# Estimating Broad-Sense Heritability using Longitudinal Secular Trend Data – Final Presentation

**Sarah Davis**

**April 9th, 2020**

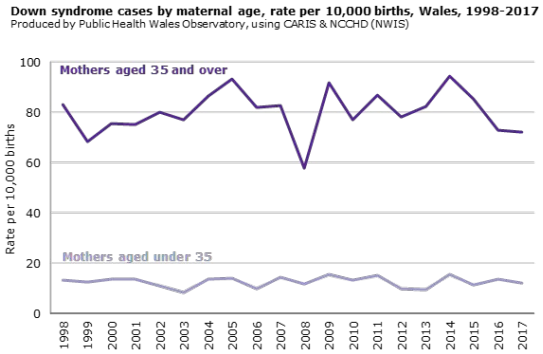
# Overview of Presentation

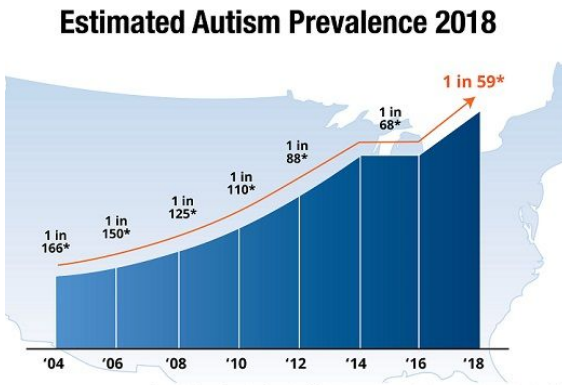
1. *Introduction*
2. *Background*
3. *Methods*
4. *Results*
5. *Discussion*
6. *Conclusion*



# Introduction: Hypothesis

- There are many proposed methods for calculating heritability, each have problems
- Given incidence/prevalence data, using the minimum and maximum values seen should give us insight into how “heritable” a trait is
  - Based on the idea that genes are generally slow-to-change



# Background

# Background: Heritability

- **Broad sense heritability ($H^2$) is defined as:**

$$H^2 = \frac{var(G)}{var(P)}$$

- **Traditional methods calculate narrow sense heritability ($h^2$) using:**
  - **Twin studies (Falconer's formula)** $h^2 = 2(r_{mz} - r_{dz})$
  - **GWAS studies** $h^2 = \sum_{i \in S} 2f_i(1 - f_i)\beta_i^2$

# Background: Missing Heritability

$$\pi_{explained} = \frac{h_S^2}{h_{all}^2}$$

$h_S^2$ = Proportion of phenotypic variance explained by a set of known variants S

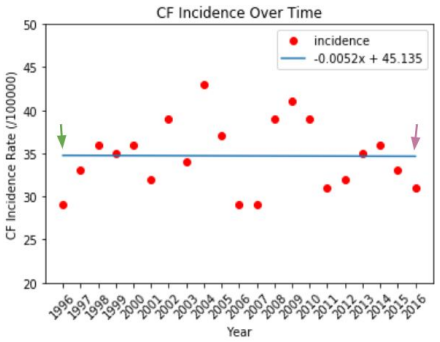
$h_{all}^2$ = Proportion of phenotypic variance explained by all variants

- What are some proposed causes of this discrepancy?
  - Failing to account for common/rare SNPs
  - Failing to account for genetic interactions
  - The microbiome being shared between relatives

# Methods

# Methods: Our $H^2$ method

- We use longitudinal incidence/prevalence data to find the minimum and maximum of the "environmental phenotypic range"
  - variety of ways to do this
- Use a simple "min/max" formula to obtain our $H^2$ estimate



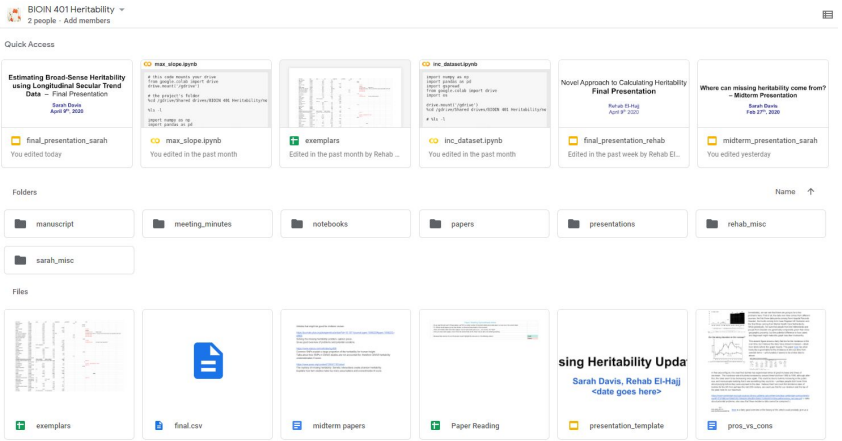

Minimum:
= -0.0052 (2016) + 45.135
= 34.66

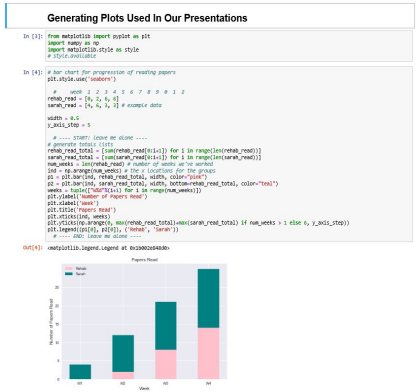
Maximum:
= -0.0052 (1996) + 45.135
= 34.76

Heritability
= Minimum / Maximum
= 34.66 / 34.76
= 0.997

# Methods: Set Up

- Google Drive
- Presentation template
- Notebook to keep track of reading (46 lines of code)

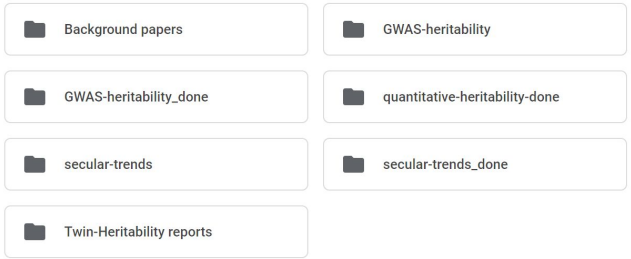


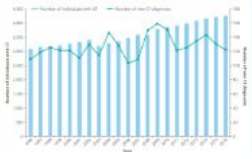






# Methods: Background Reading

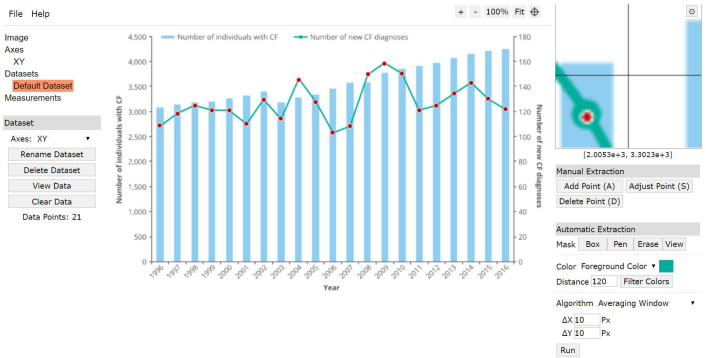
- Next step was going through the 158 papers sent to us by Dr. Wishart
- Out of these, I processed 129
  - 87 inc/prev papers
  - 12 twin heritability papers
  - 29 miscellaneous papers

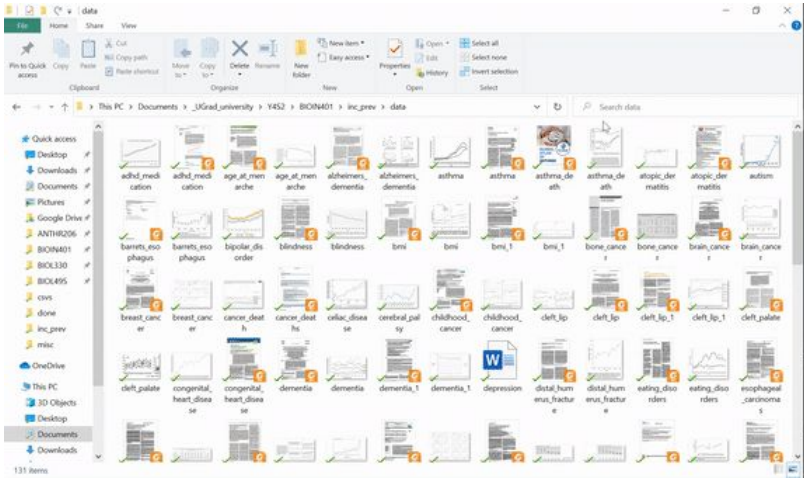



| Paper Name | Date Published | Phenomenon | Charts of Interest | Synopsis | Places Measured |
|---|---|---|---|---|---|
| THE CANADIAN CYSTIC FIBROSIS REGISTRY 2016 ANNUAL DATA REPORT | November, 2017 | Cystic Fibrosis |  | 89.2% of CF patients in Canada carry at least one copy of the most common CF-causing mutation, F508del -- growth seems to be staying constant over time | Canada |

# Methods: Data Needed

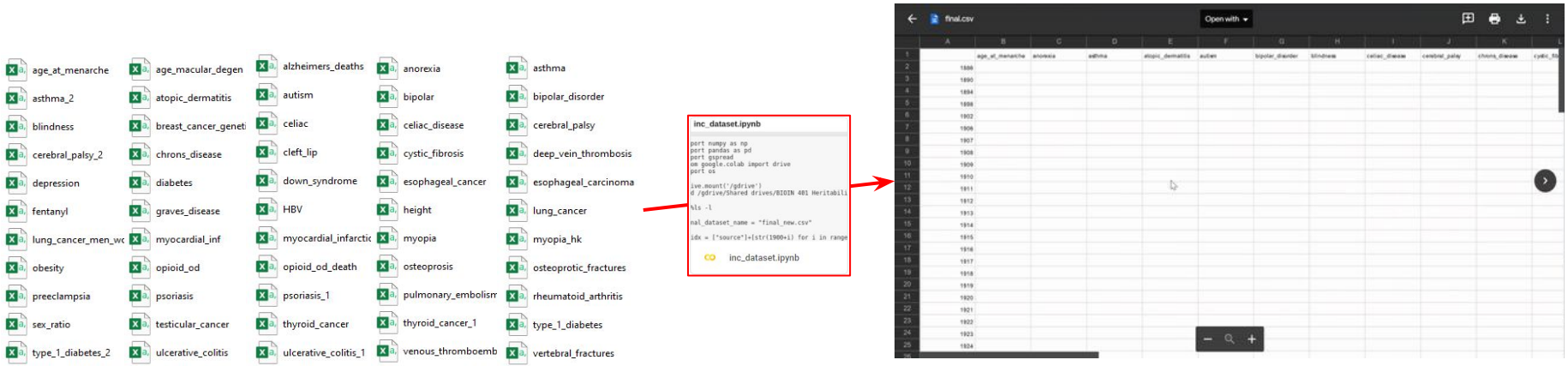
- In order to calculate our heritability estimates, we needed lots of longitudinal data
- I found 68 new papers that report longitudinal data, of these, I converted a total of 55 (using webplotdigitizer) into .csv format



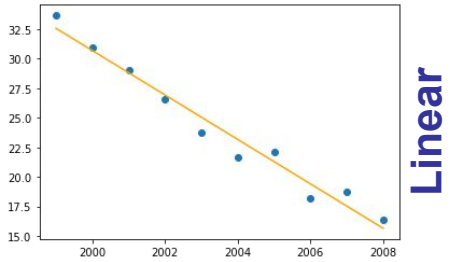
# Methods: Master Dataset

- I wrote 21 lines of code to combine all these CSVs into a “master dataset”
  - Much easier to automate heritability calculations this way

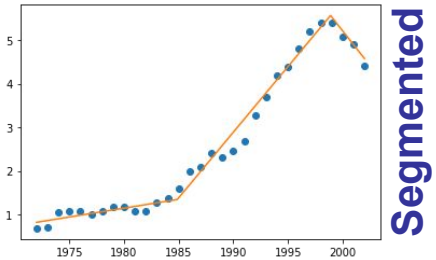


# Methods: Calculations

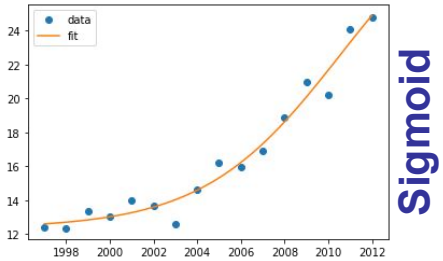
- Taking the absolute min and max of the data is sometimes not the best option (e.g., cystic fibrosis)
- I implemented three methods to capture all other possible scenarios
  - Linear regression
  - Segmented linear regression
  - Sigmoid regression
- Total: 198 lines of code




Linear

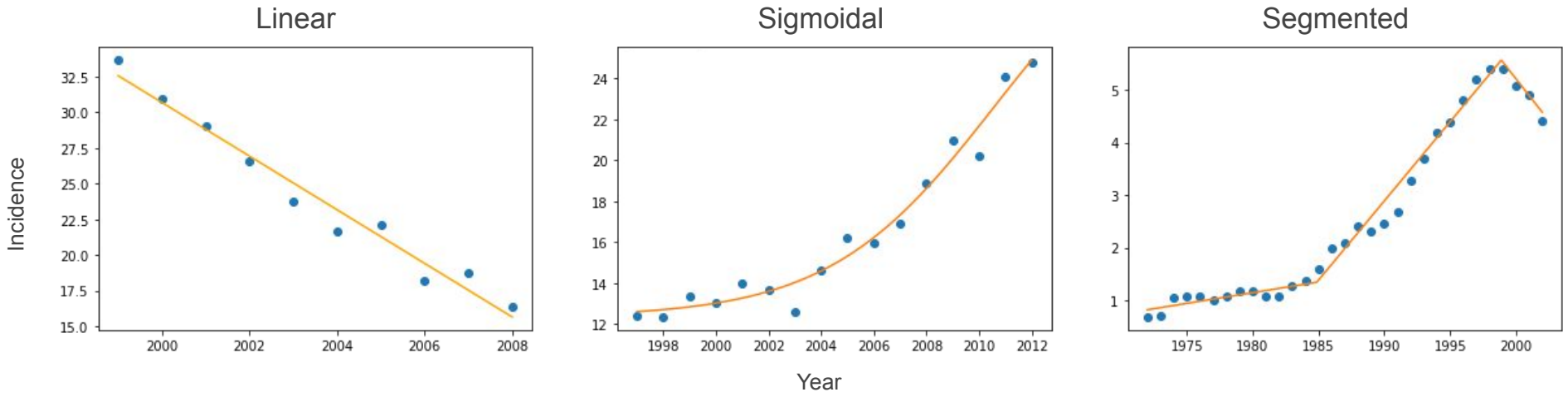



Segmented




Sigmoid

# Methods: Calculations



# Methods: Comparisons

- We needed to compare our estimates to twin and GWAS heritability coefficients
  - This involved collecting comparable GWAS heritability coefficients and twin heritability coefficients
  - I tackled the twin coefficients:

| Condition/Phenotype | Twin Heritability | Source |
|---|---|---|
| Acne (teenagers) | 0.85 | https://www.se |
| Acne (women) | 0.81 | https://www.nc |
| Acne vulgaris female | 0.39 | https://www.ca |
| Acne vulgaris male | 0.63 | https://www.ca |
| Age-related macular degeneration | 0.45 | https://www.sci |
| Age-related macular degeneration | 0.26-0.70 | https://jamanet |
| Age-related macular degeneration (hard drusen ≥20) | 0.81 | https://www.sci |
| Age-related macular degeneration (soft drusen ≥125 µm) | 0.57 | https://www.sci |
| Aggressive periodontitis (Chinese people) | 0.64 | "unpublished in |
| Agoraphobia | 0.48 | https://onlinelit |
| Alcohol Use Disorder (females) | 0.42 | https://www.nc |
| Alcohol Use Disorder (males) | 0.61 | https://www.nc |
| Alzheimer's disease | 0.79 | https://drive.go |
| Amyotrophic lateral sclerosis | 0.61 | https://jnnp.bm |
| Anorexia nervosa | 0.56 | https://jamanet |
| Anorexia nervosa | 0.74 | https://www.ca |
| Anorexia nervosa (broad definition) | 0.52 | https://www.ca |
| Anorexia nervosa (narrow definition) | 0.48 | https://www.ca |
| Anxiety disorder | 0.54 | https://www.ca |
| Asthma | 0.53 | https://www.sci |
| Asthma | 0.91 | https://erj.ersjo |
| Asthma | 0.92 | https://www.ca |
| Asthma (Denmark) | 0.73 | https://chempo |
| Asthma (females) | 0.688 | https://reader.e |
| Asthma (males) | 0.05 | https://reader.e |
| Asthma (Swedish females) | 0.62 | https://onlinelit |
| Asthma (Swedish males) | 0.76 | https://onlinelit |

| Condition/Phenotype | Twin Heritability | Source |
|---|---|---|
| Autism | 0.37 | https://onlinelibr |
| Autism | 0.91-0.93 | https://www.cam |
| Autism (females) | 0.87 | https://onlinelibr |
| Autism (males) | 0.73 | https://onlinelibr |
| Autism Spectrum Disorder | 0.38 | https://onlinelibr |
| Barrett's esophagus | 0.35 | https://academic |
| Behcet's disease | 0.41 | https://pdfs.sem |
| Bipolar disorder | 0.93 | https://ajp.psych |
| Body Mass Index (age 20) | 0.77 | https://jamanetw |
| Breast cancer | 0.31 | https://cebp.aac |
| Bronchopulmonary dysplasia | 0.82 or 0.79 depending on definition | https://pediatric |
| Bulimia nervosa | 0.61 | https://www.can |
| Celiac disease | 0.75 | https://gut.bmj.c |
| Celiac disease (Silent) | 0.57 | https://www.sci |
| Celiac disease (Symptomatic) | 0.87 | https://www.sci |
| Chronic gastroesophageal reflux | 0.43 | https://www.nct |
| Cleft lip (Danish) | 0.17 | https://journals.l |
| Cleft lip (Finnish) | 0.45 | https://journals.l |
| Colon cancer | 0.16 | https://www.sci |
| Congenital heart malformation | 0.81, 0.77, 0.90 | https://onlinelibr |
| Corneal astigmatism | 0.333 | https://iovs.arvo |
| Corneal astigmatism | 0.6 | https://iovs.arvo |
| Corneal curvature | 0.71 | https://iovs.arvo |
| Coronary artery disease (death, females) | 0.38 | http://apps.webo |
| Coronary artery disease (death, males) | 0.57 | http://apps.webo |
| Criminal Behavior (females) | 0.43 | https://www.nct |

| Condition/Phenotype | Twin Heritability | Source |
|---|---|---|
| Criminal Behavior (males) | 0.5 | https://www.nct |
| Crohn's disease | 0.75 | https://academic |
| Crohn's disease | 0.89 | https://academic |
| Drinking behavior | 0.11 | https://www.bio |
| Drug abuse (female) | 0.57 | https://www.nct |
| Drug abuse (male) | 0.54 | https://www.nct |
| Drug use | 0.45 | https://www.bio |
| Early-onset obesity (children) | 0.88 | https://onlinelibr |
| Eczema (children) | 0.9 | https://erj.ersjou |
| Eczema (females) | 0.71 | https://onlinelibr |
| Eczema (males) | 0.74 | https://onlinelibr |
| Endometriosis | 0.47 | https://www.sci |
| Epilepsy | 0.45 | http://www.neu |
| Epilepsy | 0.69 | https://onlinelibr |
| Esophageal cancer | 0.25 | https://academic |
| Febrile seizures | 0.81 | https://onlinelibr |
| Fractures (vertebral) | 0.24 | http://www.dive |
| Freckles | 0.91 | https://academic |
| Glaucoma (open angle, chronic) | 0.13 | https://onlinelibr |
| Hay fever | 0.6 | article appears n |
| Hay fever | 0.78 | https://www.car |
| Hay fever (females) | 0.7 | https://onlinelibr |
| Hay fever (males) | 0.33 | https://onlinelibr |
| Height at age 20 | 0.8 | https://jamanetw |
| Inpatient psychiatric treatment | 0.47 | https://onlinelibr |
| Macular pigment measured using AF | 0.85 | https://iovs.arvo |
| Macular Pigment measured using HFP | 0.67 | https://iovs.arvo |
| Major depressive disorder | 0.33 | https://onlinelibr |
| Male-pattern baldness | 0.81 | https://www.sci |

| Condition/Phenotype | Twin Heritability | Source |
|---|---|---|
| Migraine | 0.52 | https://thejournal |
| Obesity | 0.77 | https://www.scie |
| Panic Disorder | 0.38 | https://onlinelibr |
| Polycystic ovary syndrome | 0.79 | https://academic |
| Proliferative Diabetic Retinopathy | 0.52 | https://diabetes.d |
| Prostate cancer | 0.58 | https://cebp.aacr |
| Psoriasis | 0.66 | https://onlinelibr |
| Rectal Cancer | 0.15 | https://www.scie |
| Rheumatoid arthritis (Finnish data) | 0.65 | https://www.ncb |
| Rheumatoid arthritis (UK data) | 0.53 | https://www.ncb |
| Scoliosis | 0.38 | https://link.springe |
| Sleep Quality/Disturbance | 0.33 | https://academic |
| Social phobia | 0.39 | https://onlinelibr |
| Treatment with psychoactive medication | 0.49 | https://onlinelibr |
| Type 2 diabetes (men) | 0.69 | https://link.springe |
| Type 2 diabetes (women) | 0.79 | https://link.springe |
| Ulcerative colitis | 0.23 | https://academic |
| Ulcerative colitis | 0.67 | https://academic |
| Urticaria (females) | 0.6 | https://onlinelibr |
| Urticaria (males) | 0.54 | https://onlinelibr |
| Varicose veins | 0.86 | https://jmg.bmj.c |
| Weight | 0.78 | https://jamanetw |

- ➔ 106 Twin Heritability Measurements
- ➔ 65 Phenotypes
- ➔ 62 papers

# Methods: What else can we do?

- **Throughout this project, not only have I been putting in the manual work to get the data, but in my off time I have been busy exploring different theories**

## Last Week: Heritability or Not?

We have: $\frac{prev_{geno}\ @\ \text{time} = t}{prev_{geno}\ @\ \text{time} = 1} = 1$ and $\frac{prev_{pheno}\ @\ \text{time} = t}{prev_{pheno}\ @\ \text{time} = 1} = x$

So...

$$\frac{var_{geno}}{var_{pheno}} \propto \frac{\frac{prev_{geno}\ @\ \text{time} = t}{prev_{geno}\ @\ \text{time} = 1}}{\frac{prev_{pheno}\ @\ \text{time} = t}{prev_{pheno}\ @\ \text{time} = 1}} = \frac{1}{\frac{prev_{pheno}\ @\ \text{time}=t}{prev_{pheno}\ @\ \text{time}=1}} = \frac{prev_{pheno}\ @\ \text{time}=1}{prev_{pheno}\ @\ \text{time}=t}$$

## Previously: BMI Example

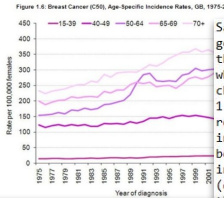
| Country Level | Town Level | Individual Level |
|---|---|---|
| Min = 20.5<br>Max = 31.9 | Min = ~13<br>Max = ~40 | Min = 7.5<br>Max = 188 |
| $H^2$ = min/max<br>$H^2$ = 20.5/31.9<br>$H^2$ = 0.64 | $H^2$ = min/max<br>$H^2$ = 13/40<br>$H^2$ = 0.33 | $H^2$ = min/max<br>$H^2$ = 7.5/188<br>$H^2$ = 0.04 |

As we go down to the individual level, out calculation relies on the (wrong) assumption that any one individual could achieve a BMI of either 7.5 or 188 given their environment, and I don't think that's true

What level should our calculation be based on?

## Last Week: Age Data

- "Most cases of breast cancer are not caused by inherited genetic factors. These cancers are associated with somatic mutations in breast cells that are acquired during a person's lifetime, and they do not cluster in families."
  - Age is an environmental factor




Calculation

Min = ~20 $H^2$ = min/max

Max = ~4

Say that the variance of flower stem length in a population of genetically identical heterozygous individuals is 10cm. Define this population as P1. Define P2 as the offspring generation, who according to classic mendelian genetics (and assuming one chromosome and no chromosomal crossover), will have a ratio of 1:2:1 homozygous dominant, heterozygous, and homozygous recessive individuals. It is not hard to see how the variance in P1 is the same as var(E), as all genetic differences are being controlled for. Now imagine these populations are left to interbreed for n-2 more generations. The variance over P1...Pn (define it to be 30cm) combined would be our var(P), as this is use the following formulas:

(E)

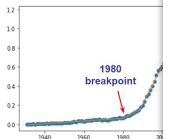
(var(P)-var(E))/var(P) = (30-10)/30 = ⅔ =

ample, the H^2 for the stem length of this

## Previously: Example with real data

$$H^2 = \frac{V_g}{V_p} = \frac{Variance_{1930 \rightarrow 1980}}{Variance_{1930 \rightarrow 2015}}$$

$$H^2 = \frac{0.000434}{0.111616} = 0.00388$$



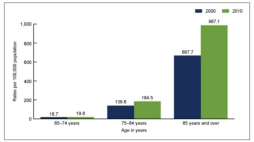

- Therefore, I calculate autism to have a heritability of ~0.4%

## Previously: $V_G$ thoug

- Are there any other ways to calculate this $V_G$ ?
- It's easy to calculate $V_E$ when the genetics are exactly the same but environments differ
  - See flower example in manuscript
- Can we calculate $V_G$ using people with different genetics but controlling for environment?
  - e.g., boarding schools, convents
  - Does this change over time??



## Last Week: More Examples




Dementia at 75-84 years

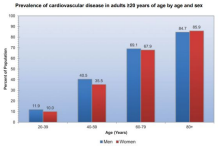
Min = 19.8
Max = 184.5

$H^2$ = min/max
$H^2$ = 19.8/184.5
$H^2$ = 0.107

Dementia at 85+ years

Min = 19.8
Max = 987.1

$H^2$ = min/max
$H^2$ = 19.8/987.1
$H^2$ = 0.0201




CVD at 40-59 years

Min = 10.0
Max = 35.5

$H^2$ = min/max
$H^2$ = 10.0/35.5
$H^2$ = 0.281

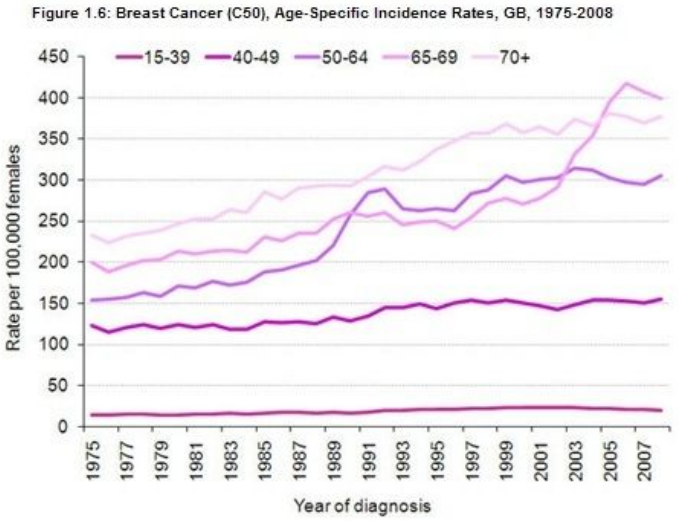
CVD at 80+ years

Min = 10.0
Max = 85.9

$H^2$ = min/max
$H^2$ = 10.0/85.9
$H^2$ = 0.116

# Methods: Age Heritability

- Often, time trends are split apart by age
- This distinction could help us estimate environmental effects, as # of years lived is a proxy for compounding environmental effects



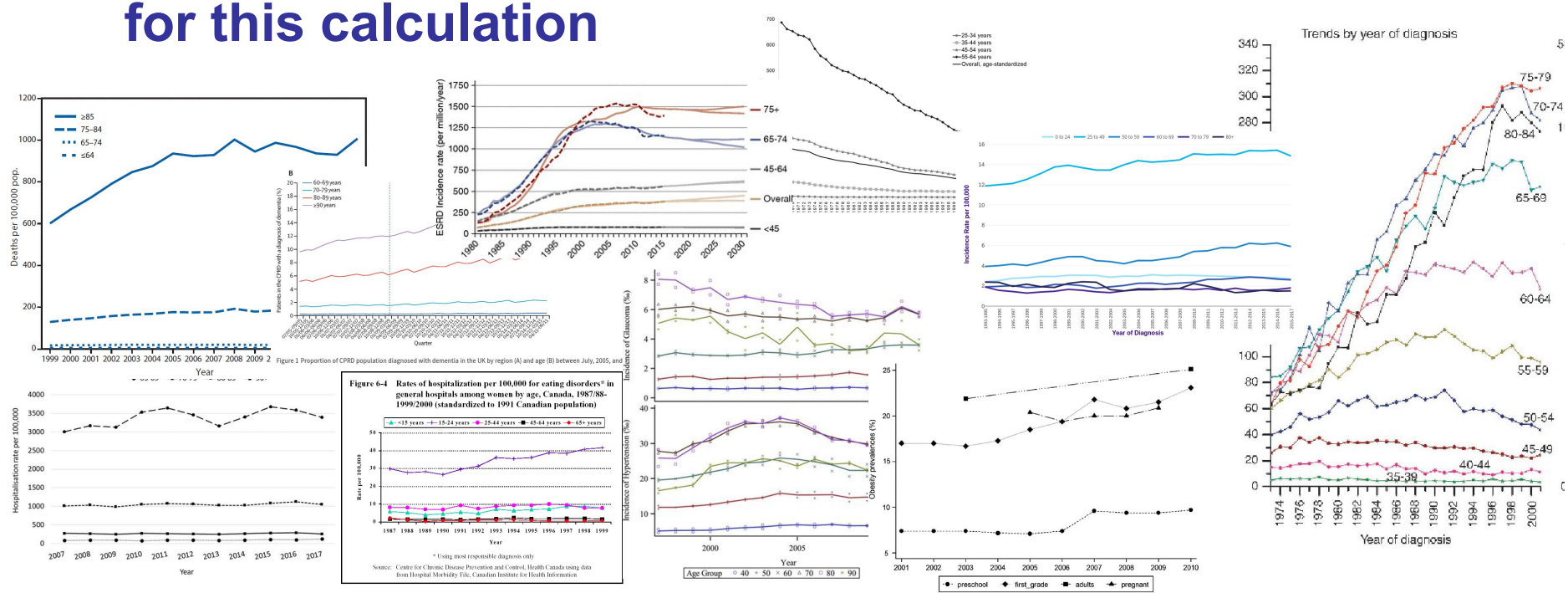

**Calculation**

Min = ~20
Max = ~420

$H^2$ = min/max
$H^2$ = 20/420
$H^2$ = 0.049

# Methods: Age Data

- On top of all the longitudinal data, I also found a total of 14 examples of longitudinal age data that can be used for this calculation



# Results

# Results: $H^2$ Estimates – Complex

| phenotype | gwas | twin | longitudinal | age | quantitative |
|---|---|---|---|---|---|
| age at menarche | 0.320 | 0.61 | – | – | 0.44 |
| alzheimers | 0.033 | 0.79 | 0.472 | 0.020 | – |
| anorexia | 0.009 | 0.58 | 0.886 | 0.018 | – |
| asthma | 0.005 | 0.74 | 0.088 | – | – |
| atopic dermatitis | 0.003 | 0.85 | 0.100 | – | – |
| autism | 0.003 | 0.65 | 0.013 | – | – |
| bipolar disorder | 0.005 | 0.93 | 0.498 | – | – |
| blood pressure (diastolic) | 0.035 | 0.49 | – | – | 0.64 |
| blood pressure (systolic) | 0.035 | 0.45 | – | – | 0.51 |
| BMI (body mass index) | 0.027 | 0.85 | – | – | 0.16 |
| breast cancer | 0.012 | 0.31 | 0.538 | 0.052 | – |
| celiac disease | 0.582 | 0.73 | 0.305 | – | – |
| cleft lip | 0.040 | 0.70 | 0.323 | – | – |
| crohn's disease | 0.260 | 0.75 | 0.340 | – | – |
| epilepsy | 0.006 | 0.45 | 0.078 | – | – |
| esophageal cancer | 0.009 | 0.25 | 0.077 | – | – |
| glaucoma | 0.004 | 0.13 | 0.563 | 0.073 | – |
| graves disease | 0.030 | 0.57 | 0.030 | – | – |
| hay fever | 0.160 | 0.78 | 0.180 | – | – |

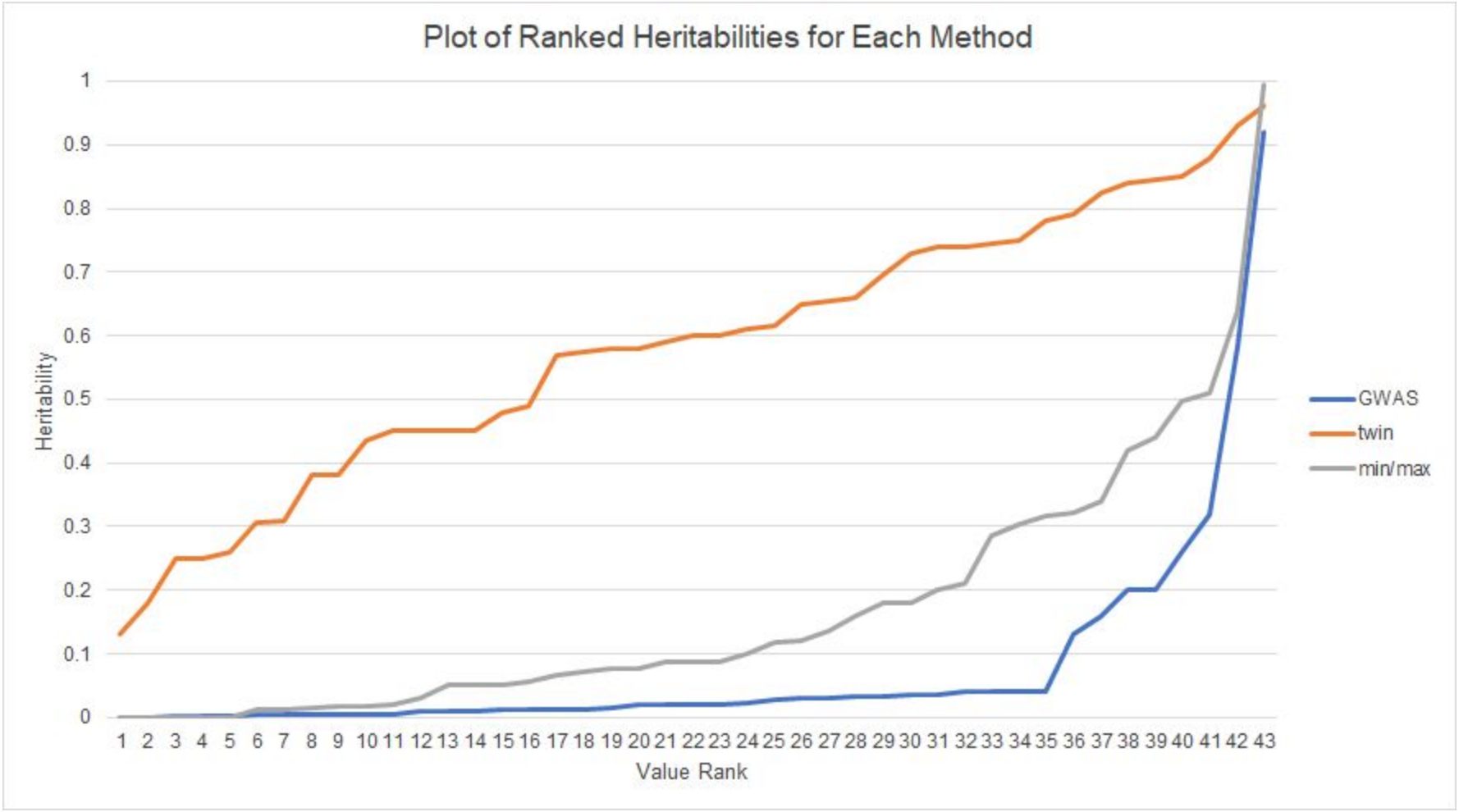
| phenotype | gwas | twin | longitudinal | age | quantitative |
|---|---|---|---|---|---|
| height | 0.200 | 0.84 | – | – | 0.21 |
| IQ | 0.016 | 0.48 | 0.180 | – | – |
| lung cancer | 0.005 | 0.18 | 0.040 | 0.016 | – |
| macular degeneration | 0.012 | 0.45 | 0.018 | – | – |
| myocardial infarction | 0.020 | 0.62 | 0.050 | 0.086 | – |
| myopia | 0.034 | 0.31 | 0.136 | – | – |
| obesity | 0.023 | 0.83 | 0.088 | 0.360 | – |
| osteoporosis | 0.001 | 0.65 | 0.000 | – | – |
| preeclampsia | n/a | 0.54 | 0.420 | – | – |
| prostate cancer | 0.003 | 0.58 | 0.000 | – | – |
| psoriasis | 0.040 | 0.66 | 0.317 | – | – |
| rheumatoid arthritis incidence | 0.030 | 0.59 | 0.201 | – | – |
| scoliosis | 0.006 | 0.38 | 0.056 | – | – |
| testicular cancer | 0.020 | 0.58 | 0.547 | 0.121 | – |
| thyroid cancer | 0.200 | 0.60 | 0.285 | – | – |
| type I diabetes | 0.020 | 0.88 | 0.118 | – | – |
| type II diabetes | 0.010 | 0.74 | 0.583 | 0.012 | – |
| ulcerative colitis | 0.040 | 0.45 | 0.052 | – | – |
| venous thromboembolism | 0.040 | 0.44 | 0.052 | – | – |

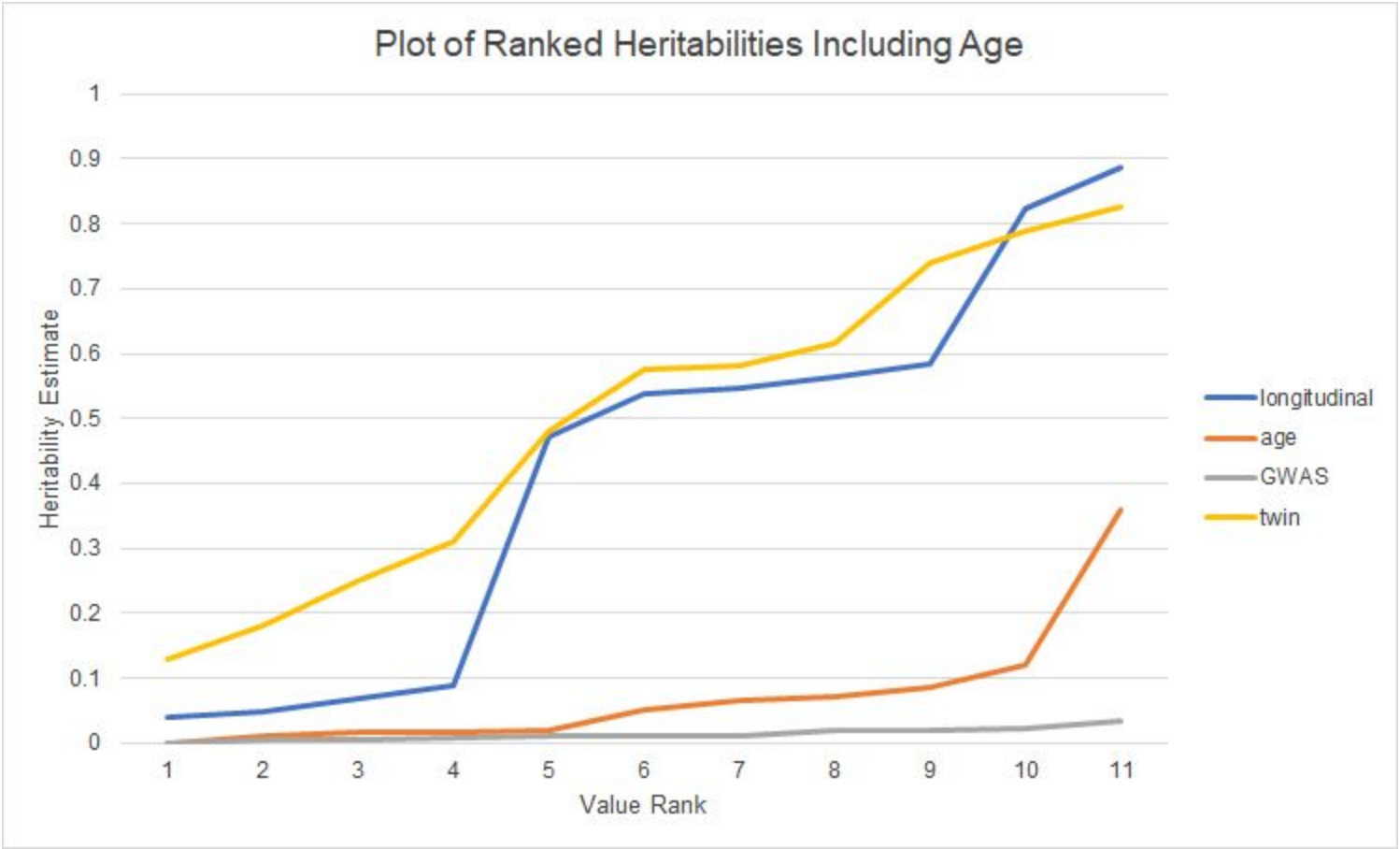
# Results: $H^2$ Estimates – Environmental

| phenotype | gwas | twin | longitudinal | age |
|---|---|---|---|---|
| hepatitis B | 0.020 | 0.60 | 0.000 | – |
| hip fracture | – | 0.48 | 0.824 | 0.000 |
| mumps | – | 0.39 | 0.000 | – |
| opioid od death | 0.000 | 0.26 | 0.000 | – |
| osteoporotic humeral fracture | – | 0.32 | 0.006 | – |
| peptic ulcers | – | 0.37 | 0.000 | – |
| poliomyelitis (polio) | – | – | 0.000 | – |
| rubella | – | 0.46 | 0.000 | – |
| smallpox | – | – | 0.000 | – |
| tuberculosis | 0.013 | 0.38 | 0.000 | – |
| left handedness | 0.012 | 0.25 | 0.070 | 0.066 |

# Results: $H^2$ Estimates – Genetic

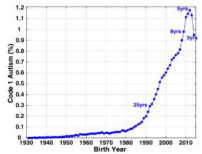
| phenotype | gwas | twin | longitudinal |
|---|---|---|---|
| black hair vs non-black hair | 0.921 | 0.96 | 0.996 |
| cystic fibrosis | CFTR, monogenic | 0.80 | 0.997 |
| down's syndrome | trisomy 21 | trisomy 21 | 0.989 |
| hereditary Haemochromatosis (HH) | HFE gene (C282Y, H63D) | HFE gene (C282Y, H63D) | 0.960 |
| sex (male) | XY chromosomes | 1.00 | 0.997 |
| sickle cell disease | HBB gene mutation | HBB gene mutation | 0.887 |


Plot of Ranked Heritabilities for Each Method
Heritability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
Value Rank
GWAS
twin
min/max


Plot of Ranked Heritabilities Including Age
Heritability Estimate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
11
Value Rank
longitudinal
age
GWAS
twin

# Results: Writing

- **Manuscript (19 pages)**
- **Pros/Cons (11 pages, ~3500 words)**
- ***H*$^2$ calculation reasoning (8 pages, ~800 words)**




**Autism**

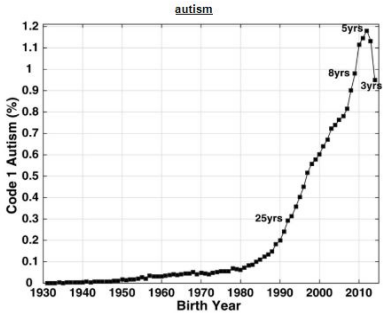
This will likely be one of the controversial topics we cover in this paper, namely because many, many people will argue that the increase in incidence is due to the shift in how autism is diagnosed. There are a few arguments as to why that is. First of all, autism was studied much better starting around the mid-20th century, therefore it became more well understood. Classic autism was first recognised in 1942 according to this source, so I need to look into how the source on the right has data from the 1930s. The data from this source are from the United States.

Most notably, the definition for autism was significantly revised in 1987, which broadened the concept of autism by adding a diagnosis at the mild end of the spectrum — pervasive developmental disorder-not otherwise specified (PDD-NOS) — and dropping the requirement for onset before 30 months. This can likely be cited as one of the causes for such a steep increase in the chart seen to the left. The DSM-IV, released in 1994 and revised in 2000, was the first edition to categorize autism as a spectrum.

There is a trend in epidemiology that goes as follows: Someone discovers a disease/disorder, then there is severe underdiagnosis because the information has not spread far enough yet, then there is a severe overdiagnosis, then afterwards there is a plateau. I would not be surprised if this is exactly what is happening here. In fact, near the most recent years, we can start to see a dip in incidence. I still believe that autism is largely an environmental phenomenon, backed up by such research as this, stating that there is a significant association between maternal immigration and the risk of having a child with an autism- spectrum disorder. It may be worthwhile to mention that autism can be caused by different things. There are certain subtypes of autism that are seemingly nearly entirely genetic. Examples of autism that have arisen from a rare or de novo mutation in a single-gene or locus include the neurodevelopmental disorders fragile X syndrome, 22q13 deletion syndrome, and 16p11.2 deletion syndrome

To wrap this all up, here is a paper, stating rather clearly that "US Department of Education Data on "Autism" Are Not Reliable for Tracking Autism Prevalence". This is pretty unfortunate -- we might have to give up the lovely chart above. Here is a nature article talking about this phenomenon.

Perhaps data from the netherlands is better to rely upon?




https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6223814/
American autism prevalence data, presumably m+f, min and max method -- min=0.001, max = 1.2 according to the paper.
We set a correction term -- that is, we assume for every child diagnosed at the minimum there were three children that were missed, and for every four children diagnosed in the present day, three are diagnosed wrongly. Given this, H^2 = 0.013

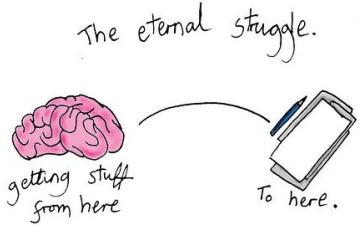
# Discussion

# Discussion

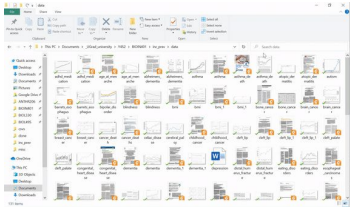
- Although this method does not measure strict “variance”, we are still measuring the fractional environmental phenotypic range
- By finding the smallest $H^2$ value possible for a phenotype and group of people, we’ll be able to make inferences about how much the environment can possibly contribute to a disease → influence treatment
- According to our hypothesis, twin studies (Falconer’s formula) overestimate heritability more than GWAS underestimates it

# Conclusions

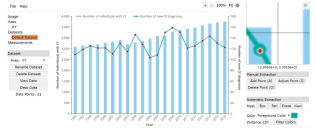
# What's left?




- What is left for this semester?
  - Final paper (Due 12:00 P.M. April 25)
- What is left for the overall project?
  - Polish manuscript to be publishable
  - Could add some more examples & fix current ones
  - Separate into the 3 different methods(?)
    - Do some more age and quantitative & estimates



# Sarah Stats



**263 papers read/processed**

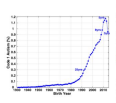
**38 pages written**

**106 twin coefficients found**

**55 longitudinal coefficients calculated**

**11 age coefficients calculated**

**265 lines of code**



**Autism**

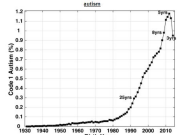
This will likely be one of the controversial topics we cover in this paper, namely because many, many people will argue that the increase in incidence is due to the shift in how autism is diagnosed. There are a few arguments as to why that is. First of all, autism was studied much better starting around the mid-20th century, therefore it became more well understood. Classic autism was first recognised in 1942 according to this source, so I need to look into how the source on the right has data from the 1930s. The data from this source are from the United States.

Most notably, the definition for autism was significantly revised in 1987, which broadened the concept of autism by adding a diagnosis at the mild end of the spectrum — pervasive developmental disorder-not otherwise specified (PDD-NOS) — and dropping the requirement for onset before 30 months. This can likely be cited as one of the causes for such a steep increase in the chart seen to the left. The DSM-IV, released in 1994 and revised in 2000, was the first edition to categorize autism as a spectrum.

There is a trend in epidemiology that goes as follows: Someone discovers a disease/disorder, then there is severe underdiagnosis because the information has not spread far enough yet, then there is a severe overdiagnosis, then afterwards there is a plateau. I would not be surprised if this is exactly what is happening here. In fact, near the most recent years, we can start to see a dip in incidence. I still believe that autism is largely an environmental phenomenon, backed up by such research as this, stating that there is a significant association between maternal immigration and the risk of having a child with an autism- spectrum disorder. It may be worthwhile to mention that autism can be caused by different things. There are certain subtypes of autism that are seemingly nearly entirely genetic. Examples of autism that have arisen from a new or de novo mutation in a single-gene or locus include the neurodevelopmental disorders fragile X syndrome, 22q13 deletion syndrome, and 16p11.2 deletion syndrome

To wrap this all up, here is a paper, stating rather clearly that "US Department of Education Data on "Autism" Are Not Reliable for Tracking Autism Prevalence". This is pretty unfortunate -- we might have to give up the lovely chart above. Here is a nature article talking about this phenomenon.

Perhaps data from the netherlands is better to rely upon?




https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5223814/

American autism prevalence data, presumably m=7, min and max method – min=0.091, max = 1.2 according to the paper.

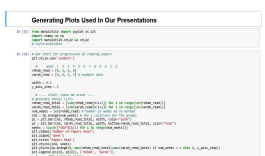
We set a correction term – that is, we assume for every child diagnosed at the minimum there were three children that were missed, and for every four children diagnosed in the present day, three are diagnosed wrongly. Gives this. $R^2 = 0.613$

| phenotype | gwas | twin | longitudinal | age |
|---|---|---|---|---|
| hepatitis B | 0.020 | 0.60 | 0.000 | -- |
| hip fracture | -- | 0.48 | 0.826 | 0.000 |
| mumps | -- | 0.39 | 0.000 | -- |
| opioid od death | 0.000 | 0.26 | 0.000 | -- |
| osteoporotic humeral fracture | -- | 0.32 | 0.006 | -- |
| peptic ulcers | -- | 0.37 | 0.000 | -- |
| poliomyelitis (polio) | -- | -- | 0.000 | -- |
| rubella | -- | 0.46 | 0.000 | -- |
| smallpox | -- | -- | 0.000 | -- |
| tuberculosis | 0.013 | 0.38 | 0.000 | -- |
| left handedness | 0.012 | 0.25 | 0.070 | 0.066 |

| phenotype | gwas | twin | longitudinal | age | quantitative |
|---|---|---|---|---|---|
| age at menarche | 0.320 | 0.61 | -- | -- | 0.66 |
| alzheimers | 0.033 | 0.79 | 0.672 | 0.020 | -- |
| anorexia | 0.009 | 0.58 | 0.886 | 0.018 | -- |
| asthma | 0.005 | 0.76 | 0.088 | -- | -- |
| atopic dermatitis | 0.003 | 0.85 | 0.000 | -- | -- |
| autism | 0.003 | 0.45 | 0.003 | -- | -- |
| bipolar disorder | 0.005 | 0.93 | 0.439 | -- | -- |
| blood pressure (diastolic) | 0.035 | 0.49 | -- | -- | 0.66 |
| blood pressure (systolic) | 0.035 | 0.45 | -- | -- | 0.51 |
| BMI (body mass index) | 0.027 | 0.85 | -- | -- | 0.16 |
| breast cancer | 0.012 | 0.31 | 0.538 | 0.052 | -- |
| celiac disease | 0.582 | 0.73 | 0.305 | -- | -- |
| cleft lip | 0.060 | 0.70 | 0.323 | -- | -- |
| crohn's disease | 0.260 | 0.75 | 0.340 | -- | -- |
| epilepsy | 0.006 | 0.65 | 0.078 | -- | -- |
| esophageal cancer | 0.009 | 0.25 | 0.077 | -- | -- |
| glaucoma | 0.006 | 0.13 | 0.563 | 0.073 | -- |
| graves disease | 0.030 | 0.57 | 0.030 | -- | -- |
| hay fever | 0.060 | 0.78 | 0.080 | -- | -- |

| phenotype | gwas | twin | longitudinal | age | quantitative |
|---|---|---|---|---|---|
| height | 0.200 | 0.84 | -- | -- | 0.21 |
| IQ | 0.036 | 0.68 | 0.180 | -- | -- |
| lung cancer | 0.005 | 0.18 | 0.060 | 0.016 | -- |
| macular degeneration | 0.012 | 0.65 | 0.018 | -- | -- |
| myocardial infarction | 0.020 | 0.62 | 0.050 | 0.086 | -- |
| myopia | 0.036 | 0.31 | 0.136 | -- | -- |
| obesity | 0.023 | 0.83 | 0.088 | 0.560 | -- |
| osteoporosis | 0.001 | 0.65 | 0.000 | -- | -- |
| preeclampsia | n/a | 0.54 | 0.420 | -- | -- |
| prostate cancer | 0.003 | 0.58 | 0.000 | -- | -- |
| psoriasis | 0.040 | 0.66 | 0.317 | -- | -- |
| rheumatoid arthritis incidence | 0.030 | 0.59 | 0.201 | -- | -- |
| scoliosis | 0.006 | 0.38 | 0.056 | -- | -- |
| testicular cancer | 0.020 | 0.58 | 0.567 | 0.121 | -- |
| thyroid cancer | 0.200 | 0.60 | 0.285 | -- | -- |
| type I diabetes | 0.020 | 0.88 | 0.218 | -- | -- |
| type II diabetes | 0.030 | 0.76 | 0.583 | 0.012 | -- |
| ulcerative colitis | 0.040 | 0.45 | 0.052 | -- | -- |
| venous thromboembolism | 0.040 | 0.44 | 0.052 | -- | -- |

| phenotype | gwas | twin | longitudinal |
|---|---|---|---|
| black hair vs non-black hair | 0.921 | 0.96 | 0.996 |
| cystic fibrosis | CFTR, monogenic | 0.80 | 0.997 |
| down's syndrome | trisomy 21 | trisomy 21 | 0.989 |
| hereditary Haemochromatosis (HH) | HFE gene (C282Y, H63D) | HFE gene (C282Y, H63D) | 0.960 |
| sex (male) | XY chromosomes | 1.00 | 0.997 |
| sickle cell disease | HBB gene mutation | HBB gene mutation | 0.887 |








# Thank you!