Mental Disorders over Time: A Dictionary-Based Approach to the Analysis of Knowledge Domains

I. Kouper[1], A. Zoss[1,2], T. Edelblute[1], M. Boyles[1], H. Ekbia[1]
[1]Indiana University
[2]Duke University

Abstract
Every decade brings changes in the perceptions of normal in mental health, as well as in how abnormal is labeled, understood, and dealt with. Neurosis, hysteria, and homosexuality are just a few examples of such changes. The shifts in terminology and classifications reflect our continuous struggle with social representations and treatment of the "other." How could we best understand mental illness categorizations and become aware of their changes over time? In this paper, we seek to address this and other questions by applying an automated dictionary-based classification approach to the analysis of relevant research literature over time. We propose to examine the domain of mental health literature with an iterative workflow that combines large-scale data, an automated classifier, and visual analytics. We report on the early results of our analysis and discuss challenges and opportunities of using the workflow in domain analysis over time.
Keywords: mental disorders; domain analysis; visual analytics; knowledge domains; concept history
doi: 10.9776/16303

Contact: inkouper@indiana.edu

1 Introduction

The notion of mental disorder is a fluid one. Every decade brings changes in what is perceived as being normal or deviant in mental health, as well as in how abnormal states are labeled, understood, and dealt with. What was called "mental retardation" a few decades ago is now called "intellectual disability," and what was understood in terms of purely emotional states is now explained in the physical language of neurons, neurotransmitters, and brain areas. The changing labels reflect the shifts in cultural attitudes and boundaries of political correctness, as well as the transient character of mental illness (Hacking, 1998).

Changes in mental illness terminology are also driven by the state of knowledge in related disciplines such as psychology, psychiatry, and neuroscience, which, in turn, influences how society views certain mental conditions. At the same time, historians and sociologists of those disciplines would argue that mental illness does not exist apart from the discourses and practices that make it exist (Borch-Jacobsen, 2009). Neurosis, hysteria, and homosexuality are just a few examples of how category descriptions can change along with stigmatization and social exclusion. The recognition, in the last fifty years, that many mental disorders (e.g., schizophrenia) have biological origins has had a humanizing effect on those who are inflicted with it. By the same token, however, the biologization has made the border between sanity and insanity more stark and less penetrable (Gold & Gold, 2014). What do these shifts imply for how we grapple with mental disorders, and how could we best understand them, become aware of them, and, perhaps, minimize their unintended consequences?

We seek to address these questions by applying a large-scale automatic dictionary-based classification to relevant research literature over time. For this, we rely on the classification systems that provide the foundations of modern sciences such as biology, chemistry, medicine, and psychiatry (Bowker & Star, 1999). The availability of large datasets, along with the application of algorithms and techniques to collect, categorize, analyze, and visualize them on various dimensions, allows us to study the ebbs and flows of disorder classes on scales that were not available until recently. In this paper, we report on the results of a study that focused on the temporal changes of major classes of mental disorders over the last few decades in the PubMed database, a freely available database of biomedical literature. In addition to the temporal trends in the literature, illustrated with a few case studies, we discuss the merits and challenges of using a curated dictionary and natural language classification in the analysis and visualization of historical information.

2 Background

Our approach to the analysis of trends in the biomedical literature derives from the domain analysis perspective in information science. Hjørland and Albrechtsen (1995) defined knowledge domains as thought or discourse communities with their own language forms and information systems that are reflections of the work of those communities. Domain analysis then maps the sociolinguistic properties of knowledge communities and examines the functions of information in them. Domain mapping has received a boost in attention in the past decade or so due to an increase in data availability and accessibility, as well as due to the development of new analytical and processing techniques (Shiffrin & Börner, 2004). Subject representations or classifications are at the core of domain analysis, which seeks to understand the contextual, dialogical, and historical character of knowledge. Understanding the shifting boundaries of naming the diseases within the research traditions is one step in doing that.

Due to the rapid increases in the volumes of information that belong to a particular domain, especially to the domain of biomedicine, classifications and natural language processing have been increasingly used for mapping domain structures and extracting information from them (Kim, Lu, & Wilbur, 2015; Krallinger & Valencia, 2005; Li, Liakata, & Rebholz-Schuhmann, 2014). Within biomedicine, domain mapping often relies on extensive controlled vocabularies developed for the purposes of search and retrieval, such as the two controlled vocabularies created and maintained by the US National Library of Medicine (NLM) – the Medical Subject Headings (MeSH) and the Unified Medical Language System (UMLS).

Both of these classifications have been successfully applied in information retrieval and other areas, but their use for knowledge domain exploration has been constrained by the variable specificity of these systems. Previous analyses of MeSH revealed some inconsistencies in how MeSH is applied as well as the differences in its breadth and depth of coverage and retroactive re-indexing of the articles (Darmoni et al., 2012; Nelson, Johnston, & Humphreys, 2001; Srinivasan & Rindflesch, 2002). MeSH descriptors can be applied either too liberally (Kastrin, Rindflesch, & Hristovski, 2014) or not exhaustively enough (Portaluppi, 2011). The UMLS system, while tracking synonymy across a variety of conceptual terms from other classifications, has been reported to appear less frequently within scientific literature (Kim et al., 2015), suggesting that the system fails to account for true natural language usage.

To overcome some of the difficulties in using classification systems, term extraction is another popular approach in the biomedical data mining literature (see, for example, Collier, Nobata, & Tsujii, 2000; Song, Kim, & Rim, 2004; Subramaniam et al., 2003). Zhou, Zhang and Hu (2006), for example, applied an approximate dictionary lookup technique to indexing abstracts of PubMed articles. The use of the approximate entry lookup, i.e., partial lookup that selects the words with maximized occurrence across concept variants, improves both recall and precision of article retrieval. Kim et al. (2015) also used natural language processing to extend existing classification systems, resulting in improvements in precision and recall. Our previous visualization work on research attention to various diseases over time applied natural language processing to article titles from the PubMed database and revealed a significant increase in publications in the category of mental and behavioral disorders between 1961 and 2012 (Zoss, Edelblute, & Kouper, 2014).

To investigate this increase further, this study focuses on the literature and classifications from this domain. While there is an abundance of literature that investigates biomedical publications overall or within certain domains, particularly genetics and biology, attention to the large-scale analysis of psychiatry and related areas is rare. We found only two papers, both from the same authors, that looked into psychiatric literature to identify co-citation networks and clusters of research attention (Wu & Duan, 2015; Wu, Long, & Duan, 2015). The use of citation networks and term clustering for domain analysis, however, can only identify communities of practice, without highlighting the changes in term usage that are so important to the exploration and destigmatization of mental disorders.

The study presented here focuses on extending classification systems with natural language processing techniques, focusing on the specific domain of mental disorders. To overcome the limitations of MeSH classifications applied to publications, we propose to extract names of mental disorders from the article titles by using a dictionary that combines mental disorder terms from three major classification systems: the retrieval-oriented MeSH and two practitioner-oriented diagnostic and billing tools: the Diagnostic and Statistical Manual of Mental Disorders produced by the American Psychiatric Association (APA) since 1952 and a chapter on mental and behavioral disorders in the International Classification of Diseases (ICD), produced by the World Health Organization since 1949.

3 Methodology

To examine historical trends in mental disorder terminology, we retrieved all available records from PubMed database in July 2014 using an Ebot utility from NLM[1] and the following MeSH terms as query: 'Mental+Disorders+[mh]' and 'MENTAL+DISORDERS+(PX)+[mh]'. Overall, 932,438 records were retrieved in XML format in batches of 10,000 articles. The XML for each article was parsed with a custom Python script to extract the following information: PubMed ID, article title, authors, author affiliations, abstract, year of publication, and subject terms (Medline Subject Headings and keywords).

The development of the dictionary of mental disorder terms involved a combination of manual and automated curation. A list of candidate terms was extracted from three existing disease classification systems: ICD-10, five editions of DSM, and the mental disorders branch F03 from the 2015 edition of MeSH ("Free 2015 ICD-10-CM Codes," 2015; US National Library of Medicine (NLM), 2015; ZoomRx, 2013). The list was narrowed by selecting the most specific terms and removing additional descriptions, such as "child onset", "unspecified", and so on. In addition to the three classifications of mental disorders, abbreviations and phobias were added from other lists (Wikipedia, 2015; HealthCentral, 2015; Culbertson, 2010). Overall, 819 entries came from DSM, 185 from ICD-10, 167 entries from MeSH, and 619 entries from other sources.

After adding name variants and morphological transformations, e.g., schizophrenic and Alzheimer, the resulting dictionary contained 2,371 unique entries. The unique entries in the dictionary were then grouped into 81 categories, using content-analytic techniques and classifications as guidance, but at the same time generalizing or specifying terms slightly differently. For example, the category "gender and sexual identify disorder" was split into two – male and female disorders – to add more distinguishing power to the dictionary. Such categorization adds flexibility, accommodates various analytical needs, and serves as a "zoom-in / zoom-out" instrument in the subsequent analysis and visualizations. The categorization provided a broader view on mental disorders with such categories as "developmental disorders", "cognitive impairment", or "intellectual disabilities", as well as a narrower view with such categories as "developmental disorders - autism, Asperger's syndrome" or "movement disorders - Huntington's disease".

The combination of disorder names and categories also helped to address the sensitive issue of inclusion and exclusion of certain conditions into the mental illness domain. For example, homosexuality has been given its own category to avoid mixing it with mental disorders and yet to allow for the investigation of how homosexuality has been treated in the psychiatric / psychological literature over time. Finally, the category "context" included the terms that appeared in mental disorder classification systems but were, rather, the causes, side effects or symptoms of disorders – for example, "child abuse", "non-compliance with treatment", or "victim of terrorism".

To extract disorder names from the titles, a Python natural language processing and classification script was written and iteratively customized. The classifier used the dictionary to analyze article titles and identify keywords and phrases that refer to mental disorders. After performing the regular natural language pre-processing steps of lowercase transformation and tokenization, the classifier identified matches as co-occurrences of dictionary entry words within a small "window" of two words longer than the length of the dictionary entry to accommodate the variance in terms' order of words – for example, "alcohol abuse" and "abuse of alcohol." The classifier retained all matches for each title to account for titles with multiple mental disorders, yielding a one-to-many classification of titles into disorder terms and categories.

The dictionary-based approach to disorder identification has been subsequently combined with visual analytics tools such as Tableau[2] and Raw[3] to enable further discovery and investigations via dedicated studies. We followed the Kohlhammer et al. (2010) framework that proposed a workflow that involves the use of data, modeling (automated analysis), and visualization to reveal trends that were hard to identify without the use of visualization. Our workflow is depicted in Figure 1.

[1] http://www.ncbi.nlm.nih.gov/Class/PowerTools/eutils/ebot/ebot.cgi
[2] http://www.tableau.com/
[3] http://raw.densitydesign.org/

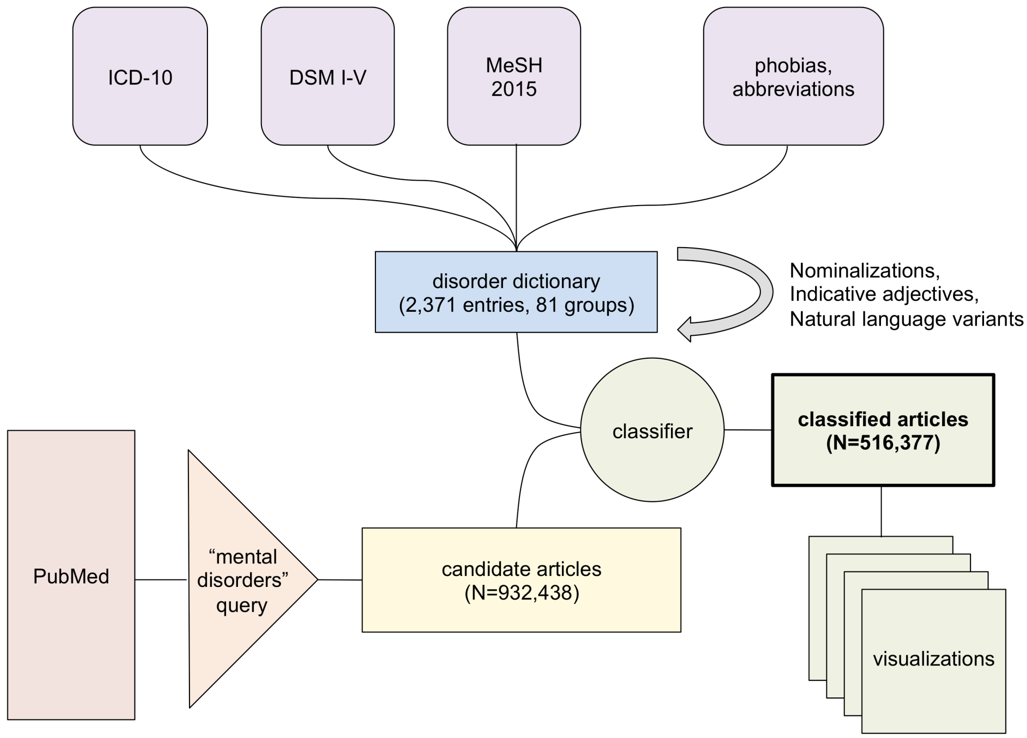

Figure 1. PubMed analysis workflow.

4 Findings

The classifier successfully identified mental disorders in 516,377 out of 932,438 articles, or approximately 55 percent. An analysis of two-word phrases appearing in the titles of articles that were not classified by the classifier demonstrated that the majority of those articles did not mention a specific mental disorder in the title, but were instead focused on the context surrounding mental disorders (e.g., “mental hospital” or “psychiatric services”), on treatments of mental disorders (e.g., “cognitive behavioral therapy”), or on broader categories of mental health (e.g., “cognitive functioning”).

4.1 Overall trends

The overall results of the classifier show that the number of papers focusing on a specific disorder follows the dramatic overall increase of mental disorder publications over the time period of 1945 to 2014 – the period during which we obtained results from the classifier. In the 1950s and early 1960s, the classifier found that between 50% and 60% of mental disorder publications focused on a specific disorder. This decreased somewhat in the late 1960s, falling to about 37% by 1973. After 1973 a steady increase in publications about specific disorders has been recorded, growing to over 62% in 2013 (the last year for which we have complete data).

The use of the two-level dictionary that contained both the specific names and categories of mental disorders allowed us to examine the larger trends in the literature as well as the more nuanced shifts in the form of “case studies.” Figure 2 provides a large-trend overview of the numbers of publications within the most frequent disorder categories over time.

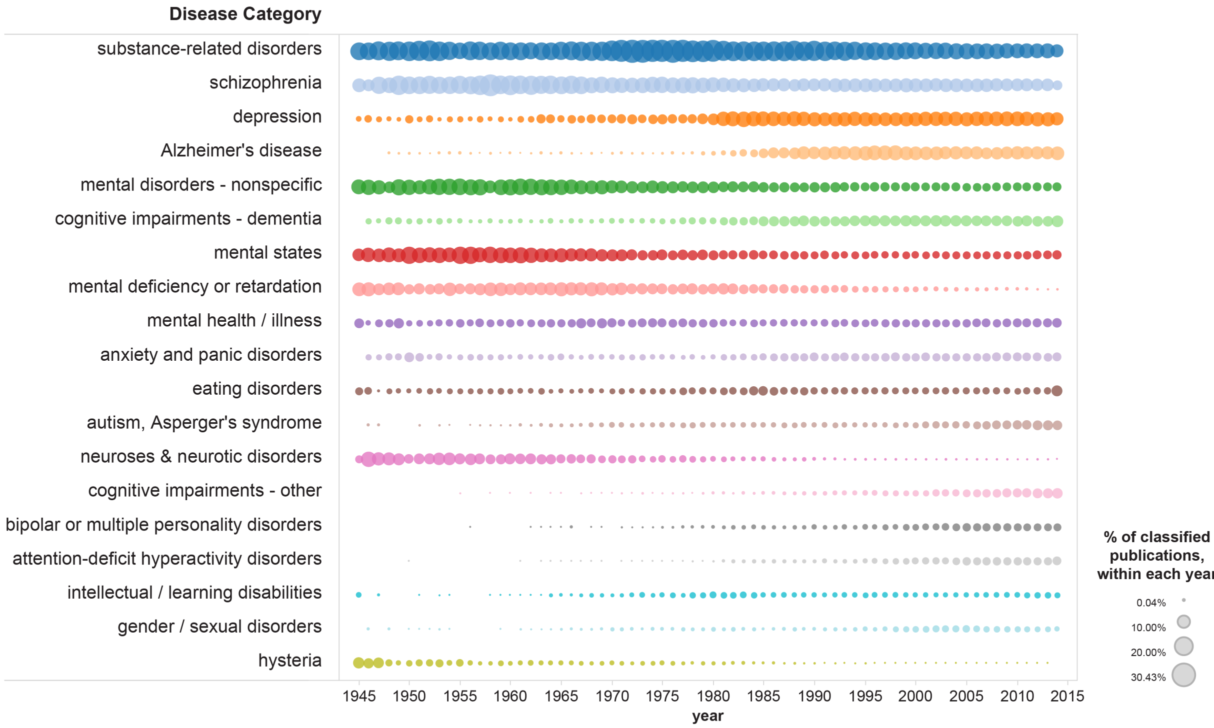

Figure 2: Publications within prominent disease categories, normalized by number of publications each year.

As Figure 2 demonstrates, two groups of disorders that have received relatively stable significant research attention over time are "substance-related disorders" and "schizophrenia," although the numbers in both groups have been decreasing gradually for some time (since the early 1970s and late 1950s, respectively). The group of substance-related disorders included such substances as alcohol, tobacco, morphine, marijuana, medical drugs and many other chemical and medical substances. Two additional groups – "mental states" and "nonspecific mental disorders," which included such explicitly named states as psychosis, delirium, hallucinations, and delusions and the references to "mental disorder" or "mental illness" in general – began at over 10% of all classified publications, but have also been decreasing over time, falling to below 5% of the publications by the 1990s.

The most dramatic increases occurred within the following groups: "depression," "Alzheimer's disease," and "cognitive impairments - dementia." Alzheimer's disease, which accounts for less than 1% of the publications before 1980, grows quickly in the 1980s to a fairly stable 8-12% of the publications after 1988. Depression has a sharp increase in the late 1970s and early 1980s, going from 3-5% of the publications to 10-13% of the publications in just a few years. Dementia research exhibits a gradual but steady increase, currently making up almost 8% of the publications. Alzheimer's disease was shown to be among the most funded research areas, receiving approximately $289 million yearly, which may have accounted for its dramatic increase in research attention (Liu, Coulet, LePendu, & Shah, 2012).

The dictionary-based approach also allowed for a detailed analysis of terminology shifts over time. By including both historical and modern terminology in the dictionary, we remediate the effects of re-writing and erasure of historical term usage in modern classification systems, which may be insignificant for the purposes of information retrieval or diagnosis and treatment of disorders, but is crucial for historical analyses of knowledge domains and the histories of ideas.

The following three case studies were selected as examples of how analyzing changes in specific term use is likely to be especially important. The "autism" case study and the use of specific terms analyzed within this concept provides empirical evidence in support of the history of this research area that was previously documented using other methods. The "intellectual disabilities" case study helps to trace the interplay between terms related to now pejorative and discarded "mental retardation" and terms related to the preferred "intellectual disabilities." With the "gender and sexual disorders" case study, the term-based analysis compares male-specific disorders to female-specific and gender-neutral disorders, thereby highlighting a disparity in research attention to one of the genders that otherwise can be easily overlooked.

4.2 Case 1: Autism

The amount of research publications about autism begins to increase at the beginning of the 1960s and then, after a small spike at the end of the 1970s - beginning of the 1980s, it plateaus for almost two decades to begin steadily growing again at the end of the 1990s (see Figure 3). By 2013 the combined literature on autism comprises almost 6% of the literature on mental disorders in PubMed.

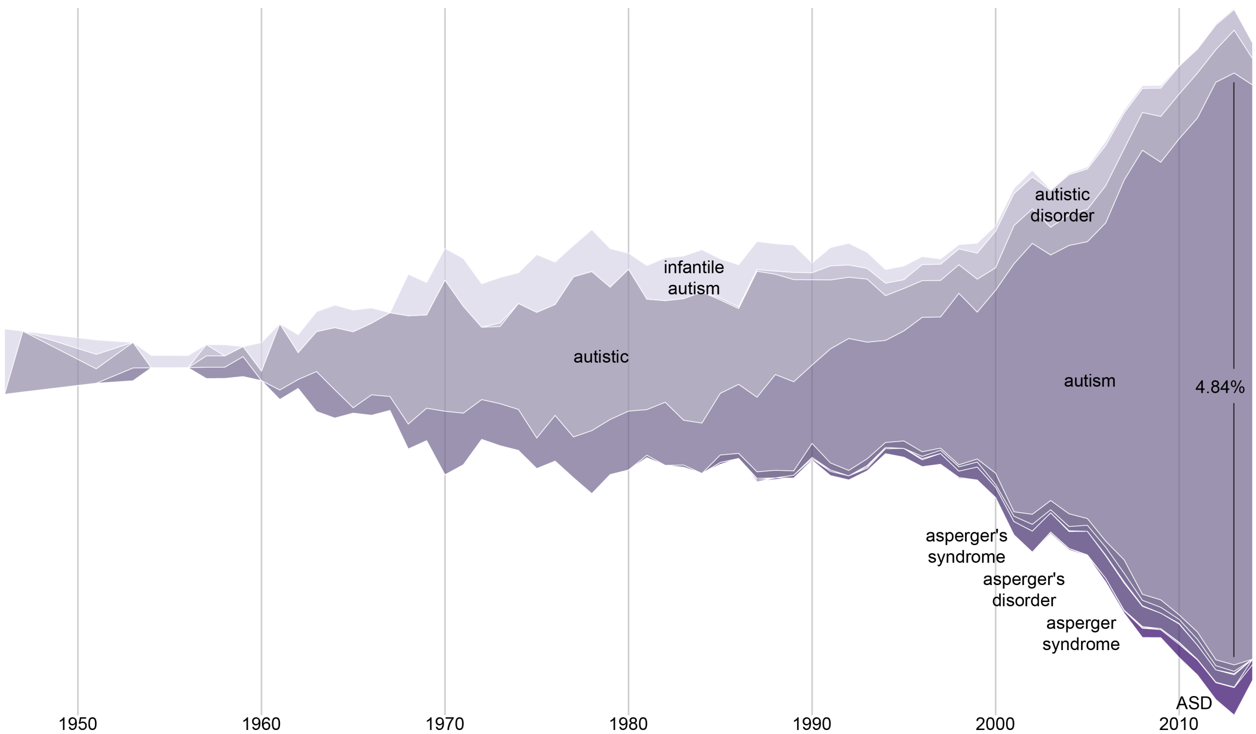

Figure 3: Research attention to autism in the PubMed database over time. Shades of purple indicate the use of specific terms. Per cent label indicates the highest peak for the single term in a single year.

As a relatively new disorder that was clinically recognized after WWII, in the earlier years autism had a rather narrow definition in the scientific literature. It did not have its own diagnostic criteria in the first edition of the DSM, referred to as "childhood schizophrenic" instead. The term "autistic" appeared in the second edition of DSM, somewhat reflecting its growing use in the literature and in practice; and the term "infantile autism" was recognized only in DSM-III (published in 1980), even though the term appeared in the titles in the 1960s and in the 1970s.

Silberman (2015), who chronicles the history of autism, attributes the expansion and subsequent growth in attention to autism to the grassroot efforts in the 1970s that sought to correct a course set earlier by the American psychiatrist Leo Kanner, who put the blame of autism on cold and withholding parents. Rather than giving in to the shame and stigma brought about by Kanner's narrow definition, parents sought treatment for their children under a much broader understanding of the disorder. Autism disorder was later brought to broad public awareness by the movie "Rain Man" in 1988, which was based on a real life story. These developments gave rise to an autism pandemic, described as "an optical illusion brought about by a sin of diagnostic parsimony" (Senior, 2015).

Our data, which shows a dramatic growth in the number of publications in the late 1990s, reflects these background developments in a vivid manner. The subsequent growth in federal funding, with an average annual increase rate of 15% between 1997 and 2006, provides part of the explanation of the sustained growth in publications in the following years (Singh, Illes, Lazzeroni, & Hallmayer, 2009).

Asperger syndrome (or disorder) was present in the DSM prior to the fifth edition, but then was replaced by "autism spectrum disorder" with a severity scale. In delineating the history of Asperger syndrome, Wing (1998) contemplated the challenges of the "specific disease model" and argued that distinguishing between syndromes and disorders is akin to classifying clouds. The small but persistent existence of this term in the publications demonstrates that researchers continue to use this term in the 2000s, surrounded by the historical uses of “autistic psychopathology” and “autism”.

4.3 Case 2: Intellectual disabilities

Our second case is an example of how changes in sensitive terminology can be traced via large scale analysis and visualizations. Intellectual disability is a condition that has been documented under different names over time, including mental degeneration, mental retardation, learning disability, and intellectual developmental disorder.

Our analysis shows that the terms that refer to mental deficiency and mental retardation account for the majority of literature on intellectual and learning disabilities (see Figure 4). The use of the term “mental retardation” spiked at the end of the 1960s to cover ~9.4% of the literature on mental disorders, and while its use has consistently declined, it is still in use into the 2000s. The term “learning disability” was somewhat in use in the literature published before the 1950s, but its use steadily increases between the mid-1960s and the 1980s and then decreases after that. The term “intellectual disability” has gained popularity from the 1990s and into the 2000s.

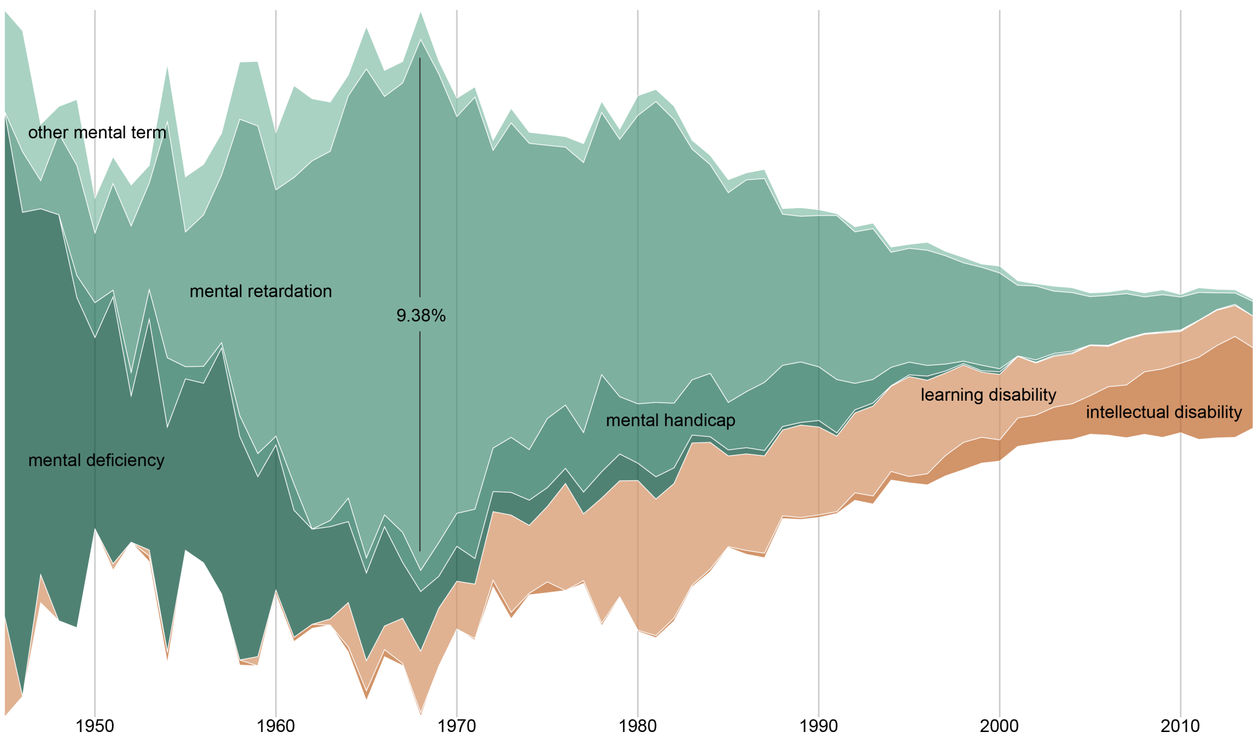

Figure 4: Research attention to intellectual disabilities. Per cent label indicates the highest peak for the single sub-category in a single year.

The shift from using the terms “mental deficiency” and “mental retardation” to “intellectual disability” has been visibly advocated for in the United States and the United Kingdom. In August 2010 a bill was passed into law in the United States that was called Rosa’s Law[4] and required replacement of the words “mental retardation” in federal laws with “intellectual disability.” The law is named after Rosa Marcellino, a girl with Down Syndrome whose parents advocated for the removal of the words “mentally retarded” from the health and education documents. The term “general learning disability” has been recommended in the UK to replace terms such as “mental handicap” or “mental retardation.” Rosa’s Law is a small part of the long series of changes that happened since the early 1900s and even before that and went from the early dark history of “cretins” and “imbeciles” through the shift to care to the humanization of people with intellectual disabilities (Richards, Brady, & Taylor, 2014).

The history of the uses of these terms illustrates the pervasiveness of the terms once they get accepted and the difficulties of changing naming conventions and developing consensus of what is pejorative or not. The mapping of the terms provided above is just the first step in a series of studies that

[4] http://www.gpo.gov/fdsys/pkg/PLAW-111publ256/pdf/PLAW-111publ256.pdf

can reveal the historical trends and further raise awareness about this issue. A brief look into some of the publication titles and their metadata, for example, indicates that the titles that still use the term "mental retardation" came from the countries other than the US and the UK or from the authors who are affiliated with non-US and non-UK institutions.

4.4 Case 3: Gender and sexual disorders

As a category, gender and sexual disorders includes disorders that are distinctly male or female as well as disorders that occur in both genders. For example, female sexual arousal disorder or erectile dysfunction are distinctly gendered, while gender dysphoria, the condition of feeling discontent with gender identity given at birth, affects both males and females. While overall the number of papers that focus on gender and sexual disorders is relatively small (~2% of all classified papers), the difference between research attention to each of the genders is striking.

Across all years, male disorders appear in 1.04% of classified papers, compared to the 0.58% of classified papers focusing on neutral gender disorders and the mere 0.12% of papers focusing on female disorders. Male erectile dysfunction received the largest number of publications, spiking around 2005 to 2.09%.

The disparity in attention to male versus female gender and sexual disorders is tied to the role of pharmaceutics in psychiatric research. Male gender and sexual disorders are easier to target in drug development and many drugs have entered the market in the past few decades. Female gender and sexual disorders have a long history of being ignored due to a variety of other reasons as well, and gender biases in psychoanalysis and other psychiatric and psychological traditions are well-documented, e.g., (Brennan, 1992). The recent controversy over the so-called "pink pill" illustrates the continuing bias in academic and policy circles, reflecting the broader cultural tendencies (Ungar, 2015).

5 Discussion and Conclusion

The knowledge domains have been mapped many times, with mostly their structures and contributor ties being explored. This work contributes to domain analyses by emphasizing a historical dimension and bridging the quantitative and qualitative aspects of conceptual work. This is not to suggest that these techniques can answer all our questions about the historical changes of domain knowledge. Visualization, particularly, often trades off aesthetic and informational qualities, creating a dilemma for big data approaches (Ekbia et al., 2015). Another tradeoff is between the precision of top-down work of curated terminology and the recall of bottom-up approaches that rely on more flexible natural language processing. Our approach provides high precision because of the dictionary and manual verification of results, which seems to be very important here, but with other approaches such as topic modeling, recall may be more important due to the need of identifying broader clusters of terminology.

The combination of dictionary-based approach, natural language processing, and visualizations employed in this paper facilitates the analysis of historical trends in knowledge domains at varying levels of detail. The trends and the cases reported here merely scratch the surface, but they demonstrate the efficacy of this integrated technique. The dual individual / category term dictionary applied to the article titles allowed us to trace changes in terminology. Visualizations provided a high-level overview of trends over time and helped us to tell several stories of specific naming labels within the latter disorder categories. Used at various levels of magnification, visualizations afford additional opportunities for analysis and discovery. In compressing numbers to shapes and colors, we discovered that visualization was essential for verifying both the quality of the original data and the output of the classifier. For example, a simple line chart at the beginning of our analysis revealed systematic gaps in the data retrieval, while frequency bar charts directed our attention to the relative prominence of different classifications.

The workflow developed within this paper and applied to a subset of the PubMed database allowed us to get tangible results that are useful in raising the awareness of the persistence of certain terms and pointing to the disparities in attention to some categories of the disorders. A subsequent crucial step in this type of work is to engage domain experts to help with deeper interpretations and further interrogation of the data. We propose to strengthen the workflow with interactive elements, ultimately rendering a more dynamic and complete analytic tool that can be applied to many domains. Improved interactive elements might include the use of overview charts that feature clickable labels that link to other, more detailed and specific visualizations. We envision a system where a user can explore multiple charts at one time, perhaps, linked together in some visual fashion. Other enhancements could include improving the classifier to feature rule-based classification that better accounts for common variations, such as plurals, synonyms, and similar yet semantically different words such as disease/disorder or addiction/dependence, and allows historians and sociologists of science and medicine to explore alternate categorizations and trends by editing or expanding the dictionary. More sophisticated processing of titles needs to be expanded into the abstract and ultimately full text of the articles.

6 References

Borch-Jacobsen, M. (2009). *Making minds and madness: from hysteria to depression.* Cambridge: Cambridge University Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Brennan, T. (1992). *The interpretation of the flesh: Freud and femininity*. New York: Routledge.

Byron, L., & Wattenberg, M. (2008). Stacked graphs–geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics, 14* (6), 1245-1252.

Collier, N., Nobata, C., & Tsujii, J.-I. (2000). Extracting the names of genes and gene products with a hidden Markov model. In *Proceedings of the 18th conference on computational linguistics - COLING '00* (pp. 201– 207). Stroudsburg, PA, USA: Association for Computational Linguistics. doi: 10.3115/990820.990850

Culbertson, F. (2010). The phobia list. Retrieved from http://phobialist.com

Darmoni, S., Soulamia, L., Letord, C., Jaulent, M.-C., Griffon, N., Thirion, B., & Névéol, A. (2012). Improving information retrieval using medical subject headings concepts: A test case on rare and chronic diseases. *Journal of Medical Library Association, 100*(3), 176-183.

Ekbia, H., Matiolli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S., & Sugimoto, C. (2015). Big data, bigger dilemmas: A critical review. *Journal of American Society for Information Science and Technology, 66*(8), 1523-1746.

Free 2015 ICD-10-CM codes. (2015). Retrieved from http://www.icd10data.com/ICD10CM

Gold, J., & Gold, I. (2014). *Suspicious minds: How culture shapes madness.* New York: Free Press.

Hacking, I. (1998). *Mad travelers: Reflections on the reality of transient mental illnesses.* Charlottesville, VA: University of Virginia Press.

Havre, S., Hetzler, B., & Nowell, L. (2000). ThemeRiver: Visualizing theme changes over time. In *Proceedings of the IEEE Symposium on Information Vizualization INFOVIS '00*. Washington, DC: IEEE Computer Society.

HealthCentral. (2015). ADHD and mental health acronyms. Retrieved from http://www.healthcentral.com/adhd/understanding-adhd-260744-5.html

Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science, 46*(6), 400-425.

Kastrin, A., Rindflesch, T. C., & Hristovski, D. (2014). Large-scale structure of a network of co-occurring MeSH terms: Statistical analysis of macroscopic properties. *PLoS ONE, 9*(7), e102188.

Kim, S., Lu, Z., & Wilbur, J. (2015). Identifying named entities from PubMed for enriching semantic categories. *BMC Bioinformatics, 16*(57), 1-10.

Kohlhammer, J., Ruppert, T., Davey, J., Mansmann, F., & Keim, D. (2010). Information visualisation and visual analytics for governance and policy modelling. In *Proceedings of the European Commission CROSSROAD workshop "A participative roadmap for ICT research in electronic governance and policy modelling"*. Retrieved from http://kops.uni-konstanz.de/bitstream/handle/123456789/13654/Kohlhammer%20etal.pdf

Krallinger, M., & Valencia, A. (2005). Text-mining and information-retrieval services for molecular biology. *Genome Biology, 6*(7), 224-227.

Li, C., Liakata, M., & Rebholz-Schuhmann, D. (2014). Biological network extraction from scientific literature: State of the art and challenges. *Briefings in Bioinformatics, 15*(5), 856-877.

Liu, Y., Coulet, A., LePendu, P., & Shah, N. H. (2012). Using ontology-based annotation to profile disease research. *Journal of the American Medical Informatics Association: JAMIA, 19*(e1), e177-86. doi: 10.1136/amiajnl-2011-000631

Nelson, S., Johnston, W., & Humphreys, B. (2001). Relationships in medical subject headings (MeSH). In I. C. Bean & R. Green (Eds.), *Relationships in the organization of knowledge* (p. 171-184). Dordrecht: Springer.

Portaluppi, F. (2011). The Medical Subject Headings thesaurus remains inaccurate and incomplete for electronic indexing and retrieval of chronobiologic references. *Biological Rhythm Research, 42*(6), 491-504. doi: 10.1080/09291016.2011.613619

Richards, S. B., Brady, M. P., & Taylor, R. L. (2014). *Cognitive and intellectual disabilities: Historical perspectives, current practices, and future directions.* New York: Taylor & Francis.

Senior, J. (2015, August 17). “Neurotribes” by Steve Silberman. *New York Times Book Review.* Retrieved from http://www.nytimes.com/2015/08/23/books/review/neurotribes-by-steve-silberman.html

Shiffrin, R. M., & Börner, K. (2004). Mapping knowledge domains. In I. R. M. Shiffrin & K. Börner (Eds.), *Mapping knowledge domains* (p. 5183-5185). Washington, DC: The National Academies Press.

Silberman, S. (2015). *Neurotribes: The legacy of autism and the future of neurodiversity*. Princeton, NJ: Avery/Penguin Random House.

Singh, J., Illes, J., Lazzeroni, L., & Hallmayer, J. (2009). Trends in US autism research funding. *Journal of Autism and Developmental Disorders, 39*(5), 788-95. doi: 10.1007/s10803-008-0685-0

Song, Y.-I., Kim, S.-B., & Rim, H.-C. (2004). Terminology indexing and reweighting methods for biomedical text retrieval. In *Proceedings of the SIGIR’04 workshop on search and discovery in bioinformatics.* Sheffield, UK: ACM.

Srinivasan, P., & Rindflesch, T. (2002). Exploring text mining from MEDLINE. In *Proceedings of the AMIA symposium* (p. 722-726).

Subramaniam, L., Mukherjea, S., Kankar, P., Srivastava, B., Batra, V., Kamesam, P., & Kothari, R. (2003). Information extraction from biomedical literature: Methodology, evaluation and an application. In *Proceedings of the ACM conference on information and knowledge management*, New Orleans, Louisiana.

Ungar, L. (2015). FDA approves world’s first pill to boost women’s libido. *USA Today.* Retrieved from http://www.usatoday.com/story/news/nation/2015/08/18/fda-acts--flibanserin/31840455/

US National Library of Medicine (NLM). (2015). Mesh tree structures - 2015, F03 - mental disorders. Retrieved from: http://www.nlm.nih.gov/mesh/2015/mesh_trees/F03.html

Wikipedia. (2015). List of abbreviations for diseases and disorders. Retrieved from https://en.wikipedia.org/wiki/List_of_abbreviations_for_diseases_and_disorders

Wing, L. (1998). The history of Asperger syndrome. In E. Schopler, G. B. Mesibov, & L. J. Kunce (Eds.), *Asperger syndrome or high-functioning autism?* (p. 11-25). New York: Plenum Press.

Wu, Y., & Duan, Z. (2015). Analysis on evolution and research focus in psychiatry field. *BMC Psychiatry, 15*, 105. doi: 10.1186/s12888-015-0482-1

Wu, Y., Long, C., & Duan, Z. (2015). Analysis on international scientific collaboration and research focus on depression field. *Chinese Medical Journal, 128*, 687-693.

Zhou, X., Zhang, X., & Hu, X. (2006). MaxMatcher: Biological concept extraction using approximate dictionary lookup. In I. Q. Yang (Ed.), *Proceedings of the 9th Pacific Rim international conference on artificial intelligence PRICAI’06* (p. 1145-1149). Berlin, Heidelberg: Springer-Verlag.

ZoomRx. (2013). Data from “Our increasing insanity: The evolution of the DSM” visualization. Retrieved from http://visualising.org/full-screen/216355

Zoss, A., Edelblute, T., & Kouper, I. (2014). Diseases across the top five languages of the PubMed database: 1961 - 2012. Poster presented at the ACM WebSci’14 conference, June 2014, Bloomington, IN. Retrieved from http://figshare.com/articles/WebSci_2014_data_visualization_challenge/1033878