ASD Prevalence and Cases by State: Supplemental Methods Details

Introduction

Our goal is to estimate the number of ASD cases and prevalence for each of the 50 US states and District of Columbia (DC) and the total ASD cases and prevalence across all states for age 18-84. In addition, we will estimate the ASD cases and prevalence by sex (male, female) for each state and DC. We have the following available information:

1. US and state population stratified by age (one year age class) and sex (male, female)
2. Deaths by state stratified by age and sex
3. Mortality rate by state stratified by age and sex

Definitions

We define the following terms for developing our estimator of ASD cases and prevalence by state and sex.

- State = i ($i = 1, 2, 3, ..., 51$) (includes Washington DC)
- Age = j (j = 3-17, 18, 19, ..., 84)
- Sex = k ($k = 0$ (female), 1 (male))
- N = population
- γ = ASD cases
- $\rho = \gamma/N$ = prevalence

- δ = deaths
- P = probability of death
- $S = 1 - P$ = probability of survival

Estimating ASD Cases by State

Our procedure for estimating the number of ASD cases across all age classes and sex by state is by calculating the

1. i^{th} state age 3-17 ASD prevalence and cases stratified by the k^{th} sex ($\rho_{i(17)k}$ and $\gamma_{i(17)k}$, respectively) using a Bayesian hierarchical model, we assume exchangeability, a two step process
2. Age 18^{+} ASD prevalence and cases by one year age increment and sex, i.e., age class 18^{+} is defined as $i = 18, 19, 20, ..., 84$
3. ASD cases for the i^{th} state and k^{th} sex by summing over all age groups
4. total number of ASD cases for the i^{th} state by summing over the males and females ASD cases
5. i^{th} state's prevalence using the total number of cases and the state's 2017 population

First, we estimate the number of cases in the age group 3-17 by modeling using the National Survey of Children's Health (NSCH) 2016, 2017, and 2018 data. Let $\rho_{i(17)k}$ be the ASD prevalence for the i^{th} state and k^{th} sex for age group 3-17, where $\rho_{i(17)k} = \gamma_{i(17)k}/N_{i(17)k}$. Given the age 3-17 prevalence we use simulation to estimate, starting with age 18, the total ASD cases for the i^{th} state and k^{th} sex using $\gamma_{i(17)k} = \rho_{i(17)k}N_{i(17)k}$. Using the age group 3-17 ASD prevalence to begin our simulation we estimate the ASD cases, and prevalence, for each subsequent age class assuming a constant mortality ratio, by sex, between the ASD cases and the general population. We estimate the number of ASD cases in state i, age group j, and sex k for the 18+ age group using the following method.

Modeling Assumptions

We assume the ASD to general population mortality rate ratio is constant across the states and age groups. We assume the ASD prevalence and mortality rates vary among males and females. Hence, the mortality rate is held constant within a sex, which results in the ASD prevalence rate changing over time. The ASD prevalence rate will decrease as age increases as we assume the mortality rate in the ASD group is higher than the general population.

Estimating Age Group 18^+ ASD Cases

We use our estimated age 3-17 prevalence by state and sex to calculate the prevalence and ASD cases by age 18^+ class. Let $\rho_{i(17)k} = \rho_{i(18)k}$ if prevalence in the age group 18 equals the age group 3-17 prevalence. Furthermore, define the prevalence for state i, age class j ($j \geq 18$), and sex k using:

$$\rho_{ijk} = \frac{\gamma_{ijk}}{N_{ijk}} \tag{1}$$

Assuming prevalence for age group 18 equals the prevalence for age 3-17 within a state and sex results in:

$$\rho_{i(18)k} = \rho_{i(17)k} \tag{2}$$

We relax the assumption of a constant ASD age class prevalence for state i, age j, and sex k using the following method. Define the state, age, and sex mortality in the general population as:

$$P_{ijk} = \frac{\delta_{ijk}}{N_{ijk}} \tag{3}$$

Define the mortality ratio for ASD to the population as

$$P^R_{ijk} = c_{ik} = \frac{P^{ASD}_{ik}}{P_{ijk}} \tag{4}$$

Note that the probability of death is held constant across age class for ASD because we lack enough information to estimate the ASD mortality rate by age class. Equation (4) implies that if the $P^R_{ijk} = 1.0$ then the adjustment factor for the number of cases for each age group is not adjusted for mortality beyond the population mortality rate. However, we expect a $P^R_{ijk} > 1.0$, from prior research, and as a result we adjust the expected number of cases by the P^R_{ijk} using the following method. Mortality rate in the population is expressed as P_{ijk}. In the absence of mortality we can express the adjusted (adjusted for no mortality) general population total as

$$N^{NM}_{ijk} = \frac{N_{ijk}}{1 - P_{ijk}} \tag{5}$$

Given the adjusted population (N^{NM}_{ijk}) we can estimate the number of ASD cases in the absence of mortality as:

$$\gamma^{NM}_{ijk} = N^{NM}_{ijk}\rho_{i(j-1)k} \tag{6}$$

Note that at this point we use the prior age class j estimate of the prevalence, starting with age 18 because age 3-17 prevalence is estimated from the model, and adjust this prevalence based on the mortality rate below. The adjusted number of ASD cases based on the ASD mortality rate for age class 18^+ is expressed as:

$$\gamma^{adj}_{ijk} = \gamma^{NM}_{ijk}(1 - P^{ASD}_{ik}) \tag{7}$$

Using substitution and given equations (5), (6), and (7) we can express the estimated number of cases more succinctly as

$$\gamma^{adj}_{ijk} = N_{ijk}\rho_{i(j-1)k}\frac{S^{ASD}_{ik}}{S^{POP}_{ijk}} \tag{8}$$

Note the adjusted prevalence is the back two terms of equation (8), i.e.,:

$$\rho^{adj}_{ijk} = \frac{\gamma^{adj}_{ijk}}{N_{ijk}} = \frac{1}{N_{ijk}}N_{ijk}\rho_{i(j-1)k}\frac{S^{ASD}_{ik}}{S^{POP}_{ijk}} = \rho_{i(j-1)k}\frac{S^{ASD}_{ik}}{S^{POP}_{ijk}} \tag{9}$$

Hence, our final prevalence estimator only depends upon the prior year's ASD prevalence, the ASD survival rate (which is constant across age classes), and the population survival rate (which varies by age class). Given this estimated prevalence we can estimate the number of ASD cases using the population for a given age class. Note that only the initial prevalence for age 3-17 estimated from the model and the ASD survival rate estimated from prior studies are random, all other terms are known quantities.

Given equation (8) we can estimate the number of ASD cases in state i over all age classes and sex k as

$$\gamma_i^{total} = \gamma_{i(17)(0)} + \gamma_{i(17)(1)} + \sum_{j=18}^{n} \gamma_{ij(0)}^{adj} + \sum_{j=18}^{n} \gamma_{ij(1)}^{adj} \tag{10}$$

$$\gamma_i^{total} = \sum_{k=0}^{1} \gamma_{i(17)k} + \sum_{j=18}^{n} \sum_{k=0}^{1} \gamma_{ijk}^{adj} \tag{11}$$

and the i^{th} state, weighted, prevalence is expressed as

$$\rho_i = \frac{\gamma_i^{total}}{N_i} \tag{12}$$

Note we can now obtain the US total ASD cases and prevalence as:

$$\rho = \sum_{i=1}^{51} \frac{\gamma_i^{total}}{N_i} \frac{N_i}{N} = \frac{1}{N} \sum_{i=1}^{51} \gamma_i^{total} \tag{13}$$

Algorithm for Estimating ASD Cases by State

Our estimators for the state and national level ASD cases and prevalences (equations 8 and 9) have two sources of uncertainty as we assume the population, population level deaths, and mortality rates are known quantities. Our source of variability are the estimated ASD prevalence for age 3-17 and ASD to population mortality rate. We account for these sources of uncertainty in our simulation. Our algorithm, assuming we will estimate the prevalence for age 3-84, proceeds as:

- We estimate the state level ASD age 3-17 mean prevalence and standard error of the prevalence on the logit scale by sex using a logistic model and R survey package and NSCH data for years 2016-2018
- Use the prevalence and SE to fit a hierarchical Bayesian model and obtain the partially-pooled β's and SE by state and sex assuming exchangeability and a two step process

- Obtain the population by state and sex for age 3-17 and each subsequent yearly age group to age 84
- Obtain the deaths by state and sex for age 3-17 and each subsequent yearly age group
- Group yearly age classes into 5 year age classes by state and sex to obtain mortality rate by state, sex, and age class
- Assume mortality rate is constant within each 5 year age group
- Draw $N(\beta, \sigma)$ from the estimated Bayesian hierarchical model and estimate the age 3-17 prevalence by state and sex
- Draw from the Bayesian hierarchical meta-analysis SMR and SE to obtain ASD mortality rate
- Use equations 8 and 9 to obtain the ASD cases and prevalence by state, sex, and age class
- Sum over all age classes and sex by state to obtain the state level ASD cases and prevalence (equations 11 and 12)
- Obtain the US prevalence and ASD cases using equation 13
- Repeat 10,000 times to obtain a mean estimate of the ASD cases and prevalence by state and overall and the 95% simulation interval bounds

Prevalence Model

The age 3-17 prevalence modeling was a two-step process. We used the NSCH data by state and sex to obtain prevalence estimates and the SE of the estimates using R survey package and incorporated the complex sampling design weights into the estimation. We then used the age 3-17 prevalence beta estimates and SE by sex and state in a Bayesian hierarchical model with the states treated as the random effect. We assumed the beta estimates were distributed normally and we fixed the variance to the estimated SE for the states. We assumed a vague prior for the pooled mean of N(0, precision=1e-6) and hyperprior for the variance using a t-distribution with four degrees of freedom and variance of 100^2. We used 10,000 samples as the burn-in, 100,000 samples after burn-in with thinning set to 10 for a total of 10,000 samples to summarize the posterior distributions. Using the posterior distribution we estimated the mean and standard error by sex and used those values in our simulation. We estimated the pooled posterior distribution for the beta and associated SE. We backed transformed to estimate the overall prevalence posterior distribution by sex and used the estimated beta and SE in our simulation.

Standardized Mortality Ratio Model

We estimated the standardized mortality ratio (SMR) by sex using the following meta-analysis method. For females and males we used five studies (Mouridsen, two studies by Pickett, Hirvikoski, Gillberg). Each study provided the observed and expected number of deaths. We used a Bayesian hierarchical Poisson model with the observed mortality as the outcome and ln(expected mortality) as the offset. We used a vague prior, N(0, precision=1e-6), for the mean pooled estimate and hyperprior for the variance using a t-distribution with four degrees of freedom and variance of 100^2. We used 10,000 samples as the burn-in, 50,000 samples after burn-in with thinning set to five for a total of 10,000 samples to summarize the posterior distributions. Using the posterior distribution of the pooled SMR we estimated the mean and standard error by sex and used those values in our simulation.

Prevalence Simulation

Our estimation of the overall prevalence by state and by sex and state for age classes 18-84 was conducted using simulation and the following process. We obtain the mortality by state and sex for age 3-17, and 18 to 84 by year. The file contains the number of deaths and population by strata (state, sex, and age). Our prevalence estimates are based on the year 2017 population so we obtained the population data by state, sex, and age strata. Next we obtain the estimated age 3-17 meta-analysis prevalence, on the β scale and associated SE by state and sex. We run 10,000 simulations using the following method. We randomly draw from a normal distribution using our estimated mean and SE meta-analysis estimates by sex for the SMR. We select a randomly drawn SMR by sex for each simulation, 10,000 drawn SMRs. Next we randomly draw 10,000 prevalence estimates using our meta-analysis estimates by state and sex. Lastly, we predict the prevalence by state, sex, and age using our equation 9. Our simulation results in 10,000 estimates for the prevalence and ASD cases for age class 18-84 by state and sex and we summarize these results using the mean and 95% simulation interval (SI). In addition we calculated the the prevalence ratio (PR) and prevalence difference (PD) for males versus females by state and overall and summarize using the mean and 95% SI.