Supplemental Material

Description of the Models Used to Estimate Congenital Heart Defect Prevalence and Case Counts for the United States in 2010

Overview of Modeling Approach

The goal of this modeling effort was to produce estimates for the prevalence, and associated case counts, of congenital heart defects (CHDs) in the 2010 U.S. population by sex and race/ethnicity within the following age classes: 0-5 years, 6-12 years, 13-17 years, 18-24 years, 25-44 years and 45 years and older. Prevalence and case count estimates are developed both for all CHDs and for those CHDs classified as severe. As with any approach based on modeling, the estimates presented in this paper are derived under a number of assumptions. A list of the specific assumptions used to derive the estimates presented here is provided in Supplemental Table 1. A primary assumption underlying our approach is that the sex- and age class-specific CHD prevalence rates among U.S. Non-Hispanic Whites (NHW) during 2010 are identical to CHD prevalence rates observed between 2000 and 2010 among corresponding sex and age groups in Québec, Canada. An analysis of these Québec CHD data has been reported by Marelli et al.[1] and the specific Québec data used to develop the estimates presented here are listed in Supplemental Table 2.

To illustrate the modeling approach, let the observed Québec CHD prevalence rate for males in age class i be designated by $PR_i^{Q\_M}$ and the corresponding Québec CHD prevalence among females by $PR_i^{Q\_F}$. The assumption on the equivalence of the Québec rates with those in the U.S. NHW population implies that

$$PR_i^{NHW\_M} = PR_i^{Q\_M} \qquad [S1]$$

and

$$PR_i^{NHW\_F} = PR_i^{Q\_F} \quad [S2]$$

where $PR_i^{NHW\_M}$ and $PR_i^{NHW\_F}$ are the estimated 2010 CHD prevalence rates in age class i among U.S. NHW males and females respectively. Sex-specific CHD prevalence rate estimates in a given age class among U.S. Non-Hispanic Black (NHB) and Hispanic populations were calculated by multiplying the appropriate NHW prevalence estimate, equations S1 or S2 depending on sex, by a race/ethnicity adjustment factor. As will be shown, this factor corresponds to the ratio of survival probabilities through age 5 between those born with a CHD and the general U.S. population for the NHB or Hispanic populations to the corresponding ratio in the NHW population. The rationale for this estimator is given in the next section. However, to illustrate the approach, let the mortality rate among persons with CHD for ages 0 to 5 be designated as $M_{0-5}^{NHB\_M\_CHD}$ for NHB males and $M_{0-5}^{NHW\_M\_CHD}$ for NHW males. Similarly, let the 0 to 5 mortality rate among the entire U.S. population be given by $M_{0-5}^{NHB\_M}$ and by $M_{0-5}^{NHW\_M}$ for NHB males and NHW males respectively. If one assumes that there are no race/ethnicity differences in the relative survival probability between those born with a CHD and the general U.S. population among those older than age 5, then an estimator for the age class i CHD prevalence rate among NHB males, $PR_i^{NHB\_M}$, is given by

$$PR_i^{NHB\_M} = PR_i^{NHW\_M} * R_{NHB\_M} \quad [S3]$$

where

$$R_{NHB\_M} = \frac{(1-M_{0-5}^{NHB\_M\_CHD})/(1-M_{0-5}^{NHB\_M})}{(1-M_{0-5}^{NHW\_M\_CHD})/(1-M_{0-5}^{NHW\_M})}$$

In equation S3, R_{NHB_M} is the race/ethnicity adjustment factor used to derive the age class i CHD prevalence rate estimate for NHB males based on the corresponding age class specific NHW male prevalence estimate. In addition, if

$$S_{0-5}^{NHB\_M\_CHD} = 1 - M_{0-5}^{NHB\_M\_CHD}$$

is defined as the survival probability though age 5 for NHB males born with a CHD with $S_{0-5}^{NHW\_M}$, $S_{0-5}^{NHW\_M\_CHD}$ and $S_{0-5}^{NHW\_M}$ defined similarly, then the NHB male race/ethnicity adjustment factor can be written as

$$R_{NHB\_M} = \frac{S_{0-5}^{NHB\_M\_CHD} \Big/ S_{0-5}^{NHB\_M}}{S_{0-5}^{NHW\_M\_CHD} \Big/ S_{0-5}^{NHW\_M}} \qquad [S4]$$

Equation S4 illustrates that the suggested race/ethnicity adjustment factor corresponds to the ratio of the probability of survival though age 5 among NHB males with CHD to the general 0 to 5 NHB male population divided by a similar ratio comparing survival probability though age 5 among NHW males. A similar argument leads to estimating the CHD prevalence within the NHB female (NHB_F) population in age class i, $PR_i^{NHB\_F}$, as

$$PR_i^{NHB\_F} = PR_i^{NHW\_F} * R_{NHB\_F} \qquad [S5]$$

with

$$R_{NHB\_F} = \frac{S_{0-5}^{NHB\_F\_CHD} \Big/ S_{0-5}^{NHB\_F}}{S_{0-5}^{NHW\_F\_CHD} \Big/ S_{0-5}^{NHW\_F}}$$

where $S_{0-5}^{NHB\_F\_CHD}$ is the survival probability through age 5 for NHB females born with a CHD, $S_{0-5}^{NHB\_F}$ is the survival probability through age 5 among the general U.S. NHB female population

and $S_{0-5}^{NHW\_F\_CHD}$ and $S_{0-5}^{NHW\_F}$ are similar survival probabilities among the U.S. NHW female population.

Prevalence among Hispanic males ($HISP\_M$) and Hispanic females ($HISP\_F$) within age class i was estimated as

$$PR_i^{HISP\_M} = PR_i^{NHW\_M} * R_{HISP\_M} \quad \text{[S6]}$$

with

$$R_{HISP\_M} = \frac{S_{0-5}^{HISP\_M\_CHD} / S_{0-5}^{HISP\_M}}{S_{0-5}^{NHW\_M\_CHD} / S_{0-5}^{NHW\_M}}$$

and

$$PR_i^{HISP\_F} = PR_i^{NHW\_F} * R_{HISP\_F} \quad \text{[S7]}$$

where

$$R_{HISP\_F} = \frac{S_{0-5}^{HISP\_F\_CHD} / S_{0-5}^{HISP\_F}}{S_{0-5}^{NHW\_F\_CHD} / S_{0-5}^{NHW\_F}}$$

Sex- and race/ethnicity-specific estimates for the survival probabilities in the general U.S. population used in equations S4 through S7, are based on the average annual number of deaths and population counts in this age class obtained from the Centers for Disease Control and Prevention's Detailed Mortality data file for the years 1999 through 2014.[2] The specific data extracted from these mortality records that were used as inputs to the modeling process are listed in Supplemental Table 3. Sex- and race-specific estimates for survival through age 5

among those born with a CHD were based on data collected from 1999 through 2007 by the Texas Birth Defects Registry. An analysis of these data has been published by Nembhard et al.[3] The specific data on the 0 to 5 survival probabilities among children born with a CHD used as inputs to the CHD prevalence estimation models presented here are listed in Supplemental Table 4.

Estimates of CHD case counts within age class, sex and race/ethnicity strata were obtained by multiplying the estimated strata-specific CHD prevalence estimates by the corresponding July 1, 2010 U.S. Census Bureau intercensal population estimate for that strata.[4]

An identical approach was used to estimate 2010 U.S. prevalence rates and case counts for severe CHDs with the exception that the outcome was limited to those types of CHDs classified in the severe category. The CHDs classified as severe include tetralogy of Fallot, truncus arteriosus, complete and congenitally corrected transposition of the great arteries, atrioventricular septal defect and univentricular hearts.

Derivation of Race/Ethnicity Prevalence Adjustment Factors for Estimation NHB and Hispanic CHD Prevalence Rates

The rational for the race/ethnicity adjustment factors given in equations S4 through S7 will be illustrated using estimation of NHB male prevalence within a given age class. Let the random variable CHD be defined so that CHD = 1 if a child is born with a CHD and CHD = 0 otherwise. Using the notation that $P[\,A\,|B,C]$ is the conditional probability of event A given events B and C have occurred, we define the CHD prevalence in age class i among NHB males as $P[CHD = 1 \mid S_i\,, NHB\_M]$ where S_i is the event that the person has survived long

enough to be in age class i and NHB_M designates that this probability corresponds to a person in the NHB male population. Similarly, $P[CHD = 1 \mid S_i, NHW\_M]$ is the CHD prevalence in age class i among NHW males. Note that

$$P[CHD = 1 \mid S_i, NHB\_M] = PR_i^{NHB\_M}$$

and

$$P[CHD = 1 \mid S_i, NHW\_M] = PR_i^{NHW\_M}$$

as defined in equations S1 and S3, but we use the conditional probability definitions of these terms here to provide a rational for the adjustment used to derive $PR_i^{NHB\_M}$ from $PR_i^{NHW\_M}$.

To obtain a prevalence estimate among NHB males in age class i, the estimated prevalence among NHW males in that age class is multiplied by the adjustment factor

$$R_{NHB\_M} = \frac{P[CHD=1 \mid S_i, NHB\_M]}{P[CHD=1 \mid S_i, NHW\_M]} \quad .$$

By the laws of conditional probability,

$$P[CHD = 1 \mid S_i, NHB\_M] = \frac{P[CHD=1 \mid NHB\_M] * P[S_i \mid CHD=1, NHB\_M]}{P[S_i \mid NHB\_M]}$$

and

$$P[CHD = 1 \mid S_i, NHW\_M] = \frac{P[CHD=1 \mid NHW\_M] * P[S_i \mid CHD=1, NHW\_M]}{P[S_i \mid NHW\_M]} \quad .$$

Therefore, the NHB male race/ethnicity adjustment factor can be written as

$$R_{NHB\_M} = \left(\frac{P[CHD=1 \mid NHB\_M]}{P[CHD=1 \mid NHW\_M]}\right) * \left(\frac{P[S_i \mid CHD=1, NHB\_M] / P[S_i \mid NHB\_M]}{P[S_i \mid CHD=1, NHW\_M] / P[S_i \mid NHW\_M]}\right) \quad .$$

If we now assume, see Supplemental Table 1, that there are no race/ethnicity differences in the birth prevalence of CHD, that is, we assume

$$P[CHD = 1 \mid NHB\_M] \;=\; P[CHD = 1 \mid NHW\_M] \qquad ,$$

then R_{NHB_M} can be rewritten as

$$R_{NHB\_M} \;=\; \frac{P[S_i \mid CHD{=}1,\ NHB\_M] \,/\, P[S_i \mid NHB\_M]}{P[S_i \mid CHD{=}1,\ NHW\_M] \,/\, P[S_i \mid NHW\_M]} \quad . \qquad \text{[S8]}$$

We now define the term $SR_i^{NHB\_M}$ as the ratio of the survival probability through age class i for NHB males born with a CHD to the corresponding survival probability in the general U.S. NHB male population. If $SR_i^{NHW\_M}$ is the corresponding ratio of survival probabilities among NHW males, then the adjustment factor for age class i given in equation S8 can be written using these newly defined terms as

$$R_{NHB\_M} = \frac{SR_{0-5}^{NHB\_M}}{SR_{0-5}^{NHW\_M}} \;*\; \prod_{j=6-12}^{i} \frac{SR_j^{NHB\_M}}{SR_j^{NHW\_M}} \qquad . \qquad \text{[S9]}$$

We assume, see Supplemental Table 1, that within sex and for all ages greater than 5 years, there are no race/ethnicity differences in the ratios of survival probabilities for those born with a CHD and the general U.S. population. In other words, we assume that, for those older than 5 years of age, the impact on survival probability due to having a CHD within a given race/ethnicity group relative to those in the general population of the same race/ethnicity is equal for NHWs, NHBs and Hispanics. In terms of equation S9, this assumption implies that

$$\frac{SR_i^{NHB\_M}}{SR_i^{NHW\_M}} \;=\; 1$$

for all i corresponding to age classes other than age class 0 to 5. Under this assumption, R_{NHB_M} reduces to

$$R_{NHB\_M} = \frac{SR_{0-5}^{NHB\_M}}{SR_{0-5}^{NHW\_M}}$$

or

$$R_{NHB\_M} = \frac{P[S_{0-5} \mid CHD{=}1,\ NHB\_M] \Big/ P[S_{i0-5} \mid NHB\_M]}{P[S_{0-5} \mid CHD{=}1,\ NHW\_M] \Big/ P[S_{0-5} \mid NHW\_M]}$$

$$= \frac{S_{0-5}^{NHB\_M\_CHD} \Big/ S_{0-5}^{NHB\_M}}{S_{0-5}^{NHW\_M\_CHD} \Big/ S_{0-5}^{NHW\_M}}$$

which equals the NHB male adjustment factor used in the estimator given in equation S4

An identical argument as that given above for NHB males can be used to justify the suggested race/ethnicity adjustment factors for the NHB female and Hispanic male and female CHD prevalence estimators given in equations S5 through S7.

Estimation of Sampling Error for the CHD Prevalence Rate and Case Count Estimates

With the exception of the census-based population counts, all inputs to the models were considered to be realizations of underlying random variables and, therefore, subject to sampling variability. We used a combination of approaches to propagate the sampling variability of the model inputs through to the final prevalence and case count estimates. Following recommendations from the National Center on Health Statistics[5], standard errors for the general U.S. population survival probabilities through age 5 were estimated under the assumption that the observed number of deaths in this age group is a sample from a Poisson distribution. Standard errors for the survival probability through age 5 for persons born with a CHD were also estimated using the Poisson assumption. However, in this case both the CHD

case counts and the number of deaths among persons with a CHD, found in Supplemental Table 4, were considered to be Poisson random variables. The standard error of one minus the ratio of deaths to CHD case counts, that is the desired survival probability through age 5 for persons born with a CHD, was then estimated using a Taylor Series approximation[6].

Given the estimated standard errors for the components of the race/ethnicity adjustment factors, that is survival probabilities though age 5 for those with a CHD and for the general population, estimated standard errors for the race/ethnicity adjustment factors given in equations S4 through S7 were derived again via Taylor Series approximation. Simulation results indicated that the sampling distribution for the race/ethnicity adjustment factors was well approximated by the normal distribution. As a result, the sampling distributions for the adjustment factors were assumed to be normal with a mean given by the estimated values for the factors based on the observed survival probabilities and variance equal to the square of the approximated standard errors.

Standard error estimates for the reported Québec CHD prevalence rates given in Supplemental Table 2 were also derived under a Poisson assumption on the reported case counts. The sampling variability associated with the Québec CHD prevalence rates was assumed to be normal with mean given by the observed rates for each age and sex group and variance given by the square of the estimated standard error.

Given the assumed sampling distributions for both the Québec prevalence rates and the race/ethnicity adjustment factors, we used a Monte Carlo process to estimate the sampling variability associated with the final U.S. CHD prevalence and case counts estimates. The Monte Carlo algorithm was comprised of the following steps

1. Sample a collection of CHD NHW prevalence rates from the assumed sampling distributions for the Québec rates within each sex and age class strata.
2. Sample a collection of race/ethnicity adjustment factors from the assumed normal sampling distribution for these variables.
3. Derive the estimated CHD prevalence for the U.S. population by multiplying the estimated NHW prevalence from Step 1 by the appropriate race/ethnicity adjustment factor sampled in Step 2.
4. Estimated CHD case counts by multiplying the estimated prevalence rates by the appropriate population size within age class, sex and race/ethnicity strata.
5. Store the resulting prevalence and case count estimates.
6. Repeat Steps 1-5 10,000 times.

At the conclusion of the Monte Carlo sampling routine, we had 10,000 estimates of both the CHD prevalence rates and the associated case counts within each age class, sex and race/ethnicity stratum. This collection of estimates was viewed as an empirical estimator for the sampling distributions for these values. Final estimates for the prevalence and case counts were defined as the mean of the Monte Carlo generated sampling distributions with 95% confidence intervals estimated by the range separating the 2.5^{th} and 97.5^{th} percentiles of the distribution of generated values.

Use of the Monte Carlo approach enabled estimation of the sampling variability of estimates reflecting combinations of results across age class, sex and race/ethnicity strata, for example, the estimated prevalence rates and case counts presented in manuscript Tables 1 and 2. To derive these estimates, generated case counts and populations were summed across appropriate strata to obtain the desired collapsed strata estimates. Prevalence within these collapsed strata was then estimated as the summed case count divided by the sum of

the appropriate population counts. This approach resulted in 10,000 values for the desired summary estimates. For example, the prevalence and case count estimates by child and adult categories and across all ages presented in manuscript Table 1 were generated by addition of estimated CHD case and population counts across appropriate strata. Again, the presented estimate is the mean of the 10,000 samples and 95% confidence intervals are defined by the 2.5^{th} and 97.5^{th} percentiles of the generated values.

References

1. Marelli AJ, Ionescu-Ittu R, Mackie AS, Guo L, Dendukuri N, Kaouache M. Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010. *Circulation.* 2014;130:749-756.
2. United States Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death, 1999-2014 on CDC WONDER Online Database, released 2015. Accessed at http://wonder.cdc.gov/ucd-icd10.html; Access date 29 January 2016
3. Nembhard WN, Xu P, Ethen MK, Fixler DE, Salemi JL, Canfield MA. Racial/ethnic disparities in timing of death during childhood among children with congenital heart defects. *Birth Defects Res A Clin Mol Teratol.* 2013;97:628-643.
4. United States Department of Health and Human Services (US DHHS), Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), Bridged-Race Population Estimates, United States July 1st resident population by state, county, age, sex, bridged-race, and Hispanic origin, on CDC WONDER On-line Database. Accessed at http://www.census.gov/popest/data/intercensal/national/files/US-EST00INT-ALLDATA.pdf; Access date 29 January 2016
5. National Center for Health Statistics. Vital Statistics of the United States; Mortality, 1999. Technical Appendix. Hyattsville, Maryland. 2004.
6. Papanicolaou, A. 2009. Taylor approximation and the delta method. Accessed at http://web.stanford.edu/class/cme308/OldWebsite/notes/TaylorAppDeltaMethod.pdf; Access date 27 April 2016

Supplemental Table 1: Assumptions Underlying the Approach Used to Estimate 2010 U.S. Congenital Heart Defect (CHD) Prevalence and Case Counts

Assumption	
1	Observed age- and sex-specific CHD prevalence from 2000 to 2010 among residents of Québec, Canada is identical to age- and sex-specific CHD prevalence in the U.S. Non-Hispanic white population during 2010.
2	No race/ethnicity differences in CHD birth prevalence in the U.S. population.
3	For those older than 5 years, there are no race/ethnicity differences in the ratio of the probability of survival to a given age among those born with a CHD to that in the general U.S. population.
4	The probabilities of survival through age 5 for persons born with a CHD observed in the Texas Birth Defects Registry for the years 1999 through 2007 are generalizable to the entire U.S. population born with a CHD over the time frame addressed in this study.
5	While survival probabilities through age 5 for persons born with a CHD differ by race/ethnicity, these probabilities do not differ by sex.
6*	Observed CHD case counts in the Québec population are random variables following a Poisson distribution. The number of deaths through age 5 in the general U.S. population is also assumed to follow a Poisson distribution. Both the number of CHD cases and the number of deaths among those cases through age 5 observed in follow-up data from the Texas Birth Defects Registry are assumed to follow Poisson distributions.

* This set of distributional assumptions was used to estimate the sampling variability associated with the CHD prevalence and case count estimates.

Supplemental Table 2: Number of Congenital Heart Defect (CHD) Cases, Population Counts and CHD Prevalence Rates by Sex, Age Class and CHD Severity Type Québec, Canada, 2000 to 2010.

Age Group	CHD Type	CHD Cases	Population Count	CHD Prevalence*	Standard Error
			Males		
0 to 5	All	2,881	247,728	11.63	0.22
	Severe	275	247,728	1.11	0.07
6 to 12	All	4,204	267,950	15.69	0.24
	Severe	467	267,950	1.74	0.08
13 to 17	All	2,907	228,591	12.72	0.24
	Severe	442	228,591	1.93	0.09
18 to 24	All	2,213	345,974	6.4	0.14
	Severe	473	345,974	1.37	0.06
24 to 44	All	6,300	1,041,269	6.05	0.08
	Severe	1,206	1,041,269	1.16	0.03
45 +	All	12,607	1,853,946	6.8	0.06
	Severe	770	1,853,946	0.42	0.01
			Females		
0 to 5	All	3,670	260,849	14.07	0.23
	Severe	391	260,849	1.5	0.08
6 to 12	All	4,072	282,016	14.44	0.23
	Severe	584	282,016	2.07	0.09
13 to 17	All	2,455	239,284	10.26	0.21
	Severe	458	239,284	1.91	0.09
18 to 24	All	1,836	360,844	5.09	0.12
	Severe	413	360,844	1.14	0.06
24 to 44	All	4,206	1,090,816	3.86	0.06
	Severe	856	1,090,816	0.78	0.03
45 +	All	12,973	1,686,599	7.69	0.07
	Severe	566	1,686,599	0.34	0.01

Source: Adapted from Marelli et al. (2014)[1]

* Prevalence per 1,000 persons

Supplemental Table 3: Average Annual Number of Deaths, Population Counts and Survival Probabilities for the U.S. Population age 0 to 5 Years by Sex and Race/Ethnicity, 1999 through 2014.

Race/Ethnicity	Average Annual Deaths	Average Annual Population	Survival Probability	Standard Error
Males				
Non-Hispanic White	9,238	7,522,663	0.999	$3.6*10^{-4}$
Non-Hispanic Black	5,081	1,827,241	0.997	$7.4*10^{-4}$
Hispanic	3,436	2,814,064	0.999	$6.0*10^{-4}$
Females				
Non-Hispanic White	7,095	7,169,659	0.999	$3.7*10^{-4}$
Non-Hispanic Black	4,023	1,768,393	0.997	$7.5*10^{-4}$
Hispanic	2,769	2,699,487	0.999	$6.1*10^{-4}$

Source: National Center for Health Statistics (Detailed Mortality File, 1999-2014)[2]

Supplemental Table 4: Observed Congenital Heart Defect (CHD) Case Counts, Number of Deaths and Survival Probabilities Among Persons Ages 0 to 5 from the Texas Birth Defects Registry, 1999 to 2007 by Race/Ethnicity and CHD Severity.

Race/Ethnicity	CHD Type	CHD Cases	Deaths	Survival Probability	Standard Error
Non-Hispanic White	All	16,980	1,474	0.91	$0.24*10^{-2}$
	Severe	1,944	461	0.76	$1.23*10^{-2}$
Non-Hispanic Black	All	5,020	646	0.87	$0.54*10^{-2}$
	Severe	557	174	0.69	$2.71*10^{-2}$
Hispanic	All	24,854	2,278	0.91	$0.20*10^{-2}$
	Severe	2,308	692	0.70	$1.30*10^{-2}$

Source: See Nembhard et al. (2013)[3]