2008 Ohio Family Health Survey

Health Policy Brief

Using Small-Area Estimation Techniques for County-level Estimates of Select Indicators from the 2008 Ohio Family Health Survey

Anirudh V.S. Ruhil, Holly Raffle, Sara Boyd, Nicole R. Yandell
Voinovich School of Leadership and Public Affairs, Ohio University

Financial and logistical constraints invariably prevent national and state surveys of health behaviors or characteristics from surveying populations of all counties, places, or other sub-national/sub-state geographies (for example, city neighborhoods). However, policy or programmatic considerations often require that reliable estimates be available for these smaller geographies. Small area estimation (SAE) techniques provide one means of deriving estimates for smaller geographies that are undersampled (or not sampled at all) in national/state surveys.

Small-area estimation techniques provide, both in theory and in practice, substantial leverage by way of enabling analysts to generate estimates for smaller geographies (counties, places, neighborhoods) that are often undersampled (or not sampled), in national/state surveys. In this report, we have provided a brief overview of these techniques, as well as a demonstration of some basic estimation techniques – both model-free and model-based, with and without spatial smoothing. We did so in the context of the 2008 Ohio Family Health Survey (OFHS). The small area estimates we derived, regardless of modeling options, depart significantly from the unconditional survey-weighted estimates. Given that the survey-weighted estimates are design-unbiased, it would be prudent to regard them as “true” estimates of each of the target response variables. Consequently, we can benchmark all other estimates reported here against the survey-weighted estimates. For the most part, and regardless of the substantive question we focus on, notice for example that the synthetic estimates are on average within ±1 - 2% of the survey-weighted estimates – well within the usual confidence intervals.

Clearly, if the goal is to generate reliable county-level estimates of diabetes, stroke, cancer, and so on – regardless of the sex or age or poverty-level of sub-populations – from the 2008 OFHS, then we recommend use of the survey-weighted estimates. This is so largely because the 2008 OFHS survey provides good coverage for virtually all counties. Another reason for the discrepancy between the model-based estimates and the direct survey-weighted estimates could be that the predictors used in the model were essentially few to begin with and even then not driven by substantive knowledge of the specific factors known to predict diabetes, obesity, and so forth. When sample coverage of the small areas is sparse, however, as is the case with the BRFSS data, then model-based estimates will by default be preferred because the surveys like the BRFSS do not cover each county in the state.

Overall, our research suggests that the OFHS makes two vital contributions. First, the OFHS is the only source of county-level information on a host of health status indicators for the state. Indeed, ideally the OFHS would run every two years, if not annually because in doing so public health agencies, policymakers, and researchers would have access to more timely, trend data at the sub-state level. Without this frequency of data, users are forced to rely either upon outdated data (such as the 2004 OFHS until the 2008 OFHS data were released) or then upon small area estimates that are both cumbersome and noisy to obtain from the BRFSS. Second, the 2004 and 2008 OFHS provide a unique opportunity to compare the accuracy of BRFSS-derived county-level estimates for selected health indicators vis-a-vis the direct survey-weighted estimates the OFHS yields. Such comparisons could illustrate the extent to which estimates derived from the various small area estimation techniques applied to BRFSS data (where few counties are sampled in any given state in any given year) approach the “true” values embodied in the OFHS, and when they fail to overlap, the causes for these failures. In our ongoing work we are undertaking this latter inquiry, comparing in particular (i) the model-free synthetic estimates, (ii) the EBLUP approach of the random-intercept mixed logit models, and (iii) the Hierarchical Bayes approach to 2008 OFHS estimates.

Table 1: Sub-Groups of Analytic Interest in the 2008 OFHS Survey

Category	All	Minority Groups		
		African-American	Hispanic	Asian
Gender	Both	Both	Both	Both
Age	0-17 18-34 35-54 55-64 65+	0-17 18+	0-17 18+	0-17 18+
Family Income*	≤ 100% 101 to ≤ 150% 151 to ≤ 200% 201 to ≤ 250% 251 to ≤ 300% 301 to ≤ 400% > 400%	≤ 100% 101 to ≤ 200% 201 to ≤ 300% 301 to ≤ 400% > 400%		
Region	Metropolitan Appalachian Rural (non-Appalachian) Suburban	Each of the 6 largest Metro Counties		

*Family income is measured in terms of Federal poverty level, and in particular, the level at which a family is considered to be living in poverty, accounting for family size.

The main policy consideration that emerges from our analyses is that local and regional community health agencies and health care providers should use data and information provided from instruments such as the Ohio Family Health Survey to shape policies and programs that address health problems and stressors at regional and local levels. When instruments like the OFHS are unavailable, and the proliferation and ease of modern statistical computing resources notwithstanding, local, regional, and state policymakers and health service providers should consider the pros and cons of employing synthetic, model-based, and spatial techniques to examine their communities.

Table 2: 2003-2004 OFHS Sample Disposition by Appalachian Cluster

Demographic	Appalachian Cluster	Minimum	Maximum
Gender			
Male	4,112	67	370
Female	7,319	152	690
Total	11,431		
Age			
18-24	466	5	39
25-34	1,141	17	114
35-44	1,744	29	194
45-54	2,331	41	237
55-64	2,463	41	209
65+	3,289	67	267
Total	11,431		
Income			
< 100%	2,266	49	150
101 – 150%	1,608	28	104
151 – 200%	1,269	18	84
201 – 300%	2,287	47	206
301% +	4,004	58	516
Total	11,431		
Imputed Race			
White/Other	10,980	227	1,012
Black/African-American	172	0	12
Hispanic	242	2	28
Asian	40	59	775
Total	11,434		