Analysis of Protein Function Overlap and Disease Mechanisms of Alzheimer's Comorbidities

Laura A. Christman[1*], Rachel A. Simon[1*], Amy Hoang[1*], Kayla M. Day[1*], Kelley Chan[1†], Jesse T. Clayton[1†], and William A. McLaughlin[1]
[1]Geisinger Commonwealth School of Medicine, Scranton, PA 18509
*Master of Biomedical Sciences Program
†Doctor of Medicine Program
Correspondence: wmclaughlin@som.geisinger.edu

Abstract

Background: Comorbidities of Alzheimer's disease may be explained in part by shared genetic components. Our aim was to characterize statistically significant co-occurring proteins and their functions in comorbid disease pairs. These were established from population-based inpatient data and overlap of involved proteins. The results are used to further describe the pathology of the comorbidities of Alzheimer's disease.

Methods: Statistically significant comorbid disease pairs of Alzheimer's disease were extracted from Medicare hospitalization data. To identify significant comorbidities based on protein overlap, mappings between diseases and proteins were first retrieved through keyword searches of Disease Ontology terms across the PubMed abstracts mapped to proteins in UniProt. DisGeNET provided additional disease to protein mappings. Quantification of the statistical significance of comorbid diseases was determined through a phi coefficient analogous to what was used in a study by Hidalgo et al.

Results: Of the 34 significant comorbidities of Alzheimer's disease based on a study by Hidalgo et al., 18 were also significant based on protein overlap. Proteins involved in these 18 comorbidities are considered for how gain or loss of functions may contribute. An example of a statistically significant overlapping protein involved in both Alzheimer's and Parkinson's disease is tyrosine-protein kinase Fyn (Fyn). A gain of function of Fyn may contribute to the pathology of both diseases and ultimately to the comorbidity via an increase in a proinflammatory response.

Conclusion: Understanding genetic associations between disease comorbidities may have important implications for understanding the disease mechanisms. By finding overlapping proteins and elucidating the function of these proteins in comorbid disease states, we further describe the underlying pathology of comorbid disease pairs of Alzheimer's disease, with a focus on inferred gain or loss of functions that may contribute.

Introduction

Comorbidities are thought to have complex genotypic and phenotypic relationships; these relationships often cannot be defined directly by a single gene (1), as cellular functions compose an interconnected cascade of processes working in tandem to sustain normal functioning (2). When there is an alteration in this cascade, a vast array of physiological malfunctions can result as comorbidities (2). The underlying etiology of comorbidities can be influenced by a multitude of determinants from the genome, proteome, metabolome, and the environment (3). The involvement of cellular networks in comorbidities have been studied previously using health care and insurance data to define genotypic and phenotypic overlap between diseases via a network approach (1, 4). Recognition of genetic overlap, or shared genetic components between comorbidities, can enhance understanding of complex phenotypes and their connections to other diseases (1). Genotypic and phenotypic relationships in comorbidities have been previously analyzed through utilization of population-based health care data (1, 2, 4, 5). Once analyzed, this data can be used to create vast disease networks which can quantify comorbid relationships and determine statistical significance (1, 4–6).

A common problem encountered throughout these studies is having a multitude of data with a limited scope; for example, protein-protein interactions can only draw conclusions about protein interactions and not genetic information (7). Accessibility of health care data has also proven to be a challenge, but a previous study has made their population-based health care data readily accessible to the public (4). To support analyses of disease to protein mappings, DisGeNET integrates data from a multitude of sources to include genotype-phenotype relationships of varying genetic or environmental origins which contribute to human disease (8).

Consider that Alzheimer's disease (AD) is a neurodegenerative disorder which may initially manifest as memory loss and difficulty speaking and can progress to more severe clinical symptoms as more neurons become damaged over time (9). It is estimated that 5.8 million people aged 65 years or older are living with AD in 2020 in the United States (10). However, as the general population ages and life spans increase, this number is expected to more than double to 13.8 million people in the United States by the year 2050 (10). As the mean age of the population increases, the percentage of those living with AD increases. Of individuals aged 65 to 74, 3% have been diagnosed with AD, 17% of individuals aged 75 to 84 and 32% of individuals aged 85 years or older share the same diagnosis (9). AD places a great physical burden on those suffering from the disease, but also places significant social and economic burdens on their families who often become their caregivers as their disease progresses. It is estimated that individuals with dementia receive $321,780 of health care and caregiving costs throughout the course of their disease, which is $184,500 greater than their cost of care if they did not have the disease (11). More evidence needs to be collected regarding the progression of AD and how the disease develops over an individual's lifespan.

Frailty	Hyperhomocysteinemia
Gait dysfunction	High serum levels of amyloid beta 40
Poor self-rated health	High sensitivity C reactive protein levels at mid-life
Low bone mineral density	Heavy smoking
Carotid atherosclerosis	High BMI at mid-life
Hypertension	Low BMI
Hypotension	History of stress
Depression	Less physical activity
Less cognitive activity	Low socioeconomic status
Diabetes mellitus type 2*	Neuroinflammation*
Low education level*	Neuroticism*

Table 1. Risk factors of AD based on systematic analysis (12). Terms denoted with * have a positive association with risk of developing AD.

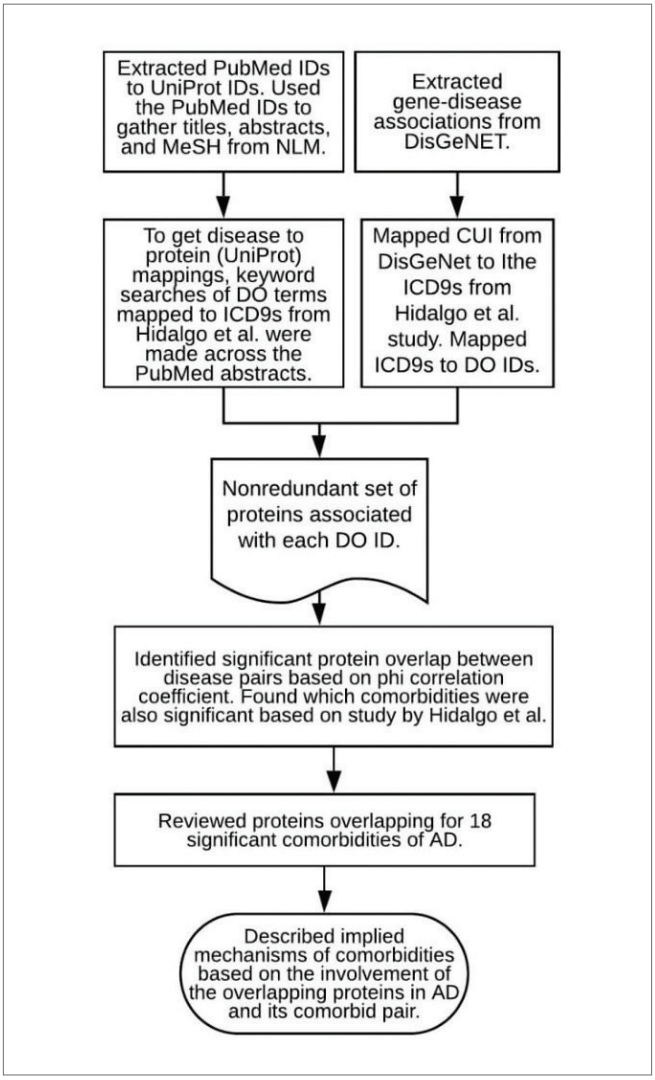

Figure 1. Flowchart detailing methodology

Many modifiable factors have been identified that may increase the risk of developing AD, including comorbidities; however the underlying biological mechanisms contributing to these processes require further elucidation and characterization. Based on systematic analysis (12), Table 1 lists the defined risk factors for AD with other conditions with a positive association with risk of developing AD. Patients diagnosed with AD and other comorbid diseases tend to have decreased cognitive function, decreased daily functioning in daily activities, and more psychiatric symptoms, increasing their health burden (13). The burden of comorbidities leads to the symptoms of AD worsening in a shorter time span due to a dynamic relationship between various disease-causing factors, not a cause-and-effect relationship (13). Patients with AD have significantly more comorbid medical conditions as documented by a greater number of emergency room visits and hospitalizations (14).

The greatest contributing risk factors for AD-related dementia include age, family history, and genetics, but AD is not a normal process of aging. Regarding family history, individuals have a greater likelihood of developing AD if they have one or more first-degree relatives with the condition (9). With regard to genetics, it has been estimated that 1% or less of AD cases are linked to mutations involved in the genes for amyloid precursor protein (APP), presenilin 1, and presenilin 2 (9). Individuals with Down syndrome are at a greater risk for developing AD, as the APP gene is encoded on chromosome 21 (9).

More research is needed to enhance understanding of other genetic mechanisms involved in the development of AD. We infer that there is a need to further quantify and describe the overlap of genes between AD and its comorbidities. Describing the function of proteins encoded by these overlapping genes may provide valuable evidence toward understanding the underlying disease mechanisms that contribute to each statistically significant comorbidity of AD.

Through analysis of population-based data (4) and the use of comorbidity etiology data from DisGeNET and based on identified mapping from UniProt, co-occurring proteins can be identified between AD and others which can improve understanding of the disease phenotype (4, 8). This study sought to determine statistically significant relationships between co-occurring proteins and their functions within associated comorbid disease pairs of AD as determined by population level data and the overlap of the involved proteins. We describe the inferred protein functional changes, for example, gain or loss of functions, that are implicated in each identified comorbidity.

Methods

Study population

Medicare claims data were extracted from a previous study by Hidalgo et al (4), which consisted of patients over the age of 65. The dataset included hospitalizations from 1990 through 1993, where 90.1% of the demographic were classified as white (4). The data utilized ICD-9 diagnostic codes, which was the primary coding system available at the time the record was extracted. Of the 32,341,347 inpatient claims retrieved from

the study, only 13,039,347 participants were considered (4, 15, 16). The number of participants in the study only included hospitalized patients. The remaining were individuals who were not hospitalized during this period (4). Additionally, the study focused only on hospitalization claims due to its consistent nature in the diagnosis of diseases (4). Despite being approximately 30 years old, this data set was used because it is the most complete and most readily available dataset containing population-based healthcare.

Disease to protein mappings, UniProt

To determine significant Alzheimer's comorbidities, the study depended on retrospective data. A visual interpretation of the methods is shown in Figure 1.

An automated extraction of 11,013,767 PubMed abstracts mapping proteins mapped to 10,353,916 UniProt IDs was done based on the corresponding reference section in the UniProt flat files (17). Based on the PubMed IDs, text fields including titles, abstracts, and medical subject headings (MeSH) were retrieved from the National Library of Medicine. Keyword searches using Disease Ontology (DO) terms (18) mapped to the ICD9 codes represented in the Hidalgo study (4) were used as text queries in a MySQL database tables containing PubMed abstract text fields mapped to UniProt accession codes. The mappings between disease ontology and ICD-9 codes are available via OBO file (18). The text search for each disease term thereby provided a list of UniProt accession codes associated with the disease.

Disease to protein mappings, DisGeNET

DisGeNET aids as an extensive resource in providing a collection of genes and variants associated with human diseases (8). We utilized three different tab separated files (19): "curated gene-disease associations," "UniProt Downloads," and "UMLS (united medical language system) CUI (concept unique identifiers) to several disease vocabularies". "Curated gene-disease associations" is a file compiling associations from UniProt and CUI IDs (listed as "diseaseId") to various other databases; however, for the purpose of this research, we extracted only UniProt data and CUI IDs and disregarded the other databases and their data. The "UniProt Downloads" file includes UniProt accession codes corresponding to its respective DisGeNET unique gene identifiers. "UMLS CUI to several disease vocabularies" contains mappings of DisGeNET genes (gene identifiers) to UniProt entries and UMLS CUIs to various disease vocabularies, including ICD-9-CM and DO (8). We utilized these files to generate disease ontology to UniProt accession codes mappings which added to those generated by analyses described above.

Data analyses

Quantitative analyses of protein and functional overlap

The formula for the phi correlation utilized for analysis of patient data is described in the Hidalgo study (4). The phi correlation is Pearson's correlation for binary variables, and a value greater than 0 is indicative of a co-occurrence that is more frequent than would be expected by chance alone (2, 4, 20). The calculation for the phi correlation is the following:

$$\varphi_{ij} = \frac{C_{ij}\,N - P_i P_j}{\sqrt{P_i P_j (N - P_i)(N - P_j)}}$$

P*i* is the number of patients in the sample with disease i. P*j* is the number of patients in the sample with disease j. C*ij* is defined as the number of patients with proteins from both diseases, and N is the total number of patients (4). Others (2) utilized the phi correlation similarly; however, instead of using prevalence of the disease (P), they replaced this variable with I for incidence (2). The desired result is a φ value greater than zero, indicating a greater co-occurrence than would be expected by chance (2).

The phi correlation provides an analogous tool for the study of overlapping proteins between comorbid diseases. The variables are modified accordingly. Modification of the formula was accomplished by defining C*ij* as proteins related to both diseases, N as number of proteins, and P*i* and P*j* as proteins in diseases i or j, respectively.

A *t*-test was used to determine statistical significance of the calculated phi values using the following equation:

$$t = \frac{\varphi\sqrt{n-2}}{\sqrt{1-\varphi^2}}$$

We then created a programmed extraction that yielded a total of 494 disease pairs through the use of Python 2.7 and 3.7.7. The generated file was a table that included the overlapping proteins associated with these pairs, allowing for determination of AD comorbidities along with associated significant overlapping proteins. From this, the disease pairs that included AD and the shared proteins associated with the pair were documented in a separate file for subsequent literature review and analysis.

With the determination of which proteins have a statistically significant overlap, it is also possible to determine the statistical significance of the pathological functions of these proteins as seen in their respective diseases. We were able to redefine the variables in order to apply the phi correlation to functional overlap. In this case, the variables were defined as: P*i* and P*j* for how many proteins in each disease that had a selected function. C*ij* is the number of total proteins with that one function that were involved in both diseases. N is therefore the total number of proteins with the selected function. The functional overlap assessment is ongoing and not yet completed.

Qualitative analyses of functional overlap using literature reviews

Two procedures were followed after the identification of the statistically significant proteins identified for each disease pair. First, utilization of extracted UniProt protein accession codes and the UniProt online database allowed for determination of the protein name. Second, a comprehensive Google Scholar search was completed based on the protein name, followed by analysis of appropriate journals to further understand its pathophysiological functions in AD and its comorbid disease pair.

For example, consider one comorbid disease pair, AD and Parkinson's disease. The first shared protein of interest was P06241, which the UniProt online database identified as tyrosine-protein kinase Fyn, also known as kinase Fyn (21). The next protein we examined was Q16620, the BDNF/NT-3 growth factors receptor or TrkB (22). Lastly, we also gathered

Dementia	Secondary Parkinson disease
Senile degradation of brain	Psychotic disorder
Parkinson's disease	Bullous pemphigoid
Cerebral degeneration	Deficiency anemia
Plasma protein metabolism disease	Cerebrovascular disease
Kidney failure	Osteoporosis
Anemia	Hypoglycemia
Major depressive disorder	Cardiovascular system disease
Glaucoma	Cerebral artery occlusion

Table 2. Diseases which were found to have statistically significant overlap based on both comorbidity and shared proteins with AD.

information on P02511, alpha-crystallin B chain, also known as CRYAB (23). The literature review conducted utilizing Google Scholar and PubMed included the following terms: Alzheimer's, Parkinson's, comorbid, comorbidity, P06241, tyrosine-protein kinase Fyn, Fyn, kinase Fyn, tyrosine-protein kinase Fyn AND Alzheimer's, tyrosine-protein kinase Fyn AND Parkinson's, Q16620, BDNF/NT-3 growth factors receptor, TrkB, P02511, BDNF/NT-3 AND Alzheimer's, BDNF/NT-3 AND Parkinson's, TrkB AND Parkinson's, alpha-crystallin B chain, TrkB AND Alzheimer's, alpha-crystallin B chain AND Alzheimer's, CYRAB AND Alzheimer's, CRYAB, and alpha-crystallin B chain AND Parkinson's.

The selected literature was chosen based on relevance to the disease pair and discussion of the function of the protein as seen in the diseases of interest. We sought evidence for whether there was a gain or loss of function as it contributes to the pathophysiology of the diseases.

Results

Quantitative results were received from running the programs that analyzed the phi correlation. As previously mentioned, a *t*-test was used to determine statistical significance of the calculated phi values. The *t*-test threshold was set to 2.576 with a confidence interval of 99%. A 2-by-2 contingency table was created to analyze significant comorbidities and significant protein overlap with 1 degree of freedom. The value N, which was the total UniProt accession codes mapped to disease terms, was equal to 1,560,091. This analysis yielded a chi-square statistic of 8,438 and the *t*-statistic yielded a *p*-value less than 0.001. Out of 235,302 identified disease pairs, 60,639 disease pairs were linked to DOID terms, ICD-9-CM codes, and PubMed abstracts.

Out of these 60,639 disease pairs, 58,685 disease pairs were not statistically significant by both patient data and protein data. There were 374 disease pairs determined to be not statistically significant by patient data, but significant by protein data. Conversely, 1,145 disease pairs were determined to be statistically significant by patient data, but not by protein data. There were 435 disease pairs which had statistically significant overlap based on both comorbidity and shared proteins.

Overlapping UniProt accession code	Protein name	Gain or loss of function	Role in PD	Role in AD
P06241	Tyrosine-protein kinase Fyn (Fyn)	Gain of function	Fyn regulates proinflammatory signaling and mediates microglial neuroinflammatory processes in PD.	Fyn is associated with both amyloid beta accumulation as well as tauopathy. It is explored as a therapeutic target for AD.
Q16620	BDNF/NT3 growth factors receptor (TrkB)	Loss of Function	Loss of function of TrkB is linked to degeneration of neurons; and an increase in TrkB may inhibit neurodegeneration.	TrkB is found to be decreased in AD brains. Its decrease can cause signaling dysfunction in the nervous system.
P02511	Alpha-crystallin B chain	Gain of function	Alpha-crystallin B chain has increased expression in the substantia nigra of the PD brain.	Alpha-crystallin B chain is associated with amyloid deposition, and aberrant expression may affect proper protein folding.

Table 3. Application of qualitative analysis to describe the functions of protein etiological factors found to be in common between Alzheimer's disease and Parkinson's disease.

Out of 435 disease pairs, there were 18 disease pairs which had statistically significant overlap based on both comorbidity and shared proteins with AD. These diseases are listed in Table 2. From the quantitative analysis conducted, AD and Parkinson's disease were found to be a significant comorbid pair. The quantitative results revealed several proteins that were shared between the two diseases. Of those, literature analysis confirmed the association of the following proteins: tyrosine-protein kinase Fyn (Fyn kinase), BDNF/NT3 growth factors receptor, and alpha-crystallin B chain (CRYAB). Table 3 provides a brief overview of the analysis of the function of these proteins as seen in Parkinson's disease and AD.

Tyrosine-protein kinase Fyn (Fyn), UniProt accession P06241, was an overlapping protein in the comorbid disease pair. Fyn is a non-receptor tyrosine kinase that plays many different roles, including processes such as cell growth and survival, immune response, cell signaling, cytoskeletal remodeling, and axon guidance (21). Fyn kinase is currently being explored as a possible therapeutic target for AD (24). Instead of focusing on the accumulation of amyloid beta plaques for treatment, Fyn is a downstream effector in the cascade. Soluble forms of amyloid beta protein bind to cellular prion protein on the surface of neurons, which then begins a cascade that ends at Fyn kinase (25). Fyn is also known to be associated with microtubule-associated protein tau, a protein known to be involved in AD pathology (25, 26). There are also studies exploring Fyn kinase's role in Parkinson's disease, one of which found that Fyn is a regulator of proinflammatory signaling (27). The study showed that Fyn is needed in order for proinflammatory responses, such as cytokine release, and that it may be upregulated in chronic inflammation. Overall, the study had demonstrated that, as an upstream signaling regulator, Fyn mediates microglial neuroinflammatory processes in Parkinson's disease (27).

Another shared protein between Parkinson's and AD was the BDNF/NT3 growth factors receptor (TrkB), UniProt Accession Q16620. This is a receptor tyrosine kinase; the BDNF/NT3 growth factors receptor functions as a regulator of neuron survival, growth, and differentiation, as well as synapse formation and plasticity. It may also have a part in signaling and communication between neurons and glia (22). BDNF signaling is known to have effects on gene expression and protein synthesis and to modulate neurotransmission (28). Alpha-synuclein aggregation is characteristic in Parkinson's disease; alpha-synuclein has been found to selectively interact with TrkB and inhibit its signaling, leading to neuronal cell death and further contributing to the pathology of the disease (29). Furthermore, in animal models of Parkinson's disease, it was found that BDNF prevents degeneration of dopaminergic neurons and can improve both motor performance and neurotransmission (30). An increase in BDNF and TrkB was shown to inhibit neurodegeneration and was also associated with gene expression related to neuronal proliferation and survival (30). Studies have shown that BDNF and TrkB both significantly decrease in AD (31). One study suggests that a reduction in TrkB can cause signaling dysfunctions in early AD (31). The mechanism at play may be linked to the deprivation of TrkB, causing an increase in inflammatory cytokines (32).

Commonly, AD and Parkinson's disease involve alpha-crystallin B chain, UniProt accession P02511. This heat-shock protein acts like a chaperone and is noted to prevent proteins from aggregating in stress conditions (23). The gene for this protein is known as CRYAB; a study has found that knockdown of CYRAB may be useful as a therapy against proteinopathies, with Parkinson's disease as an example (33). Alpha-crystallin B chain has also been found to have increased expression in the substantia nigra in Parkinson's disease, and may play a role in glial pathology during dopaminergic neuron degeneration (34). Another study found an association between alpha-crystallin B chain with the deposition of amyloid plaques in AD (35). It is one of many small heat shock proteins that have been noted to be linked to pathological lesions found in AD (36).

Dicussion

Characterization of disease comorbidities often involves genetic components between diseases and neglects the functional overlap of proteins that are involved (37). Considering that various risk factors may increase prevalence of AD, studying its functional characterization can aid in identifying its molecular mechanisms (9). Contrasted with previous studies, our work proposes that functional measures could provide additional information that could possibly be missed if only genetic overlap was considered (37).

Subsequent to the identification of statistically significant comorbidities of AD using the quantitative analyses of the protein overlap, a discussion of the functions of the overlapping proteins was considered using reviews of the literature. We contrast the function of the overlapping proteins as describing the non-diseased state, its putative role in AD, its putative role in the disease comorbid with AD, and its putative role specificity in the comorbid condition. Comparison of the role of the protein was conducted to determine how its potential change in function, either a gain or loss of function, contributed to the comorbidity. For example, loss of function or gain of function mutations associated with a protein may result in an increased risk for each disease separately and for the comorbid condition.

We focused on one comorbid disease pair: AD and Parkinson's disease. Three overlapping proteins were identified: tyrosine-protein kinase Fyn, BDNF/NT3 growth factors receptor (TrkB), and alpha-crystallin B chain. Based on the quantitative results when comparing the two diseases, the overlapping proteins indicate that there is an association between the two conditions. The qualitative review of the literature describing the protein factors was found to support the quantitative results. Insight is formed regarding the molecular etiology of each neurodegenerative disorder as well as for the comorbidity, which happens in a significant number of individuals.

Conclusion

As we continue with the research, we will utilize a more comprehensive set of comorbidities of AD based on the population studies which is provided in tabular form and accessible via a web interface. We aim to compile more extensive data by including recent protein analysis from UniProt and better-defined diagnosis of AD, which may be achieved through ICD-10 diagnostic codes.

Although our study represents an extensive analysis of functional overlap between co-occurring proteins of AD comorbidities, there are a few limitations that should be noted. First, the study population was restricted to only Medicare patients (4). This limited the dataset, as the age range was bound to individuals 65 and older, which may exclude early-onset AD cases. Second, the dataset only considered inpatient diagnostic codes and disregarded outpatient codes. The lack of attention to outpatient claims could indicate that the claims data only recognizes moderate to severe cases, thus excluding the milder cases. Third, the extracted dataset adopted ICD-9 as its primary diagnostic code, because the dataset was generated prior to the implementation of ICD-10. The ICD-10 coding system classifies diseases that are consistent with current clinical practices and medical advances (38). There is a greater level of detail enabling greater specificity that is aligned with ICD-10 diagnostic codes (39). Fourth, the study population predominantly included white patients only, making up about 90% of the entire study. The lack of racial diversity limits the appearance of different comorbidity pairs and protein function overlap within those pairs. Fifth, the Medicare data included records from multiple hospitals. Removal of personal identifiers of the patients made it difficult to define how AD was diagnosed in these individuals. For example, why were these patients diagnosed with AD as opposed to dementia, vascular dementia, or frontotemporal dementia? Additionally, were these patients diagnosed based on postmortem autopsy? Lastly, prevalence of AD in the United States' general population was taken from the 2010 Census Bureau. With the 2020 Census currently taking place, there may be inaccuracies included in the numbers, thus not being representative in the future.

Acknowledgments

Laura Christman, Kayla Day, Amy Hoang, Rachel Simon, and William McLaughlin wrote the manuscript. Laura Christman, Kayla Day, Amy Hoang, Rachel Simon, William McLaughlin, Paul J. DePietro, Jesse Clayton, and Kelley Chan helped develop the computer programs. We thank Clinton Sonier for helping to organize the data and programs. Funding was provided in part by the Marquardt Foundation for Alzheimer's Research.

Disclosures

The authors and other contributors to this study have no conflicts of interest.

References

1. Rzhetsky A, Wajngurt D, Park N, Zheng T. Probing genetic overlap among complex human phenotypes. *PNAS*. 2007;104(28):11694–11699.
2. Park J, Lee D-S, Christakis NA, Barabási A-L. The impact of cellular networks on disease comorbidity. *Mol Sys Biol*. 2009;5(1):262.
3. Loscalzo J, Kohane I, Barabasi A-L. Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. *Mol Sys Biol*. 2007;3(1):124.
4. Hidalgo CA, Blumm N, Barabási A-L, Christakis NA. A Dynamic Network Approach for the Study of Human Phenotypes. *PLoS Comput Biol*. 2009;5(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2661364/
5. Kim JH, Son KY, Shin DW, Kim SH, Yun JW, Shin JH, et al. Network analysis of human diseases using Korean nationwide claims data. *J Biomed Inform*. 2016;61:276–282.
6. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. *PNAS*. 2007;104(21):8685–8690.
7. Qiu Y-Q, Zhang S, Zhang X-S, Chen L. Detecting disease associated modules and prioritizing active genes based on high throughput data. *BMC Bioinformatics*. 2010;11(1):26.
8. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. 2017;45(D1):D833–839.
9. Alzheimer's Association. 2020 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2020;16(3):391–460.
10. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology*. 2013;80(19):1778–1783.
11. Jutkowitz E, Kane RL, Gaugler JE, MacLehose RF, Dowd B, Kuntz KM. Societal and Family Lifetime Cost of Dementia: Implications for Policy. *J Am Geriatr Soc*. 2017;65(10):2169–2175.
12. Xu W, Tan L, Wang H-F, Jiang T, Tan M-S, Tan L, et al. Meta-analysis of modifiable risk factors for Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 2015;86(12):1299–1306.
13. Haaksma ML, Vilela LR, Marengoni A, Calderón-Larrañaga A, Leoutsakos J-MS, Olde Rikkert MGM, et al. Comorbidity and progression of late onset Alzheimer's disease: A systematic review. *PLoS One*. 2017;12(5). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5417646/
14. Zhao Y, Kuo T-C, Weir S, Kramer MS, Ash AS. Healthcare costs and utilization for Medicare beneficiaries with Alzheimer's. *BMC Health Serv Res*. 2008;8:108.
15. Lauderdale DS, Furner SE, Miles TP, Goldberg J. Epidemiologic Uses of Medicare Data. *Epidemiol Rev*. 1993;15(2):319–327.
16. Mitchell JB, Bubolz T, Paul JE, Pashos CL, Escarce JJ, Muhlbaier LH, et al. Using Medicare claims for outcomes research. *Medical Care*. 1994;32(7):JS38–51.
17. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506-515.)
18. Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res*. 2019;47(Database issue):D955–962.
19. DisGeNET Downloads [Internet]. DisGeNET. Available from: https://www.disgenet.org/downloads
20. Cramér H. Mathematical methods of statistics. *Princeton University Press*; 1999: 43.
21. Tyrosine-protein kinase Fyn [Internet]. *UniProt Consortium: European Bioinformatics Institute - Protein Information Resource*. 2020. Available from: https://www.uniprot.org/uniprot/P06241
22. BDNF/NT-3 growth factors receptor [Internet]. *UniProt Consortium: European Bioinformatics Institute - Protein Information Resource*. 2020. Available from: https://www.uniprot.org/uniprot/Q16620
23. Alpha-crystallin B chain [Internet]. *UniProt Consortium: European Bioinformatics Institute - Protein Information Resource*. 2020. Available from: https://www.uniprot.org/uniprot/P02511
24. Nygaard HB, van Dyck CH, Strittmatter SM. Fyn kinase inhibition as a novel therapy for Alzheimer's disease. *Alzheimers Res Ther*. 2014;6(1):8.
25. Nygaard HB. Targeting Fyn kinase in Alzheimer's disease. *Biol psychiatry*. 2018;83(4):369-376.
26. Shirazi SK, Wood JG. The protein tyrosine kinase, fyn, in Alzheimer's disease pathology. *Neuroreport*. 1993;4(4):435-437.
27. Panicker N, Saminathan H, Jin H, Neal M, Harischandra DS, Gordon R, Kanthasamy K, et al. Fyn kinase regulates microglial neuroinflammatory responses in cell culture and animal models of Parkinson's disease. *J Neurosci*. 2015;35(27):10058-10077.

28. Mercado NM, Collier TJ, Sortwell CE, Steece-Collier K. BDNF in the aged brain: translational implications for Parkinson's disease. *Austin Neurol Neurosci*. 2017;2(2).
29. Kang SS, Zhang Z, Liu X, Manfredsson FP, Benskey MJ, Cao X, et al. TrkB neurotrophic activities are blocked by α-synuclein, triggering dopaminergic cell death in Parkinson's disease. *Proc Natl Acad Sci USA*. 2017;114(40):10773-10778.
30. Palasz E, Wysocka A, Gasiorowska A, Chalimoniuk M, Niewiadomski W, Niewiadomska G. BDNF as a promising therapeutic agent in Parkinson's disease. *Int J Mol Sci*. 2020;21(3):1170.
31. Devi L, Ohno M. TrkB reduction exacerbates Alzheimer's disease-like signaling aberrations and memory deficits without affecting β-amyloidosis in 5XFAD mice. *Transl Psychiatry*. 2015;5(5):e562.
32. Wang ZH, Xiang J, Liu X, Yu SP, Manfredsson FP, Sandoval IM, et al. Deficiency in BDNF/TrkB Neurotrophic Activity Stimulates δ-Secretase by Upregulating C/EBPβ in Alzheimer's Disease. *Cell Rep*. 2019;28(3):655-669.
33. Lu SZ, Guo YS, Liang PZ, Zhang SZ, Yin S, Yin YQ, et al. Suppression of astrocytic autophagy by αB-crystallin contributes to α-synuclein inclusion formation. *Transl Neurodegener*. 2019;8(1):3.
34. Liu Y, Zhou Q, Tang M, Fu N, Shao W, Zhang S, et al. Upregulation of alphaB-crystallin expression in the substantia nigra of patients with Parkinson's disease. *Neurobiol Aging*. 2015;36(4):1686-1691.
35. Renkawek K, Voorter CE, Bosman GJ, Van Workum FP, De Jong WW. Expression of αB-crystallin in Alzheimer's disease. *Acta Neuropathol*. 1994;87(2):155-160.
36. Wilhelmus MM, Boelens WC, Otte-Höller I, Kamps B, Kusters B, Maat-Schieman ML, et al. Small heat shock protein HspB8: its distribution in Alzheimer's disease brains and its inhibition of amyloid-β protein aggregation and cerebrovascular amyloid-β toxicity. *Acta Neuropathol*. 2006;111(2):139-149.
37. Rubio-Perez C, Guney E, Aguilar D, Piñero J, Garcia-Garcia J, Iadarola B, et al. Genetic and functional characterization of disease associations explains comorbidity. *Scientific Reports*. 2017;7(1):1–14.
38. Cartwright DJ. ICD-9-CM to ICD-10-CM Codes: What? Why? How?. *Adv Wound Care* (New Rochelle). 2013;2(10):588–592.
39. ICD - ICD-10-CM - International Classification of Diseases,(ICD-10-CM/PCS transition [Internet]. 2019. Available from: https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm