What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971-2006

Katherine M. Keyes, MPH
Rebecca L. Utz, PhD
Whitney Robinson, PhD
Guohua Li, MD DrPH

Analysts often use different conceptual definitions of a cohort effect, and therefore different statistical methods, which lead to differing empirical results. A definition often used in sociology assumes that cohorts have unique characteristics confounded by age and period effects, whereas epidemiologists often conceive that period and age effects interact to produce cohort effects. The present study aims to illustrate these differences by estimating age, period, and cohort (APC) effects on obesity prevalence in the U.S. from 1971-2006 using both conceptual approaches. Data were drawn from seven cross-sectional waves of the National Health and Nutrition Examination Survey. Obesity was defined as BMI≥30 for adults and ≥95th percentile for children under 20. APC effects were estimated using the classic constraint-based method (first-order effects estimated and interpreted), the Holford method (first-order effects estimated but second-order effects interpreted), and median polish method (second-order effects are estimated and interpreted). Results indicated that all methods report significant age and period effects, with lower obesity prevalence in early life as well as increasing prevalence in successive surveys. Positive cohort effects for more recently born cohorts emerged based on the constraint-based model; when cohort effects were considered second-order estimates, no significant effects emerged. First-order estimates of age-period-cohort effects are often criticized because of their reliance on arbitrary constraints, but may be conceptually meaningful for sociological research questions. Second-order estimates are statistically estimable and produce conceptually meaningful results for epidemiological research questions. Age-period-cohort analysts should explicitly state the definition of a cohort effect under consideration. Our analyses suggest that the prevalence of obesity in the U.S. in the latter part of the 20th century rose across all birth cohorts, in the manner expected based on estimated age and period effects; the absence or presence of cohort effects depends on the conceptual definition and therefore statistical method used.

American Public Health Association
Philadelphia, PA
November 2009

Table 1. Constraint-based generalized log-linear approach: age-period-cohort effects on obesity prevalence in the United States, 1971-2006 (N=91,755)

Age *(2-74 years)*			
	Risk Ratio	95% Confidence Interval	
Age	**1.08**	**1.07**	**1.09**
Age-Squared	**0.99938**	**0.99932**	**0.99944**

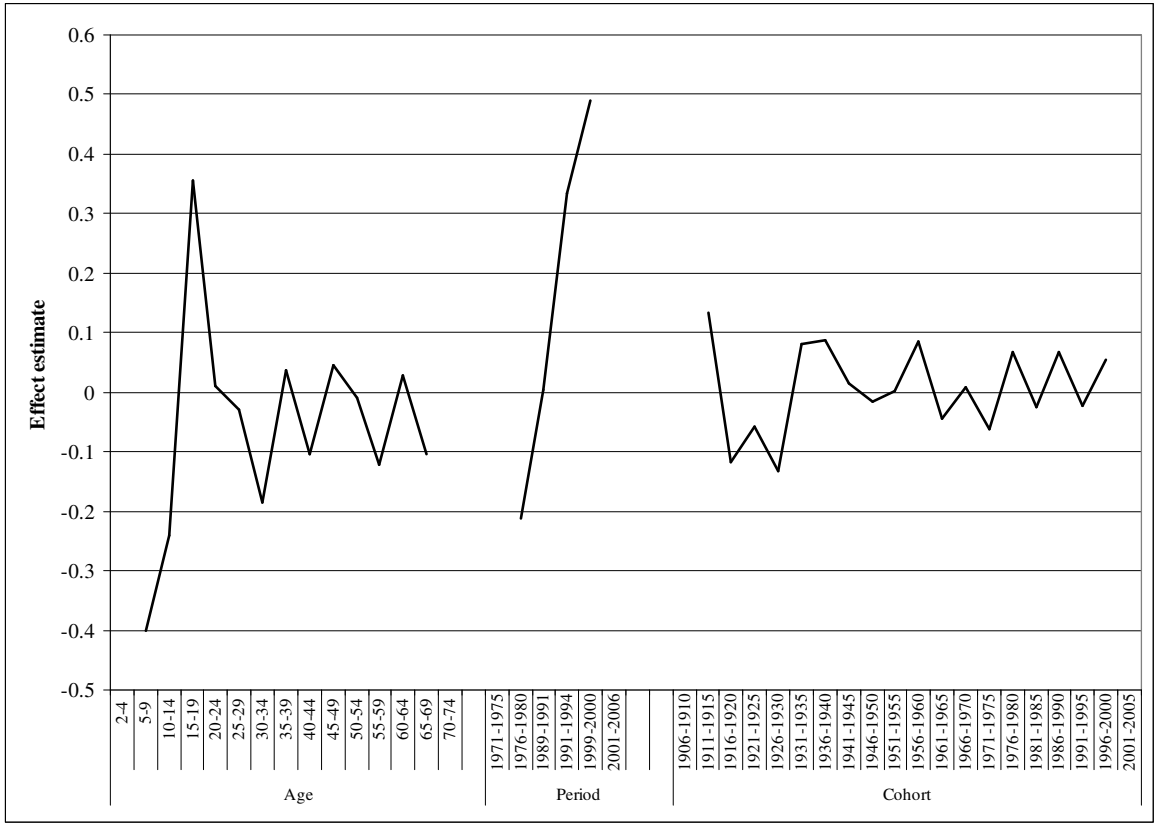
Period *(1971-2004)*			
	Risk Ratio	95% Confidence Interval	
Pre-1980	*Reference*		
1989-1991	**1.71**	**1.51**	**1.93**
1991-1994	**2.08**	**1.80**	**2.40**
1999-2000	**2.33**	**1.91**	**2.85**
2001-2006	**2.19**	**1.75**	**2.74**

Cohort *(1901-2005)*			
	Risk Ratio	95% Confidence Interval	
1901-1905	**0.74**	**0.53**	**1.03**
1906-1910	**0.72**	**0.53**	**0.97**
1911-1915	**0.78**	**0.59**	**1.03**
1916-1920	**0.74**	**0.59**	**0.93**
1921-1925	**0.77**	**0.63**	**0.93**
1926-1930	**0.82**	**0.70**	**0.96**
1931-1935	**0.85**	**0.75**	**0.97**
1936-1940	1.03	0.92	1.15
1941-1945	*Reference*		
1946-1950	0.99	0.89	1.11
1951-1955	0.98	0.86	1.12
1956-1960	1.09	0.93	1.27
1961-1965	1.17	0.97	1.41
1966-1970	**1.27**	**1.01**	**1.59**
1971-1975	**1.37**	**1.06**	**1.77**
1976-1980	**1.59**	**1.18**	**2.14**
1981-1985	**2.05**	**1.46**	**2.86**
1986-1990	**1.97**	**1.36**	**2.85**
1991-1995	**2.65**	**1.76**	**3.99**
1996-2000	**2.03**	**1.29**	**3.21**
2001-2005	**1.63**	**0.98**	**2.71**

Constant 0.02
X^2 (df) 5840 (26)

Bold = statistically significant at the p<0.05 level

Figure 1. Holford approach: estimated curvature[+] of age-period-cohort effects on obesity prevalence in the United States, 1971-2006 (N=91,755)



[+] Curvatures can be interpreted as the change (increase or decrease) in the underlying linear slope of age, period, and cohort. Curvatures for the earliest and latest ages/periods/cohorts are not estimated because they have only one adjacent parameter.

Table 3. Median polish approach: age-period-cohort effects on obesity prevalence in the United States, 1971-2006 (N=91,755)

Age *(2-74 years)*			
	Risk ratio	95% confidence interval	
2-4	**0.23**	**0.16**	**0.34**
5-9	**0.55**	**0.37**	**0.81**
10-14	**0.57**	**0.38**	**0.84**
15-19	**0.55**	**0.38**	**0.82**
20-24	**0.57**	**0.39**	**0.84**
25-29	0.76	0.51	1.12
30-34	*Reference*		
35-39	0.99	0.67	1.46
40-44	1.14	0.77	1.68
45-49	1.09	0.74	1.61
50-54	1.28	0.87	1.89
55-59	1.41	0.95	2.08
60-64	1.26	0.85	1.86
65-69	1.22	0.83	1.81
70-74	1.05	0.71	1.55

Period *(1971-2004)*			
	Risk ratio	95% confidence interval	
1971-1975	*Reference*		
1976-1980	1.02	0.72	1.45
1981-1985	1.27	0.89	1.81
1989-1991	**1.47**	**1.03**	**2.09**
1991-1994	**1.78**	**1.25**	**2.54**
1999-2000	**2.32**	**1.63**	**3.30**
2001-2006	**2.47**	**1.73**	**3.51**

Cohort *(1901-2005)*			
	Risk ratio	95% confidence interval	
1901-1905	1.23	0.99	1.49
1906-1910	1.09	0.94	1.26
1911-1915	1.09	0.97	1.24
1916-1920	1.03	0.92	1.15
1921-1925	1.01	0.91	1.12
1926-1930	1.02	0.92	1.12
1931-1935	1.09	0.99	1.20
1936-1940	1.03	0.94	1.14
1941-1945	*Reference*		
1946-1950	1.02	0.93	1.13
1951-1955	0.95	0.87	1.05
1956-1960	0.98	0.89	1.08
1961-1965	0.91	0.83	1.01
1966-1970	0.98	0.89	1.08
1971-1975	1.01	0.92	1.12
1976-1980	1.07	0.97	1.18
1981-1985	1.14	0.99	1.27
1986-1990	1.13	0.98	1.26
1991-1995	1.10	0.97	1.24
1996-2000	0.98	0.85	1.13
2001-2005	1.03	0.85	1.24

Bold = statistically significant at the p<0.05 level