

IDENTIFICATION OF QUESTIONABLE EXCLUSION CRITERIA IN MENTAL DISORDER CLINICAL TRIALS USING A MEDICAL ENCYCLOPEDIA*

HANDONG MA

*Department of Biomedical Informatics, Columbia University, 622 West 168 Street, PH-20
New York, NY, 10032, USA*

Email: handongma.work@gmail.com

CHUNHUA WENG

*Department of Biomedical Informatics, Columbia University, 622 West 168 Street, PH-20
New York, NY, 10032, USA*

Email: cw2384@cumc.columbia.edu

Precision medicine requires precise evidence-based practice and precise definition of the patients included in clinical studies for evidence generalization. Clinical research exclusion criteria define confounder patient characteristics for exclusion from a study. However, unnecessary exclusion criteria can weaken patient representativeness of study designs and generalizability of study results. This paper presents a method for identifying questionable exclusion criteria for 38 mental disorders. We extracted common eligibility features (CEFs) from all trials on these disorders from ClinicalTrials.gov. Network Analysis showed scale-free property of the CEF network, indicating uneven usage frequencies among CEFs. By comparing these CEFs' term frequencies in clinical trials' exclusion criteria and in the PubMed Medical Encyclopedia for matching conditions, we identified unjustified potential overuse of exclusion CEFs in mental disorder trials. Then we discussed the limitations in current exclusion criteria designs and made recommendations for achieving more patient-centered exclusion criteria definitions.

* This study was funded by National Library of Medicine grant R01LM009886 (Bridging the semantic gap between clinical research eligibility criteria and clinical data).

1. Introduction

Randomized controlled trials (RCT) produce high-quality evidence but often lack patient representativeness of the real-world population. Clinical research eligibility criteria define the characteristics of a research volunteer for study inclusion or exclusion. Typically, exclusion reasons relate to age, gender, ethnicity, complex comorbidities, conflicting interventions, or patient preference¹. Although exclusion criteria do not bias the comparison between intervention and control groups, which reflects a trial's internal validity, exclusion criteria can impair the external validity of a trial^{2,3}. It has been shown in various disease domains that clinical trial participants are often not representative of the real-world patient population to which an RCT is intended to apply, and that the lack of patient representativeness has impaired the generalizability of clinical trials^{3,4}.

Thus, it is imperative to develop methods for justifying the exclusion criteria in clinical trials. However, this task is fraught with challenges. First, many eligibility criteria are vague and complex¹ and cannot be easily represented in a computable format that allows for automated screening of unjustifiable exclusion criteria⁵. Second, clinical researchers often do not have a sufficiently precise picture of the real-world patient population to make informed decisions about exclusion criteria. Although the wide adoption of Electronic Health Record (EHR) make this idea more promising than ever⁶⁻⁹, aggregating EHR data to profile the real-world patient population is a nontrivial exercise, due to common data fragmentation and data quality problems¹⁰. Therefore, it is worthwhile to explore alternatives to the EHR-based data-driven approach, especially through combining different data sources in order to increase patient representativeness of clinical trial eligibility criteria. This paper presents the feasibility of such a knowledge-based approach, using PubMed Health Medical Encyclopedia knowledge. PubMed Health Medical Encyclopedia (hereinafter, PubMed Encyclopedia) is a service created by the National Center for Biotechnology Information (NCBI), and made accessible by the U.S. National Library of Medicine (NLM), to provide summaries of diseases and conditions¹¹. Such a meta-analysis with automatic data-mining methods across different data sources provides us new insights into clinical trial design and can inform precise evidence-based practice.

2. Methods

We chose mental disorder clinical trials for a proof of principle but the method should generalize to other fields of medicine. We hypothesized that the occurrence of a term in PubMed Encyclopedia for a symptom, a medication, or a chemical compound could be used to indicate its relevance to the mental disorder (condition) under consideration. For each term in each mental disorder, we compared the term frequencies in the exclusion criteria of all the clinical trials on that condition in ClinicalTrials.gov and the term's occurrence in PubMed Encyclopedia. On this basis we identified terms that occur frequently in both exclusion criteria and PubMed. We further hypothesized that a term with a certain level of frequency of use in PubMed Health Encyclopedia about a mental disorder should be deemed relevant to that disorder. Thus, its frequent use in excluding patients with this trait from clinical trials on that disorder could be questionable.

We built an exclusion criteria network including all mental disorders based on the method from Boland and Weng et al.'s previous work¹². Using that network, we identified the common exclusion criteria for mental disorders and assessed their appropriateness of use. We identified clinical trials for 84 mental disorders in the category of "Behaviors and Mental Disorders" in ClinicalTrials.gov. For each condition, using our published tag-mining algorithm¹³, we extracted all common eligibility features (CEFs) that each occurred in at least 3% of all clinical trials related to each condition in ClinicalTrials.gov. This method is capable of automatically deriving frequent UMLS tags from clinical text using part-of-speech (POS) tagger, N-grams model, and UMLS unique concept identifier. For example, we found the UMLS concept "ethanol", which belongs to the "organic chemical - pharmacologic substance" semantic type, occurred in 74.7% of the alcoholism clinical trials while occurred in only 26.8% of depression trials. For each mental disorder, we were able to generate a list of UMLS concepts with their frequencies of use in inclusion and exclusion criteria section.

We calculated the frequencies of use aggregated across all mental disorders for inclusion and exclusion purposes, respectively, for each of these CEFs. We also analyzed the frequency distribution of these CEFs by their UMLS semantic types. We constructed a two-mode network for all the 84 mental disorders and their top 20 CEFs, based on the disorder-CEF associations. Then we projected this network to a one-mode network based on CEFs using the Newman (2001) method (tnet)¹⁴, a classic method used in detecting communities in networks. The process worked by selecting one set of nodes (i.e., CEFs), and linking two nodes, if they were connected to the same node in the other set of nodes (i.e., conditions). For each mental disorder, we analyzed the distribution of the degree of all the nodes in this network to assess the usage of the CEFs in the mental disorder trials. Since most CEFs occurred equally in inclusion and exclusion criteria section, we used a mutual information filter to identify distinctive CEFs, regardless inclusion or exclusion, because Mutual Information is one of the commonly used quantities that measure independence between variables. We calculated the Mutual Information (MI) for each CEF. The formula is as follows (1):

$$I(U; C) = \sum_{e_t \in \{0,1\}} \sum_{e_c \in \{0,1\}} P(U = e_t, C = e_c) \log_2 \frac{P(U=e_t, C=e_c)}{P(U=e_t)P(C=e_c)}, \quad (1)$$

For each mental disorder, U is a random variable indicating the presence (number 1) or absence (number 0) of a CEF in every eligibility criterion (e_t) and C is a random variable representing the inclusion (number 0) or exclusion (number 1) status of the eligibility criterion (e_c). We used additive smoothing to make sure CEF unique to only one section were included in the analysis. Since we aimed to target the most informative CEFs used as exclusion criteria, we chose the CEFs with positive MI scores in exclusion criteria as candidates for future comparisons. The cutoff of MI score retained the CEFs that are more frequently used as exclusion criteria rather than inclusion criteria. We used these CEFs to represent the common confounder patient characteristics excluded by clinical trials on each condition.

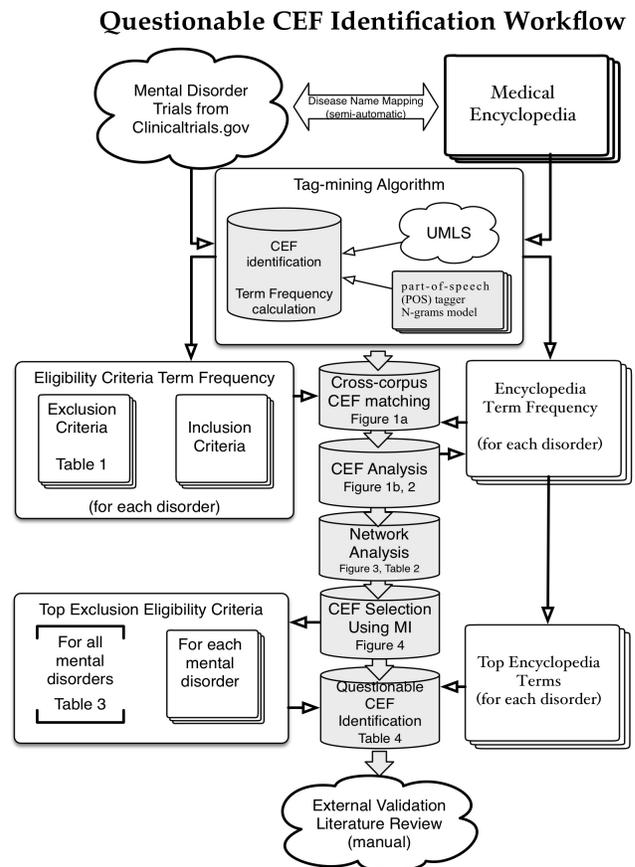
To generate the PubMed dataset, due to the heterogeneous condition names, we used a semi-automatic method to match the condition names in PubMed Health with the condition names in ClinicalTrials.gov. For example, Alzheimer disease in ClinicalTrials.gov was manually matched with

Alzheimer's disease in PubMed Health database. A total of 38 mental disorders were matched and manually validated. We processed the PubMed Encyclopedia's website content¹¹ for each matched mental disorder and used the same tag-mining algorithm¹³ to extract all the terms for risk factors, causes, symptoms, signs, exams and tests, treatment options, and complications, and obtained their aggregated frequencies across all 38 mental disorders. For each of the 38 identified mental disorders, we aligned and ranked their CEF terms by their frequencies in ClinicalTrials.gov and their occurrences in PubMed Health Medical Encyclopedia, respectively, and compared their relative importance in each content source according to the ranks. Questionable CEFs were identified with high frequencies of use in both clinical trials and PubMed Encyclopedia. The entire workflow was shown in Fig. 1(a).

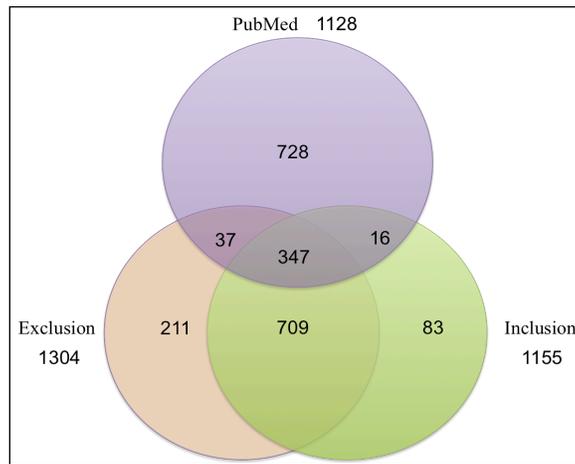
3. Results

3.1. CEF overlap between exclusion and inclusion criteria for mental disorder trials

We extracted 1304 exclusion CEFs and 1155 inclusion CEFs for all of the clinical trials for mental disorders (Fig. 1 (a)). A total of 1403 unique CEFs were identified, 1056 of which were present in both inclusion and exclusion criteria. The large overlap necessitated CEF selection to identify the most informative exclusion criteria. The top three frequent semantic types were *disease or syndrome*, *pharmacological substance*, and *finding*, respectively (Fig. 1(b)). Table 1 shows the top 10 most frequently used CEFs for *Alzheimer's disease*.

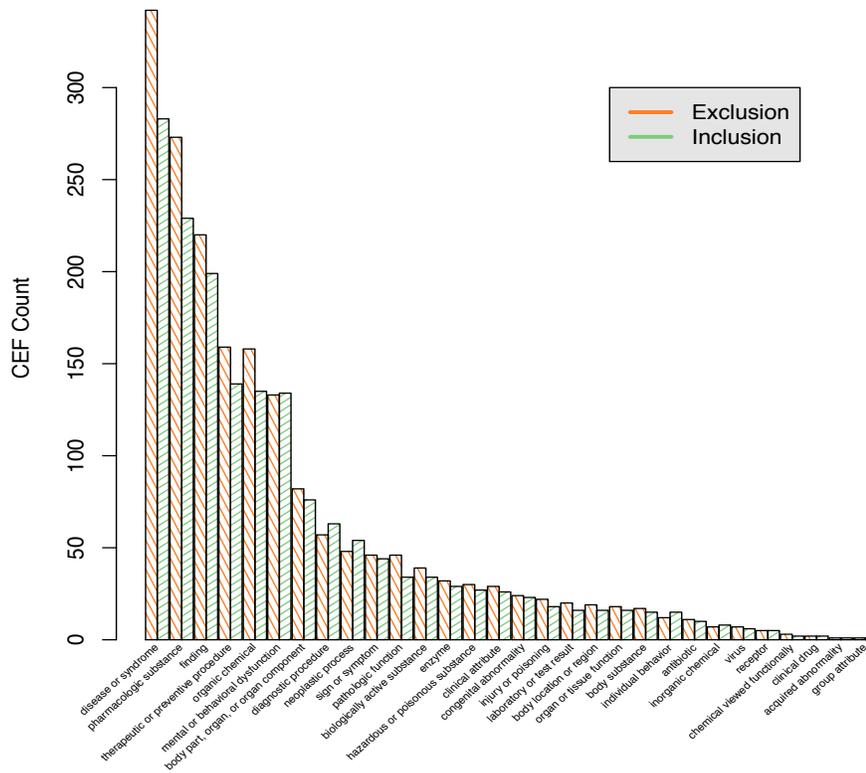


(a)



(b)

CEF distribution in UMLS semantic types



(c)

Fig 1. (a) Workflow of Identifying Questionable CEF (b) CEF Overlap between Exclusion and Inclusion CEF (c) Semantic Components between Exclusion and Inclusion CEF

Table 1. The top 10 most used exclusion CEFs for Alzheimer's diseases trials

Mostly Used Exclusion CEF	Frequency	UMLS Semantic Type
Mental disorders	29%	Mental or behavioral dysfunction
Allergy severity - severe	25%	Finding
Ethanol	23%	Organic chemical; pharmacologic substance
Depressed mood	23%	Finding; mental or behavioral dysfunction
Unstable status	21%	Finding
Cerebrovascular accident	21%	Disease or syndrome; therapeutic or preventive procedure
Magnetic resonance imaging	20%	Diagnostic procedure
Active brand of pseudoephedrine-triprolidine	17%	Organic chemical; pharmacologic substance
Pharmaceutical preparations	16%	Pharmacologic substance
Substance abuse problem	16%	Mental or behavioral dysfunction

3.2. CEF distribution among mental disorders trials

Out of the total 1403 unique CEFs, very few were used in a large number of clinical trials, while most of other CEFs were unique to one or several disorders (Fig. 2 (a)). On average, a CEF was present in 7.49 mental disorders for exclusion purposes and 6.28 for inclusion purposes. Each condition had 125.1 exclusion CEFs and 104.8 inclusion CEFs. Most of the frequently used CEFs were general factors for mental disorders (such as hypersensitivity or pharmacologic substance). Some were regularly used in clinical trials on other conditions (such as excluding gravidity, unstable states and allergy severity - severe). The top five mental disorders with the most exclusion CEFs were: *Restless Legs Syndrome* (226), *Substance Withdrawal Syndrome* (192), *Pick Disease of the Brain* (186), *Tic Disorders* (184) and *Front-temporal Dementia* (180) (Fig. 2 (b)).

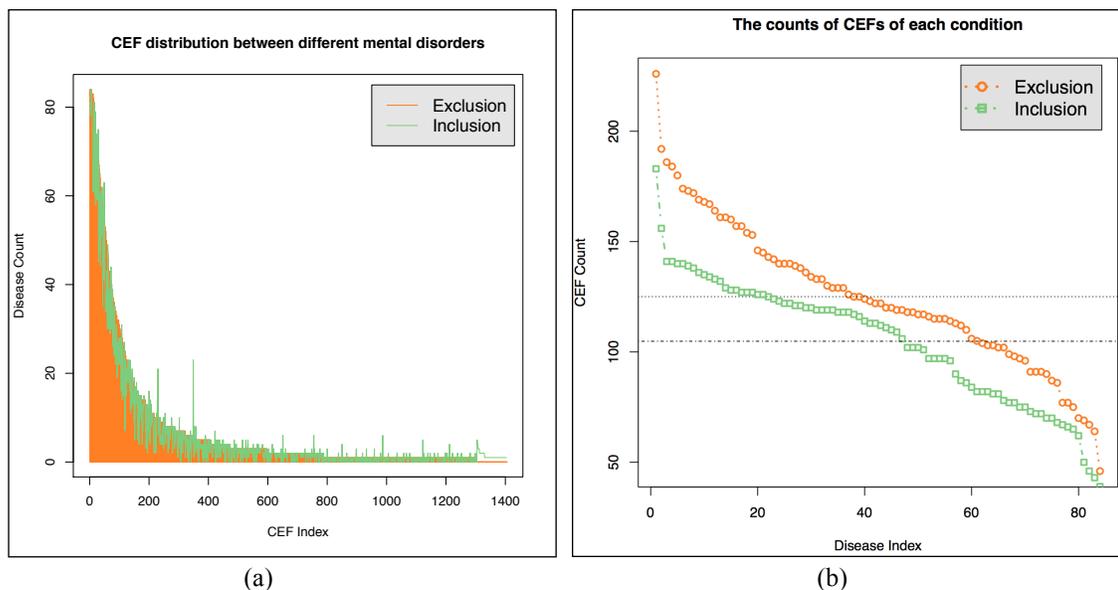


Fig 2. CEF distribution between mental disorders (a) CEFs are indexed and ranked based on disease count in exclusion section. (b) Diseases are indexed and ranked based on exclusion CEF count.

3.3. Network construction and analysis for CEFs in mental disorder trials

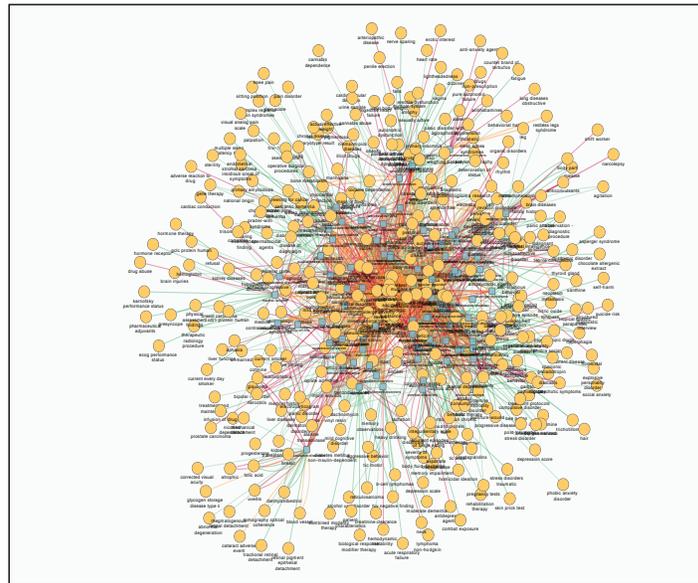
We built a two-mode network for all mental disorders and CEFs based on the Disease-CEF

linkages (Fig. 3(a)). In this network, there were two groups of nodes, the diseases and CEFs. The top 20 CEFs for each mental disorder were represented as orange ellipses and mental disorders were represented by blue round rectangles. The diseases were connected with different sets of CEFs and could be clustered based on the similarities of those connections. The edges were weighted as the frequency for a CEF to be associated with a mental disorder. Edges were red (for CEFs used only for exclusion), or orange (for CEFs used for both inclusion and exclusion), while edges of inclusion CEFs were green and dark green, respectively. In the network, we identified some hubs with higher degrees than other nodes, which indicated that a small portion of CEFs was frequently used for patient selection in most mental disorder trials. For example, diseases like *amnesia* and *bipolar disorder* shared more common CEFs with other mental disorders compared to diseases such as *associative disease* and *restless legs syndrome*, etc. In the network, most of the related disorders were clustered together using their CEF similarities such as *panic disorder* and *phobic disorder*, *Tourette syndrome* and *Tic disorder*. From the network, we also found some of the mental disorders, while not pathologically related, shared similar CEF sets. We also projected each of the three two-mode networks (Inclusion, Exclusion and PubMed) into one-mode network based on CEFs. Our analysis showed that three networks all display features of a scale-free network (Fig. 3(b)), which was similar to that of many of the real-world giant networks. The attributes of the network are listed in Table 2.

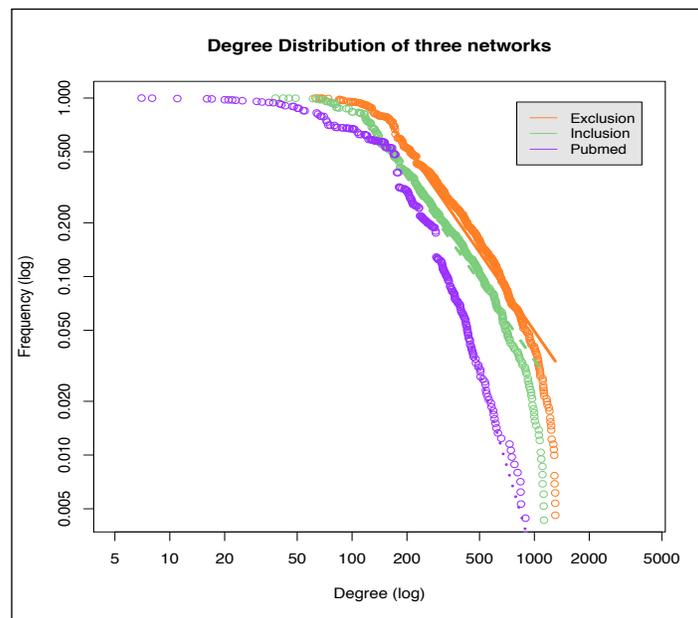
Table 2. Attributes of the One-Mode CEF Network

Data	CEF in network	Degree	One mode degree	Closeness	Betweenness
Inclusion	1155	7.60 ± 0.82	241.26 ± 12.00	6.3e-03 ± 3.6e-05	1656.99 ± 1685.7
Exclusion	1304	8.06 ± 0.84	309.10 ± 13.68	4.1e-03 ± 2.3e-05	1351.62 ± 935.4
PubMed	1128	2.27 ± 0.15	175.48 ± 8.04	8.6e-03 ± 3.9e-05	1895.01 ± 1407.6

* 95% confidence interval used



(a)



(b)

Figure 3. Network Structure (a) and Degree Distribution (b) between Mental Disorders and CEFs (b)
Regression lines are plotted as solid or dashed lines.

3.4. CEF selection using Mutual Information (MI)

Some CEFs were equally used in inclusion and exclusion criteria. Using mutual information, we

discarded CEFs with equal or higher occurrences in the inclusion section than in the exclusion criteria. Of a total of 1403 unique CEFs, only 632 had an MI value greater than 0, 568 had an MI value of 0 and 203 had an MI value below zero. The bigger the MI value is, the more frequently the CEF is for exclusion uses. To preserve all informative CEFs to match with the PubMed dataset, we selected all CEFs with a MI scores greater than 0 for further analysis. The benchmark analysis for CEFs and their MI distributions are in Fig. 4. Through this selection step, many common but non-discriminative CEFs were discarded, such as *pharmacologic substance*, *physical assessment findings*, and *intravenous infusion procedures*. In contrast, discriminative CEFs (e.g., *suicidal*, *unstable status*, *psychotic disorders*) were retained. However, it should be noted that some discriminative CEFs (such as *pregnancy tests*, *multiple endocrine neoplasia*, etc.) might be missed by this selection step.

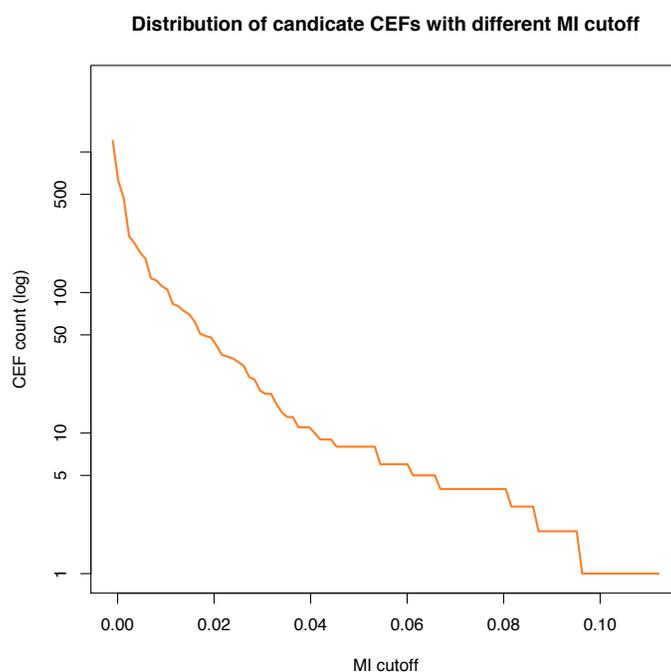


Figure 4. Mutual Information Score Distribution for CEFs

3.5. Aggregated cross-condition occurrence comparison for retained CEFs

We contrasted the aggregated occurrences of exclusion CEFs (N=1422) across the 38 matched disorders with partial results displayed in Table 3. For example, ethanol was a CEF present in all 38 mental disorders' exclusion criteria, implying 100% prevalence, and was present in the PubMed descriptions for 21 disorders (55.2%).

Table 3. A contrast of exclusion and PubMed occurrences across 38 mental disorders for top exclusion CEFs

Exclusion CEFs	Exclusion	PubMed	UMLS Semantic Types
Pharmaceutical preparations	38	38	Pharmacologic substance
Ethanol	38	21	Organic chemical; pharmacologic substance
Depressed mood	38	8	Finding; mental or behavioral dysfunction
Psychotic disorders	38	3	Mental or behavioral dysfunction
Hypersensitivity	37	3	Clinical attribute; finding; pathologic function
Hepatic	36	4	Body location or region
Antipsychotic agents	35	9	Pharmacologic substance
Unipolar depression	31	4	Mental or behavioral dysfunction
Anti-depressive agents	30	12	Pharmacologic substance
Screening for cancer	30	6	Diagnostic procedure
Benzodiazepines	30	5	Organic chemical; pharmacologic substance

The average CEF prevalence among the 38 mental disorders in the exclusion criteria and PubMed were 7.33 and 1.86, respectively, so that CEFs occurred less often in PubMed than in exclusion criteria. Among the top exclusion CEFs for the 38 mental disorders, we found some candidate CEFs that simultaneously had frequent PubMed occurrences (i.e., questionable CEFs), such as *ethanol*, *malignant neoplasms*, *anti-depressive agents*, and *depressed mood*.

We also analyzed the condition-specific CEF rankings between PubMed and exclusion criteria and identified questionable CEFs that had high PubMed rankings. Some example questionable CEFs for specific sleep disorder are listed in Table 4. *Hepatic* is associated with *sleep disorder* according to PubMed Health but was frequently used for excluding patients from 6.82% of sleep disorder clinical trials. Another questionable CEF is *sleep apnea syndromes*, whose frequency in exclusion criteria of all sleep disorder trials was as high as 24.6%, was ranked as top two relevant PubMed description for sleep disorder; therefore, we should be cautious when frequently using it as exclusion criteria. Another example is hypersensitivity (to treatment), which is common in the real-world population but is frequently excluded in randomized controlled trials (i.e., frequently as high as 13.6%).

Table 4. Questionable CEFs for excluding patients in sleep disorder clinical trials

Questionable CEF	Frequency	PubMed Rank	UMLS Semantic Type
Hepatic	6.82%	1	Body location or region
Sleep apnea syndromes	24.6%	2	Disease or syndrome
Sleep apnea obstructive	12.5%	3	Biologically active substance; disease or syndrome
Narcolepsy	8.05%	3	Disease or syndrome
Caffeine	5.10%	3	Organic chemical; pharmacologic substance
Hypersensitivity	13.6%	4	Clinical attribute; finding; pathologic function
Malignant neoplasms	9.52%	4	Finding; neoplastic process
Psychotic disorders	7.56%	4	Mental or behavioral dysfunction

4. Discussion

We investigated the exclusion criteria commonly used in mental disorder trials. The top four UMLS semantic types that contained the most questionable CEFs were *pharmacologic substance*, *mental or behavioral dysfunction*, *disease or syndrome*, and *finding*. Although some exclusion criteria of these semantic types have been used for years, their use in exclusion remains unexplained especially given their high prevalence among the real-world patients, most of who have several mental comorbidities or take multiple medications concurrently. Most of the drugs are for treating depressed mood or alcohol consumption or are anti-depressive drugs. If we exclude patients with those traits, we may generate a “pure” but not “typical”¹⁵ test population, which may weaken the generalizability of these trials.

For a single mental disorder, the method proposed herein also detected several questionable CEFs. A recent study shows at least 50% of bipolar patient populations are excluded by at least one major exclusion criterion¹⁵. Using our method, we not only identified most of the exclusion criteria for bipolar disorder aggregated from previous studies (*drug abuse*, *alcohol abuse*, *significant medical conditions*, *pregnancy or lactation*, *suicidal risk* and *psychotropic medications*), but also retrieved information about which medical condition or medication was frequently used to exclude patients. *Ethanol*, *antipsychotic agents*, and *antidepressive agents* were questionable for excluding patients. This prediction corresponds to previous findings¹⁵ that drug and alcohol abuse represent the most exclusion for bipolar trials, and provides more details for locating questionable exclusion criteria.

Although this study only focused on mental disorders for the detailed analysis, this pipeline can be easily applied to other disease domains. Most parts of the analyzing pipeline are fully automatic. The clinical trial eligibility criteria and medical encyclopedia for other diseases exist in similar format as used in this study, and can be processed in a large scale. However, considering the possible uniqueness of mental disorder domain, it is necessary to clarify the predictive power of this pipeline on a larger scale and different clinical settings, especially given poorly matched corpuses between clinical trial eligibility criteria and disease encyclopedia.

Several findings of this study shed light on future eligibility criteria designs. First, the scale-free feature of disease-CEF network suggests that a small number of exclusion criteria can be standardized and reused for most mental disorder trials. Second, trials for different conditions shared similar exclusion criteria, implying that some cohort selection criteria can be reused across conditions with little modification. Third, the power of exclusion for a single clinical trial should be quantified to avoid sampling biases in clinical trial designs.

5. Conclusion

This study demonstrates the promising value of applying a knowledge-based approach to assessing the patient-centeredness of clinical trial exclusion criteria by linking different data sources, including ClinicalTrials.gov and PubMed Medical Encyclopedia. In the future, proactive analyses like this could be conducted during clinical research designs to optimize clinical research eligibility

criteria design and study participant selection to better achieve precise evidence definition⁹.

6. Acknowledgments

We thank Dr. Riccardo Miotto for sharing eTACT methods for n-gram extraction.

References

1. Ross J, Tu S, Carini S, Sim I. Analysis of eligibility criteria complexity in clinical trials. *AMIA Summits on Translational Science Proceedings*. 2010;2010:46.
2. Friedman LM, Furberg C, DeMets DL. *Fundamentals of clinical trials*. Vol 4: Springer; 2010.
3. Elting LS, Cooksley C, Bekele BN, et al. Generalizability of cancer clinical trial results: prognostic differences between participants and nonparticipants. *Cancer*. 2006;106(11):2452-2458.
4. Heiat A, Gross CP, Krumholz HM. Representation of the elderly, women, and minorities in heart failure clinical trials. *Archives of Internal Medicine*. 2002;162(15).
5. Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. *Journal of biomedical informatics*. 2010;43(3):451-467.
6. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*. 2012;13(6):395-405.
7. Hoffman S, Podgurski A. Improving health care outcomes through personalized comparisons of treatment effectiveness based on electronic health records. *The Journal of Law, Medicine & Ethics*. 2011;39(3):425-436.
8. Weng C, Li Y, Ryan P, et al. A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records. *Applied clinical informatics*. 2014;5(2):463.
9. Weng C. Optimizing Clinical Research Participant Selection with Informatics. *Trends in Pharmacological Sciences*. 2015:In Press.
10. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*. 2013;20(1):144-151.
11. Ozdas A, Shiavi RG, Silverman SE, Silverman MK, Wilkes DM. Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. *IEEE Trans Biomed Eng*. 2004;51(9):1530-1540.
12. Boland MR, Miotto R, Weng C. A Method for Probing Disease Relatedness Using Common Clinical Eligibility Criteria. *Studies in health technology and informatics*. 2013;192:481.
13. Miotto R, Weng C. Unsupervised mining of frequent tags for clinical eligibility text indexing. *Journal of biomedical informatics*. 2013;46(6):1145-1151.
14. Newman ME. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E*. 2001;64(1):016132.
15. Hoertel N, Le Strat Y, Lavaud P, Dubertret C, Limosin F. Generalizability of clinical trial results for bipolar disorder to community samples: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of clinical psychiatry*. 2013;74(3):265-270.