. The codelists were then screened manually. The codes then were all identified into three categories: "current vaper", "ex-vaper" and "non-vaper" according to the meaning and key words included in the terms. The codelists were then matched with Medcodeid by using the CPRD code browser. This study compares codelists with existing codelists in a systematic review. However, some SNOMED CT codes do not match Medcodeid codes, and these codes are still included in the list. Finally, the codelist was sent to a respiratory physician for review, and the final version of the code list was edited according to the clinician's advice.

A total of 23 SNOMED CT Concept ID codes and 91 SNOMED CT Description ID codes were found during the search, all of which could be categorized into a relevant vaping status. The code list is available online at https://github.com/NHLI-Respiratory-Epi. For excel spreadsheet version of codelist to identify vaping status, see Supplementary Table 9.

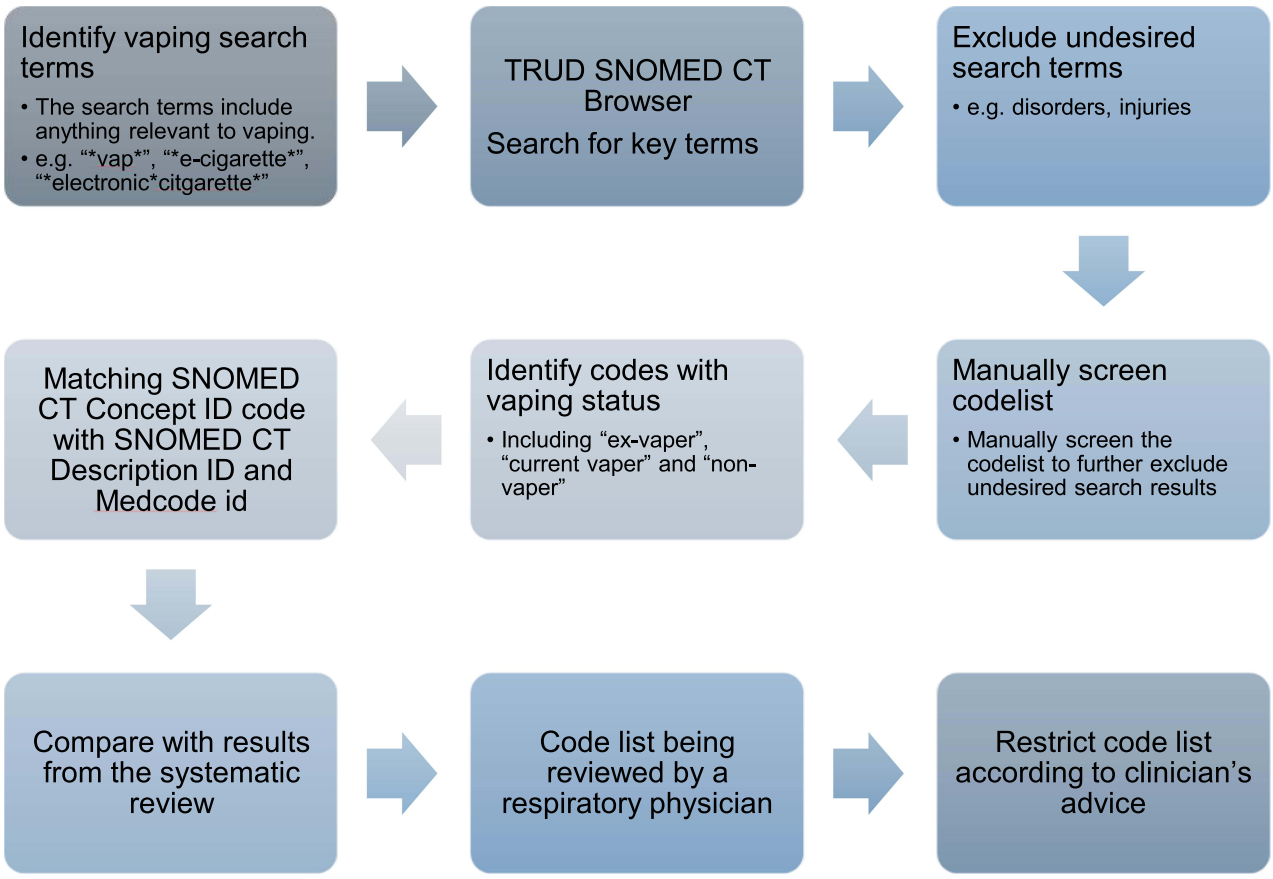

Figure 6 Flow chart of creating vaping codelist.

Discussion

We systematically reviewed the literature to identify smoking and vaporing codelists. There was a difference between the codelists found in publications and those found in codelist repositories. The codes found in publications were more likely to be listed in the main text, whereas the codelists found in codelist repositories were more comprehensive and consistent with each other. One reason for this could be the difference in focus. The focus of publications is on reporting the codes used, rather than evaluating the validity of codelists. Most publications do not include codelists for variables apart from exposures and outcomes, and smoking status, mostly as a covariate, is more likely to be neglected. However, the accuracy of defining covariates can make a difference to the study results.[20]

Secondly, ICD codelists used in publications are shorter and can be easily listed in the main text, but it is not feasible to list a SNOMED codelist or Readcode codelist in the main text. Furthermore, most publications reviewed in this systematic review were American studies, and ICD codes are commonly used in America. Most publications have used the ICD coding system. ICD coding systems are used in secondary care settings, while other coding systems are used in primary care settings. This explains why ICD coding systems have fewer details and subcategories for smoking and vaping behaviors. While the codelist repositories are all studies using the UK coding system, Readcode and SNOMED CT are commonly used in UK primary care settings; they are common in repositories from UK institutions. Different countries may have codelist repositories relevant to their healthcare systems.

Another finding during the review was the difference in codelists owing to differences in their focus. For example, all the smoking-related codelists reviewed have slightly different focuses: some focus more on the smoking amount per day, and some may include pregnancy smoking and secondhand smoke exposure. However, not all codes in the codelists can identify the smoking status of patients. Vaping became a new phenomenon that became popular after 2010. There were

far fewer codes relevant to vaping, and all codes were found among smoking codelists. Therefore, the codelist developed for vaping status can be used in future studies.

The second part of this work has formed codelists to identify smoking and vaping statuses. The codelist used to identify smoking status was based on the results of a systematic review and an existing codelist. This codelist was compared with 114 codes found in the NHS Primary Care Domain Reference Set Portal[26] to record smoking status; 97 codes were matched, and only 17 codes from the NHS Refsets did not match the codes in this codelist. The 17 codes that did not match were those that did not refer to a clear smoking status. From the comparison, the similarity between the current codelist and NHS refsets is high. This codelist can be used to identify the smoking status of patients in future studies. The codelist to identify the vaping status was formed based on the TRUD SNOMED CT dictionary, which identifies all SNOMED CT codes that are currently being used to represent the vaping status. No NHS Refsets for vaping status were available for comparison.

Most publications (90.55% of smoking-related publications and 94.67% of relevant publications) did not publish codelists on smoking/vaping status. According to the REporting of studies Conducted using Observational Routinely collected Data (RECORD) checklist[27] and STrengthening the Reporting of Observational studies in Epidemiology (STROBE) checklist,[28] observational studies using routinely collected data should report their codelists for classifying each variable. However, many publications only included codelists for the main exposure and outcome variables, or only sent codelists on request.

Strengths and Limitations

This study systematically reviewed codelists used in publications and repositories for smoking and vaping. This review also derived a SNOMED CT codelist to identify smoking status and developed a SNOMED CT codelist to identify the vaping status of patients, which may benefit future studies.

One limitation of this study is the availability of code lists. Many publications found during the systematic review did not have codelists available to the public; therefore, it was difficult to evaluate codelists from these studies. Second, this review focuses on the frequency of codes used in each study but does not evaluate the quality of the codelists being used or conduct a formal validation study.

Conclusion

Although smoking is a widely studied health behavior, the clinical codes used to identify smoking status vary among the studies. This study compared the codelists used to identify smoking status and developed a recommended set of SNOMED CT codes to form a codelist. Vaping is a new phenomenon that is becoming increasingly popular. To date, there are few codelists regarding vaping status found in publications and codelist repositories. This study produced a SNOMED CT codelist to identify vaping status based on the NHS SNOMED CT dictionary. Transparency among codelists is essential for trustworthy, reproducible, EHR-based research. All codelists produced in this study are publicly available at https://github.com/NHLI-Respiratory-Epi.

Acknowledgments

The citation counts of the publications and code list repositories reviewed in this study are listed in Supplementary Tables 10.

Disclosure

Professor Jennifer Quint reports grants and/or personal fees from UKRI, AZ, NIHR, HDR UK, BI, Insmed, and GSK, outside the submitted work. The authors report no other conflicts of interest in this work.

References

1. Bruijnzeel AW. The unhealthy association between smoking, vaping, and other drug use. *Nicotine Tob Res*. 2022;24(8):1139–1140. doi:10.1093/ntr/ntac130

2. Hernández-Pérez A, García-Gómez L, Robles-Hernández R, et al. Addiction to tobacco smoking and vaping. *Rev Invest Clin*. 2023;75(3):158–168. doi:10.24875/ric.23000117
3. Salvi S. Tobacco smoking and environmental risk factors for chronic obstructive pulmonary disease. *Clinics Chest Med*. 2014;35(1):17–27. doi:10.1016/j.ccm.2013.09.011
4. Wheaton AG, Liu Y, Croft JB, et al. Chronic obstructive pulmonary disease and smoking status - United States, 2017. *MMWR Morb Mortal Wkly Rep*. 2019;68(24):533–538. doi:10.15585/mmwr.mm6824a1
5. NHS infom. Tobacco. Available from: https://www.nhsinform.scot/healthy-living/stopping-smoking/reasons-to-stop/tobacco/. Accessed November 03, 2023.
6. NHS Digital. Statistics on smoking, England 2020: part 1: smoking-related ill health and mortality. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-smoking/statistics-on-smoking-england-2020/part-1-smoking-related-ill-health-and-mortality. Accessed October 30, 2024.
7. McNeill A, Brose L, Calder R, Simonavicius E, Robson D. Vaping in England: an evidence update including vaping for smoking cessation, February 2021. *Public Health England*. 2021;1–247.
8. Thirión-Romero I, Pérez-Padilla R, Zabert G, Barrientos-Gutiérrez I. Respiratory Impact of electronic cigarettes and "Low-Risk" tobacco. *Rev Invest Clin*. 2019;71(1):17–27. doi:10.24875/ric.18002616
9. Traboulsi H, Cherian M, Abou Rjeili M, et al. Inhalation toxicology of vaping products and implications for pulmonary health. *Int J Mol Sci*. 2020;21(10):3495. doi:10.3390/ijms21103495
10. Morjaria JB, Campagna D, Caci G, O'Leary R, Polosa R. Health impact of e-cigarettes and heated tobacco products in chronic obstructive pulmonary disease: current and emerging evidence. *Expert Rev Respir Med*. 2022;16(11–12):1213–1226. doi:10.1080/17476348.2023.2167716
11. Oriakhi M. Vaping: an emerging health hazard. *Cureus*. 2020;12(3). doi:10.7759/cureus.7421
12. Rothnie KJ, Müllerová H, Thomas SL, et al. Recording of hospitalizations for acute exacerbations of COPD in UK electronic health care records. *Clin Epidemiol*. 2016;Volume 8:771–782. doi:10.2147/CLEP.S117867
13. Massen GM, Stone PW, Kwok HH, et al. Review of codelists used to define hypertension in electronic health records and development of a codelist for research. *Open Heart*. 2024;11(1):e002640. doi:10.1136/openhrt-2024-002640
14. National Center for Health Statistics. ICD-10-CM. Available from: https://www.cdc.gov/nchs/icd/icd-10-cm/index.html. Accessed October 30, 2024.
15. NHS Digital. Read Codes. Available from: https://digital.nhs.uk/services/terminology-and-classifications/read-codes. Accessed October 30, 2024.
16. Booth N. What are the read codes? *Health Libr Rev*. 1994;11(3):177–182. doi:10.1046/j.1365-2532.1994.1130177.x
17. Donnelly K. SNOMED-CT: the advanced terminology and coding system for eHealth. *Stud Health Technol Inform*. 2006;121:279.
18. NHS Digital. SNOMED CT. Available from: https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct. Accessed October 30, 2024.
19. 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(6):e99825. doi:10.1371/journal.pone.0099825
20. Stone PW, Osen M, Ellis A, Coaker R, Quint JK. Prevalence of chronic obstructive pulmonary disease in England from 2000 to 2019. *Int J COPD*. 2023;18:1565–1574. doi:10.2147/COPD.S411739
21. Bhattarai N, Charlton J, Rudisill C, Gulliford MC. Coding, recording and incidence of different forms of coronary heart disease in primary care. *PLoS One*. 2012;7(1):e29776. doi:10.1371/journal.pone.0029776
22. MacRae C, Whittaker H, Mukherjee M, et al. Deriving a standardised recommended respiratory disease codelist repository for future research. *Pragmatic Observational Res*. 2022;Volume 13:1–8. doi:10.2147/POR.S353400
23. Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas IJ. Validation of asthma recording in electronic health records: a systematic review. *Clin Epidemiol*. 2017;9:643–656. doi:10.2147/clep.S143718
24. Jayatunga W, Stone P, Aldridge RW, Quint JK, George J. Code sets for respiratory symptoms in electronic health records research: a systematic review protocol. *BMJ open*. 2019;9(3):e025965. doi:10.1136/bmjopen-2018-025965
25. NHS TRUD. SNOMED CT UK monolith edition, RF2: snapshot. Available from: https://isd.digital.nhs.uk/trud/users/authenticated/filters/0/categories/26/items/1799/releases. Accessed November 08, 2024.
26. NHS Refsets. Primary care domain - reference sets (PCD refsets). Available from: https://app.powerbi.com/view?r=eyJrIjoiNDRmYjEwMzQtZGE3MS00ZGE5LTgwMTUtNjQ2NGE1NTZiYmEzIiwidCI6IjM3YzM1NGIyLTg1YjAtNDdmNS1iMjIyLTA3YjQ4ZDc3NGVlMyJ9. Accessed March 10, 2025.
27. RECORD. RECORD Checklist. Available from: https://www.record-statement.org/checklist.php. Accessed November 11, 2024.
28. STROBE. STROBE checklists. Available from: https://www.strobe-statement.org/checklists/. Accessed November 11, 2024.