

1 **Exposure to autoimmune disorders increases Alzheimer’s disease risk in a multi-**  
2 **site electronic health record analysis**

3 Grace D. Ramey<sup>1,2</sup>, Alice Tang<sup>2,3</sup>, Thanaphong Phongpreecha<sup>4,5,6</sup>, Monica M. Yang<sup>7</sup>, Sarah R.  
4 Woldemariam<sup>2</sup>, Tomiko T. Oskotsky<sup>2,8</sup>, Thomas J. Montine<sup>4</sup>, Isabel Allen<sup>9</sup>, Zachary A. Miller<sup>10</sup>,  
5 Nima Aghaeepour<sup>5,6,11</sup>, John A. Capra<sup>2,9,\*</sup>, Marina Sirota<sup>2,8,9,\*</sup>

- 6  
7 1 Biological and Medical Informatics PhD Program, UCSF, San Francisco, CA  
8 2 Bakar Computational Health Sciences Institute, UCSF, San Francisco, CA  
9 3 School of Medicine, UCSF, San Francisco, CA  
10 4 Department of Pathology, Stanford University, Palo Alto, CA  
11 5 Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University, Palo  
12 Alto, CA  
13 6 Department of Biomedical Data Science, Stanford University, Palo Alto, CA  
14 7 Department of Medicine, Division of Rheumatology, UCSF, San Francisco, CA  
15 8 Department of Pediatrics, UCSF, San Francisco, CA  
16 9 Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA  
17 10 Memory and Aging Center, UCSF, San Francisco, CA  
18 11 Department of Pediatrics, Stanford University, Palo Alto, CA

19  
20 \*Corresponding authors information:  
21 John A. Capra  
22 UCSF Bakar Computational Health Sciences Institute, San Francisco, CA 94143  
23 tony@capralab.org

24  
25 Marina Sirota  
26 UCSF Bakar Computational Health Sciences Institute, San Francisco, CA 94143  
27 Marina.Sirota@ucsf.edu

28  
29 Conflict of interest: The author has declared that no conflict of interest exists.

## 30 **Abstract**

31 Molecular studies of Alzheimer’s disease (AD) implicate potential links between autoimmunity  
32 and AD, but the underlying clinical relationships between these conditions remain poorly  
33 understood. Electronic health records (EHRs) provide an opportunity to determine the clinical  
34 risk relationship between autoimmune disorders and AD and understand whether specific  
35 disorders and disorder subtypes affect AD risk at the phenotypic level in human populations. We  
36 evaluated relationships between 26 autoimmune disorders and AD across retrospective  
37 observational case-control and cohort study designs in the EHR systems at UCSF and Stanford.  
38 We quantified overall and sex-specific AD risk effects that these autoimmune disorders confer.  
39 We identified significantly increased AD risk in autoimmune disorder patients in both study  
40 designs at UCSF and at Stanford. This pattern was driven by specific autoimmunity subtypes  
41 including endocrine, gastrointestinal, dermatologic, and musculoskeletal disorders. We also  
42 observed increased AD risk from autoimmunity in both women and men, but women with  
43 autoimmune disorders continued to have a higher AD prevalence than men, indicating  
44 persistent sex-specificity. This study identifies autoimmune disorders as strong risk factors for  
45 AD that validate across several study designs and EHR databases. It sets the foundation for  
46 exploring how underlying autoimmune mechanisms increase AD risk and contribute to AD  
47 pathogenesis.

48

## 49 **Introduction**

50 Alzheimer’s disease (AD) is a debilitating neurodegenerative disease that is accompanied by  
51 enormous social and economic burdens, and its prevalence is increasing due to the growing  
52 aging population worldwide (1,2). AD is characterized biologically by amyloid plaques and tau  
53 deposition in the brain, while clinical syndromic diagnoses, such as specific forms of progressive  
54 memory loss, have evolved with the development of better AD diagnostic tests and biomarkers

55 (3,4). Treatments that slow cognitive decline have been a large focus of AD research over the  
56 past several years (5), and understanding of underlying risks and pathogenesis in order to treat  
57 the disease at an earlier stage is especially important considering that current treatments are  
58 still unable to fully rescue normal cognition (6).

59 Many prior molecular and genetic studies suggest a potential role of the immune system and  
60 chronic inflammation in AD pathogenesis (7-9). Indeed, over half of the genetic variants associated  
61 with AD to date are primarily expressed in immune cells (10). Furthermore, several studies point to  
62 immune pathways like the NLRP3 inflammasome (11) and complement system (12-14) becoming  
63 dysregulated in AD experimental animal and human models. However, the extent of contribution of  
64 immune system dysfunction to AD remains poorly understood at the clinical phenotype level in  
65 diverse human populations. Autoimmune disorders are one potential source of chronic immune  
66 dysregulation, and their clinical risk relationship with neurodegenerative diseases like AD has yet to  
67 be fully characterized. Furthermore, autoimmune disorders exhibit a similar sex disparity to AD  
68 (15,16), affecting women more so than men (17-19), suggesting a potential relationship between  
69 biological mechanisms and clinical manifestations that has yet to be quantified. Therefore, studying  
70 risk relationships in individuals with autoimmune disorders and AD will provide a powerful way  
71 to understand the role of autoimmunity as a risk factor for AD overall and across sexes.

72 With advances in curation of real-world datasets (20), such as electronic health records  
73 (EHRs), there is a great opportunity to investigate clinical risk relationships between many  
74 autoimmune disorders and AD. The large sample sizes that EHRs provide also allow for robust  
75 analyses that can be stratified in a sex- and disease-specific manner and validated across  
76 hospital sites. Here, we examine risk associations between AD and 26 different autoimmune  
77 disorders in the UCSF EHR system, and we further stratify our analyses by sex to holistically  
78 understand the biological effects of immune dysfunction on AD pathogenesis at the phenotypic  
79 level. We show that there is a clear and strong risk association between autoimmune disorders  
80 and AD overall and in both men and women, but that women with autoimmune disorders

81 continue to have the highest AD prevalence compared to men with autoimmune disorders. We  
82 show evidence for increased AD risk effects from particular autoimmune disorders and disorder  
83 subtypes, and we further investigate the timing of AD onset in autoimmune disorders patients.  
84 Finally, we provide robust validation of our risk associations from the Stanford EHR system to  
85 demonstrate stability of risk signals across different study designs and mitigate potential  
86 confounding factors.

87

## 88 **Results**

89 We selected patients with autoimmune disorders and/or AD for case-control and cohort study  
90 designs from the UCSF and Stanford EHR databases, which contain information on over 5  
91 million and 3.8 million patients, respectively. For patients with autoimmune disorders, we  
92 identified individuals with each of 26 different autoimmune disorders of interest (Table S1), and  
93 type 1 diabetes, rheumatoid arthritis, autoimmune thyroiditis, and inflammatory bowel disease  
94 were among the most prevalent in the study groups. The case-control study groups consisted of  
95 7,812 individuals (3,906 AD patients and 3,906 non-AD controls, Fig 1, S1) from UCSF and  
96 13,292 individuals (6,646 AD patients and 6,646 non-AD controls, Fig 1, S1) from Stanford. The  
97 cohort study groups consisted of 27,630 individuals (13,815 autoimmune disorder patients and  
98 13,815 non-autoimmune controls, Fig 1, S1) from UCSF and 260,516 individuals (130,258  
99 autoimmune disorder patients and 130,258 non-autoimmune controls, Fig 1, S1) from Stanford.  
100 In the UCSF data set, the mean ( $\pm$ SD) lifespan in the case-control group was 80.05 ( $\pm$ 6.74)  
101 years for AD patients and 80.07 ( $\pm$ 6.73) years in non-AD controls, while in the cohort study  
102 group, the mean lifespan for autoimmunity patients was 69.25 ( $\pm$ 13.21) years and 69.10  
103 ( $\pm$ 13.38) years for non-autoimmune controls. Because individuals had a higher proportion of  
104 censored death information in the Stanford data set, these people were assessed by birth year  
105 rather than lifespan, and we verified that this difference would not significantly influence results

106 (see Methods and Sensitivity Analyses). In the Stanford data set, the mean birth year ( $\pm$ SD) was  
107 1934.53 ( $\pm$ 10.11) for AD patients and 1934.50 ( $\pm$ 10.07) for non-AD controls in the case-control  
108 study group, while in the cohort study group, the mean birth year both for autoimmunity patients  
109 and non-autoimmune controls was 1968.38 ( $\pm$ 21.83) (Table 1). Women made up a majority of  
110 each of our study groups, representing 61.9% (case-control) and 57.9% (cohort) of individuals in  
111 the UCSF study groups, and 62.3% (case-control) and 64.9% (cohort) of individuals in the  
112 Stanford study groups. Further demographic information across study designs and EHR data  
113 sets is listed in Table 1.

114 We compared the risk of being diagnosed with AD in patients with autoimmune disorders  
115 compared to non-autoimmune controls across case-control and cohort study designs in both the  
116 discovery (UCSF) and validation (Stanford) data sets. We evaluated the risk of AD in the study  
117 groups both overall and in a sex-stratified manner across multiple levels of autoimmune disorder  
118 stratifications to determine specific autoimmune drivers of AD risk.

119

120 ***Autoimmune disorder diagnoses are significantly associated with increased risk of AD***  
121 ***overall and within sex-specific groups***

122 In the case-control study groups, overall, individuals with autoimmune disorders had  
123 significantly higher odds of an AD diagnosis compared to non-autoimmune controls in both  
124 discovery (OR = 1.7, 95% CI 1.4-2.0,  $p = 2.5e-9$ , Fig 2A) and validation (OR = 1.4, 95% CI 1.2-  
125 1.6,  $p = 1.4e-7$ , Fig 2A) data sets. We observed even larger AD risk associations with  
126 autoimmune disorders in the cohort study groups, where odds ratios were 2.0 (95% CI 1.7-2.4,  
127  $p = 7.0e-15$ , Fig 2A) and 1.6 (95% CI 1.4-1.9,  $p = 1.6e-13$ , Fig 2A) for UCSF and Stanford,  
128 respectively. This consistent elevated risk across study designs and EHR systems highlights a  
129 strong connection between autoimmunity and AD.

130 We divided our discovery and validation data into female- and male-only subsets to  
131 determine if AD risk remained elevated within sex-specific groups and learn if the overall risk

132 association was primarily driven by women. In the female-only subsets of the case-control study  
133 designs, we observed significantly greater odds of an AD diagnosis in women with autoimmune  
134 disorders compared to control women at both UCSF (OR = 1.8, 95% CI 1.4-2.2,  $p = 4.1e-8$ , Fig  
135 2B) and Stanford (OR = 1.3, 95% CI 1.1-1.5,  $p = 3.0e-3$ , Fig 2B). Similarly, among women in the  
136 cohort study groups, we observed significantly greater AD risk in women with autoimmune  
137 disorders at UCSF (OR = 1.8, 95% CI 1.5-2.2,  $p = 4.9e-9$ , Fig 2B) and at Stanford (OR = 1.4,  
138 95% CI 1.2-1.7,  $p = 1.2e-5$ , Fig 2B). There were also strong associations between autoimmunity  
139 and AD in the male-only subsets. In the case-control study groups, men with autoimmune  
140 disorders had significantly greater AD risk compared to control men across UCSF and Stanford  
141 studies (UCSF OR = 1.5, 95% CI 1.1-2.0,  $p = 0.014$ ; Stanford OR = 1.8, 95% CI 1.4-2.3,  $p =$   
142  $8.2e-7$ ; Fig 2B). The results in the cohort study groups agreed (UCSF OR = 2.5, 95% CI 1.7-3.6,  
143  $p=1.3e-7$ ; Stanford OR = 2.3, 95% CI 1.8-3.0,  $p = 2.0e-11$ ; Fig 2B), highlighting higher risk in  
144 men with autoimmune disorders compared to control men across both EHR systems. The  
145 increased AD risk observed in both men and women with autoimmune disorders suggests that  
146 increased AD risk conferred by autoimmunity is not driven solely by one sex.

147  
148 ***AD prevalence is increased in the presence of autoimmune disorders, but women remain***  
149 ***the most affected***

150 While autoimmunity associates with AD risk within both sexes, we next tested whether it  
151 elevated risk more in one sex than the other. Furthermore, we wanted to determine if the  
152 presence of autoimmunity diminishes the well-documented AD sex-disparity wherein women  
153 tend to develop AD more often than men. To address these questions, we conducted an AD  
154 prevalence analysis within the cohort study groups of our discovery and validation data sets.  
155 Due to the smaller number of men compared to women in our data, we conducted 1:1 matching  
156 of women to men based on demographic variables (see Methods) and computed AD prevalence  
157 across sex and autoimmunity stratifications. In our discovery data set, women with autoimmune

158 disorders (N = 5,821) had the highest AD prevalence at 3.0%, followed by men with  
159 autoimmune disorders (N = 5,821) at 1.9%, control women (N = 5,821) at 1.7%, and finally  
160 control men (N = 5,821) at 0.79% (Fig 2C). While the absolute prevalence values were lower at  
161 Stanford, likely due to younger patients being included in the cohort study group as a result of  
162 censored age information (see Methods), there was a similar hierarchy in the validation data  
163 set, where women with autoimmune disorders (N = 45,743) had the highest AD prevalence at  
164 0.47%, followed by men with autoimmune disorders (N = 45,743) at 0.43%, control women (N =  
165 45,743) at 0.31%, and control men (45,743) at 0.19%. As expected, the higher prevalence in  
166 control women compared to control men (UCSF corrected  $p = 6.2e-5$ ; Stanford corrected  $p =$   
167  $1.1e-3$ ; Fig 2C) corroborates well-documented sex-disparities in AD. We then discovered that  
168 women with autoimmune disorders exhibited a higher AD prevalence than men with  
169 autoimmune disorders in the UCSF data set (corrected  $p = 9.9e-4$ ). The AD prevalence  
170 difference between autoimmunity patients of different sexes was roughly equal in magnitude to  
171 the difference between control patients of different sexes (1.1% versus 0.91%), suggesting that  
172 sex-disparities in AD remain present, even with autoimmunity conferring greater risk in both  
173 sexes. The intersex comparison in autoimmune patients in the Stanford data set was not  
174 significant ( $p = 0.4$ ), but women with autoimmune disorders did exhibit a slightly higher AD  
175 prevalence than the corresponding men, likely indicating that women continue to bear the most  
176 risk for AD even when a significant immune perturbation like autoimmunity is at play.

177

178 ***Specific autoimmune disorder subtypes are associated with greater AD risk, driven by***  
179 ***individual disorders***

180 Next, to determine if particular classes of autoimmune disorders are associated with AD risk  
181 more than others, we divided the 26 autoimmune disorders into subtypes based on the organ  
182 system each one primarily affects. This resulted in eight disease subtype categories:  
183 musculoskeletal, gastrointestinal, dermatologic, systemic, vascular, hematologic, neurologic,

184 and endocrine (Fig 3A). We computed the AD odds ratios for individuals with these disease  
185 subtypes in each of the case-control and cohort study designs across UCSF and Stanford data  
186 sets. In the overall case-control and cohort study designs of the UCSF data set, autoimmune  
187 disorders in the gastrointestinal (case-control OR = 6.1, 95% CI 2.7-14.6, corrected  $p=7.2e-6$ ;  
188 cohort OR = 2.5, 95% CI 1.4-4.7, corrected  $p=6.5e-3$ , Fig 3B), hematologic (case-control OR =  
189 18.1, 95% CI 4.7-84.8, corrected  $p=2.3e-6$ ; cohort OR = 4.5, 95% CI 1.8-13.4, corrected  
190  $p=3.6e-3$ , Fig 3B), endocrine (case-control OR = 2.8, 95% CI 1.8-4.4, corrected  $p=6.0e-6$ ;  
191 cohort OR = 1.9, 95% CI 1.4-2.6, corrected  $p=2.2e-4$ , Fig 3B), musculoskeletal (case-control  
192 OR = 2.5, 95% CI 1.7-3.9, corrected  $p=2.7e-5$ ; cohort OR = 2.0, 95% CI 1.5-2.8, corrected  
193  $p=4.7e-5$ ), and dermatologic (case-control OR = 3.1, 95% CI 1.6-6.1, corrected  $p=1.2e-3$ ; cohort  
194 OR = 2.4, 95% CI 1.4-4.0, corrected  $p=2.6e-3$ ) categories associated with increased risk for AD.  
195 Systemic, vascular, and neurologic disease subtypes did not significantly associate with  
196 increased AD, potentially because we were underpowered to detect risk associations for these  
197 particular subtypes (Table S2).

198 Several of the autoimmune disorder groups that were significantly associated with AD risk in  
199 the UCSF data set were validated by the Stanford data set in one or both study designs.  
200 Gastrointestinal (case-control OR = 2.5, 95% CI 1.5-4.2, corrected  $p=1.6e-3$ ; cohort OR = 2.0,  
201 95% CI 1.4-3.0, corrected  $p=1.1e-3$ , Fig 3B), endocrine (cohort OR = 1.9, 95% CI 1.3-2.6,  
202 corrected  $p=1.4e-3$ , Fig 3B), musculoskeletal (case-control OR = 1.7, 95% CI 1.3-2.3,  
203 corrected  $p=1.6e-3$ ; cohort OR = 1.5, 95% CI 1.2-1.8, corrected  $p=1.0e-3$ , Fig 3B), and  
204 dermatologic (case-control OR = 2.9, 95% CI 1.8-4.7, corrected  $p=1.7e-5$ ; cohort OR = 1.6, 95%  
205 CI 1.2-2.2, corrected  $p=2.6e-2$ , Fig 3B) disorders all conferred significantly more AD risk in  
206 autoimmune patients compared to non-autoimmune controls in the Stanford EHR system.

207 Next, we investigated if specific autoimmune disorders were driving the larger subtype and  
208 overall risk associations to determine if particular disorders were greater risk factors than others.  
209 Within the gastrointestinal category, inflammatory bowel disease was consistently associated



210 with AD risk across UCSF (case-control OR = 5.0, 95% CI 2.1-12.9, corrected  $p=4.7e-4$ ; cohort  
211 OR = 2.1, 95% CI 1.1-4.0, corrected  $p=0.050$ , Fig 3B, S3) and Stanford (case-control OR = 4.0,  
212 95% CI 2.1-7.9, corrected  $p = 3.7e-5$ ; cohort OR = 2.3, 95% CI 1.4-3.7, corrected  $p = 1.1e-3$ ,  
213 Fig 3B, S3) study groups. Other disorders that conferred increased AD risk in at least two study  
214 groups across data sets included autoimmune thyroiditis, type 1 diabetes, and rheumatoid  
215 arthritis. Autoimmune thyroiditis, in the endocrine category, exhibited an odds ratio of 3.0 (95%  
216 CI 1.4-6.4, corrected  $p = 9.0e-3$ , Fig 3B, S3) in the case-control study group and 2.3 (95% CI  
217 1.4-4.1, corrected  $p = 4.2e-3$ , Fig 3B, S3) in the cohort study group at UCSF, in addition to  
218 being a risk factor in the cohort study group of the Stanford data set (OR = 2.4, 95% CI 1.3-4.7,  
219 corrected  $p = 1.1e-2$ , Fig 3B, S3). Also in the endocrine category, type 1 diabetes was  
220 significantly associated with increased AD risk in the UCSF case-control study group (OR = 2.8,  
221 95% CI 1.6-4.9, corrected  $p=8.6e-4$ , Fig 3B, S3) while also associating with nominally significant  
222 risk in the UCSF cohort study group (OR = 1.6, 95% CI 1.1-2.4, *uncorrected*  $p=0.02$ , Fig 3B,  
223 S3). Furthermore, type 1 diabetes was a significant AD risk factor in the Stanford cohort study  
224 group (OR = 1.8, 95% CI 1.2-2.7, corrected  $p=2.4e-2$ , Fig 3B, S3).

225 Validated drivers of risk like autoimmune thyroiditis and type 1 diabetes highlight a potential  
226 link between immune-mediated endocrine dysfunction and later AD pathogenesis. In addition to  
227 endocrine disorders, rheumatoid arthritis was the primary disease that conferred risk in the  
228 musculoskeletal disease category, and it was a significant risk factor in both UCSF study groups  
229 (case-control OR = 2.2, 95% CI 1.3-3.7, corrected  $p=1.1e-2$ ; cohort OR = 2.1, 95% CI 1.4-3.2,  
230 corrected  $p=8.8e-4$ , Fig 3B, S3) as well as in the Stanford cohort study group (OR = 1.4, 95% CI  
231 1.1-1.9, corrected  $p = 3.9e-2$ , Fig 3B, S3), with nominal significance in the Stanford case-control  
232 study group (OR = 1.5, 95% CI 1.0-2.3, *uncorrected*  $p = 2.9e-2$ , Fig 3B, S3).

233

234 ***AD risk from disease subtypes exhibits variable sex-specific effects***

235 Next, we stratified the disease subtype and individual disease analyses by sex to determine if  
236 any particular risk association was driven more by one sex compared to the other. Some  
237 disease subtypes were fairly constant in exhibiting what appeared to be sex-specific risk  
238 signals. For example, the endocrine category of disorders was predominantly significant among  
239 women, exhibiting a significant AD risk effect in the female UCSF study groups (case-control  
240 OR = 2.7, 95% CI 1.6-4.7, corrected p = 8.4e-4; cohort OR = 1.8, 95% CI 1.3-2.7, corrected p =  
241 8.7e-3, Fig S2) and the female Stanford cohort study group (cohort OR = 1.8, 95% CI 1.2-2.8,  
242 corrected p = 0.04). In the dermatologic category, male-specific AD risk was higher than that of  
243 the female-specific comparison in the majority of study groups. In the UCSF cohort study group,  
244 men with dermatologic autoimmune disorders exhibited an odds ratio of 1.5 (95% CI 1.4-9.3,  
245 corrected p=0.04, Fig S2), while women exhibited an odds ratio of 2.0 that was insignificant  
246 after multiple testing correction (95% CI 1.1-3.8, corrected p=0.20, Fig S2). Similarly, in both  
247 Stanford study groups, men with dermatologic autoimmune disorders were at greater AD risk  
248 than controls (case-control OR = 7.8, 95% CI 3.4-18.5, corrected p=4.3e-7; cohort OR = 2.6,  
249 95% CI 1.5-4.7, corrected p=1.9e-3, Fig S2) compared to women (case-control OR = 1.6, 95%  
250 CI 0.9-3.0, corrected p = 0.94; cohort OR = 1.2, 95% CI 0.8-1.8, corrected p = 3.5, Fig S2).  
251 Interestingly, in the case-control study group at UCSF, we also observed increased female-  
252 specific AD risk conferred by dermatologic autoimmune disorders (OR = 6.4, 95% CI 2.4-18.3,  
253 corrected p=4.4e-4, Fig S2), perhaps indicating variable sex-specific effects for this category of  
254 diseases. Similarly, we observed strong female-specific AD risk conferred by gastrointestinal  
255 conditions in our discovery data set (case-control OR = 11.8, 95% CI 4.2-37.1, corrected  
256 p=1.0e-6; cohort OR = 2.9, 95% CI 1.5-6.0, corrected p=6.5e-3, Fig S2), only to detect strong  
257 male-specific AD risk from these conditions in our validation data set (case-control OR = 3.4,  
258 95% CI 1.4-8.5, corrected p=4.0e-2; cohort OR = 3.3, 95% CI 1.6-7.5, corrected p=4.8e-3, Fig  
259 S2). These findings warrant more investigation into the factors influencing the direction of sex-  
260 specificity for these disorder classes.

261 Clear sex-specific differences in the risk conferred by individual autoimmune disorders were  
262 harder to elucidate given a lack of power after stratifying by both sex and disorder (Table S2).  
263 We nonetheless were able to identify significant female-specific AD risk conferred by  
264 autoimmune thyroiditis in the cohort study groups of our two data sets (UCSF cohort OR = 2.4,  
265 95% CI 1.3-4.5, corrected  $p=1.5e-2$ ; Stanford cohort OR = 2.5, 95% CI 1.2-5.4, corrected  
266  $p=3.1e-2$ , Fig S4), likely driving the female-specific risk in the endocrine category of diseases.  
267 Several other diseases increased AD risk significantly across some study groups, but not across  
268 others (Fig S4) or exhibited different directions of sex-specific effects across study groups (Fig  
269 S4). This again highlights the need for more studies of sex differences in the interaction  
270 between the autoimmunity and AD in the future.

271

## 272 ***Sensitivity Analyses***

273 We performed several sensitivity analyses to address the potential presence of confounders in  
274 EHR data and test the robustness of the risk associations we identified. The sensitivity analyses  
275 were performed in the discovery data set alone, as we wanted to verify the risk associations we  
276 saw in the smaller UCSF study groups before performing further validations in the Stanford  
277 study groups. Our first sensitivity analysis was to implement an age cutoff in our study designs  
278 to quantify AD risk due to autoimmunity in two older sub-populations of individuals. The first  
279 cutoff only included individuals in the study groups that had reached an age of 65 years or older.  
280 When comparing autoimmune patients to non-autoimmune controls in this >65 data set, we  
281 observed an AD odds ratio of 1.6 (95% CI 1.4-2.0,  $p=9.7e-9$ ) in the case-control study group  
282 and 2.0 (95% CI 1.6-2.4,  $p=1.7e-14$ ) in the cohort study group (Table S3). Raising the cutoff to  
283 80 years of age or older also resulted in a strong AD risk association with autoimmune disorders  
284 (case-control OR = 1.7, 95% CI 1.4-2.2,  $p=1.7e-6$ ; cohort OR = 2.4, 95% CI 1.8-3.1,  $p=1.8e-12$ ,  
285 Table S3), indicating that our signal was robust to the age range of patients in our discovery  
286 data set.

287 Next we performed a sensitivity analysis to address possible confounder and collider  
288 effects that healthcare utilization can cause in EHR systems, since these effects can exaggerate  
289 or attenuate differences between exposure and outcome groups. In addition to matching on  
290 demographic variables (see Methods), we matched individuals based on the similarity of the  
291 length of time between their first and last hospital visit date. After conducting this matching and  
292 recomputing odds ratios in each study group, autoimmune disorders still associated with  
293 increased risk for AD (case-control OR = 1.3, 95% CI 1.1-1.5,  $p = 6.0e-3$ ; cohort OR = 1.4, 95%  
294 CI 1.2-1.7,  $p = 5.2e-6$ , Table S3). Matching on each patient's frequency of visits per year also  
295 resulted in consistent increased risk associations between autoimmune disorders and AD (case-  
296 control OR = 1.3, 95% CI 1.1-1.6,  $p = 3.7e-4$ ; cohort OR = 1.5, 95% CI 1.3-1.7,  $p = 2.0e-6$ ,  
297 Table S3), highlighting the connection between autoimmunity and AD at the clinical level even  
298 when adjusting for different healthcare utilization measures.

299 In the final sensitivity analysis, we recomputed odds ratios in each study group without  
300 including age at death as a variable in matching, to verify that leaving it out of the matching  
301 criteria did not significantly alter results (see Matching and Finalization of Study Groups in  
302 Methods). This resulted in an AD odds ratio of 1.3 (95% CI 1.1-1.5,  $p = 7.1e-4$ ) in the case-  
303 control study group and 2.1 (95% CI 1.8-2.5,  $p = 4.7e-17$ ) in the cohort study group (Table S3),  
304 indicating that leaving total lifespan information out of the matching criteria preserves risk signal.  
305 Several disease subtypes and individual disease associations seen in the main UCSF analysis  
306 were present in many, if not all, of these sensitivity conditions, including the increased AD risk  
307 due to endocrine, gastrointestinal, and musculoskeletal disorders driven by inflammatory bowel  
308 disease, type 1 diabetes, autoimmune thyroiditis, and rheumatoid arthritis (Table S3).

309

### 310 ***Sex, but not autoimmunity, associates with accelerated AD onset over time***

311 In addition to examining the presence or absence of AD in autoimmunity patients through our  
312 risk analyses, we also tested whether autoimmune disorders influence the timing of AD onset.

313 We hypothesized that autoimmune disorders would decrease the age at which people are  
314 diagnosed with AD, potentially due to the early presence of chronic inflammation that  
315 autoimmune disorders bring about. We constructed new longitudinal cohorts for this analysis  
316 consisting only of AD patients with and without autoimmune disorders. This resulted in 292  
317 autoimmune patients with 292 matched controls at UCSF (N = 584 total AD patients, Fig 4A),  
318 and 392 autoimmune patients with 392 matched controls at Stanford (N = 784 total AD patients,  
319 Fig 4A). We first compared the distributions of AD diagnosis age among patients. In the UCSF  
320 longitudinal cohort, the average age at which autoimmune patients were diagnosed with AD was  
321 75.6 years, compared to the controls at 76.5 years (Fig S5A). In the Stanford longitudinal  
322 cohort, the average AD diagnosis age was 81.8 years in autoimmune patients compared to 82.4  
323 years in controls (Fig S5A). While the AD diagnosis age was lower in autoimmune patients in  
324 each data set, the differences between age distributions were not significant in either case  
325 (UCSF  $p = 0.11$ , Stanford  $p = 0.17$ , Fig S5A), likely because of relatively small sample size  
326 (Table S2). The 0.6- to 1-year difference in diagnosis age we observed is nonetheless striking,  
327 given the often rapid symptomatic decline (21) of individuals with AD. Even being diagnosed  
328 with AD half a year earlier could be extremely impactful for patients and their quality of life.

329 We also observed differences in AD onset age when we stratified individuals both by sex and  
330 autoimmune disorder. Starting in the UCSF longitudinal cohort, women with autoimmune  
331 disorders exhibited significantly younger AD onset ages (mean 75.0 years, Fig 4B) compared to  
332 men with autoimmune disorders (mean 76.7 years,  $p = 0.017$ , Fig 4B). Similarly, control women  
333 exhibited significantly younger AD onset ages (mean 76.0 years, Fig 4B) compared to control  
334 men (mean 77.7 years,  $p = 0.011$ , Fig 4B). Comparing autoimmune patients to matched  
335 controls within each sex did not reveal younger onset ages (female  $p = 0.19$ , male  $p = 0.28$ , Fig  
336 4B), likely indicating that sex plays a larger role in the timing of AD onset. Results from the  
337 Stanford longitudinal cohort all agreed directionally with the UCSF results, however with  
338 insignificant results (Fig 4B).

339 Finally, we performed a survival analysis to determine if the risk of developing AD in different  
340 stratifications varied over time. We used a Cox proportional hazard model for the analysis with  
341 sex and autoimmunity presence as covariates. Having an autoimmune disorder did not  
342 significantly alter the hazard of individuals developing AD over time (UCSF Hazard Ratio (HR) =  
343 1.1, 95% CI 1.0-1.3,  $p = 0.16$ ; Stanford HR = 1.0, 95% CI 0.9-1.2,  $p = 0.5$ , Fig S5B),  
344 corroborating the results from the distributional analysis. The hazard effects of sex were unclear  
345 given disagreement between UCSF and Stanford data sets. Being male in the UCSF  
346 longitudinal cohort trended toward being protective (HR = 0.85, 95% CI 0.7-1.0,  $p = 0.064$ , Fig  
347 S5B), while in the Stanford longitudinal cohort it was significantly associated with risk of  
348 developing AD earlier (HR = 1.2, 95% CI 1.0-1.4,  $p = 0.02$ , Fig S5B). Better powered analyses  
349 into sex as an influential variable in the timing of AD onset are needed to resolve this mismatch.

350

## 351 Discussion

352 We have shown that autoimmune disorders are associated with increased risk of being  
353 diagnosed with AD in case-control and cohort study designs across two different EHR systems,  
354 suggesting that autoimmunity is a significant risk factor for AD. We observed a clear and strong  
355 increased risk signal in our study groups, both overall and in female- and male-specific subsets  
356 of our data, indicating that autoimmunity increases AD risk regardless of sex. Interestingly, in  
357 our AD prevalence analysis, women with autoimmune disorders exhibited increased AD  
358 prevalence compared to men with autoimmune disorders, indicating that both autoimmunity and  
359 sex play crucial roles in AD risk.

360 We observed specific subtypes of autoimmune disorders that conferred AD risk in patients.  
361 These particular classes of disorders suggest the presence of shared pathophysiology between  
362 particular autoimmunity subtypes and AD. For example, metabolic dysfunction has been  
363 highlighted in previous AD pathogenesis studies (22), so it is possible that metabolic pathways

364 exist that link endocrine autoimmune disorders, which greatly affect metabolism, and AD  
365 pathogenesis. Similarly, there have been documented links between changes in the gut-brain  
366 axis (23) and microbiome (24) that are associated with AD in previous work, which could relate  
367 AD pathogenesis to pathways involved in gastrointestinal autoimmune disorders. Depending on  
368 the autoimmune disorder subtype and individual autoimmune disorders we analyzed, we also  
369 observed some sex-specific risk associations, highlighting that autoimmune disorders that fall  
370 under the category of certain physiological systems may manifest differently in women  
371 compared to men. This can provide more insight into the specific mechanisms that may go awry  
372 to cause or exacerbate AD in particular sexes.

373 Our findings validate and provide new insights on previous work. Recently, Miller et al. (25)  
374 found increased prevalence of inflammatory bowel disease in a small cohort of late-onset AD  
375 cases which we have validated in much larger clinical data sets. Furthermore, a study of the  
376 Swedish National Patient Register (26) highlighted increased incidence rates of several  
377 autoimmune disorders including hypothyroidism/thyroiditis, type 1 diabetes, Addison's disease,  
378 Sjögren's syndrome, and pernicious anemia in dementia patients. We were able to corroborate  
379 several of these signals with our analysis; we also found type 1 diabetes and a form of  
380 thyroiditis as significant modifiers of risk. We further built upon these findings by analyzing  
381 disease subtypes to identify particular physiological systems that may be more involved in AD  
382 pathogenesis, and we conducted extensive age of onset and sex-specific analyses to robustly  
383 characterize AD risk in autoimmune individuals. Finally, a recent study (27) in individuals of the  
384 UK Biobank indicated increased hazard ratios for four major autoimmune disorders (rheumatoid  
385 arthritis, multiple sclerosis, psoriasis, and inflammatory bowel disease) in a longitudinal cohort.  
386 Our results expand on this study by investigating several more autoimmune disorders and sex  
387 as a biological variable in the risk analyses. Compared to the individuals in the UK Biobank and  
388 Swedish studies, our population of patients was also more diverse, suggesting generalizability

389 of the increased risk association between autoimmune disorders and AD across diverse  
390 populations of humans.

391 By using two different study designs (case-control and cohort) in both a discovery and  
392 validation data set in all of our analyses, we were further able to show robust associations while  
393 combating common problems of selection bias (28), data inaccuracies (29), and confounding  
394 that can be common in EHR studies. Our workflow can be further applied not only to  
395 autoimmune disorders and their risk relationship to other neurological diseases, but to any two  
396 disease types of interest that can be captured in an EHR data system.

397 The current study has several limitations that should be considered when evaluating our  
398 results. First, the groupings used for the disease subtype analysis are imperfect. Since the exact  
399 underlying mechanism of many of the autoimmune disorders we investigated remains unknown,  
400 grouping by physiological system enabled us to study these diseases in aggregate, but it  
401 reflects only one dimension along which each autoimmune disorder is related to the others.  
402 Another caveat is the presence of censored death information in the validation data set. This  
403 may have resulted in the incorporation of younger patients in the Stanford study groups who  
404 had yet to develop either AD or a later-onset autoimmune disorder, and as such, this may have  
405 deflated the odds ratios in the Stanford study groups. It is possible that this led to less  
406 agreement in risk associations between UCSF and Stanford particularly for the subtype and  
407 individual disorder analyses. While we matched individuals based on lifespan and/or birth year  
408 within each of our study groups, future work could include age-matching between discovery and  
409 validation data sets to enhance similarity of comparisons even further. Finally, the stratified  
410 analyses we conducted were less likely to yield consistent significant results across sites and  
411 study designs. Much of this was likely the result of reduced power due to the small sample sizes  
412 in the stratified cohorts. This also occasionally yielded very high odds ratios and large  
413 confidence intervals in our analyses. Nonetheless, many of these analyses highlight promising  
414 specific hypotheses for further validation and molecular characterization.



415 In summary, autoimmune disorders are strong risk factors for AD that act across sexes. This  
416 study illustrates the usefulness of EHRs for cross-trait analyses, and it also informs further  
417 mechanistic hypotheses about the exact molecular interfaces that may go awry in the interaction  
418 of the immune and nervous systems to promote AD pathogenesis. Further risk factor analyses  
419 for debilitating neurological conditions such as AD will empower clinicians to inform patients of  
420 their risk profiles for different diseases. Ultimately, deeper understanding of these connections  
421 between risk and disease can empower patients themselves to make lifestyle changes or take  
422 relevant treatments that can help avoid or delay disease. Our results highlight several future  
423 directions for further understanding of the risk relationship between autoimmune disorders and  
424 AD. First, quantifying how AD risk varies based on differing levels of autoimmune disorder  
425 severity and duration is needed. We hypothesize that more severe forms of autoimmune  
426 disorders may confer the most AD risk, and perhaps treatments to alleviate more severe  
427 disorders may decrease risk. Additionally, taking into account the chronic nature and onset age  
428 of many autoimmune disorders may shed more light on the temporal dynamics of the two traits  
429 interacting throughout the lifespan. Finally, integrating clinical risk analyses with other data  
430 modalities, such as genetics or proteomics will provide more molecular insight into the link  
431 between AD and autoimmune disorders and help to fully elucidate the basis for the strong risk  
432 seen at the phenotypic level in human populations.

433

## 434 **Methods**

### 435 **Sex as a Biological Variable**

436 We accounted for sex as a biological variable by performing all analyses in female- and male-  
437 specific subsets of our original study groups. This allowed us to test if any risk associations  
438 differed by sex, and if particular autoimmunity subtypes or individual disorders were driving AD  
439 risk in a particular sex. We filtered out individuals with an unknown sex from our UCSF study

440 groups. Due to differences in encoding sex within the UCSF and Stanford EHR systems, there  
441 were a small number of individuals (0.02% of the total) with an unknown sex that were included  
442 in the Stanford cohort data set. These individuals were not included in any sex-specific  
443 analyses.

444

## 445 **Study Group Identification**

### 446 *Autoimmune Disorder, Alzheimer's Disease, and Healthy Control Patient Identification*

447 We identified 26 autoimmune disorders of interest (Table S1) for our study based on prior  
448 literature and prevalence in the general population. Individuals with each autoimmune disorder  
449 were identified by string-matching autoimmune disorder names (Table S4) with billing concepts.  
450 Concepts were subsequently standardized for use in the UCSF EHR database (30) which is  
451 based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model  
452 (CDM) and primarily uses Systemized Nomenclature of Medicine (SNOMED) concept  
453 encodings. All standardized billing concepts were examined by UCSF rheumatologists to  
454 confirm validity and relevance of each concept to each autoimmune disorder of interest. We  
455 compiled a final list of 878 autoimmune billing concepts (Table S5). We identified patients who  
456 had these concepts present in their UCSF medical record to construct our discovery data set,  
457 and used identical concepts to identify patients from the Stanford EHR database for our  
458 validation data set. We identified patients with AD in a similar manner to autoimmune disorder  
459 patients after string-matching AD terms to concepts (Table S6) and checking for billing concept  
460 occurrence in each patient record in the UCSF and Stanford systems. We only included AD  
461 billing concepts related to late-onset or sporadic AD, as early-onset AD is thought to have  
462 distinct etiology and stronger genetic components (31). Additional demographic data from the  
463 UCSF and Stanford EHR systems was collected on patients including date of birth, date of  
464 death (if available), self-reported race, self-reported ethnicity, and sex. We also gathered

465 healthcare utilization statistics on patients, including number of doctor’s visits, total number of  
466 unique diagnoses, and first and last medical visit date.

467 We identified healthy control individuals without autoimmune disorders to compare to the  
468 autoimmune patient groups from the UCSF and Stanford databases. To do this, we searched  
469 for patients *without* any of the 878 finalized autoimmune disorder billing concepts present in  
470 their records, and we additionally removed individuals from the healthy control group who had  
471 concepts that were similar to any of the 878 finalized autoimmune concepts. For example, an  
472 individual who had the billing concept “family history of Celiac disease” in their EHR but did not  
473 have a more specific Celiac disease billing concept identifying a personal diagnosis of Celiac  
474 disease would have been excluded. Similarly, several billing concepts that were too general to  
475 pertain specifically to an autoimmune disorder (e.g. kidney disease) but that represented a  
476 serious condition were excluded from the healthy controls. For our cohort study design, we  
477 matched the healthy controls to autoimmune disorder patients based on criteria further  
478 described in the *Matching and Finalization of Study Groups* section of the Methods.

479 To identify a population of non-AD healthy controls to compare to our AD patient group, we  
480 searched for patients *without* any AD billing concepts in their records and matched them with  
481 AD cases based on demographic factors. To clarify some terminology, the control individuals  
482 that were matched to autoimmune disorder cases will be referred to as the “non-autoimmune  
483 controls” going forward, whereas the control individuals that were matched to AD cases will be  
484 referred to as the “non-AD controls”. These are two separate groups of controls, but they may  
485 have overlap in individuals as someone without both an autoimmune disorder and AD might be  
486 in both control groups. Additionally, someone in one of the disease groups may be in a control  
487 group for the other disease. For example, an AD patient might show up in the non-autoimmune  
488 control group, since the requirement to be in that group is not having an autoimmune disorder.

489 *Data Cleaning and Quality Control*

490 Our data cleaning pipeline involved several steps. First, we performed quality control to remove  
491 any individuals with missing demographic information in the self-reported race, self-reported  
492 ethnicity, sex, and birth year categories from consideration for our disease and healthy control  
493 groups in the UCSF data set. Due to differences in encoding some of this demographic  
494 information between UCSF and Stanford EHR systems, a small number of individuals with  
495 unknown demographic values were included in the Stanford data set (see Sex as a Biological  
496 Variable), but each individual with an “unknown” demographic field was similarly matched with  
497 another Stanford individual with an “unknown” value, removing any issues comparing people  
498 without matching information. In the UCSF study groups, we also required individuals to have a  
499 valid reported age at death, as this allowed us to compare individuals by total lifespan. We  
500 restricted individuals to be between 30 and 90 years of age at their death in these groups. In the  
501 Stanford study groups, there was a greater degree of censored death information such that  
502 including lifespan information in matching would have extremely limited our study group sizes,  
503 so we did not enforce this constraint. We verified that leaving out lifespan as a matching  
504 variable in the Stanford groups did not significantly alter overall risk signals, so we felt confident  
505 leaving it out of our validation study group criteria (See Sensitivity Analyses in Results). It is  
506 likely that, because of censored data being included in the Stanford data sets, the increased risk  
507 associations we saw in our analyses would be even stronger in real life (see Discussion). In  
508 addition to quality control on demographics, we also removed individuals who had zero hospital  
509 visits or whose first and last visit dates were the same.

510 We next determined 1) who in the autoimmune disorder patient groups and corresponding  
511 non-autoimmune control groups had an AD diagnosis, and 2) who in the AD patient groups and  
512 corresponding non-AD control groups had an autoimmune disorder diagnosis. Within our  
513 autoimmune/AD disease groups and respective healthy control groups, different individuals  
514 were demarcated with their assigned “person ID” following OMOP conventions. We then  
515 determined which person IDs of one group overlapped with the person IDs of another group.

516 For example, to discover which autoimmune patients and non-autoimmune controls had an AD  
517 diagnosis, we determined which of the person IDs of our AD patient group overlapped with the  
518 person IDs of the autoimmune and non-autoimmune groups. In a similar manner, we  
519 determined which AD patients and which non-AD controls had an autoimmune diagnosis by  
520 overlapping the person IDs of our autoimmune patient group with the AD and non-AD person  
521 IDs.

522 We also identified the relative dates of AD and autoimmunity diagnoses for patients. To  
523 focus on the effect of autoimmunity on AD, we did not consider individuals who had an AD  
524 diagnosis prior to their autoimmune disorder diagnosis. Specifically in the UCSF data set, we  
525 computed several metrics for each individual to aid the matching of autoimmune or AD patients  
526 to their respective controls when performing different sensitivity analyses. These metrics  
527 included each individual's age when they died, the total length of the UCSF EHR record (last  
528 visit date - first visit date), and the frequency of doctor's visits per year

529 To understand which autoimmune disorder subtypes might be driving AD risk, we grouped  
530 the 26 autoimmune disorders of interest into 8 distinct subtype categories (Table S1 and Fig 3A)  
531 and assigned patients into categories based on which autoimmune disorder(s) they had. If a  
532 patient had multiple autoimmune disorders (UCSF N = 2,378 patients, Stanford N = 37,332  
533 patients) across different subtype categories, they were counted once within each subtype risk  
534 association analysis, and therefore could be represented in more than one analysis.

#### 535 Study Designs, Matching, and Finalization of Study Groups

536 To mitigate selection bias and ensure robustness of results, we examined the risk of receiving  
537 an AD diagnosis following an autoimmune disorder diagnosis using two study designs in each of  
538 our EHR data sets: a retrospective case-control study with AD patients and non-AD matched  
539 controls, and a retrospective cohort study with autoimmune disorder patients and non-  
540 autoimmune matched controls. We performed 1:1 matching of patients to controls for each  
541 study group using propensity scoring on each individual's birth year, sex, self-reported race, and

542 self-reported ethnicity. We also matched on lifespan in the UCSF study groups (See Data  
543 Cleaning). We enforced exact matches between patients and controls in the categories of sex,  
544 self-reported race, and self-reported ethnicity. In our main analysis and throughout follow-up  
545 sensitivity analyses, we ensured high-quality matching by verifying that the average absolute  
546 standardized mean error between each matched pair was less than 0.1. We conducted final  
547 quality control in both study groups of our main and sensitivity analyses by removing any  
548 matched pairs of individuals where a control individual was diagnosed with AD prior to the  
549 matched disease case being diagnosed with an autoimmune disorder. All matching was  
550 performed using the nearest neighbor method of the MatchIt package (4.5.3) in R. After  
551 matching and cleaning, we were left with four “study groups”: a case-control and cohort study  
552 group from UCSF, and a case-control and cohort study group from Stanford.

553

#### 554 **AD Risk Analysis in Case-Control and Cohort Study Groups**

555 In order to enable comparison across the different study designs, we computed odds ratios to  
556 quantify the risk of being diagnosed with AD in autoimmune disorder patients compared to non-  
557 autoimmune controls. Across each of the study groups, we computed odds ratios at three  
558 levels: 1) across all autoimmune disorders combined, 2) across autoimmunity subtypes, and 3)  
559 across individual autoimmune disorders. At each of these three levels, we repeated the analysis  
560 in a sex-stratified manner to explore if risk was sex-specific. All odds ratios were computed  
561 using a Fisher exact test on contingency tables of AD/autoimmunity patients, and we used the  
562 `oddsratio.fisher` function of the `epitools` package (version 0.5-10.1) in R to compute these  
563 statistics.

564 To account for multiple hypothesis testing, we used Bonferroni corrections for each odds  
565 ratio analysis. For the disease subtype risk analysis, we corrected p-values within each overall  
566 or sex stratification group (e.g., a correction factor of 8 for the 8 disease subtype comparisons).  
567 For the specific disorder risk analysis, we performed a within-disorder-subtype p-value

568 correction, to evaluate which specific conditions within an autoimmune disorder subtype group  
569 had significant subtype effects. This meant the correction factor for a particular disorder  
570 comparison was determined by the number of conditions within that particular disorder's  
571 subtype category (e.g., p-values for each autoimmune disorder in the endocrine subgroup were  
572 corrected by a factor of 4, due to the endocrine subgroup being comprised of 4 diseases).

573

#### 574 **AD Prevalence Calculations in Cohort Study Groups**

575 We calculated the prevalence of AD in patients with autoimmune disorders compared to  
576 matched controls in a sex-stratified manner. To ensure that a difference in prevalence was not  
577 due to underlying demographic differences between female and male individuals, we took the  
578 cohort study groups across UCSF and Stanford and matched the smaller sample size of male  
579 individuals to female individuals based on birth year, self-reported race, and self-reported  
580 ethnicity, while matching additionally on lifespan in the UCSF study groups. Again, we used  
581 exact matching on self-reported race and self-reported ethnicity while using propensity score  
582 matching on the remaining variables. We then calculated the percentage of people with AD in  
583 each stratification: female individuals with autoimmune disorders, male individuals with  
584 autoimmune disorders, female non-autoimmune control individuals, and male non-autoimmune  
585 control individuals. To obtain a 95% confidence interval for each prevalence statistic, we  
586 bootstrapped the data 1,000 times. We again used a Fisher's exact test to compute significance  
587 of prevalence differences among stratifications.

588

#### 589 **Longitudinal AD Onset Analysis**

590 To understand the effect of autoimmunity on the risk of AD diagnosis over time, we conducted a  
591 longitudinal age of onset analysis. For this, we constructed new longitudinal study groups from  
592 the UCSF and Stanford EHR systems consisting of individuals with both an autoimmune  
593 disorder and AD from our cohort study designs. These longitudinal cohorts also passed our data

594 quality control pipeline. We matched the autoimmunity patients with AD to non-autoimmune  
595 individuals with AD from the larger UCSF and Stanford background databases. The same  
596 variables used in matching for the main odds ratio analysis were also used here. The age of AD  
597 onset for each individual in each longitudinal study group was then calculated by taking the  
598 difference in time between a person’s birth year and the first appearance of an AD billing  
599 concept in the person’s medical record. The Mann–Whitney U test and Cox proportional hazard  
600 modeling conducted in these study groups was performed using the stats (v4.1.3) and  
601 survival/survminer (v3.5-5/0.4.9) packages, respectively, in R.

602

### 603 **Statistics**

604 We used Fisher’s exact test for all odds ratio calculations and for comparisons in our AD  
605 prevalence analysis. We also used a Mann-Whitney U test for a non-parametric comparison of  
606 distributions in our longitudinal AD onset analysis. Bonferroni corrections were applied to p  
607 values throughout this study; these are denoted by the use of “corrected” or “adjusted” p values.

608

### 609 **Study Approval**

610 All analysis of University of California, San Francisco and Stanford University electronic health  
611 record data was performed under the approval of the Institutional Review Boards from  
612 University of California, San Francisco and Stanford University, respectively. All clinical data  
613 were de-identified and written informed consent was waived by the institutions.

614

### 615 **Data Availability**

616 Individual patient data is not publicly available due to patient data sharing privacy. Code not  
617 limited by patient data sharing permissions can be found at  
618 [https://github.com/gramey02/AD\\_AID\\_Project](https://github.com/gramey02/AD_AID_Project). All patient and demographic data curation from  
619 the UCSF and Stanford EHR systems was performed using Microsoft SQL server and the DBI



620 (v1.1.3) and odbc (1.3.4) packages in R. Discovery data was last curated from the UCSF OMOP  
621 database on August 4th, 2023, and validation data was last curated from the Stanford OMOP  
622 database on December 12th, 2023. Data cleaning, matching, and analysis steps were  
623 conducted using R version 4.1.3, and plots were created with the ggplot2 package (v3.4.2).  
624 Values for all data points in graphs are reported in the Supporting Data Values file.

625

## 626 **Author Contributions**

627 A.T., Z.M., M.S, and G.D.R. conceptualized this research project. G.D.R. investigated and  
628 curated all UCSF data and performed all UCSF statistical and computational analysis. T.P.  
629 performed all Stanford validation steps, including data curation and statistical/computational  
630 analysis. M.Y. provided valuable clinical input on EHR methodology and clinical billing codes.  
631 G.D.R. and A.T. wrote the manuscript, and G.D.R. created visualizations. J.A.C. and M.S.  
632 provided mentorship, review/editing of writing, and funding for the work. I.A., N.A., T.M., T.O.,  
633 and S.W. provided expertise and manuscript feedback.

634

## 635 **Acknowledgements**

636 We would like to thank all of the members of J.A.C.'s, M.S.'s, and N.A.'s labs for all of their  
637 feedback throughout this work. We acknowledge our funding sources, including NIA  
638 R01AG060393, NIAMS P30 AR070155, F30 Fellowship 1F30AG079504-01, and the UCSF  
639 Discovery Fellows. We would also like to acknowledge the use of the [UCSF Information](#)  
640 [Commons](#) and [UCSF Research Analysis Environment](#) computational research platforms.  
641 Through these platforms, the project was supported by the National Center for Advancing  
642 Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1  
643 TR001872. Its contents are solely the responsibility of the authors and do not necessarily  
644 represent the official views of the NIH.

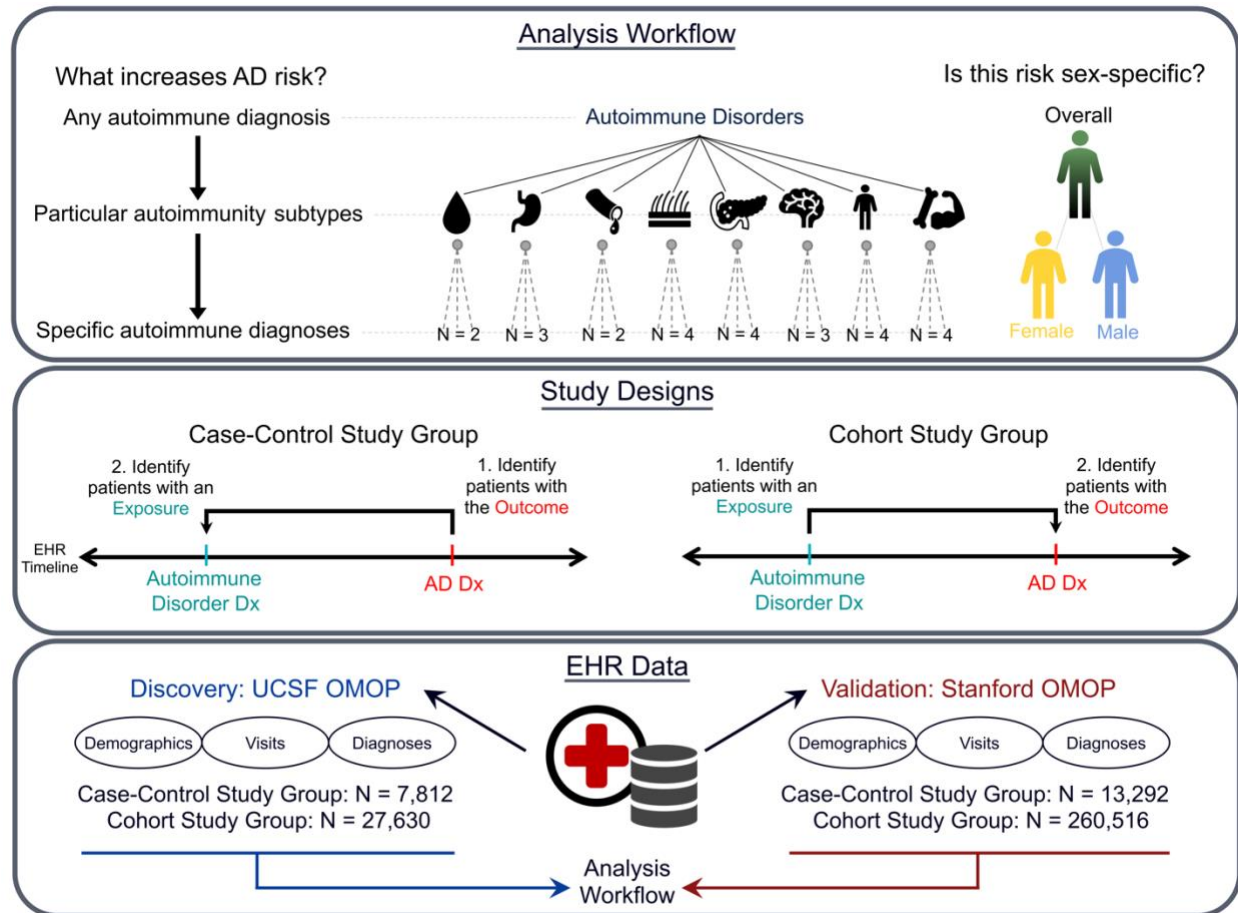
## 645 References

- 646 1. 2024 Alzheimer's disease facts and figures. *Alzheimers Dement J Alzheimers Assoc* .  
647 2024;20(5). <https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf>
- 648 2. 2023 Alzheimer's disease facts and figures. *Alzheimers Dement J Alzheimers Assoc*.  
649 2023;19(4):1598-1695. doi:10.1002/alz.13016
- 650 3. Jack CR, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's  
651 disease. *Alzheimers Dement J Alzheimers Assoc*. 2018;14(4):535-562.  
652 doi:10.1016/j.jalz.2018.02.018
- 653 4. Revised Criteria for Diagnosis and Staging of Alzheimer's | AAIC. Revised Criteria for  
654 Diagnosis and Staging of Alzheimer's | AAIC. Accessed March 24, 2024.  
655 <https://aaic.alz.org/diagnostic-criteria.asp>
- 656 5. Van Dyck CH, et al. Lecanemab in Early Alzheimer's Disease. *N Engl J Med*.  
657 2023;388(1):9-21. doi:10.1056/NEJMoa2212948
- 658 6. Huang LK, et al. Clinical trials of new drugs for Alzheimer disease: a 2020–2023 update. *J*  
659 *Biomed Sci*. 2023;30(1):83. doi:10.1186/s12929-023-00976-6
- 660 7. Newcombe EA, et al. Inflammation: the link between comorbidities, genetics, and  
661 Alzheimer's disease. *J Neuroinflammation*. 2018;15:276. doi:10.1186/s12974-018-1313-3
- 662 8. Princiotta Cariddi L, et al. Alzheimer's Disease: From Immune Homeostasis to  
663 Neuroinflammatory Condition. *Int J Mol Sci*. 2022;23(21):13008.  
664 doi:10.3390/ijms232113008
- 665 9. Laurent C, et al. Hippocampal T cell infiltration promotes neuroinflammation and cognitive  
666 decline in a mouse model of tauopathy. *Brain*. 2017;140(1):184-200.  
667 doi:10.1093/brain/aww270
- 668 10. Chen X, Holtzman DM. Emerging roles of innate and adaptive immunity in Alzheimer's  
669 disease. *Immunity*. 2022;55(12):2236-2254. doi:10.1016/j.immuni.2022.10.016
- 670 11. Ahmed ME, et al. Co-Localization of Glia Maturation Factor with NLRP3 Inflammasome and  
671 Autophagosome Markers in Human Alzheimer's Disease Brain. *J Alzheimers Dis*.  
672 2017;60(3):1143-1160. doi:10.3233/JAD-170634
- 673 12. Rogers J, et al. Complement activation by beta-amyloid in Alzheimer disease. *Proc Natl*  
674 *Acad Sci U S A*. 1992;89(21):10016-10020.
- 675 13. Walker DG, McGeer PL. Complement gene expression in human brain: comparison  
676 between normal and Alzheimer disease cases. *Mol Brain Res*. 1992;14(1):109-116.  
677 doi:10.1016/0169-328X(92)90017-6
- 678 14. Berg A, et al. Reduced removal of synaptic terminals from axotomized spinal motoneurons  
679 in the absence of complement C3. *Exp Neurol*. 2012;237(1):8-17.  
680 doi:10.1016/j.expneurol.2012.06.008

- 681 15. Dou DR, et al. Xist ribonucleoproteins promote female sex-biased autoimmunity. *Cell*.  
682 2024;187(3):733-749.e16. doi:10.1016/j.cell.2023.12.037
- 683 16. Ortona E, et al. Sex-based differences in autoimmune diseases. *Ann Ist Super Sanita*.  
684 2016;52(2):205-212. doi:10.4415/ANN\_16\_02\_12
- 685 17. Belonwu SA, et al. Sex-Stratified Single-Cell RNA-Seq Analysis Identifies Sex-Specific and  
686 Cell Type-Specific Transcriptional Responses in Alzheimer's Disease Across Two Brain  
687 Regions. *Mol Neurobiol*. 2022;59(1):276-293. doi:10.1007/s12035-021-02591-8
- 688 18. Lopez-Lee C, et al. Sex Differences in Neurodegeneration: The Role of the Immune System  
689 in Humans. *Biol Psychiatry*. 2022;91(1):72-80. doi:10.1016/j.biopsych.2021.01.002
- 690 19. Natri H, et al. The Pregnancy Pickle: Evolved Immune Compensation Due to Pregnancy  
691 Underlies Sex Differences in Human Diseases. *Trends Genet*. 2019;35(7):478-488.  
692 doi:10.1016/j.tig.2019.04.008
- 693 20. Stang PE, et al. Advancing the Science for Active Surveillance: Rationale and Design for the  
694 Observational Medical Outcomes Partnership. *Ann Intern Med*. 2010;153(9):600-606.  
695 doi:10.7326/0003-4819-153-9-201011020-00010
- 696 21. Stages of Alzheimer's. Alzheimer's Disease and Dementia. Accessed March 24, 2024.  
697 <https://alz.org/alzheimers-dementia/stages>
- 698 22. Caspersen C, et al. Mitochondrial A $\beta$ : a potential focal point for neuronal metabolic  
699 dysfunction in Alzheimer's disease. *FASEB J*. 2005;19(14):2040-2041. doi:10.1096/fj.05-  
700 3735fje
- 701 23. Marizzoni M, et al. A peripheral signature of Alzheimer's disease featuring microbiota-gut-  
702 brain axis markers. *Alzheimers Res Ther*. 2023;15(1):101. doi:10.1186/s13195-023-01218-5
- 703 24. Ferreiro AL, et al. Gut microbiome composition may be an indicator of preclinical  
704 Alzheimer's disease. *Sci Transl Med*. 2023;15(700):eabo2984.  
705 doi:10.1126/scitranslmed.abo2984
- 706 25. Miller ZA, et al. Left-handedness, learning disability, autoimmune disease, and seizure  
707 history influence age at onset and phenotypical targeting of Alzheimer's disease [preprint].  
708 <https://doi.org/10.1101/2022.12.17.22283307>. Posted on medRxiv December 20, 2022.
- 709 26. Li X, et al. Dementia and Alzheimer's disease risks in patients with autoimmune disorders:  
710 Autoimmune disorders and dementia. *Geriatr Gerontol Int*. 2018;18(9):1350-1355.  
711 doi:10.1111/ggi.13488
- 712 27. Huang J, et al. Inflammatory Diseases, Inflammatory Biomarkers, and Alzheimer Disease.  
713 *Neurology*. 2023;100(6):e568-e581. doi:10.1212/WNL.0000000000201489
- 714 28. Beesley LJ, Mukherjee B. Statistical inference for association studies using electronic health  
715 records: handling both selection bias and outcome misclassification. *Biometrics*.  
716 2022;78(1):214-226. doi:10.1111/biom.13400

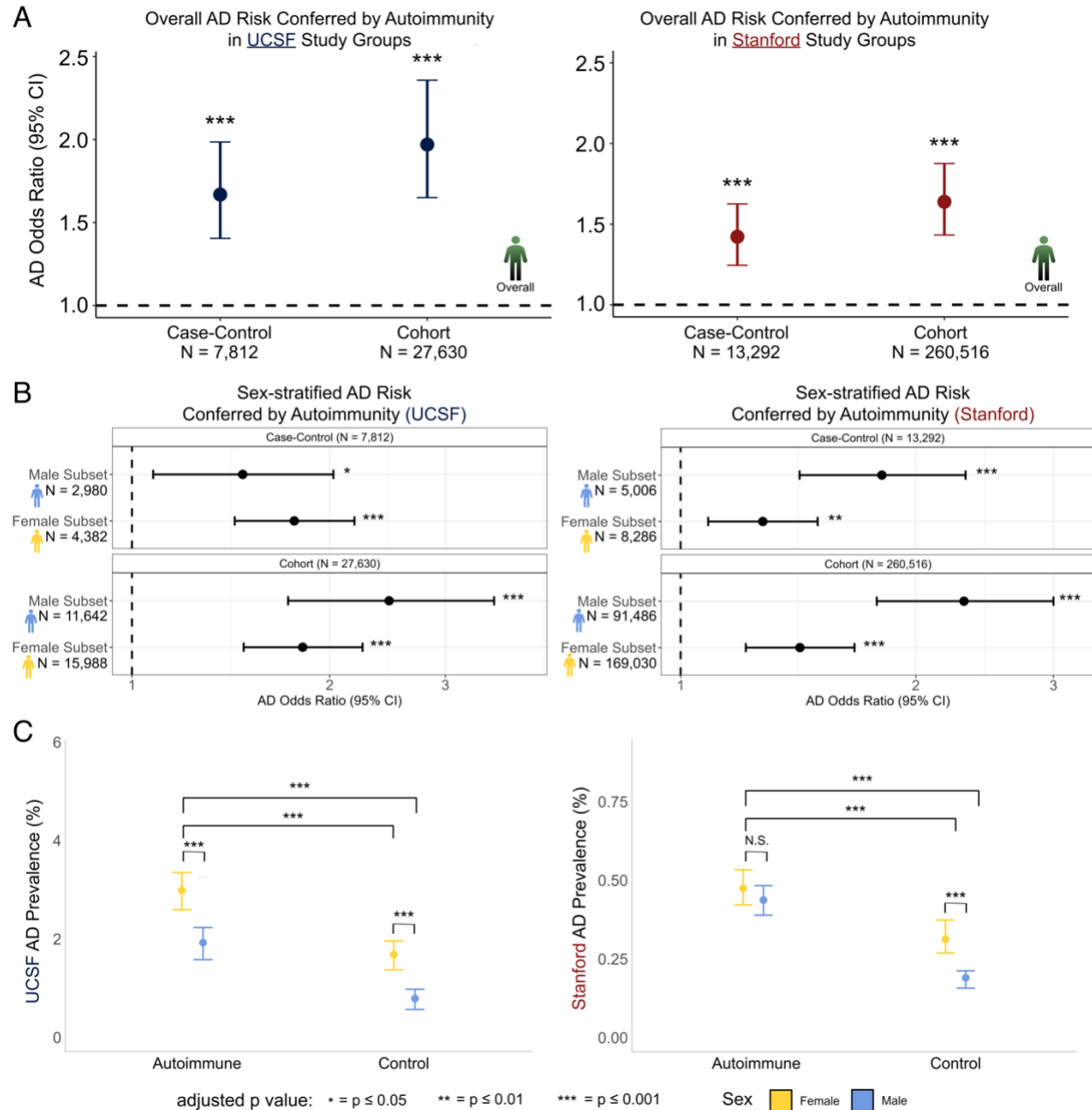
- 717 29. Peskoe SB, et al. Adjusting for selection bias due to missing data in electronic health  
718 records-based research. *Stat Methods Med Res.* 2021;30(10):2221-2238.  
719 doi:10.1177/09622802211027601
- 720 30. University of California, San Francisco, Academic Research Systems (2022). UCSF DeID  
721 CDW-OMOP. 2023-August. University of California, San Francisco. Dataset. Available  
722 through <https://data.ucsf.edu/research/deid-data>.
- 723 31. Wingo TS, et al. Autosomal recessive causes likely in early-onset Alzheimer disease. *Arch*  
724 *Neurol.* 2012;69(1):59-64. doi:10.1001/archneurol.2011.221

725 **Figures & Figure Legends**



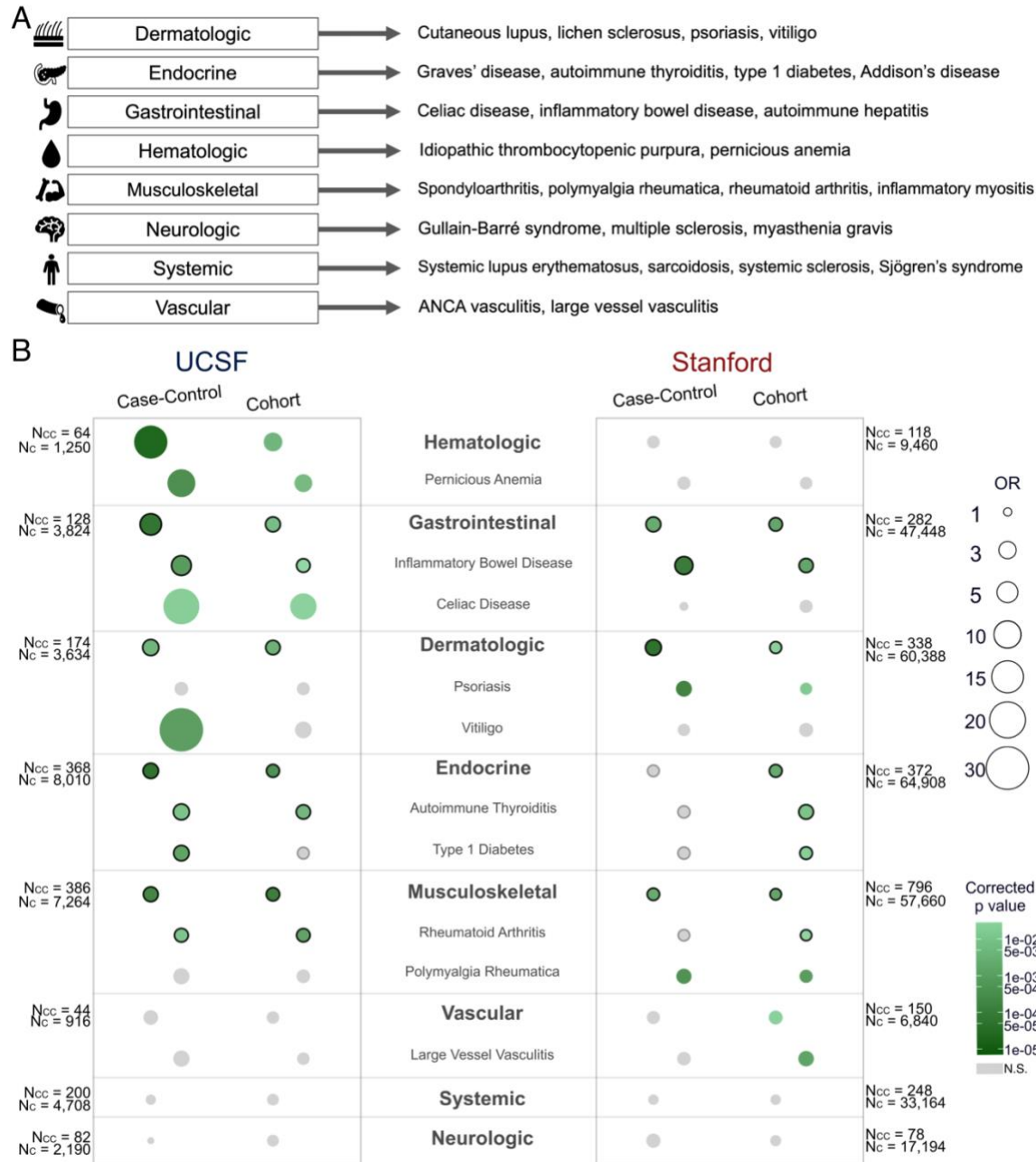
726

727 **Figure 1. Risk analysis framework and observational study designs.** We take a top-down hierarchical  
 728 approach to calculate AD risk in autoimmune disorder patients based on two large electronic health record  
 729 (EHR) data sets. We compute odds ratios to assess if any autoimmune disorder diagnosis, a particular  
 730 autoimmune subtype diagnosis, or a specific autoimmune disorder diagnosis increases AD risk. We used  
 731 both a case-control and cohort study design to ensure robustness and reduce biases. For the case-control  
 732 study design, we first identified patients with the *outcome* of interest (an AD diagnosis, red) and then  
 733 determined which of the AD patients also had an autoimmune diagnosis (blue). We matched the AD  
 734 patients to non-AD controls using propensity score matching and gathered information on other  
 735 demographic variables for cases and controls. For the cohort study design, we identified patients with the  
 736 *exposure* first (an autoimmunity diagnosis, blue) and determined which of the exposed patients also had  
 737 an AD diagnosis (red). Demographic information was extracted and propensity score matching was used  
 738 to match autoimmune cases to non-autoimmune controls. We used these study structures to analyze data  
 739 from both the UCSF (discovery) and Stanford (validation) EHR databases. Dx = diagnosis



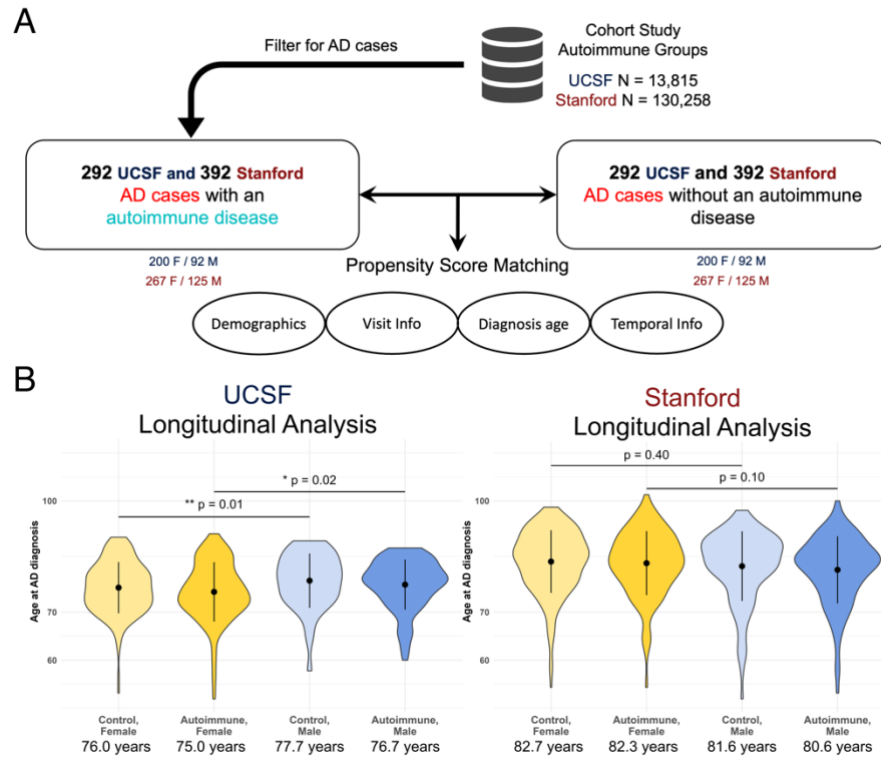
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750

**Figure 2. Autoimmune disorders are associated with increased AD risk across study designs and EHR data sets.** **A)** Odds ratios quantifying AD risk in autoimmune patients versus non-autoimmune controls. We observed increased odds of AD across both the UCSF (left) and Stanford (right) data sets, and across case-control and cohort study groups within each data set, robustly highlighting greater AD risk conferred by autoimmunity. **B)** AD risk in the female- and male-only subsets of each data set. Increased AD risk was present in both sexes in each data set. **C)** AD prevalence calculated in the cohort study designs in different sex and disorder strata. Confidence intervals were obtained by bootstrapping the data. AD prevalence was higher in women with autoimmune disorders compared to all other groups in both the UCSF (left) and Stanford (right) data.



751  
752  
753  
754  
755  
756  
757  
758  
759  
760

**Figure 3. Specific autoimmune disorder subtypes and individual diseases associate with increased AD risk. A)** Individual autoimmune disorders can be grouped into subtypes based on physiological symptomatology. We used eight subtype groups in our analysis. **B)** Odds ratios quantifying AD risk in patients with different autoimmune disorder subtypes and specific autoimmune disorders compared to controls. Disorder subtypes are in bold, and the specific disorders that fall into each subtype category are listed below. Only individual autoimmune disorders that were statistically significant are pictured. The larger and darker the circle is, the greater the effect size and significance of the odds ratio, respectively. N.S. = Not Significant after multiple testing correction



761  
762  
763

**Figure 4. AD onset is accelerated in women, with potential acceleration**

764 **from autoimmune disorders. A)** Study design for the longitudinal AD onset  
765 analysis, conducted using data from the UCSF and Stanford EHR data sets.

766 **B)** Distributions of AD diagnosis age among individuals with and without  
767 autoimmune diseases in the longitudinal cohorts, stratified by sex. Numbers  
768 below stratifications are the mean age at diagnosis, and the black summary  
769 lines highlight the mean and standard deviation of each distribution. Within  
770 the autoimmune and control subgroups, women were diagnosed with AD at  
771 a younger age than men, indicating that sex plays a large role in the age of  
772 AD onset.



773 **Tables**

774 **Table 1. UCSF and Stanford Study Group Demographics**

UCSF					Stanford				
Case-Control Study Design					Case-Control Study Design				
	Non-AD Controls	AD Patients	p value	SMD		Non-AD Controls	AD Patients	p value	SMD
N	3,906	3,906			N	6,646	6,646		
Sex (%)			1	<0.001	Sex (%)			1	<0.001
Female	2416 (61.9)	2416 (61.9)			Female	4,143 (62.3)	4,143 (62.3)		
Male	1,490 (38.1)	1,490 (38.1)			Male	2,503 (37.7)	2,503 (37.7)		
Unknown	0 (0.0)	0 (0.0)			Unknown	0 (0.0)	0 (0.0)		
Self-reported race (%)			1	<0.001	Self-reported race (%)			1	<0.001
Asian	531 (13.6)	531 (13.6)			Asian	952 (14.3)	952 (14.3)		
Black or African American	232 (5.9)	232 (5.9)			Black or African American	315 (4.7)	315 (4.7)		
Native American or Alaska Native	4 (0.1)	4 (0.1)			Native American or Alaska Native	9 (0.1)	9 (0.1)		
Native Hawaiian or Other Pacific Islander	174 (4.5)	174 (4.5)			Native Hawaiian or Other Pacific Islander	39 (0.6)	39 (0.6)		
White	2,634 (67.4)	2,634 (67.4)			White	4,182 (62.9)	4,182 (62.9)		
Other	331 (8.5)	331 (8.5)			Other	813 (12.2)	813 (12.2)		
Unknown/Decline to State	0 (0.0)	0 (0.0)			Unknown/Decline to State	336 (5.1)	336 (5.1)		
Self-reported ethnicity (%)			1	<0.001	Self-reported ethnicity (%)			1	<0.001
Not Hispanic or Latino	3,672 (94.0)	3,672 (94.0)			Not Hispanic or Latino	5,732 (86.2)	5,732 (86.2)		
Hispanic or Latino	234 (6.0)	234 (6.0)			Hispanic or Latino	503 (7.6)	503 (7.6)		
Unknown/Decline to State	0 (0.0)	0 (0.0)			Unknown/Decline to State	411 (6.2)	411 (6.2)		
Birth year (mean (SD))	1935.52 (5.93)	1935.52 (5.92)	1	<0.001	Birth year (mean (SD))	1934.53 (10.11)	1934.50 (10.07)	0.827	0.004
Lifespan, years (mean (SD))	80.05 (6.74)	80.07 (6.73)	0.945	0.002	Lifespan, years (mean(SD))	NA	NA		
Autoimmune disease status = True (%)	235 (6.0)	377 (9.7)	<0.001	0.136	Autoimmune disease status = True (%)	415 (6.2)	575 (8.7)	<0.001	0.092
Cohort Study Design					Cohort Study Design				
	Non-Autoimmune Controls	Autoimmune Disorder Patients	p value	SMD		Non-Autoimmune Controls	Autoimmune Disorder Patients	p value	SMD
N	13,815	13,815			N	130,258	130,258		
Sex (%)			1	<0.001	Sex = Male (%)			1	<0.001
Female	7,994 (57.9)	7,994 (57.9)			Female	84,515 (64.9)	84,515 (64.9)		
Male	5,821 (42.1)	5,821 (42.1)			Male	45,717 (35.1)	45,717 (35.1)		
Unknown/Decline to State	0 (0.0)	0 (0.0)			Unknown/Decline to State	26 (0.0)	26 (0.0)		
Self-reported race (%)			1	<0.001	Self-reported race (%)			1	<0.001
Asian	1,241 (9.0)	1,241 (9.0)			Asian	19,022 (14.6)	19,022 (14.6)		
Black or African American	1,195 (8.7)	1,195 (8.7)			Black or African American	5,036 (3.9)	5,036 (3.9)		
Native American or Alaska Native	85 (0.6)	85 (0.6)			Native American or Alaska Native	559 (0.4)	559 (0.4)		
Native Hawaiian or Other Pacific Islander	541 (3.9)	541 (3.9)			Native Hawaiian or Other Pacific Islander	1,087 (0.8)	1,087 (0.8)		
White	8,916 (64.5)	8,916 (64.5)			White	72,757 (55.9)	72,757 (55.9)		
Other	1,837 (13.3)	1,837 (13.3)			Other	22,111 (17.0)	22,111 (17.0)		
Unknown/Decline to State	0 (0.0)	0 (0.0)			Unknown/Decline to State	9,686 (7.4)	9,686 (7.4)		
Self-reported ethnicity (%)			1	<0.001	Self-reported ethnicity (%)			1	<0.001
Not Hispanic or Latino	12,339 (89.3)	12,339 (89.3)			Not Hispanic or Latino	101,442 (77.9)	101,442 (77.9)		
Hispanic or Latino	1,476 (10.7)	1,476 (10.7)			Hispanic or Latino	17,816 (13.7)	17,816 (13.7)		
Unknown/Decline to State	0 (0.0)	0 (0.0)			Unknown/Decline to State	11,000 (8.4)	11,000 (8.4)		
Birth year (mean (SD))	1946.49 (13.43)	1946.55 (13.51)	0.705	0.005	Birth year (mean (SD))	1968.38 (21.83)	1968.38 (21.83)	0.724	0.001
Lifespan, years (mean (SD))	69.25 (13.21)	69.10 (13.38)	0.327	0.012	Lifespan, years (mean(SD))	NA	NA		
AD status = True (%)	195 (1.4)	379 (2.7)	<0.001	0.093	AD status = True (%)	354	579	<0.001	0.029

776  
777  
778  
779

Reports of 'NA' for lifespan indicate censored death information in the Stanford study groups. SMD = Standardized Mean Difference, SD = Standard deviation