Case Identification of Depression in Inpatient Electronic Medical Records: Systematic Review

Allison Grothman, Kendra G. Tickner, William J. Ma, Oliver D. Slater-Kinghorn, Elliot A. Martin, Danielle A. Southern, Hude Quan

Submitted to: JMIR Medical Informatics
on: June 14, 2023

Table of Contents

Case Identification of Depression in Inpatient Electronic Medical Records: Systematic Review

Allison Grothman[1]; Kendra G. Tickner[1]; William J. Ma[1]; Oliver D. Slater-Kinghorn[1]; Elliot A. Martin[2, 1] PhD; Danielle A. Southern[1, 3] MSc; Hude Quan[1, 3] MD, PhD

[1]Centre for Health Informatics Cumming School of Medicine University of Calgary Calgary CA
[2]Alberta Health Services Calgary CA
[3]Department of Community Health Sciences Cumming School of Medicine University of Calgary Calgary CA

Corresponding Author:
Elliot A. Martin PhD
Centre for Health Informatics
Cumming School of Medicine
University of Calgary
CWPH Building
3280 Hospital Drive NW
Calgary
CA

Abstract

Background: Electronic medical records (EMRs) contain large amounts of detailed clinical information. Using chart review to identify conditions within large quantities of EMRs can be time-consuming and inefficient. EMR-based phenotyping utilizing machine learning (ML) and natural language processing (NLP) algorithms is a continually developing area of study that holds potential for numerous mental health disorders.

Objective: This review outlines and evaluates the current state of EMR-based case phenotyping for depression.

Methods: A systematic review of EMR-based algorithms for phenotyping depression was completed. This research encompassed studies published from January 2000 to May 2023. The search involved three databases: Embase, MEDLINE, and APA PsycInfo. This was done using selected keywords that fall into three categories: terms connected with EMR, terms connected to case identification, and terms pertaining to depression. This document adheres to the Preferred Reporting Items for Systematic Reviews and Meta-analyses Extension for Scoping Reviews (PRISMA) guidelines.

Results: A total of 20 articles were assessed and summarized in the review. Most of these studies were undertaken in the United States, accounting for 75% (15/20). This was followed by the United Kingdom with 15% (3/20) and Spain with 10% (2/20). Both data-driven and clinical rule-based methodologies were identified. The development of EMR-based phenotypes and algorithms is indicative of the data accessibility permitted by each respective health system, leading to varying performance levels among different algorithms.

Conclusions: It is crucial to understand the commonalities and disparities in health systems, data gathering techniques, data release procedures, and current clinical routes for successful algorithm development strategies. There have been several propositions for strategies that aid in defining cases based on phenotype.

(JMIR Preprints 14/06/2023:49781)
DOI: https://doi.org/10.2196/preprints.49781

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Case Identification of Depression in Inpatient Electronic Medical Records: Systematic Review

Allison Grothman[1], email: allison.grothman@ucalgary.ca

William J Ma[1], email: william.ma1@ucalgary.ca

Kendra G Tickner[1], email: kendra.tickner@ucalgary.ca

Elliot A Martin PhD[1,2], email: eamartin@ucalgary.ca

Danielle A Southern MSc[1,3], email: dasouthe@ucalgary.ca

Hude Quan MD PhD[1,3], email: hquan@ucalgary.ca

1. Centre for Health Informatics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
2. Alberta Health Services, Calgary, Alberta, Canada
3. Department of Community Health Science, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

All authors can be reached at:

Centre for Health Informatics

Cumming School of Medicine

University of Calgary

CWPH Building, 3280 Hospital Drive NW

Calgary, Alberta, T2N 4Z6

Canada

Corresponding Author: Dr. Elliot Martin

Telephone: (403) 220-2779

Fax: (403) 210-9744

Email: eamartin@ucalgary.ca

Keywords

electronic medical records; EMR phenotyping; depression; health services research; precision public health

Abstract

Background

Electronic medical records (EMRs) contain large amounts of detailed clinical information. Using chart review to identify conditions within large quantities of EMRs can be time-consuming and inefficient. EMR-based phenotyping utilizing machine learning (ML) and natural language processing (NLP) algorithms is a continually developing area of study that holds potential for numerous mental health disorders.

Objective

This review evaluates the current state of EMR-based case identification for depression and provides guidance on using current algorithms and constructing new ones.

Methods

A systematic review of EMR-based algorithms for phenotyping depression was completed. This research encompassed studies published from January 2000 to May 2023. The search involved three databases: Embase, MEDLINE, and APA PsycInfo. This was done using selected keywords that fall into three categories: terms connected with EMR, terms connected to case identification, and terms pertaining to depression. This document adheres to the Preferred Reporting Items for Systematic Reviews and Meta-analyses Extension for Scoping Reviews (PRISMA) guidelines.

Results

A total of 20 articles were assessed and summarized in the review. Most of these studies were undertaken in the United States, accounting for 75% (15/20). The United Kingdom followed this with 15% (3/20) and Spain with 10% (2/20). Both data-driven and clinical rule-based methodologies

were identified. The development of EMR-based phenotypes and algorithms indicates the data accessibility permitted by each health system, leading to varying performance levels among different algorithms.

Conclusions

Better utilization of structured and unstructured EMR components through techniques such as ML and NLP has the potential to improve depression phenotyping. However, more validation must be carried out to have confidence in depression case identification algorithms in general.

Introduction

Background

Depression is a significant factor contributing to the global burden of disease. It contributes significantly to the cost of healthcare services, with depression treatment services costing an average of $550 per patient in Alberta, Canada, in the 2007/08 fiscal year [1]. Depression also carries a significantly higher mortality rate [2]. Surveillance of depression in the population is necessary to understand the needs of patients and allocate limited resources where they are most needed. This surveillance will ultimately allow healthcare professionals to make more targeted decisions when implementing population-level interventions.

Electronic medical records (EMR) are a digitized collection of patient records documented by medical professionals. They contain various types of patient information, including test results, demographic data, and information about medication orders, recorded in structured data fields and free-text data, such as discharge summaries and nurses' notes [3-5]. EMRs were designed to aid individual patient care but are increasingly used for other purposes, such as research and gathering data for precision public health efforts, as they are compiled in large data warehouses [6-9]. An area that will be instrumental in applying EMR to public health is case phenotyping, which is developing case definitions to identify positive cases of a disorder in EMR data.

Accurate case identification in EMRs is an area where more research needs to be conducted. This is especially true for case identification of psychiatric disorders. Previous reviews of phenotyping algorithms for psychiatric disorders only considered primary care databases as their setting [10,11]. However, these are very different from inpatient EMR systems. For one, hospital inpatients are more likely to identify errors and omissions than patients in outpatient care or primary care [12]. EMR data has been used in research for psychiatric patients in various specific inpatient use cases, including assessing patient safety events in psychiatric inpatient units [13]. Research has also shown that hospitals with electronic psychiatric EMRs had lower readmission rates for psychiatric patients compared to hospitals without electronic records. Similarly, hospitals where psychiatric records were accessible to non-psychiatric physicians had lower 14 and 30-day readmission rates [14]. In 2015, patients with a mental health diagnosis made up over 11% of hospital separations and 25% of hospital days [15]. Accurate case identification for inpatient stays for this at-risk population can help to identify what treatments have been most successful more efficiently than traditional research methods and could work in personalizing care for a more successful treatment plan.

Objectives

This study aims to provide an overview of existing algorithms for depression case identification in inpatient EMR. It will examine the performance of the algorithms and how they were constructed to provide guidance to those wishing to use an existing algorithm or to construct new ones.

Methods

Identifying Relevant Literature

This review follows the methodology outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) 2020 Statement [16]. First, we used the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes for depression provided by Elixhauser et al. (1998) [17] to identify relevant terms, then developed a Boolean algorithm using these terms, as well as terms related to EMR, and terms related to case identification (Multimedia Appendix 1). Finally, we searched the following three databases: Embase (1974 to May 2023), Ovid MEDLINE (1946 to May 2023), and APA PsycInfo (1806 to May 2023) for peer-reviewed papers and exported the results of the search to a reference manager program (Zotero).

Selecting Studies

Identified papers were screened in two stages. First, titles and abstracts were screened by two reviewers working independently to determine whether they met our established eligibility criteria. Articles were included if they were retrieved by the Boolean search and presented a case definition, involved depression and EMR, were published between January 2000 and May 2023 and were written in English. We excluded papers that only used administrative databases, as this study is focused on case phenotyping using EMR data. Next, full articles were reviewed for all abstracts that both reviewers identified as eligible. This review was done by two reviewers working independently. To be included, studies had to use EMR for phenotyping and use inpatient data, and the case definition developed had to be for depression. The inpatient data source requirement was added because of differences in coding standards between primary care and inpatient settings. Disagreements at either screening stage were resolved by consensus, and if necessary, a third reviewer was consulted. We searched the references of all included articles for additional eligible articles, which we then screened using the same criteria. The search was designed to include all papers that utilized an algorithm phenotyping for depression with an EMR. The two most common methods were NLP and ML, which included but were not limited to. The search terms used to identify this category were not specific to a type of algorithm or method of case identification, as the purpose was to include a broad range of variations in phenotypic methodology (Multimedia Appendix 1).

Extracting Data

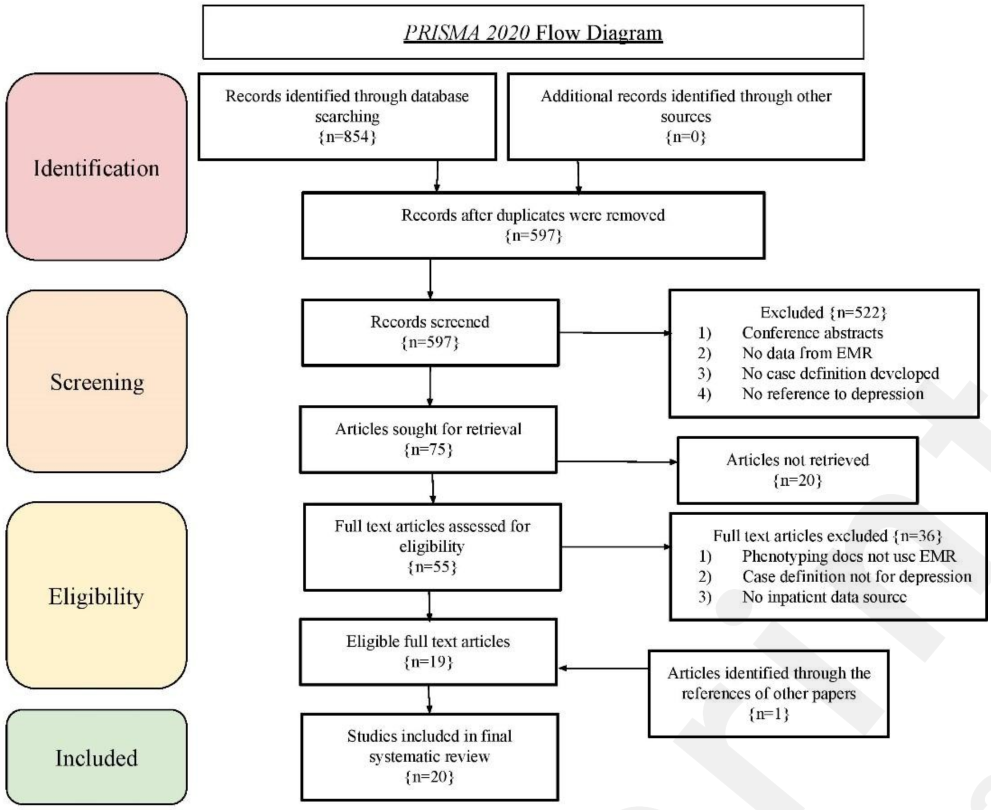
We adapted an existing data extraction form (Multimedia Appendix 4, Lee et al. (2021)) [18] to collect the results of our review. Data was extracted by one reviewer and then confirmed by a second reviewer. Components we extracted include study characteristics (country, year, inpatient or outpatient setting), specific data source and details of the data, validation methodology (e.g., chart review), as well as detailed descriptions of the phenotype developed, the methods used, and the purpose for the case definition. We recorded the performance of the developed algorithms as reported in each study. We recorded the elements of EMR used, whether other databases or diagnostic codes were used and whether AI techniques (ML, NLP) were used as binary variables. Finally, based on the study's primary objective, we classified each study into one of three categories (algorithm development, outcome analysis, and comorbidity analysis).

Results

Article Screening

The database search returned a total of 854 articles. After 257 duplicates were removed, 597 abstracts remained. Then, 522 abstracts were excluded in the title and abstract screening, leaving 75 articles for full-article review. Of these, 20 articles could not be retrieved, and 36 were excluded based on the exclusion criteria. The 19 remaining articles met all eligibility criteria and were included in the review. One additional article was identified for inclusion from the references of the included articles, resulting in 20 articles for the review [19-38]. The PRISMA flow diagram illustrating these steps is shown in Figure 1.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

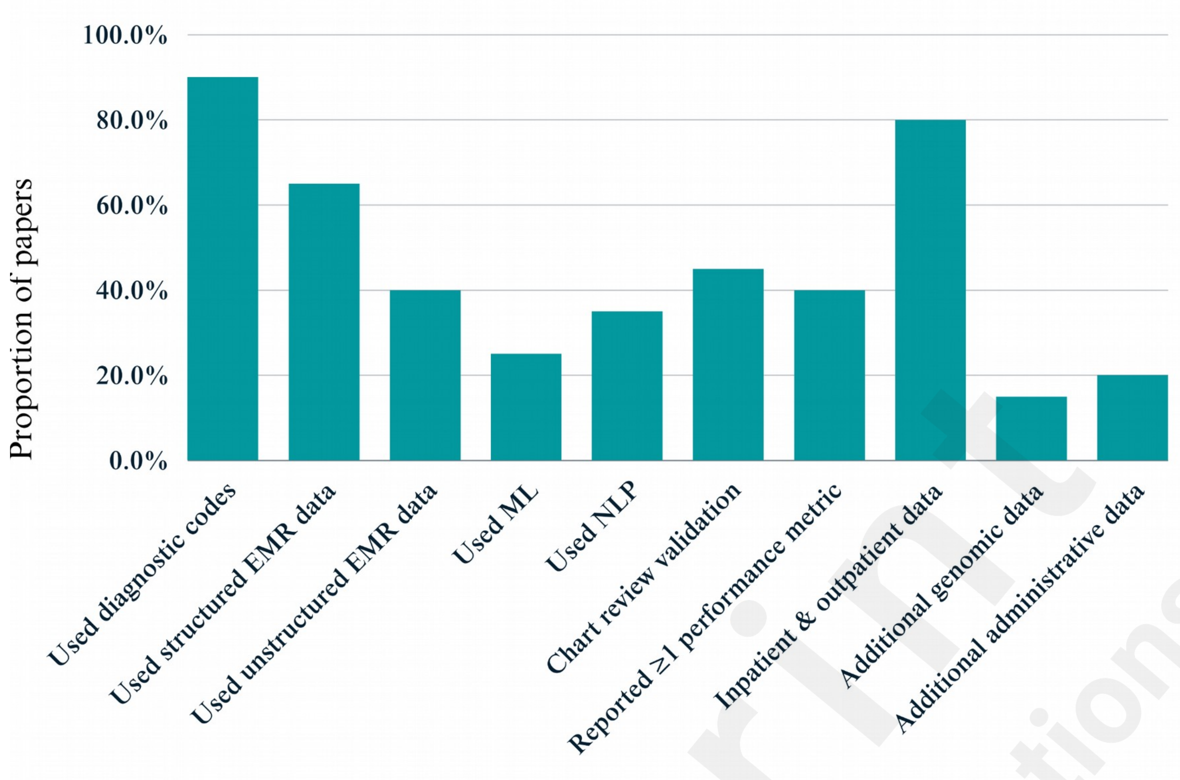


Characterizing the Identified Literature

Of the 20 studies we identified, the majority occurred in the United States (15/20, 75%). The remaining studies were from the United Kingdom (3/20, 15%) and Spain (2/20, 10%). All the studies were published in 2005 or later.

A summary of our main findings is provided in Figure 2.

Figure 2. Summary of findings



Most studies looked at inpatient and outpatient data (16/20, 80%), while fewer focused solely on inpatient data (4/20, 20%). A few studies (4/20, 20%) linked EMR data to administrative databases. These studies used structured fields of EMR and diagnostic codes found in administrative databases. They occurred in three countries (USA, UK, and Spain) and were all published in 2020 or later. Another three studies (3/20, 15%) linked EMR to genomic data (the Partners HealthCare Biobank, USA; the Michigan Genomics Initiative, USA; and the pediatric biorepository database of the Center for Applied Genomics at the Children's Hospital of Philadelphia, USA). This linkage was conducted in an epidemiological analysis study to find genetic associations between conditions. These characteristics are shown in Table 1.

Table 1. Characteristics of Included Articles.

Article Reference	Country	EMR Setting	Additional Data Sources Used
Dashti H.S. et al. (2018)	United States	Inpatient/Outpatient	Genomic data
Dorr D.A et al. (2022)	United States	Inpatient/Outpatient	None

Edgcomb J.B. et al. (2021)	United States	Inpatient/Outpatient	None
Estiri H. et al. (2021)	United States	Inpatient/Outpatient	None
Fang Y. et al. (2022)	United States	Inpatient/Outpatient	Genomic data
Fernandes A.C. et al. (2018)	United Kingdom	Inpatient/Outpatient	None
Goulet J.L. et al. (2005)	United States	Inpatient/Outpatient	None
Hong C. et al. (2021)	United States	Inpatient/Outpatient	Administrative data
Ingram W.M. et al. (2020)	United States	Inpatient/Outpatient	None
Khapre S. et al. (2021)	United Kingdom	Inpatient/Outpatient	Administrative data
Mar J. et al. (2020)	Spain	Inpatient/Outpatient	Administrative data
Mason A. et al. (2022)	United Kingdom	Inpatient/Outpatient	None
Mayer M. et al. (2017)	Spain	Inpatient/Outpatient	None
McCoy T.H. et al. (2018)	United States	Inpatient	None
Parthipan A. et al. (2019)	United States	Inpatient	None
Perlis R.H. et al. (2011)	United States	Inpatient/Outpatient	None
Slaby I. et al. (2022)	United States	Inpatient	Genomic data
Tvaryanas A.P. et al. (2014)	United States	Inpatient/Outpatient	None
Yusufov M. et al. (2022)	United States	Inpatient/Outpatient	Administrative data
Zhou L. et al. (2015)	United States	Inpatient	None

Most of the identified studies (18/20, 90%) used diagnostic codes in their case definition for depression. The most common codes employed were ICD-9, followed by ICD-10. In many studies, the diagnostic code case definitions were combined with structured data elements, such as patient demographics (sex, age, etc.), laboratory results, medications, and procedures. For example, procedures were coded with Current Procedural Terminology (CPT) codes and other types of classifications. Structured EMR data was used in 13/20 studies (65%). Fewer studies (8/20, 40%) incorporated unstructured data elements, such as clinical notes. To analyze these elements, some studies used standardized vocabularies, such as the Unified Medical Language System (UMLS), to develop lists of keywords. Most studies using unstructured data employed Natural Language Processing (NLP) techniques to analyze the free-text data in unstructured EMR fields (7/20, 35%). NLP is commonly used on free-text medical data to transform the data into a structured format that can be processed using statistical techniques and Machine Learning (ML). A quarter of the identified studies (5/20, 25%) used ML to develop phenotyping algorithms. ML models included logistic regression, random forest, and propositional rule learners. Table 2 contains details about the algorithms defined in each study.

Table 2. Summary of Algorithms.

ticle Reference	Diagnostic Codes?	EMR – Structured?	EMR – Unstructured?	ML?	NLP?	Validation Methodology	Sensitivity	Specificity	PPV	AUC
shti H.S. et al. 018)	No	Yes	Yes	Yes	Yes	Chart review	0.81		0.90	
rr D.A. et al. 022)	Yes	Yes	No	No	No	Not specified				
gcomb J.B. et al. 021)	Yes	No	No	No	No	Not specified				
tiri H. et al. 021)	Yes	No	No	No	No	Not specified				
ng Y. et al. (2022)	Yes	Yes	No	No	No	Not specified				
rnandes A.C. et al. 018)	Yes	Yes	No	No	No	Not specified				
ulet J.L. et al. 005)	Yes	No	No	No	No	Chart review	0.45	0.90		
ng C. et al. 021)*	Yes	Yes	No	Yes	No	Chart review				0.83
gram W.M. et al. 020)	Yes	Yes	No	No	No	Convergent validity				
apre S. et al. 021)	Yes	Yes	Yes	No	Yes	Not specified				
ar J. et al. (2020)	Yes	Yes	Yes	Yes	No	Chart review				0.80

ason A. et al. (2022)	Yes	No	No	No	No	Not specified				
ayer M. et al. (2017)	Yes	Yes	No	No	No	Not specified				
cCoy T.H. et al. (2018)	Yes	No	No	No	No	Not specified				
rthipan A. et al. (2019)	Yes	Yes	Yes	No	Yes	Chart review				
rlis R.H. et al. (2011)	Yes	Yes	Yes	No	No	Chart review	0.39	0.95	0.78	0.87
aby I. et al. (2022)	Yes	Yes	Yes	No	Yes	Chart review			0.95	
aryanas A.P. et al. (2014)	Yes	No	No	No	No	Not specified				
sufov M. et al. (2022)	Yes	Yes	Yes	No	Yes	Chart review	0.85	0.95		
ou L. et al. (2015)	No	No	Yes	Yes	Yes	Chart review	0.87	0.92		

*AUCPRC and F-score were only available for Hong C. et al. (2021). The best algorithm in that article had an AUCPRC of 0.90 and an F-score of 0.81.

Only 9 studies (45%) conducted a chart review to produce a reference standard to which to compare phenotyping results. Since most of the identified studies (14/20, 70%) were conducted with a larger goal of which phenotyping depression was a small part, many did not provide much information on the methods of their phenotyping. Most studies did not report any metrics measuring the diagnostic accuracy of developed phenotyping algorithms. The 6 metrics reported were sensitivity, specificity, positive predictive value (PPV), area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AUCPRC), and F-score. No studies reported negative predictive value (NPV). These metrics are displayed in Table 2.

We classified each study into one of three general purposes: algorithm development, comorbidity analysis, and outcome analysis. A small percentage of the identified studies (6/20, 30%) were conducted for algorithm development. These studies did not look at applications of the phenotyping algorithms developed; instead, they focused on phenotyping methods and algorithm performance. The rest of the studies used a case definition for depression as a step toward a larger goal. For 9 of these studies (9/20, 45%), this goal was outcome analysis or analyzing the effect of depression on patient outcomes, such as mortality, suicide attempts, and psychotherapy receipt. For the remaining studies (5/20, 25%), the goal was comorbidity analysis, examining the prevalence of depression as a comorbidity of other conditions. The comorbidities studied included HIV, hepatitis C, and cancer. Figure 3 shows the percentage of identified studies conducted for each objective. Outcome analysis studies have become more prevalent in recent years. Six were published between 2020 and 2022, up from 3 between 2000 and 2019. In addition, algorithms used for depression phenotyping in EMR have become more prevalent since 2017. This information is shown in Figure 4.

Figure 3. Percentage of studies by objective.

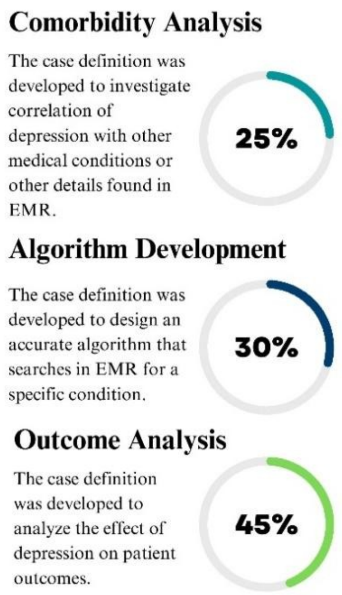
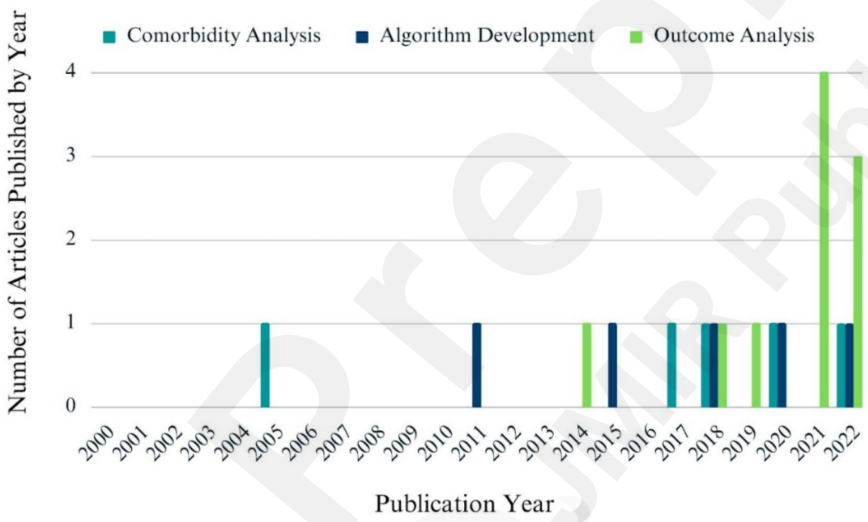


Figure 4. Objective of studies by year.



Discussion

Principal Results

In this review, we found 20 papers describing phenotyping algorithms for depression in inpatient EMR data. Most of these algorithms were case definitions using diagnostic codes, specifically ICD-9. This reflects that ICD codes are commonly used for billing purposes in the United States and are the most frequently used diagnostic codes in EMR worldwide [39]. ICD coded data is thus widely available, making it a practical choice when developing a case definition. However, case definitions using diagnostic codes achieved worse sensitivity than algorithms that only used other fields of EMR. Many algorithms also used structured EMR data [19, 20, 23, 24, 26-29, 31, 33-35, 37], but fewer used unstructured data [19, 28, 29, 33-35, 37, 38]. NLP and ML techniques were used by a minority of algorithms (NLP [19, 28, 33, 35, 37, 38], ML [19, 26, 29, 38]). These types of ML applications are relatively new and are receiving much attention from researchers [40]. The algorithms that used ML performed well on all the metrics they reported (Sensitivity 0.81-0.87, Specificity 0.82, PPV 0.90, AUC 0.80-0.83). This suggests that the information in free-text EMR data is valuable for developing accurate phenotyping algorithms. It also supports the effectiveness of ML techniques for phenotyping of depression. This is likely an area that will be explored further in future research.

Many of the papers we found did not include a chart review. If algorithms are not validated against a reference standard such as a chart review, their accuracy remains unknown. Most papers also did not report metrics measuring the validity of the algorithms developed. This limits the potential of these algorithms for application in precision healthcare. Conducting validation studies on the algorithms presented in these papers would make them more rigorous. Of the papers that did report metrics few reported sensitivity, specificity, and PPV together. This could result in skewed interpretations of phenotype performance, as a high sensitivity may come at the cost of a low PPV (or vice versa) for instance.

Based on the validity reported in these papers, EMR appears promising as a phenotyping tool for depression; however, few studies have reported metrics of diagnostic accuracy of EMR algorithms, especially comprehensive metrics to fully assess performance. Future validation studies conducted on existing case definitions would be valuable in establishing their validity and bringing these types of phenotyping algorithms to the attention of medical professionals and public health analysts. ML and NLP are small but growing areas within phenotyping research. More work could be done using these techniques on the unstructured fields in EMR, alone or in combination with other fields. Finally, as most of the studies we found were performed in the United States on US EMR data, it is to be determined how generalizable the identified case definitions are to data recorded in other jurisdictions. Both the standards of care and the methods of reporting diagnoses vary widely between healthcare systems, which could result in an algorithm only being valid in the region in which it was developed. There is a need for further research validating existing case definitions across healthcare regions or creating new case definitions specific to the EMR systems of other countries.

Limitations

Some relevant articles may have been missed as we only searched three databases. It is also possible that our search terms were not sufficiently broad to return every pertinent article. We also only considered peer-reviewed papers, not grey literature. However, we developed our search strategy in consultation with librarians and experts in the field with experience performing systematic reviews. For these reasons, we believe our search was sufficient to find articles for the review.

Conclusions

We examined current algorithms for phenotyping depression in inpatient EMR. This is an area in which more research needs to be performed. It is difficult to accurately identify cases of depression in EMR data because depression is inconsistently coded, as there is some subjectivity in its diagnosis. Diagnostic codes are primarily used in the algorithms we found, but ML on free-text data has recently achieved promising results. Most of the algorithms were developed in the United States; how well they will perform on data from other jurisdictions is yet to be known. In addition, many identified algorithms have yet to be validated against a reference standard, or their performance was not reported. To be useful for public health research, case definitions must be validated; this is an area in which future work is needed. From this study, we conclude that EMR has the potential to provide valuable insight into the indicators of depression, as well as its prevalence, common comorbidities, and associated outcomes. Future research into applying ML and NLP techniques on unstructured EMR data and studies to ascertain the validity and generalizability of existing phenotyping algorithms will be valuable in establishing EMR-based case phenotyping as a reliable tool in precision public health.

Acknowledgments

We are grateful to Natalie Wiebe for her help developing the search strategy, and to Dr. Seungwon Lee for creating the data extraction form.

Funding

This work is supported by a Foundation Grant, led by Hude Quan, through the Canadian Institutes of Health Research.

https://support.jmir.org/hc/en-us/articles/360015982471-How-should-the-Acknowledgments-section-be-formatted-

Conflicts of Interest

None declared.

Abbreviations

AI: artificial intelligence

AUC: area under the receiver operating characteristic curve

AUCPRC: area under the precision-recall curve

CPT: Current Procedural Terminology

EMR: electronic medical records

ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification

ML: machine learning

NLP: natural language processing

NPV: negative predictive value

PPV: positive predictive value

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

UMLS: Unified Medical Language System

Multimedia Appendix 1: Developed search terms

https://docs.google.com/document/d/1y0bL0EqpLJwse_m3kedMwpR6DhFId2mp1U_Jca1p5xE/edit?usp=sharing

Multimedia Appendix 2: Summary spreadsheet of identified articles

https://docs.google.com/spreadsheets/d/1KRjqGs4397tEh73Cg7QE9pieTPgeEU4jI4ePgTjzc4U/edit?usp=sharing

References

1. Slomp M, Jacobs P, Ohinmaa A, Bland R, Block R, Dewa CS, Wang C. The Distribution of Mental Health Service Costs for Depression in the Alberta Population. Can J Psychiatry SAGE Publications Inc; 2012 Sep 1;57(9):564–569. doi: 10.1177/070674371205700907
2. Chiu M, Vigod S, Rahman F, Wilton AS, Lebenbaum M, Kurdyak P. Mortality risk associated with psychological distress and major depression: A population-based cohort study. Journal of Affective Disorders 2018 Jul 1;234:117–123. doi: 10.1016/j.jad.2018.02.075
3. Offerman SR, Rauchwerger AS, Nishijima DK, Ballard DW, Chettipally UK, Vinson DR, Reed ME, Holmes JF. Use of an electronic medical record "dotphrase" data template for a prospective head injury study. West J Emerg Med. 2013 Mar;14(2):109-13. doi: 10.5811/westjem.2012.11.13400
4. Cohen S, Jannot AS, Iserin L, Bonnet D, Burgun A, Escudié JB. Accuracy of claim data in the identification and classification of adults with congenital heart diseases in electronic medical records. Arch Cardiovasc Dis. 2019 Jan;112(1):31-43. doi: 10.1016/j.acvd.2018.07.002
5. Greiver M, Barnsley J, Glazier RH, Harvey BJ, Moineddin R. Measuring data reliability for preventive services in electronic medical records. BMC Health Serv Res. 2012 May 14;12:116. doi: 10.1186/1472-6963-12-116
6. Perlis RH, Iosifescu DV, Castro VM, Murphy SN, Gainer VS, Minnier J, Cai T, Goryachev S, Zeng Q, Gallagher PJ, Fava M, Weilburg JB, Churchill SE, Kohane IS, Smoller JW. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. Psychol Med. 2012 Jan;42(1):41-50. doi: 10.1017/S0033291711000997.
7. LaFleur J, McAdam-Marx C, Alder SS, Sheng X, Asche CV, Nebeker J, Brixner DI, Silverman SL. Clinical risk factors for fracture among postmenopausal patients at risk for

fracture: a historical cohort study using electronic medical record data. J Bone Miner Metab. 2011 Mar;29(2):193-200. doi: 10.1007/s00774-010-0207-y

8. Patel RC, Amorim G, Jakait B, Shepherd BE, Mocello AR, Musick B, Bernard C, Onono M, Bukusi EA, Wools-Kaloustian K, Cohen CR, Yiannoutsos CT; Implant/Efavirenz Study Group and the East Africa IeDEA Regional Consortium. Pregnancies among women living with HIV using contraceptives and antiretroviral therapy in western Kenya: a retrospective, cohort study. BMC Med. 2021 Aug 13;19(1):178. doi: 10.1186/s12916-021-02043-z

9. Canfell OJ, Kodiyattu Z, Eakin E, Burton-Jones A, Wong I, Macaulay C, Sullivan C. Real-world data for precision public health of noncommunicable diseases: a scoping review. BMC Public Health. 2022 Nov 24;22(1):2166. doi: 10.1186/s12889-022-14452-7

10. Carreira H, Williams R, Strongman H, Bhaskaran K. Identification of mental health and quality of life outcomes in primary care databases in the UK: a systematic review. BMJ Open British Medical Journal Publishing Group; 2019 Jul 1;9(7):e029227. PMID:31270119

11. Larvin H, Peckham E, Prady SL. Case-finding for common mental disorders in primary care using routinely collected data: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2019 Oct 1;54(10):1161–1175. doi: 10.1007/s00127-019-01744-4

12. Wang B, Kristiansen E, Fagerlund AJ, Zanaboni P, Hägglund M, Bärkås A, Kujala S, Cajander Å, Blease C, Kharko A, Huvila I, Kane B, Johansen MA. Users' Experiences With Online Access to Electronic Health Records in Mental and Somatic Health Care: Cross-Sectional Study. J Med Internet Res. 2023 Dec 25;25:e47840. doi: 10.2196/47840.

13. Marcus SC, Hermann RC, Frankel MR, Cullen SW. Safety of Psychiatric Inpatients at the Veterans Health Administration. Psychiatr Serv. 2018 Feb 1;69(2):204-210. doi: 10.1176/appi.ps.201700224

14. Kozubal DE, Samus QM, Bakare AA, Trecker CC, Wong HW, Guo H, Cheng J, Allen

PX, Mayer LS, Jamison KR, Kaplin AI. Separate may not be equal: a preliminary investigation of clinical correlates of electronic psychiatric record accessibility in academic medical centers. Int J Med Inform. 2013 Apr;82(4):260-7. doi: 10.1016/j.ijmedinf.2012.11.007.

15. https://www150.statcan.gc.ca/n1/pub/82-003-x/2012004/article/11761-eng.htm#:~:text=These%20patients%20accounted%20for%2025.5,mental%20diagnosis%20(Table%201).

16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLOS Medicine Public Library of Science; 2021 Mar 29;18(3):e1003583. doi: 10.1371/journal.pmed.1003583

17. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data: Medical Care 1998 Jan;36(1):8–27. doi: 10.1097/00005650-199801000-00004

18. Lee S, Doktorchik C, Martin EA, D'Souza AG, Eastwood C, Shaheen AA, Naugler C, Lee J, Quan H. Electronic Medical Record–Based Case Phenotyping for the Charlson Conditions: Scoping Review. JMIR Med Inform 2021 Feb 1;9(2):e23934. doi: 10.2196/23934

19. Dashti H.S., Redline S., Saxena R. Polygenic risk score identifies associations between sleep duration and diseases determined from an electronic medical record biobank. Sleep United States: Oxford University Press; 2019;42(3):zsy247. doi: 10.1093/sleep/zsy247

20. Dorr D.A., Quinones A.R., King T., Wei M.Y., White K., Bejan C.A. Prediction of Future Health Care Utilization Through Note-extracted Psychosocial Factors. Med Care United States: Lippincott Williams and Wilkins; 2022;60(8):570–578. doi: 10.1097/MLR.0000000000001742

21. Edgcomb J.B., Thiruvalluru R., Pathak J., Brooks J.O. Machine Learning to Differentiate Risk of Suicide Attempt and Self-harm After General Medical Hospitalization of Women With Mental Illness. Med Care United States: NLM (Medline); 2021;59((Edgcomb) Semel Institute for Neuroscience&Human Behavior, David Geffen School of Medicine at UCLA, Los Angeles, CA):S58–S64. doi: 10.1097/MLR.0000000000001467

22. Estiri H., Strasser Z.H., Klann J.G., Naseri P., Wagholikar K.B., Murphy S.N. Predicting COVID-19 mortality with electronic medical records. npj Digit Med United Kingdom: Nature Research; 2021;4(1):15. doi: 10.1038/s41746-021-00383-x

23. Fang Y., Fritsche L.G., Mukherjee B., Sen S., Richmond-Rakerd L.S. Polygenic Liability to Depression Is Associated With Multiple Medical Conditions in the Electronic Health Record: Phenome-wide Association Study of 46,782 Individuals. Biol Psychiatry United States: Elsevier Inc.; 2022;92(12):923–931. doi: 10.1016/j.biopsych.2022.06.004

24. Fernandes A.C., Chandran D., Khondoker M., Dewey M., Shetty H., Dutta R., Stewart R. Demographic and clinical factors associated with different antidepressant treatments: a retrospective cohort study design in a UK psychiatric healthcare setting. BMJ Open United Kingdom: NLM (Medline); 2018;8(9):e022170. doi: 10.1136/bmjopen-2018-022170

25. Goulet J.L., Fultz S.L., McGinnis K.A., Justice A.C. Relative prevalence of comorbidities and treatment contraindications in HIV-mono-infected and HIV/HCV-co-infected veterans. AIDS United Kingdom: Lippincott Williams and Wilkins (250 Waterloo Road, London SE1 8RD, United Kingdom); 2005;19(SUPPL. 3):S99–S105. doi: 10.1097/01.aids.0000192077.11067.e5

26. Hong C., Rush E., Liu M., Zhou D., Sun J., Sonabend A., Castro V.M., Schubert P., Panickan V.A., Costa L., He Z., Link N., Hauser R., Gaziano J.M., Murphy S.N., Ostrouchov G., Ho Y.-L., Begoli E., Lu J., Cho K., Liao K.P., Cai T. Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data.

npj Digit Med United Kingdom: Nature Research; 2021;4(1):151. doi: 10.1038/s41746-021-00519-z

27. Ingram W.M., Baker A.M., Bauer C.R., Brown J.P., Goes F.S., Larson S., Zandi P.P. Defining major depressive disorder cohorts using the EHR: Multiple phenotypes based on ICD-9 codes and medication orders. Neurol Psychiatry Brain Res Netherlands: Elsevier GmbH; 2020;36((Ingram, Zandi) Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States):18–26. doi: 10.1016/j.npbr.2020.02.002

28. Khapre S., Stewart R., Taylor C. An evaluation of symptom domains in the 2 years before pregnancy as predictors of relapse in the perinatal period in women with severe mental illness. Eur Psychiatry United Kingdom: NLM (Medline); 2021;64(1):e26. doi: 10.1192/j.eurpsy.2021.18

29. Mar J., Gorostiza A., Ibarrondo O., Cernuda C., Arrospide A., Iruin A., Larranaga I., Tainta M., Ezpeleta E., Alberdi A. Validation of Random Forest Machine Learning Models to Predict Dementia-Related Neuropsychiatric Symptoms in Real-World Data. J Alzheimer's Dis Netherlands: IOS Press BV (E-mail: sales@iospress.com); 2020;77(2):855–864. doi: 10.3233/JAD-200345

30. Mason A., Irving J., Pritchard M., Sanyal J., Colling C., Chandran D., Stewart R. Association between depressive symptoms and cognitive-behavioural therapy receipt within a psychosis sample: a cross-sectional study. BMJ Open United Kingdom: NLM (Medline); 2022;12(5):e051873. doi: 10.1136/bmjopen-2021-051873

31. Mayer MA, Gutierrez-Sacristan A, Leis A, Pe&#241 DL, A S, Sanz F, Furlong LI. Using Electronic Health Records to Assess Depression and Cancer Comorbidities. Informatics for Health: Connected Citizen-Led Wellness and Population Health IOS Press; 2017;236–240. doi: 10.3233/978-1-61499-753-5-236

32. McCoy T.H., Yu S., Hart K.L., Castro V.M., Brown H.E., Rosenquist J.N., Doyle A.E., Vuijk

P.J., Cai T., Perlis R.H. High Throughput Phenotyping for Dimensional Psychopathology in Electronic Health Records. Biol Psychiatry United States: Elsevier USA; 2018;83(12):997–1004. doi: 10.1016/j.biopsych.2018.01.011

33. Parthipan A., Banerjee I., Humphreys K., Asch S.M., Curtin C., Carroll I., Hernandez-Boussard T. Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. PLoS ONE United States: Public Library of Science (E-mail: plos@plos.org); 2019;14(2):e0210575. doi: 10.1371/journal.pone.0210575

34. Perlis R.H., Iosifescu D.V., Castro V.M., Murphy S.N., Gainer V.S., Minnier J., Cai T., Goryachev S., Zeng Q., Gallagher P.J., Fava M., Weilburg J.B., Churchill S.E., Kohane I.S., Smoller J.W. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. Psychol Med United Kingdom; 2012;42(1):41–50. PMID:21682950

35. Slaby I., Hain H.S., Abrams D., Mentch F.D., Glessner J.T., Sleiman P.M.A., Hakonarson H. An electronic health record (EHR) phenotype algorithm to identify patients with attention deficit hyperactivity disorders (ADHD) and psychiatric comorbidities. J Neurodevelopmental Disord United Kingdom: BioMed Central Ltd; 2022;14(1):37. doi: 10.1186/s11689-022-09447-9

36. Tvaryanas AP, Maupin GM. Risk of incident mental health conditions among critical care air transport team members. Ben-Ezra B-E Beninati, Bliese, Brewin, Gibbons, Gibbons, Gibbons, Hoge, Hoge, Jekel, Jones, Kerasiotis, Kolkow, Maguen, Milliken, Pavlin, Rice, Rona, Shea, Thomas, editor. Aviation, Space, and Environmental Medicine US: Aerospace Medical Assn; 2014;85(1):30–38. doi: 10.3357/ASEM.3782.2014

37. Yusufov M., Pirl W.F., Braun I., Tulsky J.A., Lindvall C. Natural Language Processing for Computer-Assisted Chart Review to Assess Documentation of Substance use and Psychopathology in Heart Failure Patients Awaiting Cardiac Resynchronization Therapy. J

Pain Symptom Manage United States: Elsevier Inc.; 2022;64(4):400–409. doi: 10.1016/j.jpainsymman.2022.06.007

38. Zhou L., Baughman A.W., Lei V.J., Lai K.H., Navathe A.S., Chang F., Sordo M., Topaz M., Zhong F., Murrali M., Navathe S., Rocha R.A. Identifying Patients with Depression Using Free-text Clinical Documents. Stud Health Technol Inform Netherlands; 2015;216((Zhou, Lei, Chang, Sordo, Rocha) Clinical Informatics, Partners eCare, Partners Healthcare Inc. Boston, MA, USA):629–633. PMID:26262127

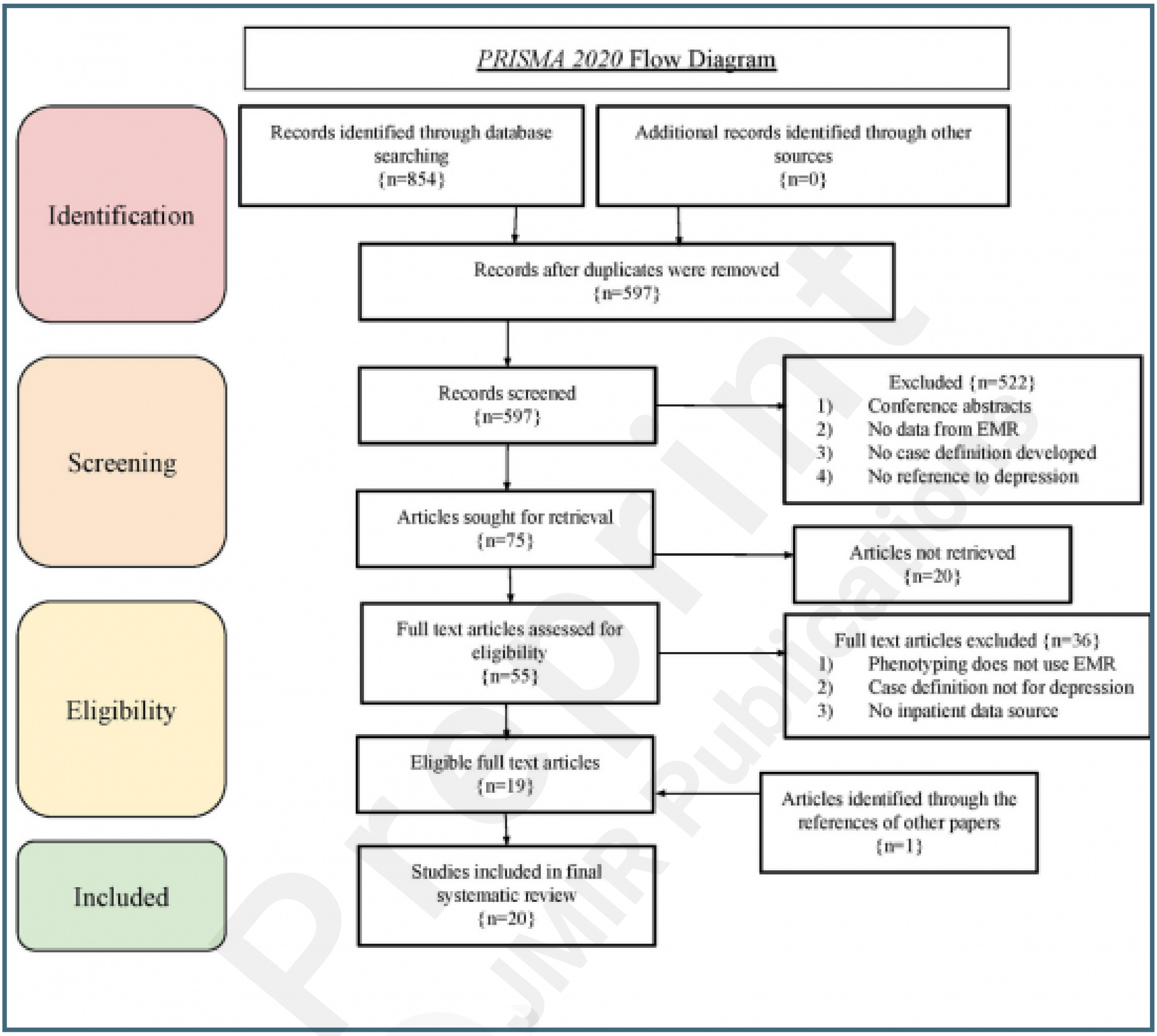
39. O'Malley KJ, Cook KF, Price MD, et al. Measuring diagnoses: ICD code accuracy. Health Serv Res. 2005;40(5 Pt 2):1620–39.

40. Le Glaz A, Haralambous Y, Kim-Dufor DH, et al. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *J Med Internet Res*. 2021;23(5):e15708. Published 2021 May 4. doi:10.2196/15708

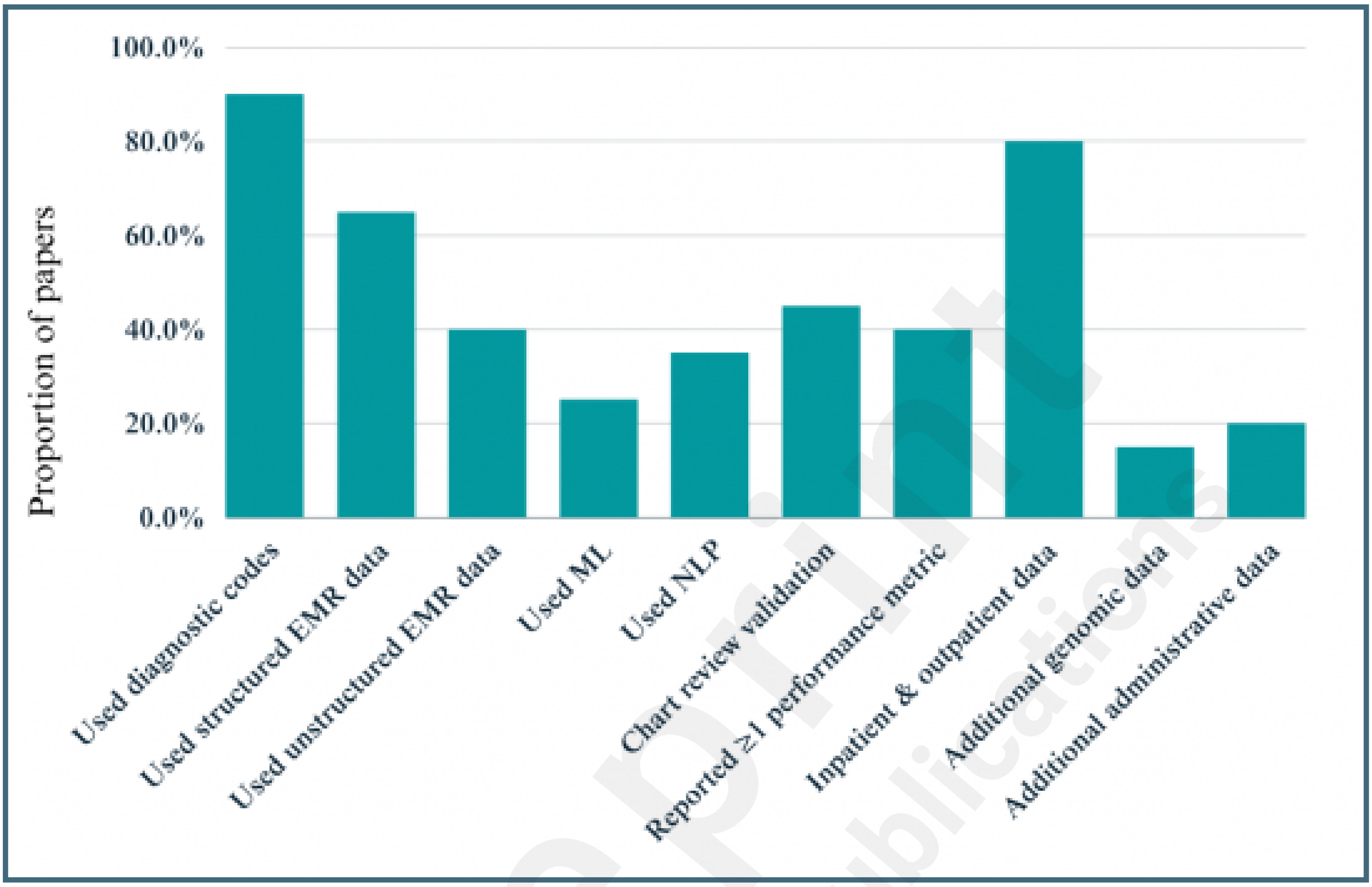
Supplementary Files

Figures

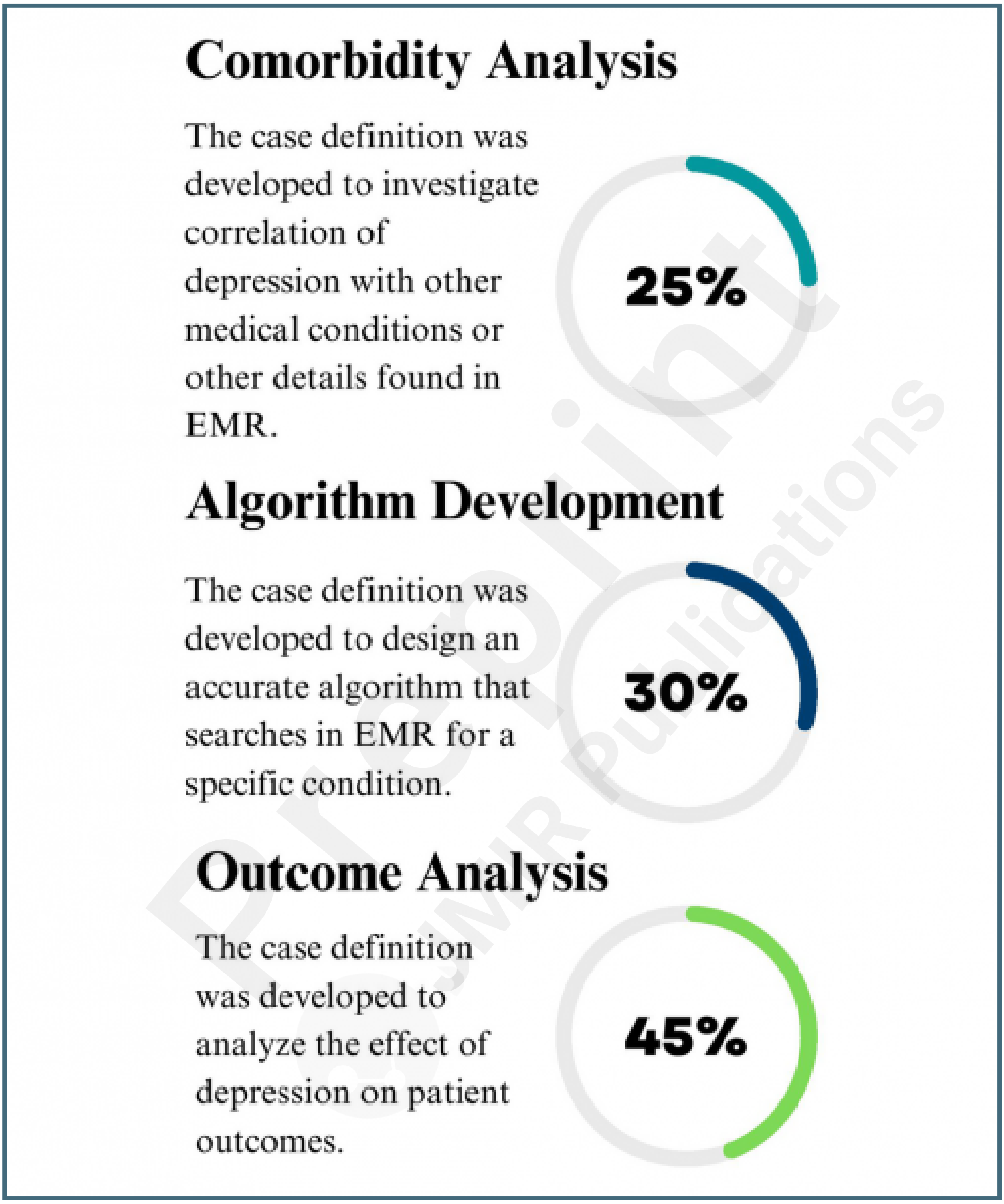
Preferred Reporting Items for Systematic reviews and Meta-Analyses flow diagram.



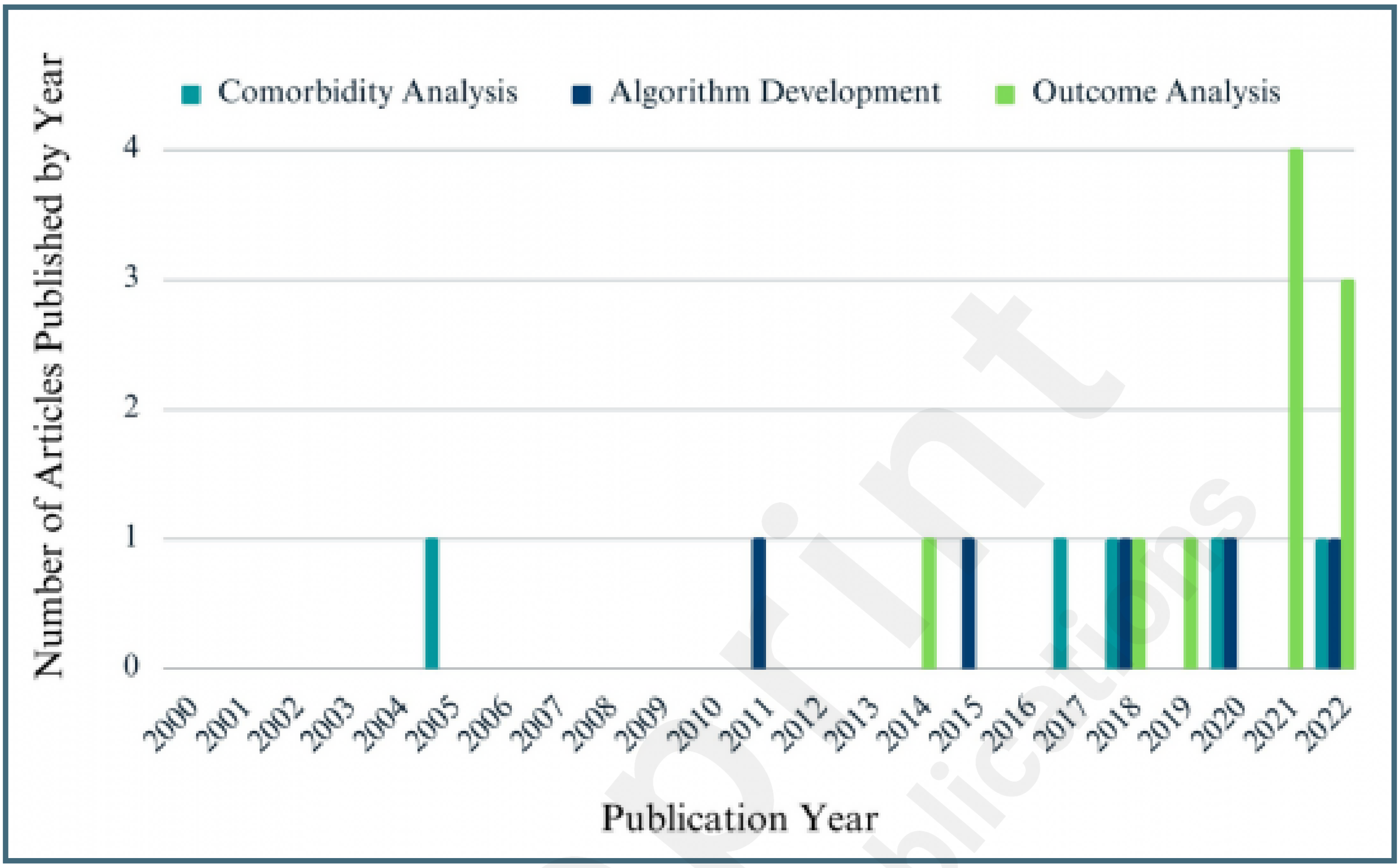
Summary of findings.



Percentage of studies by objective.



Objective of studies by year.



Multimedia Appendixes

Developed search terms.
URL: http://asset.jmir.pub/assets/6de2c1e2c1ccb360bb73226042f9361b.doc

Summary spreadsheet of identified articles.
URL: http://asset.jmir.pub/assets/f50be4b75d6b4dcbd89d1cadb70b4bbc.xls

CONSORT (or other) checklists

PRISMA checklist.
URL: http://asset.jmir.pub/assets/aefd18316f9b8ed3957826cc579f83d7.pdf