Association of type 1 diabetes with month of birth among US youth: The SEARCH for Diabetes in Youth Study

Online-Only Appendix

Item 1: Elimination of biases related to registration protocols

The SEARCH database includes prevalent cases of diabetes diagnosed during calendar year 2001 and incident cases diagnosed between the beginning of 2002 and the end of 2006. All cases had to be less than 20 at the end of the year in which they were diagnosed. For this analysis, we excluded SEARCH cases that were diagnosed after their birthday in 2006 so that all participants had an equal chance of being diagnosed by the end of 2006 (and all birth months would be represented with equal likelihood). For example, to be eligible for SEARCH in conformity with the cutoff date of 31 December, 2006, a child born on 1 January, 1996, would have had 132 months in which to be diagnosed with diabetes, while a child born on 1 December, 1996, would have had only 121 months to be diagnosed. We also excluded those who were diagnosed at age 19 years, since our protocol denied registration to any participants who became 20 years old by the end of the calendar year in which they were diagnosed. Thus, those whom we had registered with a diagnosis at age 19 years were more likely to have been born earlier in a year.

Item II: Sources of US birth information and methods for computing monthly ratios

Birth numbers were obtained from the *Vital Statistics of the United States* (1) for years 1982–1993, *Monthly Vital Statistics Reports* (2) for 1994–1996, and *National Vital Statistics Reports* (3) for 1997–2005. These sources provided national monthly birth numbers for the total population, as well as monthly birth numbers tabulated separately for whites and blacks. During 1982–1988, race was defined by the race of the child. Subsequently, race has been defined by the race of the mother. For each month, national birth numbers were calculated for "other" racial groups by subtracting the sum for whites and blacks from the total. From the national birth numbers we calculated the race-specific proportion of births for a given month and year as the total number of children of a specific race born in the given month and year divided by the total number of children of that race born in the given year.

The birth-month distribution for any particular subset of the SEARCH population (e.g., sex, race, birth cohort) was obtained by summing the number of births for a given month across the time period of interest and then dividing by the total number of births across the time period of interest. The expected number of births was obtained by taking the mean of the US proportions for the particular month of interest for the same subset of the SEARCH population. Each month's ratio of the observed to expected proportion of births is a measure of the excess (or deficit) in the number of births relative to what would be expected if SEARCH births were distributed monthly as they were distributed among all US infants of the same racial group(s) and birth year(s). These ratios would be one if there were no excess or deficit in a given month. For ease of interpretation, we subtracted one from the ratios and multiplied the results by 100 to get the percentage excess (or deficit) with reference to the US experience.

Item III: Data-smoothing process – methods and examples

For each month, we calculated a smoothed estimate from the excess (deficit) values using a kernel smoother that is a weighted average of the observed month with those within ± three months. The particular weights used were from a binomial smoother of power 6 (4) that use weights: $\frac{1}{64}$ (1, 6, 15, 20, 15, 6, 1) over a 12-month rotating span, i.e.:

$$\text{Smoothed } y_m = \frac{1}{64}(y_{m-3}+6y_{m-2}+15y_{m-1}+20y_m+15y_{m+1}+6y_{m+2}+y_{m+3}).$$

Note that since months of a year are cyclical, months 13, 14, 15, -2, -1, and 0 are equivalent to months 1, 2, 3, 10, 11, and 12, respectively.

The smoothing function we used can also be written as:

$Y_{SM} = AY$, where

$$A = \frac{1}{64}\begin{pmatrix} 20\ 15\ 6\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 6\ 15 \\ 15\ 20\ 15\ 6\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 6 \\ 6\ 15\ 20\ 15\ 6\ 1\ 0\ 0\ 0\ 0\ 0\ 1 \\ 1\ 6\ 15\ 20\ 15\ 6\ 1\ 0\ 0\ 0\ 0\ 0 \\ . \\ . \\ 15\ 6\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 6\ 15\ 20 \end{pmatrix},\quad Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ . \\ . \\ y_{12} \end{pmatrix}, \text{ and}$$

Y is the vector of estimated excesses or deficits. Since the y_i's come from a multinomial distribution, the variance matrix for Y is given by:

$\sum_Y = (I P – P P')/N$, where I is the 12x12 identity matrix, P is the 12x1 vector of p_i's, and N is the sample size in the subset of interest.

The variance matrix of the smoothed estimates is given by $\sum_{SM} = A\sum_Y A'$.

The diagonal elements of $\sum_{SM}$, the variances of the smoothed y_i's, are smaller than the diagonal elements of $\sum_Y$. To test the hypothesis that the proportion of diabetic births per month equals the proportion of US births per month, we used the $\chi 2$ test statistic:

$\chi 2 = (Y - P)' \sum^{-}_{Y} (Y - P)$,

where $\sum^{-}_{Y}$ is the generalized inverse matrix of $\sum_Y$. You get the same $\chi 2$ test statistic value whether you base it on the observed Y's or the smoothed estimate Y_{SM} since:

$\chi 2 = Y'_{SM} \sum^{-}_{SM} Y_{SM} = Y'A'A'^{-1}\sum^{-}_{Y}A^{-1}AY = Y'\sum^{-}_{Y}Y$.

The same smoothing process was replicated for analyses stratified by sex, racial groups, age at diabetes diagnosis, birth cohorts, and the geographic regions in which SEARCH participants resided. For each smoothed curve we identified the birth month with greatest diabetes excess (peak) and the month with least diabetes excess (nadir), and we computed the relative risk of diabetes associated with this contrast between the identified peak and nadir birth months (RR(max) on Figures 1–4 and A1-A3).

As an example, consider the SEARCH youth with type 1 diabetes. The numbers of these participants born in each month are shown in the Table below. The observed proportions (p_i) is simply calculated as the observed number of births in the month divided by the total number of births (N=9737). The SE's of the proportions, SE(p_i), were calculated as the square root of $p_i*(1-p_i)/N$. The expected proportion, e_i, is the mean of the race and year specific US monthly proportions. The ratio is simply the observed divided by the expected proportions, and the percent excess is 100x(1-ratio). These are the values plotted in Figure 1 with filled circles. The SE of the excess for month i is 100*SE(p_i)/e_i. The smoothed estimates are calculated using the function described above. For example, the smoothed estimate for May is given by:

$1/64[(-8.6216) + 6(1.8839) + 15(1.2477) + 20(8.4544) + 15(0.5249) + 6(3.8104) + (-2.3885)] = 3.4193$. Estimates for the other months are calculated similarly, and these are the values plotted in Figure 1 with open circles. The standard errors of the smoothed estimates are calculated as described above. These standard errors are shown in Figure 1 to illustrate the uncertainty of the smoothed estimates.

Table. Observed and expected proportions of births by month for participants with type 1 diabetes

Month	n	Observed Proportion	SD(p)	Expected Proportion	Ratio	Percent Excess	SE of Excess
1	774	0.079491	.002741318	0.080722	0.98475	-1.5253	3.39600
2	673	0.069118	.002570572	0.075639	0.91378	-8.6216	3.39847
3	830	0.085242	.002829873	0.083666	1.01884	1.8839	3.38236
4	793	0.081442	.002771818	0.080438	1.01248	1.2477	3.44589
5	889	0.091301	.002919011	0.084184	1.08454	8.4544	3.46742
6	813	0.083496	.002803414	0.083060	1.00525	0.5249	3.37517
7	889	0.091301	.002919011	0.087950	1.03810	3.8104	3.31894
8	843	0.086577	.002849867	0.088696	0.97611	-2.3885	3.21309
9	911	0.093561	.002951232	0.087153	1.07352	7.3518	3.38625
10	760	0.078053	.002718533	0.085149	0.91666	-8.3337	3.19268
11	759	0.077950	.002716895	0.080222	0.97168	-2.8318	3.38673
12	803	0.082469	.002787680	0.083122	0.99215	-0.7852	3.35374

One of our analytic tasks was to estimate the maximum degree to which one month's birth excess exceeded another month's birth deficit. This should be the greatest monthly relative risk of diabetes that our manuscript reported as "RR(max)". If we had reported RR(max) based on the unsmoothed monthly data alone we would have calculated the difference between +8.5% for May and -8.6% for October (see Figure 1), or approximately 1.19. Such a high estimate of RR(max) would have been the result of choosing isolated data points vulnerable to random exaggeration. In selecting the degree of smoothing to use in estimating a maximum value of a function such as RR(max), we balanced the positive (random) bias caused by selecting the simple observed maximum versus the negative (dampened) bias caused by excessive smoothing. Our task in using smoothing techniques was to choose the right binomial power that dampens the bias just enough to come up with approximately unbiased estimates.

Let Y_{max} (Y_{min}) be the maximum (minimum) of the 12 values y_i, i=1, to 12 of Y. By definition, the maximum (minimum) of the smoothed values will be less (greater) than or equal to the maximum (minimum) of the observed values. This does not necessarily imply that the difference between the maximum and minimum of the smoothed values will underestimate the range of the months effects. This is true because the observed maximum (minimum) of (y_1, y_2, …, y_{12}) is an over (under) estimate, i.e. positively (negatively) biased estimate, of the maximum (minimum) of (μ_1, μ_2, …, μ_{12}). Based on simulations of data generated from a sine curve where the ratio of the height to SE of the observations was the same as we estimated, the binomial smoother of power 4 did not smooth enough and gave positively biased estimates of the maximum and the binomial smoother of power 6 produced estimates that where approximately unbiased or slightly conservative, depending on the size of the subgroup being estimated.

Figure A1 Smoothed estimates (± SE) of relative diabetes prevalence associated with month of birth among US youth with type 1 diabetes, by relative latitude of residence (3 northern sites or more southern sites) and racial group (white or nonwhite). **Panel A** includes 1,427 nonwhite participants from the north (circles; 352 blacks and 1,075 others) and 1,576 from the south (triangles; 547 blacks and 1, 029 others). **Panel B** includes 5,378 white participants from the north (circles) and 1,356 from the south (triangles).

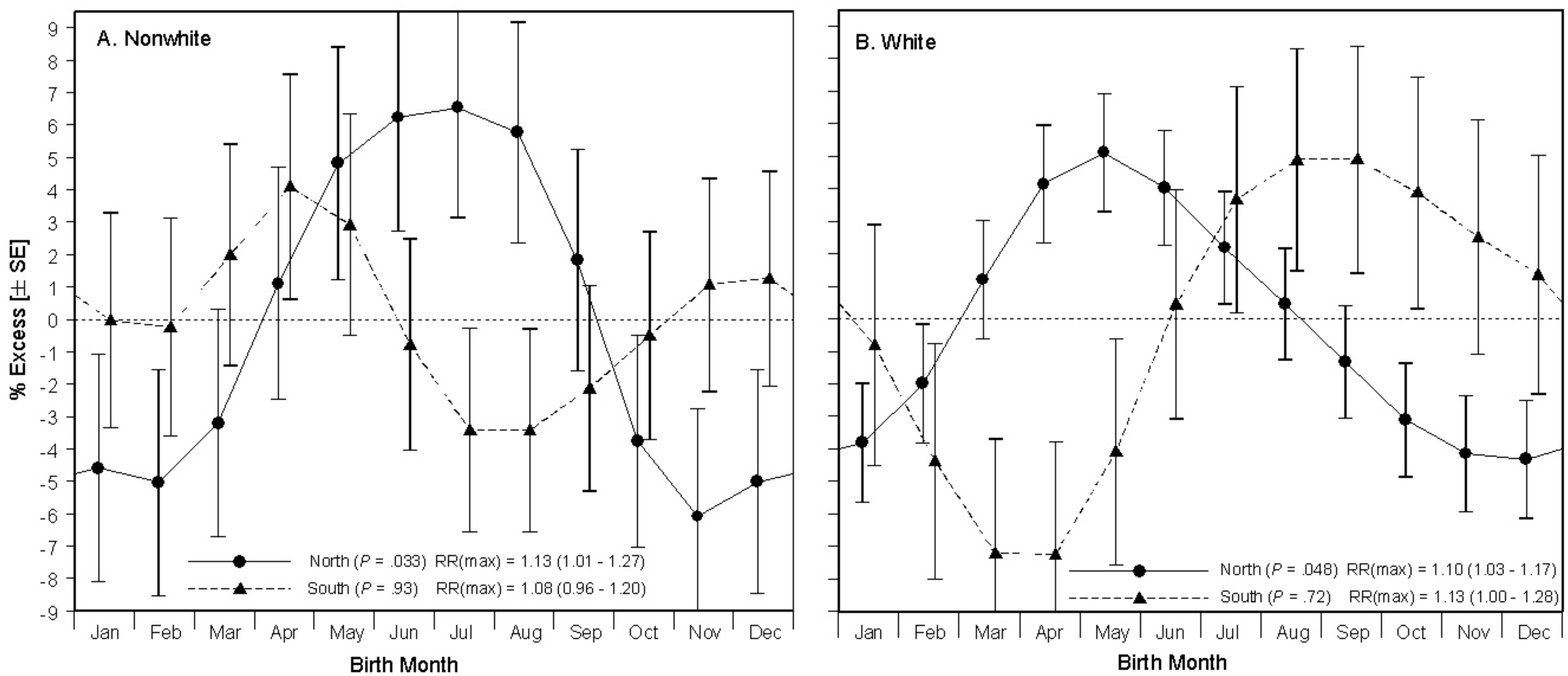


Figure A2 Smoothed estimates (± SE) of relative diabetes prevalence associated with month of birth among US youth with type 1 diabetes (restricted to the earlier cohort born in 1982-1992) diagnosed at age 0–9 years (circles, *n* = 2495) and at age 10–18 years (triangles, *n* = 3379). Age at diagnosis was unavailable for 271 participants.

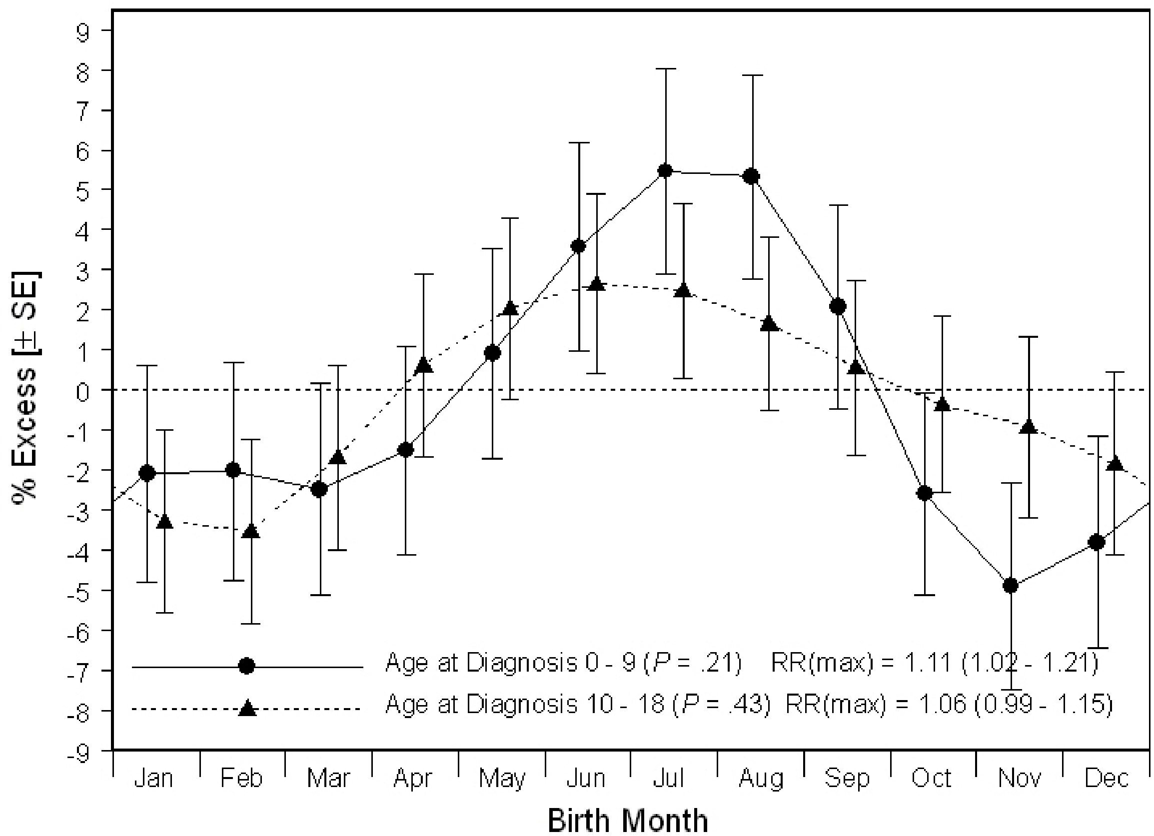
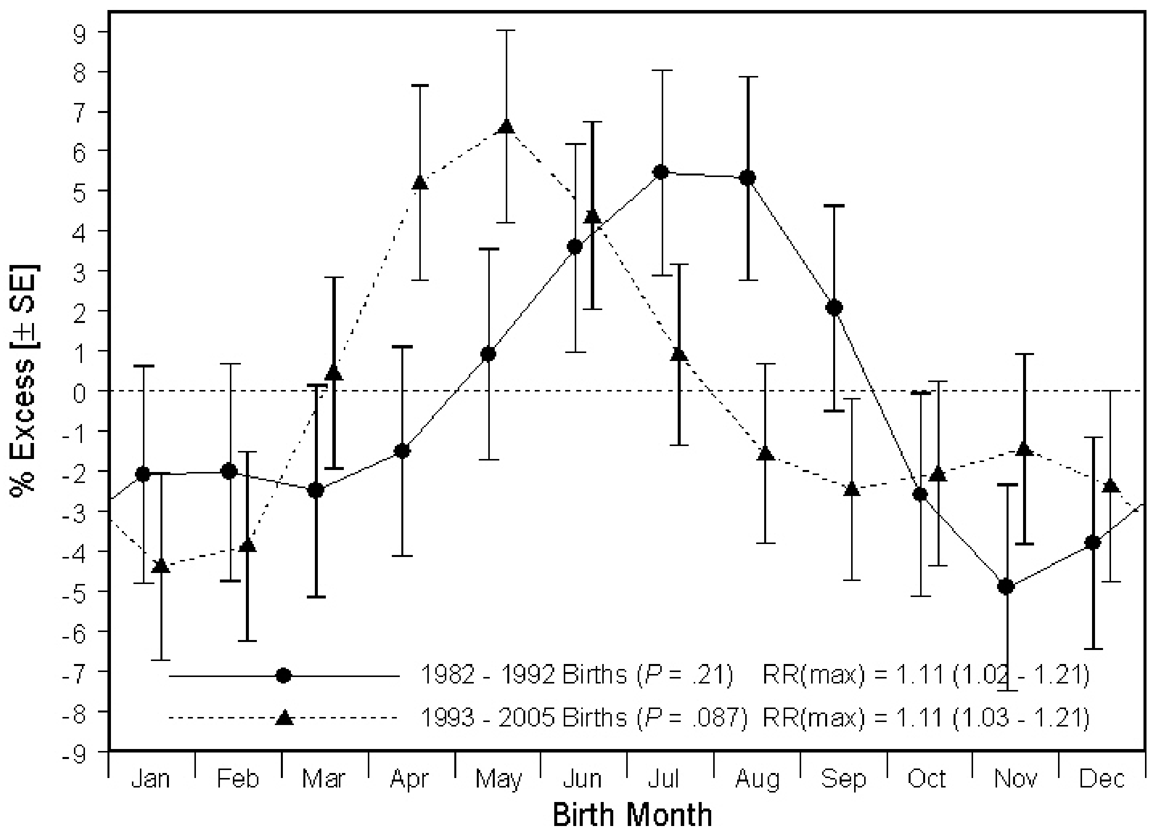


Figure A3 Smoothed estimates (± SE) of relative diabetes prevalence associated with month of birth among US youth with type 1 diabetes (restricted to those diagnosed at age 0–9 years) who were born in 1982–1992 (circles, *n* = 2,495) or in 1993–2005 (triangles, *n* = 3,102).



References

1. National Center for Health Statistics: Vital Statistics of the United States, 1983, Volume I, Natality. [article online], 1987. Available from http://www.cdc.gov/nchs/data/vsus/vsus_1983_1.pdf. Accessed 6 April 2009
2. Ventura SJ, Martin JA, Mathews TJ, Clarke SC: Advance Report of Final Natality Statistics - 1994. [article online], 1996. Available from http://www.cdc.gov/nchs/data/mvsr/supp/mv44_11s.pdf. Accessed 6 April 2009
3. Ventura SJ, Martin JA, Curtin SC, Mathews TJ: Births: Final Data for 1997. [article online], 2003. Available from http://www.cdc.gov/nchs/data/nvsr/nvsr47/nvs47_18.pdf. Accessed 6 April 2009
4. Aubury M, Luk W: Binomial filters. *J VLSI Signal Process* 12:35-50, 1996