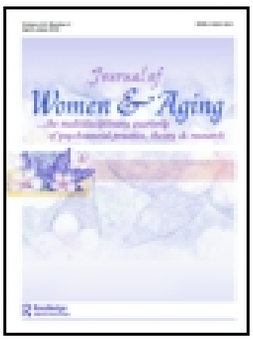

Journal of Women & Aging

ISSN: 0895-2841 (Print) 1540-7322 (Online) Journal homepage: http://www.tandfonline.com/loi/wjwa20

Sex-specific patterns and differences in dementia and Alzheimer's disease using informatics approaches

Jay Geronimo Ronquillo, Merritt Rachel Baer & William T. Lester

To cite this article: Jay Geronimo Ronquillo, Merritt Rachel Baer & William T. Lester (2016): Sex-specific patterns and differences in dementia and Alzheimer's disease using informatics approaches, Journal of Women & Aging, DOI: 10.1080/08952841.2015.1018038

To link to this article: http://dx.doi.org/10.1080/08952841.2015.1018038

Published online: 22 Apr 2016.

JOURNAL OF WOMEN & AGING
http://dx.doi.org/10.1080/08952841.2015.1018038

Routledge
Taylor & Francis Group

Sex-specific patterns and differences in dementia and Alzheimer's disease using informatics approaches

Jay Geronimo Ronquillo[a], Merritt Rachel Baer[a], and William T. Lester[b]

[a]Grinformatics, Haymarket, Virginia, USA; [b]Laboratory of Computer Science, Massachusetts General Hospital, and Harvard Medical School, Boston, Massachusetts, USA

ABSTRACT

The National Institutes of Health Office of Research on Women's Health recently highlighted the critical need for explicitly addressing sex differences in biomedical research, including Alzheimer's disease and dementia. The purpose of our study was to perform a sex-stratified analysis of cognitive impairment using diverse medical, clinical, and genetic factors of unprecedented scale and scope by applying informatics approaches to three large Alzheimer's databases. Analyses suggested females were 1.5 times more likely than males to have a documented diagnosis of probable Alzheimer's disease, and several other factors fell along sex-specific lines and were possibly associated with severity of cognitive impairment.

KEYWORDS

Alzheimer's disease; biomedical informatics; data analytics; dementia; sex and gender

Introduction

Alzheimer's disease (AD) is a serious form of dementia that affects more than 5 million people in the United States, with the majority older than 65 years and nearly twice as many women affected as men (Alzheimer's Association, 2013). By 2050, the number of elderly affected by AD is predicted to increase to 13.8 million, with health care costs increasing more than fivefold to $1.2 trillion, representing a significant public health burden to hospitals, patients, and their families (Alzheimer's Association, 2013).

The National Institutes of Health (NIH) and the NIH Office of Research on Women's Health have recently highlighted the critical need for explicitly addressing male and female differences in clinical and biomedical research (Clayton & Collins, 2014; Collins & Tabak, 2014). Sex plays a critical but only partially explored role in the etiology, diagnosis, and prognosis of AD (Carter, Resnick, Mallampalli, & Kalbarczyk, 2012). Better understanding these sex-specific differences and patterns would address an important barrier to more personalized health care for the growing elderly population (Ballard et al., 2011; Olgiati et al., 2013; Spies et al., 2013; Wei, Visweswaran, & Cooper, 2011). Looking at several modalities simultaneously, however, requires bringing large neurodegenerative data sets together for analysis (Bateman et al., 2012; Kohane, Drazen, & Campion, 2012; Ogishima et al., 2013; Romero et al., 2009; Wei et al., 2011). To our knowledge, the technology to handle "big data" for aging and dementia research and still characterize important male/female differences does not exist (Groves, Kayyali, Knott, & Van Kuiken, 2013; Khachaturian, Meranus, Kukull, & Khachaturian, 2013; Kohane et al., 2012).

While an exact description remains elusive, data analytics for clinical research characterizes information along three main dimensions: (1) *volume* refers to the increasingly massive amounts of data that must be organized, stored, and processed; (2) *variety* relates to the diverse formats that must be understood to quickly navigate complex information, and (3) *velocity* describes the rapid

CONTACT Jay Geronimo Ronquillo jgr@grinformatics.com Grinformatics, 5501 Merchants View Square #190, Haymarket, VA 20169.

analytical needs in the clinical setting (Bourne, 2014; Khachaturian et al., 2013; Kohane et al., 2012; McAfee & Brynjolfsson, 2012). There are major research efforts underway to improve the assessment, diagnosis, and treatment of dementia and AD. Clinical, medical, and genetic factors have partly explained the risk of developing cognitive impairment, but more effective prediction requires an integrated assessment of these modalities (Ballard et al., 2011; Spies et al., 2013; Wei et al., 2011). The purpose of this study was to (1) leverage informatics approaches to integrate large neurodegenerative data sets containing diverse data for males and females with cognitive impairment (mild cognitive impairment, dementia, and/or Alzheimer's disease); and (2) analyze sex-specific differences in an unprecedented number of clinical, medical, and genetic factors and discuss their implications for the health of men and women.

Methods

Study population

The population of cognitively impaired patients originated from three large clinical research data sets containing de-identified patient data relevant for the evaluation of cognitive impairment: the Alzheimer's Disease Neuroimaging Initiative (ADNI), the National Alzheimer's Coordinating Center (NACC), and the Coalition Against Major Diseases (CAMD) (Alzheimer's Disease Neuroimaging Initiative, n.d.; C-Path Online Data Repository (CODR), 2013; National Alzheimer's Coordinating Center, 2013).

The CAMD, one of seven consortia of the Critical Path Institute, was formed in 2008 by the Critical Path Institute in collaboration with the Engelberg Center for Health Care Reform at the Brookings Institution. The Coalition brings together patient groups, biopharmaceutical companies, scientists from academia, the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), the National Institute of Neurological Disorders and Stroke (NINDS), and the National Institute on Aging (NIA). The CAMD data have been volunteered by over 200 scientists of member companies and nonmember organizations (C-Path Online Data Repository (CODR), 2013). CAMD data were downloaded on July 11, 2013.

The NACC database was created using clinical and neuropathological research data from Alzheimer's Disease Centers throughout the United States, in collaboration with both the Alzheimer's Disease Genetics Consortium and the National Cell Repository for Alzheimer's Disease (National Alzheimer's Coordinating Center, 2013). The NACC database contains data from the Uniform Data Set and from 34 past and present Alzheimer's Disease Centers between 1984 and September 2013.

The ADNI was launched in 2003 by the NIA, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the FDA, private pharmaceutical companies, and nonprofit organizations, as a $60 million, 5-year public-private partnership. Its primary goal has been to test for markers intended to aid researchers and clinicians to develop new treatments and improve the efficiency of clinical trials. Subjects from three protocols (ADNI, ADNI-GO, ADNI-2) have been recruited from over 50 sites across the U.S. and Canada (Alzheimer's Disease Neuroimaging Initiative, n.d.). ADNI data were downloaded on October 10, 2013.

Data collection, definition, and classification

An informatics data pipeline in the Python programming language was created to extract existing fields, derive relevant factors, and standardize raw neurodegenerative data from ADNI, CAMD, and NACC into a single population database. Patients were categorized as having mild cognitive impairment (MCI) or dementia based on original baseline classifications, and patients with dementia were further subclassified as probable AD if documentation of a probable diagnosis was found. Relevant factors commonly associated with AD in the published literature were derived from health or medical history

tables: age, sex, history of alcohol abuse, history of smoking (current or former), Mini-Mental Status Exam (MMSE) scores, apolipoprotein E (APOE) genotype, obesity, diabetes, hypertension, hypercholesterolemia, and stroke (Ballard et al., 2011; Editors of *The Lancet Neurology*, 2010). A positive finding was defined as the presence of a relevant clinical term or abbreviation in the context of a prior diagnosis, medical history, or treatment for one of the aforementioned conditions. The APOE genotype was categorized as the presence or absence of any ε4 alleles from the appropriate subject characteristics or genetic results table. Positive obesity status was categorized as the presence of a body mass index (BMI) or calculated BMI ≥30 kg/m^2 using one of the following formulas: (1) weight (kg)/height squared (m^2) or (2) [weight (pounds)/height squared (inches^2)] × 703 (NOEIE Panel, 1998). For unstructured data elements, automated extraction and classification using natural language processing techniques was followed by manual review and confirmation of all positive and negative findings.

Statistical analysis

Summary statistics were collected for continuous data as means with standard deviations (*SD*) and medians with interquartile ranges (IQR), and for categorical data as frequencies and percentages. Two-sample *t*-tests were used to compare sex differences for continuous variables, while differences in the distribution of categorical variables (age group by sex, race/ethnic group by sex, characteristics by sex) were evaluated using chi-square tests.

For the patient subpopulation with dementia, the probability of documented probable AD was modeled with logistic regression using candidate predictor variables representing important clinical, demographic, medical, or genetic factors commonly associated with AD: age, sex, history of alcohol abuse, history of smoking, family history of dementia, diabetes, hypercholesterolemia, hypertension, stroke, abnormal MMSE (score < 27), obesity, and positive APOE ε4 status (Ballard et al., 2011; Carter et al., 2012; Editors of *The Lancet Neurology*, 2010; Sabia et al., 2014). A P value < .05 was considered significant for all analyses, which were performed using R version 3.0.2 (R Foundation for Statistical Computing) and Microsoft Excel Version 14.2.0 (Redmond, WA).

Results

General patient characteristics

There were a total of 24,270 patients with MCI or dementia, with 12,737 (52.5%) females and 11,533 (47.5%) males (Table 1). Overall, 8,138 (33.5%) patients were classified with MCI and 16,132 (66.5%) with dementia, of which 12,505 (77.5%) had documentation of probable AD. Patients had a mean age of 73.3 ± 9.5 years (median 74.0, IQR 67.0–80.0), with 19,696 (81.1%) over the age of 65.

As shown in Table 1, women with cognitive impairment were older in age distribution as well as overall mean age compared to males (p < .001 for both). While nearly 88% of males were White, approximately 82% of females were White (2,282 or 17.9% of females were minorities). There were 2.2 times more females affected by dementia than MCI, and only 1.8 times more males affected by dementia than MCI. Compared to males, a larger fraction of females with dementia had documentation of probable AD (81.9% vs. 72.4%). There were a total of 1,676 (13.2%) reported female deaths in contrast to 2,223 (19.3%) males. A greater proportion of males with dementia than MCI died (23.7% vs. 11.3%), and similarly more females with dementia died (15.8% vs. 7.5%). However, among dementia patients with probable AD, 980 (18.2%) males and 920 (12.9%) females died, compared to dementia patients without probable AD, of whom 779 (38.0%) males and 453 (28.7%) females died.

Sex-specific differences in cognitively impaired patients

Stratification by sex showed differences in a variety of relevant factors and modalities. The fraction of demented males with a family history of dementia and abnormal baseline MMSE was 6.4 and 52.1

Table 1. Demographics of cognitively impaired patient population.

Demographic	Female (*N* = 12,737)	Male (*N* = 11,533)	*p* value
Age group (years), no (%)			
20–29	4 (0)	3 (0)	
30–39	12 (0.1)	21 (0.2)	
40–49	96 (0.8)	118 (1.0)	
50–59	1,011 (7.9)	903 (7.8)	
60–69	2,867 (22.5)	2,709 (23.5)	< .001
70–79	5,018 (39.4)	4,782 (41.5)	
80–89	3,333 (26.2)	2,796 (24.2)	
90+	304 (2.4)	180 (1.6)	
Unknown	92 (0.7)	21 (0.2)	
Age (years), mean ± *SD*	73.5 ± 9.6	73.0 ± 9.4	< .001
Age (years), median (IQR)	73 (66–80)	74 (67–80)	
Race or ethnic group, No (%)			
White	10,153 (79.7)	10,142 (87.9)	
Black	1,464 (11.5)	679 (5.9)	
American Indian	68 (0.5)	47 (0.4)	
Asian or Pacific Islander	379 (3.0)	312 (2.7)	< .001
Other	371 (2.9)	200 (1.7)	
Unknown	302 (2.4)	153 (1.3)	

Note. *SD* = standard deviation; IQR = interquartile range.

Table 2. Important characteristics stratified by severity of cognitive impairment and sex.

	Mild cognitive impairment*		Dementia*	
Characteristic, no (%)	Male (*N* = 4,110)	Female (*N* = 4,028)	Male (*N* = 7423)	Female (*N* = 8,709)
Age over 65	3,388 (82.4)	3,230 (80.2)	5,945 (80.1)	7,130 (81.9)
Alcohol abuse	295 (7.2)	108 (2.7)	596 (8.0)	214 (2.5)
Diabetes	663 (16.1)	561 (13.9)	995 (13.4)	1,053(12.1)
Family history of dementia	1,807 (44.0)	1,776 (44.1)	3,739 (50.4)	4,166 (47.8)
Hypercholesterolemia	2,369 (57.6)	2,147 (53.3)	3,575 (48.2)	3,715 (42.7)
Hypertension	2,341 (57.0)	2,261 (56.1)	3,679 (49.6)	4,493 (51.6)
Abnormal MMSE	1,332 (32.4)	1,261 (31.3)	6,271 (84.5)	7,672 (88.1)
Obese	783 (19.1)	817 (20.3)	1,092 (14.7)	1276 (14.7)
Positive APOE ε4	1,091 (26.5)	998 (24.8)	2,274 (30.6)	2,609 (30.0)
Smoking history	1,862 (45.3)	1,433 (35.6)	2,569 (34.6)	1,959 (22.5)
Stroke	280 (6.8)	254 (6.3)	558 (7.5)	602 (6.9)

Note. *Indicates statistical significance $p < .05$ (male vs. female); MMSE = Mini-Mental Status Exam; APOE = apolipoprotein E.

percentage points higher respectively than those with MCI (Table 2). Conversely, males with dementia had lower relative differences in hypercholesterolemia (by 9.5 percentage points), hypertension (7.4 points), and smoking (10.7 points) than males with MCI. Similarly for females with dementia, the fraction of those with a family history of dementia and abnormal MMSE was 3.7 and 56.8 percentage points higher respectively than their MCI counterparts. Finally, females with dementia had a percentage difference in hypercholesterolemia, hypertension, and smoking that was lower by 10.6, 4.5, and 13.1 points respectively compared to females with MCI.

Comparing factors between males and females showed similar patterns for MCI and dementia, although the relative magnitudes differed. For MCI, a larger fraction of males had a history of alcohol abuse (4.5 percentage points higher), hypercholesterolemia (4.3 points), and smoking (9.7 points) than females. Similarly, a larger fraction of males with dementia had alcohol (5.6 percentage points higher), hypercholesterolemia (5.5 points), and smoking (12.1 points) histories than their female counterparts.

Subanalysis of probable AD associations with demographic, medical, clinical, and genetic factors

Development of a logistic regression model for patients with dementia showed that seven factors were significantly associated with higher or lower odds of documented probable AD (Table 3).

Table 3. Multivariable covariates for documented probable AD in dementia population.

Covariate	OR	95% CI	*p* value
Age*	1.061	1.054–1.069	< .001
Male*	0.639	0.557–0.732	< .001
Alcohol abuse*	0.702	0.548–0.902	.005
Smoking history	0.992	0.866–1.137	.911
Obese	0.933	0.792–1.101	.407
Family history of dementia*	1.272	1.113–1.454	< .001
Diabetes	0.996	0.815–1.222	.972
Hypercholesterolemia	0.971	0.843–1.118	.681
Hypertension	0.935	0.809–1.080	.36
Stroke*	0.413	0.326–0.520	< .001
Abnormal MMSE*	1.989	1.673–2.364	< .001
Positive APOE ε4*	2.194	1.919–2.511	< .001

Note. *Indicates statistical significance $p < .05$; OR = odds ratio; CI = confidence interval; MMSE = Mini-Mental Status Exam; APOE = apolipoprotein E.

Females with dementia were 1.565 times more likely than males to have a documented AD diagnosis ($p < .001$). Controlling for sex and other factors, history of alcohol abuse decreased odds by 0.702 ($p < .005$), while stroke decreased probable AD odds by 0.413 ($p < .001$). In contrast, a positive family history of dementia increased odds of probable AD 1.272 times, and positive APOE ε4 increased odds by 2.194 ($p < .001$ for all). No significant associations were detected for smoking history, obesity, diabetes, hypercholesterolemia, or hypertension ($p > .05$ for all).

Discussion

This study shows that informatics tools are capable of meeting "big" data needs as well as addressing important NIH requirements and initiatives to clearly distinguish differences between males and females in aging and dementia research. To our knowledge, this study is one of the largest integrated modality analyses of sex-specific patterns and differences in cognitively impaired individuals to date (Hannawi & Smirnakis, 2013; Khachaturian et al., 2013).

The results suggest several sex-specific differences associated with the level of cognitive function. For both MCI and dementia, males had a consistently larger proportional history of alcohol use and smoking than females, although intrasex differences were relatively small. While earlier studies have identified potential links with cognitive impairment, the full impact of these modifiable behavioral factors remains to be characterized (Beydoun et al., 2013; Edge, 2010). Interestingly, for both males and females, hypercholesterolemia and hypertension were more common in patients with MCI than dementia, and a larger fraction of deaths in the dementia population occurred in patients without probable AD than with probable AD. While this may suggest an inverse relationship between some medical conditions and the severity of cognitive impairment, it may also possibly be due to survivorship bias. These findings highlight the complex relationships between different sex-related factors and comorbidities in a large population of cognitively impaired individuals (Beydoun et al., 2013; Carter et al., 2012).

Several models for AD and dementia have investigated the impact of sex along with a variety of other factors (Bateman et al., 2012; Ferrari et al., 2013; Petersen, 2010; Spies et al., 2013). Consistent with prior studies, our regression analysis found several factors associated with higher odds of documented probable AD, including increasing age, positive family history, abnormal MMSE, and positive APOE ε4 results (Alzheimer's Association, 2013; Ballard et al., 2011; Bateman et al., 2012; Beydoun et al., 2013). According to our analysis, females were 1.5 times more likely to have a documented diagnosis of probable AD than males, further highlighting potential sex differences associated with dementia and AD. The association of alcohol use with lower odds of probable AD may seem counterintuitive, but some studies have supported a potentially protective sex-specific effect of alcohol on cognitive decline (Sabia et al., 2014). Finally, our study highlights another

potential application of these analyses: identifying risk factors that anticipate provider behavior (i.e., their likelihood of documenting a probable AD diagnosis). Because the gold standard for AD diagnosis is biopsy during autopsy, a clinical diagnosis is currently the only way a patient may possibly know if they have AD instead of another cause of dementia such as Lewy bodies, frontotemporal lobar degeneration, or vascular dementia (Karantzoulis & Galvin, 2011). As health information technology enables more accurate data to be collected and documented, the ability to predict physician behavior may be just as important to patient care as the diagnoses themselves (Jha, 2011; Jha et al., 2009; Morrison, Fernando, Kalra, Cresswell, & Sheikh, 2013; Ronquillo, 2012). For example, electronic medical record and population health data could be used to create sophisticated clinical decision support tools to help physicians rapidly identify optimal treatments targeted to the unique clinical situation of each male and female patient being assessed for Alzheimer's disease at the point of care (Jha, 2011; Kohane et al., 2012).

Our study had several limitations. First, the results may not be generalizable, given the focus on research patients in clinical studies of dementia emphasizing Alzheimer's disease. As a result, there is the potential for selection bias as well as underreporting of certain factors (e.g., deaths) in this particular population. However, we believe the study breadth and depth provide insight into relevant sex-related differences for an important patient population. Second, several important modalities were not available for all three data sets at the time of this writing; specifically, the inclusion of imaging and cerebrospinal fluid biomarkers would have significantly enhanced the analysis and are planned for future studies. Finally, the use of de-identified data sets means a potential concern for missing or overlapping data and an inability to recheck or reconfirm previously collected data. However, our study focused on reliable data sources (ADNI, NACC, CAMD) and accepted informatics methods for extracting information from existing results, and it is unlikely any issues would have substantially changed our conclusions (Ronquillo, Li, & Lester, 2012). As the era of big data for clinical research moves forward, powerful analytics tools that can handle imperfect and large "real-world" data sets, quickly assess important clinical patterns, and identify relevant sex-specific differences will play a central role in providing individualized care for the aging patient population (Ohno-Machado, 2014; Ronquillo, 2012).

Conclusion

In summary, sex-specific differences in dementia are well established, and our study builds on this previous work by performing a sex-stratified analysis with diverse medical, clinical, and genetic factors of unprecedented scale and scope by leveraging data analytics and informatics approaches. We found that several significant factors fell along sex-specific lines and were possibly associated with different levels of cognitive impairment. Future studies can extend our informatics approaches to larger data sets and further drive health innovation and discovery for the aging population. As research efforts further intensify for age-related conditions such as dementia and Alzheimer's disease, we have a unique opportunity to capitalize on the increasing availability of data, sophisticated informatics tools, and public readiness to solve critical challenges facing health care today.

Acknowledgments

Data used in the preparation of this article were obtained from the CAMD. As such, the investigators within CAMD contributed to the design and implementation of the CAMD database and/or provided data but did not participate in the analysis of the data or the writing of this report. Data used in preparation of this article were further obtained from the ADNI database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Funding

All authors included in the manuscript provided substantial contribution to conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the completed manuscript. JGR had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

The NACC database was funded by NIA Grant U01 AG016976. The ADNI database was funded by NIH Grant U01 AG024904.

ADNI is funded by the NIA, NIBIB, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Steven Ferris, PhD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG016570 (PI David Teplow, PhD), P50 AG005131 (PI Douglas Galasko, MD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar López, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P50 AG005136 (PI Thomas Montine, MD, PhD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), and P50 AG005681 (PI John Morris, MD).

References

Alzheimer's Association. (2013). 2013 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *9*(2), 19–45.

Alzheimer's Disease Neuroimaging Initiative. (n.d.).Retrieved from http://adni.loni.usc.edu/

Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., & Jones, E. (2011). Alzheimer's disease. *The Lancet*, *377*(9770), 1019–1031. doi:10.1016/S0140-6736(10)61349-9

Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., & Morris, J. C. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *The New England Journal of Medicine*, *367*(9), 795–804. doi:10.1056/NEJMoa1202753

Beydoun, M. A., Beydoun, H. A., Kaufman, J. S., An, Y., Resnick, S. M., O'Brien, R., … Zonderman, A. B. (2013). Apolipoprotein E ε4 allele interacts with sex and cognitive status to influence all-cause and cause-specific mortality in U.S. older adults. *Journal of the American Geriatrics Society*, *61*(4), 525–534. doi:10.1111/jgs.12156

Bourne, P. E. (2014). What Big Data means to me. *Journal of the American Medical Informatics Association : JAMIA*, *21*(2), 194–194. doi:10.1136/amiajnl-2014-002651

Carter, C. L., Resnick, E. M., Mallampalli, M., & Kalbarczyk, A. (2012). Sex and gender differences in Alzheimer's disease: Recommendations for future research. *Journal of Women's Health*, *21*(10), 1018–1023. doi:10.1089/jwh.2012.3789

Clayton, J. A., & Collins, F. S. (2014). Policy: NIH to balance sex in cell and animal studies. *Nature*, *509*(7500), 282–283. doi:10.1038/509282a

Collins, F. S., & Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. *Nature*, *505*(7485), 612–613. doi:10.1038/505612a

C-Path Online Data Repository (CODR). (2013). Retrieved from http://c-path.org/programs/camd/

Editors of *The Lancet Neurology*. (2010). Alzheimer's disease prevention: A reality check. *The Lancet Neurology*, 9(7), 643. doi:10.1016/S14744422(10)70145-9

Ferrari, R., Dawoodi, S., Raju, M., Thumma, A., Hynan, L. S., Maasumi, S. H., … Momeni, P. (2013). Androgen receptor gene and sex-specific Alzheimer's disease. *Neurobiology of Aging*, *34*(8), 2077.e19–2077.e20. doi:10.1016/j.neurobiolaging.2013.02.017

Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The "big data" revolution in health care: Accelerating value and innovation. *McKinsey Quarterly*, (January), 1–22.

Hannawi, Y., & Smirnakis, S. M. (2013). Emerging subspecialties: Neuroinformatics. *Neurology*, *80*(15), e166–e168. doi:10.1212/WNL.0b013e31828c2f2e

Jha, A. K. (2011). The promise of electronic records: Around the corner or down the road? *Journal of the American Medical Association*, *306*(8), 880–881. doi:10.1001/jama.2011.1219

Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., … Blumenthal, D. (2009). Use of electronic health records in U.S. hospitals. *The New England Journal of Medicine*, *360*, 1628–1638. doi:10.1056/NEJMsa0900592

Karantzoulis, S., & Galvin, J. E. (2011). Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Review of Neurotherapeutics*, *11*(11), 1579–1591. doi:10.1586/ern.11.155

Khachaturian, A. S., Meranus, D. H., Kukull, W. A., & Khachaturian, Z. S. (2013). Big data, aging, and dementia: Pathways for international harmonization on data sharing. *Alzheimer's & Dementia*, *9*(5), SS61–SS62. doi:10.1016/j.jalz.2013.09.001

Kohane, I. S., Drazen, J. M., & Campion, E. W. (2012). A glimpse of the next 100 years in medicine. *The New England Journal of Medicine*, *367*(26), 2538–2539. doi:10.1056/NEJMe1213371

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60–66.

Morrison, Z., Fernando, B., Kalra, D., Cresswell, K., & Sheikh, A. (2013). National evaluation of the benefits and risks of greater structuring and coding of the electronic health record: Exploratory qualitative investigation. *Journal of the American Medical Informatics Association: JAMIA*, *31*(3), 492–500. doi:10.1136/amiajnl-2013-001666

National Alzheimer's Coordinating Center. (2013). Retrieved from http://www.alz.washington.edu/WEB/researcher_home.html

NOEIE Panel. (1998). *Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults*. Bethesda, MD. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK2003/

Ogishima, S., Mizuno, S., Kikuchi, M., Miyashita, A., Kuwano, R., Tanaka, H., & Nakaya, J. (2013). A map of Alzheimer's disease-signaling pathways: A hope for drug target discovery. *Clinical Pharmacology and Therapeutics*, *93*(5), 399–401. doi:10.1038/clpt.2013.37

Ohno-Machado, L. (2014). NIH's Big Data to Knowledge initiative and the advancement of biomedical informatics. *Journal of the American Medical Informatics Association: JAMIA*, *21*(2), 193. doi:10.1136/amiajnl-2014-002666

Olgiati, P., Politis, A., Albani, D., Rodilossi, S., Polito, L., Zisaki, A., … Serretti, A. (2013). Effects of SORL1 gene on Alzheimer's disease. Focus on gender, neuro-psychiatric symptoms and pro-inflammatory cytokines. *Current Alzheimer Research*, *10*, 154–164. doi:10.2174/1567205011310020005

Petersen, R. C. (2010). Alzheimer's disease: Progress in prediction. *The Lancet Neurology*, *9*(1), 4–5. doi:10.1016/S1474-4422(09)70330-8

Romero, K., De Mars, M., Frank, D., Anthony, M., Neville, J., Kirby, L., … Woosley, R. L. (2009). The coalition against major diseases: Developing tools for an integrated drug development process for Alzheimer's and Parkinson's diseases. *Clinical Pharmacology and Therapeutics*, *86*(4), 365–367. doi:10.1038/clpt.2009.165

Ronquillo, J. G. (2012). How the electronic health record will change the future of health care. *The Yale Journal of Biology and Medicine*, *85*(3), 379–386. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3447201&tool=pmcentrez&rendertype=abstract

Ronquillo, J. G., Li, C., & Lester, W. T. (2012). Genetic testing behavior and reporting patterns in electronic medical records for physicians trained in a primary care specialty or subspecialty. *Journal of the American Medical Informatics Association*, *19*(4), 570–574. doi:10.1136/amiajnl-2011-000621

Sabia, S., Elbaz, A., Britton, A., Bell, S., Dugravot, A., Shipley, M., … Singh-Manoux, A. (2014). Alcohol consumption and cognitive decline in early old age. *Neurology*, *82*, 332–339. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3447201

Spies, P. E., Claassen, J. A. H. R., Peer, P. G. M., Blankenstein, M. A., Teunissen, C. E., Scheltens, P., … Verbeek, M. M. (2013). A prediction model to calculate probability of Alzheimer's disease using cerebrospinal fluid biomarkers. *Alzheimer's & Dementia*, *9*(3), 262–268. doi:10.1016/j.jalz.2012.01.010

Wei, W., Visweswaran, S., & Cooper, G. F. (2011). The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data. *Journal of the American Medical Informatics Association: JAMIA*, *18* (4), 370–375. doi:10.1136/amiajnl-2011-000101