Using National Health Interview Survey Data to Detect Signals of Natural or Intentional Outbreaks

Abera Wouhib, Doug Williams, Myron J. Katzoff, PhD, and Joe Fred Gonzalez, Jr., National Center for Health Statistics (NCHS), 3311 Toledo Road, Room 3124, Hyattsville, MD 20782

KEY WORDS: symptoms, baseline estimates

1. Introduction

The National Health Interview Survey (NHIS) administers face-to-face interviews to a nationally representative sample of households. Each week a probability sample of the civilian noninstitutionalized population of the United States is interviewed by personnel of the U.S. Bureau of the Census. Information is obtained about the health and other characteristics of each member of the household. Beginning with a new NHIS design in 1997, the NHIS contains a Basic Module every year. The Basic Module is divided into three components, the Family Core, the Sample Adult Core, and the Sample Child Core. Information on these components is collected for 50 weeks and the first two weeks of each year are used for training interviewers. The purpose of this paper is to establish a national baseline for detecting signals of natural or intentional outbreaks by estimating health characteristics based on data from two-week intervals during which information is collected in the Sample Adult and Sample Child Cores.

In the last decade natural outbreaks, such as Human Immunodeficiency Virus (HIV) and Severe Acute Respiratory Syndrome (SARS), and other natural epidemics have had serious global effects. Moreover, the threat of intentional releases of biological and chemical agents has become more real than ever before.

- How can we detect such events as early as possible?
- How can we discern the difference between intentional outbreaks and natural outbreaks if both display similar symptoms and share several common characteristics at the initial stage at a given time and place?

The simplest and most straight forward method to detect signals for natural or intentional outbreaks with symptoms similar to the common cold and flu is to identify an unusually high volume of symptoms and to compare to baseline estimates for the same period and place. Therefore, computing baseline estimates by time intervals as short as two weeks, by age group and regions would enable us to answer the questions above as quickly as possible. This could serve as a stepping stone in handling the crisis if a positive signal is detected.

Baseline estimates were computed from data collected during 1997 – 2001. They were developed mainly from public use data on two questions in the sample adult and sample child components. The following two questions were asked using a two-week reference period.

The next questions are about your recent health during the TWO WEEKS outlined on that calendar.

a- Did you have a head cold or a chest cold that started during those TWO WEEKS?

b- Did you have a stomach or intestinal illness with vomiting or diarrhea that started during those TWO WEEKS?

As mentioned earlier, the above two questions appeared in the Sampled Adult Core (age ≥ 18) and the Sampled Child Core (age ≤ 17).

2. The Data

"YES" or "NO" responses are obtained for the two questions above. The sample data are displayed by data component for each question for those who responded "YES". The time period studied was the two-week period starting from the second two-weeks up to the last (26^{th}) two-week period of each year, except for 1997, when data was collected for the entire year. Figure 1 shows the distribution of the unweighted proportions of the samples, as it was collected, by two-week period of each year for the Sample Child Core from 1997 – 2001.

Figure 1 reflects seasonal effects since head colds and chest colds are closely associated with influenza. As expected, the data show high proportions of symptoms during the winter season and low, during the summer. The five plots are very similar, and the data are seasonal for each year between 1997 and 2001.

Figure 1

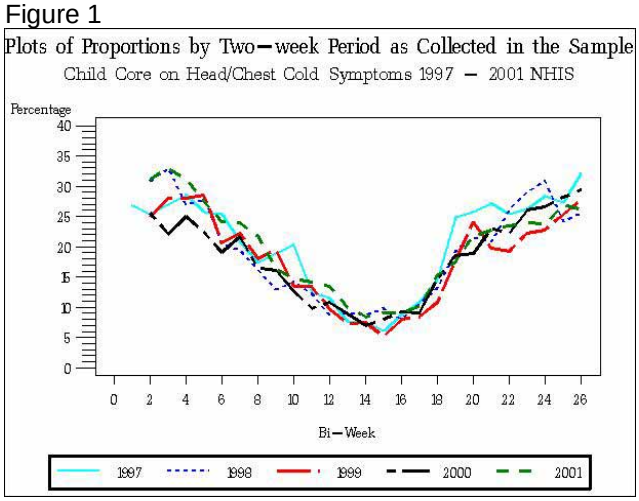
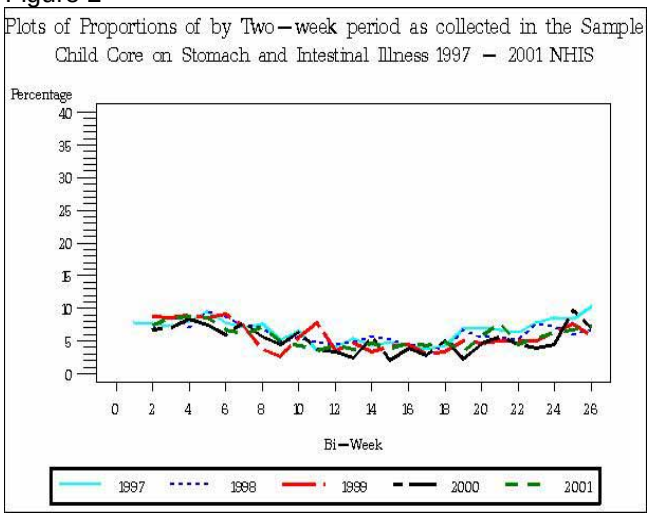


Figure 2 below shows the proportion of children in the sample having symptoms of vomiting, stomach and intestinal illness. The plot for these symptoms is more stable year round than that for the head and chest cold symptoms of Figure 1 even though the percentage of such symptoms is slightly higher in winter as compared to summer.

Figure 2



In a similar fashion, the adult component of the data is presented in Figures 3 and 4. Although the sample sizes are larger in the adult component than in the child component, the trends in the percentages are quite similar. The percentage of head cold and chest cold is displayed in Figure 3, whereas for vomiting, stomach and intestinal illness is shown in Figure 4.

Figure 3

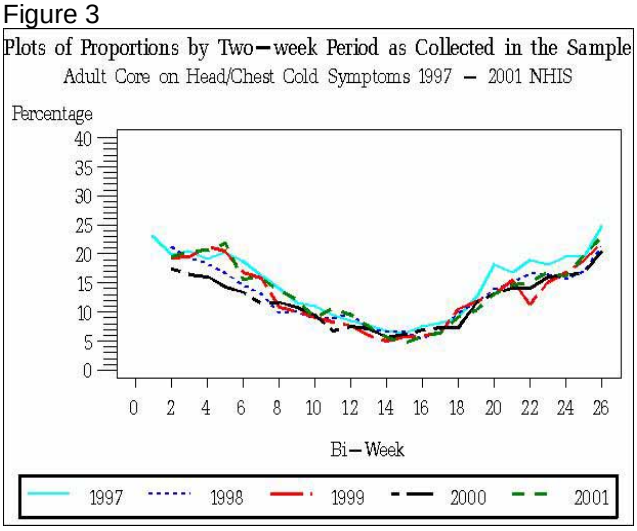
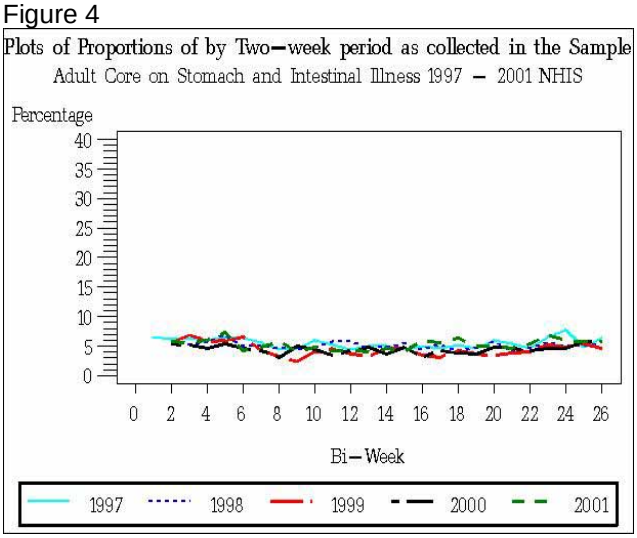


Figure 4



3. Weight

The sample for the NHIS is redesigned every decade using population data from the most recent decennial census. The sample is chosen in such a way that each person in the covered population has a known non-zero probability of selection. These probabilities of selection, along with the adjustments for nonresponse and post-stratification, are reflected in the sample weights that are provided in the data files for the sample adult and sample child questionnaires. The weights were adjusted to national population (civilian, noninstitutional) estimates by age, sex, and race/ethnicity, based on projections from the 1990 U.S. Census.

4. Analysis

Before analyzing the data, the responses from the two questions were combined as "YES" if at least one response was "YES". All other responses including "I do not know", "I am not sure" or refusal were considered as "No". The analysis was done on the basis of the two-week reference period used to collect the data for a total of 25 two-week periods for each year. The percentage of children and adults having the symptoms stated above were estimated by age group and region in the country. The estimate was computed for each reference period based on those who responded "YES" for at least one of the questions indicating exposure to one of the symptoms explained above. In order to produce reliable estimates, information collected from 1997-2001 was combined by two-week reference period. SAS proc SURVEYMEANS was used to estimate the percentages and their standard errors.

The proportion of children and adults who responded "YES" to at least one of the questions indicating exposure to the symptoms was estimated by six age groups: three age groups for adults and three age groups for children. The Three age groups for adults were 18-44, 45-64, and 65 + whereas for children they were 0-2, 3-7, and 8-17.

The results were displayed in *Figure 5* for children by age group, in *Figure **6*** for adults by age group and in *Figure 7* combined.

Figure 5

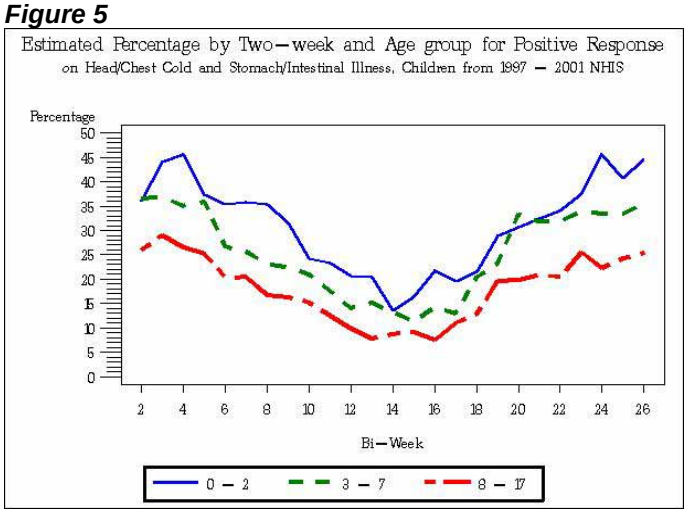


Figure 6

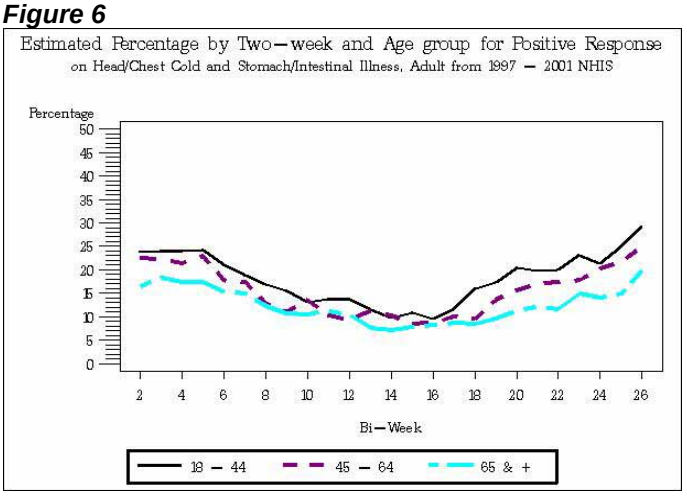
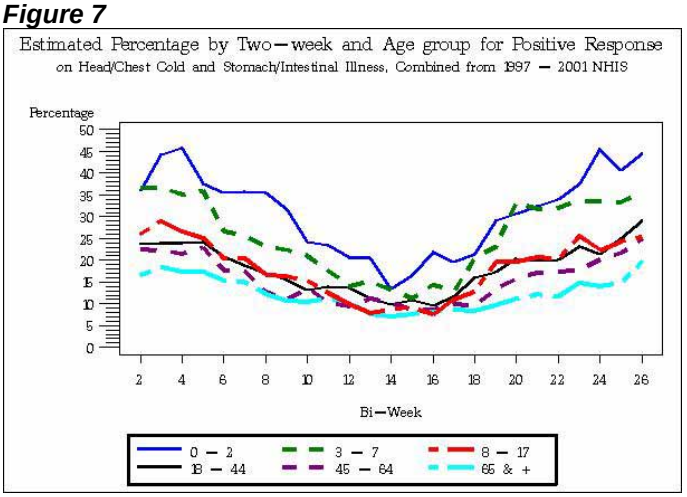


Figure 7



The proportion of children and adults showing symptoms were also estimated separately by four regions in the U.S. The four regions of the country are: West, South, Northeast and Midwest. Figures 8 - 9 show the result by two week period for children and adults, respectively. Here region 1: Northeast, Region 2: Midwest, Region 3: South, and Region 4: West. (See Appendix D for details of the regions on the U.S. map)

Figure 8

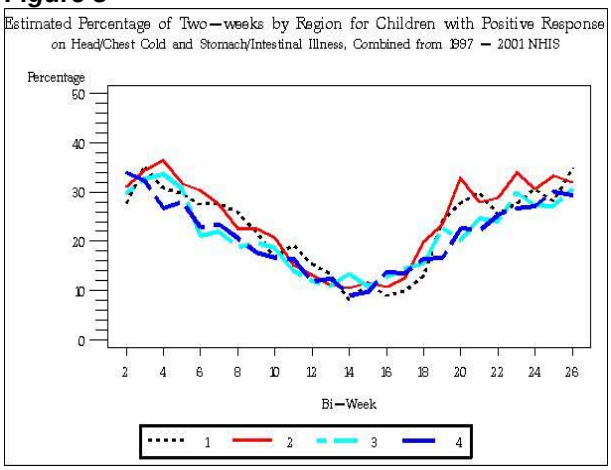
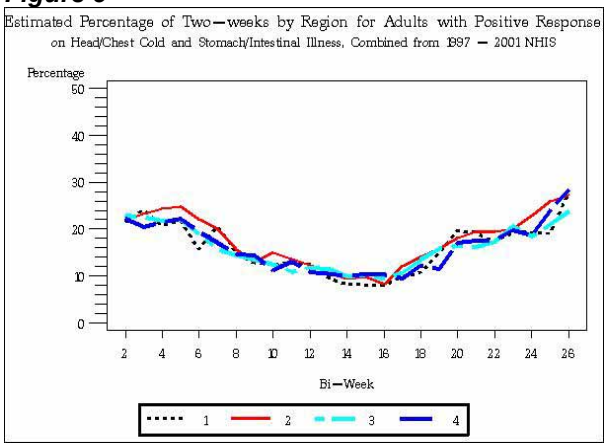


Figure 9



As displayed in the appendices, the standard errors in estimating these proportions are small enough to produce the Coefficients of Variation (CV) less than 10% for almost all estimates.

Generally the results show higher proportions for the beginning and ending months of a year, which implies high proportion of exposure in winter as compared to summer for outbreaks of this type associated with symptoms mentioned above. This pattern is similar across the board for all age groups. However, the proportions for children are higher than for adults. The regional results do not indicate the direction or the source of the outbreaks, but they consistency display higher exposure to children than adults. The five year estimates could be considered as a baseline for future studies in case catastrophic events take place as a result of natural phenomena or intentional acts. Since the data were collected in a relatively normal period (no extreme epidemics reported via surveillance systems), its use as baseline for future reference would be useful.

The estimates for the proportions are attached as appendices. The tables for estimated proportions, $\hat{P}$ with their associated standard errors and sample sizes are displayed in Appendix A. It has the estimated proportions for three age groups for children based on positive response to at least one symptom of the questions explained at the beginning of this paper for the year 1997 to 2001. The table also displays the same results computed in three age groups of adults for 1997 to 2001. Similar tables are also attached for regional estimates for adults and children as Appendix B and C, respectively.

5. Conclusion and Future Work

Children are more susceptible to these symptoms than adults, especially the younger age groups. Also, regardless of the age group or region, the results show higher proportions in winter than summer, which is consistent with a high rate of exposure to flu and other viruses in winter.

The five-year estimates could be considered as a baseline for future in case catastrophic events as a result of natural phenomena or intentional acts. Since the data were collected during relatively normal periods, its use as baseline for future reference would be important

Larger sample sizes would be necessary for reasonable estimates at state/county level, in this case further demographic and spatial analysis could be produced as a vital part of the baseline estimates

The results of this investigation are limited as they were arrived at via single dimensional analysis. Our next investigation may involve spatial analysis in addition to temporal analysis for the complete picture. In the future, data collected for more than five years will be used in order to produce reliable estimates by more detailed demographic and social classifications. Also additional information such as surveillance and over the counter drug sales data must be incorporated in the future model to produce more precise baseline estimates.

References

1 Cochran WG. Sampling Techniques (3rd ed) New York: Wiley.1977

2 1997-2001 National Health Interview Survey, CD-ROM, Series 10, NCHS, Hyattsville, MD

APPENDIX A: Sample Size, Estimated Proportion and Standard Errors, by Two-week Period and Age Group for Children and Adults, NHIS 1997-2001

Two-week	Children Age 0 - 2			Children Age 3 - 7			Children Age 8 -17			Adults Age 18 - 44			Adults Age 45 - 64			Adults Age 65 and Above		
Period	Sample	Prop.	Std Error	Sample	Prop.	Std Error	Sample	Prop.	Std Error	Sample	Prop.	Std Error	Sample	Prop.	Std Error	Sample	Prop.	Std Error
2	517	0.3597	0.018	742	0.3668	0.0146	517	0.3597	0.018	3403	0.237	0.0047	1921	0.2253	0.008	1241	0.166	0.0085
3	491	0.441	0.0164	719	0.3673	0.011	491	0.441	0.0164	3551	0.24	0.0061	1943	0.2224	0.0084	1226	0.1838	0.0094
4	527	0.4573	0.0188	720	0.3509	0.0137	527	0.4573	0.0188	3502	0.239	0.0078	1972	0.215	0.0076	1326	0.1741	0.0072
5	494	0.3744	0.0179	791	0.3597	0.0115	494	0.3744	0.0179	3533	0.242	0.0069	1851	0.2292	0.0079	1305	0.1748	0.0072
6	519	0.3544	0.0128	725	0.2669	0.0148	519	0.3544	0.0128	3499	0.21	0.0063	1808	0.1775	0.0079	1168	0.1528	0.0076
7	511	0.3584	0.0175	778	0.256	0.0106	511	0.3584	0.0175	3408	0.189	0.0057	1751	0.1742	0.0071	1122	0.1497	0.0079
8	482	0.3542	0.0218	723	0.2315	0.011	482	0.3542	0.0218	3538	0.169	0.0052	1949	0.1274	0.0071	1319	0.123	0.0081
9	478	0.3155	0.0133	766	0.224	0.0148	478	0.3155	0.0133	3533	0.154	0.0062	1898	0.1098	0.0075	1309	0.1071	0.0065
10	523	0.2415	0.0155	774	0.2099	0.0113	523	0.2415	0.0155	3546	0.131	0.0061	1934	0.1357	0.0063	1274	0.1049	0.008
11	480	0.2336	0.0148	753	0.1757	0.0076	480	0.2336	0.0148	3236	0.138	0.0044	1792	0.1027	0.0059	1305	0.1124	0.0061
12	510	0.2065	0.0128	699	0.1403	0.0083	510	0.2065	0.0128	3310	0.137	0.0046	1864	0.094	0.0053	1285	0.104	0.0055
13	435	0.2053	0.0126	728	0.151	0.0119	435	0.2053	0.0126	3259	0.115	0.0043	1877	0.1117	0.0058	1325	0.0767	0.0066
14	508	0.135	0.0108	759	0.1318	0.0107	508	0.135	0.0108	3418	0.097	0.0041	1935	0.103	0.0064	1249	0.072	0.006
15	554	0.1639	0.0097	759	0.1125	0.0083	554	0.1639	0.0097	3587	0.109	0.0044	1905	0.0853	0.0052	1170	0.0774	0.0065
16	495	0.2179	0.0092	750	0.1427	0.0097	495	0.2179	0.0092	3338	0.096	0.0037	1913	0.0864	0.005	1296	0.0827	0.0067
17	494	0.1948	0.0137	747	0.1296	0.0096	494	0.1948	0.0137	3319	0.116	0.0046	1890	0.1004	0.0073	1318	0.0866	0.0057
18	496	0.2148	0.0167	776	0.2041	0.0121	496	0.2148	0.0167	3575	0.159	0.0055	1881	0.095	0.0051	1357	0.0839	0.0054
19	443	0.2895	0.0163	707	0.2312	0.0122	443	0.2895	0.0163	3345	0.174	0.0058	1947	0.1359	0.0069	1391	0.0962	0.0051
20	440	0.3057	0.0198	764	0.3336	0.0128	440	0.3057	0.0198	3503	0.205	0.0068	2017	0.1567	0.0075	1177	0.1118	0.0077
21	510	0.3238	0.0204	709	0.318	0.0136	510	0.3238	0.0204	3466	0.198	0.0068	1894	0.1719	0.0082	1241	0.1215	0.0076
22	506	0.3403	0.015	724	0.3194	0.0144	506	0.3403	0.015	3387	0.201	0.0063	1958	0.1739	0.0072	1197	0.1169	0.007
23	464	0.3747	0.0162	726	0.3374	0.0111	464	0.3747	0.0162	3436	0.231	0.0079	1952	0.1777	0.0075	1254	0.1488	0.0077
24	465	0.4552	0.0132	721	0.3346	0.0128	465	0.4552	0.0132	3255	0.214	0.0052	1957	0.2028	0.0073	1275	0.1402	0.0073
25	448	0.4074	0.0142	675	0.3341	0.0123	448	0.4074	0.0142	3289	0.249	0.0066	1888	0.2163	0.009	1194	0.1484	0.0077
26	472	0.4447	0.0209	643	0.3549	0.0115	472	0.4447	0.0209	3082	0.29	0.0066	1834	0.25	0.0081	1161	0.1974	0.0076

Appendix B: Sample Size, Estimated Proportion and Standard Errors by Region and Two-week Period for Adults (18 Years and Older), NHIS 1997-2001

Two Wk.	*Region 1*			*Region 2*			*Region 3*			*Region 4*		
	sam. size	*prop*	*std. err*	*sam. size*	*prop*	*std. err*	*sam. size*	*prop*	*std. err*	*sam. size*	*prop*	*std. err*
2	496	0.2772	0.0222	488	0.3097	0.0111	935	0.2974	0.0102	767	0.3396	0.0074
3	486	0.3533	0.0119	533	0.3436	0.0081	951	0.3280	0.0080	699	0.3224	0.0168
4	456	0.3081	0.0176	510	0.3643	0.0082	986	0.3363	0.0145	796	0.2676	0.0072
5	486	0.2968	0.0134	596	0.3160	0.0093	1037	0.3068	0.0125	679	0.2805	0.0137
6	419	0.2753	0.0097	519	0.3045	0.0149	1033	0.2110	0.0120	693	0.2286	0.0106
7	470	0.2759	0.0175	501	0.2736	0.0066	935	0.2189	0.0085	766	0.2341	0.0127
8	502	0.2593	0.0135	575	0.2256	0.0100	1030	0.1883	0.0111	595	0.2062	0.0090
9	509	0.2186	0.0193	483	0.2253	0.0125	1067	0.1965	0.0116	759	0.1774	0.0093
10	565	0.1676	0.0070	621	0.2070	0.0085	1000	0.1857	0.0103	659	0.1658	0.0142
11	513	0.1925	0.0106	588	0.1518	0.0091	938	0.1396	0.0106	540	0.1641	0.0129
12	443	0.1544	0.0089	583	0.1316	0.0049	880	0.1193	0.0055	725	0.1176	0.0067
13	497	0.1334	0.0138	517	0.1118	0.0086	967	0.1100	0.0085	680	0.1238	0.0082
14	499	0.0804	0.0084	601	0.1059	0.0069	987	0.1340	0.0054	724	0.0909	0.0079
15	470	0.1129	0.0051	615	0.1167	0.0072	1092	0.1086	0.0087	710	0.0971	0.0100
16	513	0.0882	0.0102	556	0.1066	0.0136	887	0.1270	0.0065	729	0.1377	0.0059
17	471	0.0985	0.0074	598	0.1243	0.0064	960	0.1451	0.0097	713	0.1352	0.0095
18	517	0.1295	0.0076	658	0.1994	0.0123	993	0.1550	0.0126	615	0.1648	0.0108
19	475	0.2421	0.0104	608	0.2349	0.0128	894	0.2265	0.0113	580	0.1663	0.0106
20	492	0.2774	0.0128	570	0.3272	0.0121	986	0.2024	0.0132	693	0.2268	0.0107
21	486	0.2990	0.0060	576	0.2791	0.0073	887	0.2465	0.0103	697	0.2217	0.0196
22	501	0.2543	0.0133	638	0.2874	0.0169	948	0.2401	0.0123	627	0.2537	0.0144
23	492	0.2762	0.0094	593	0.3387	0.0137	921	0.2993	0.0129	613	0.2666	0.0145
24	448	0.3061	0.0142	660	0.3066	0.0086	1015	0.2728	0.0106	620	0.2719	0.0105
25	454	0.2809	0.0085	554	0.3334	0.0135	941	0.2719	0.0107	622	0.3005	0.0068
26	391	0.3479	0.0107	555	0.3204	0.0118	920	0.3064	0.0093	666	0.2927	0.0094

Appendix C: Sample Size, Estimated Proportion and Standard Errors by Region and Two-week Period for Children (0 - 17 years), NHIS 1997-2001

Two Wk.	*Region 1*			*Region 2*			*Region 3*			*Region 4*		
	sam. size	*prop*	*std. err*	*sam. size*	*prop*	*std. err*	*sam. size*	*prop*	*std. err*	*sam. size*	*prop*	*std. err*
2	1273	0.2142	0.0115	1392	0.2191	.0044	2314	0.2283	.0066	1586	0.2224	0.0079
3	1314	0.2408	0.0091	1480	0.2314	.0030	2469	0.2250	.0048	1457	0.2040	0.0137
4	1282	0.2084	0.0119	1434	0.2429	.0074	2444	0.2180	.0071	1640	0.2147	0.0062
5	1141	0.2175	0.0064	1558	0.2481	.0062	2510	0.2193	.0080	1480	0.2214	0.0125
6	1104	0.1554	0.0038	1368	0.2201	.0045	2520	0.1900	.0064	1483	0.1948	0.0026
7	1157	0.2057	0.0063	1403	0.2013	.0025	2293	0.1561	.0054	1428	0.1702	0.0092
8	1236	0.1509	0.0079	1674	0.1567	.0060	2482	0.1438	.0053	1414	0.1470	0.0066
9	1310	0.1269	0.0051	1361	0.1302	.0069	2383	0.1343	.0049	1686	0.1435	0.0074
10	1375	0.1219	0.0063	1590	0.1488	.0059	2464	0.1247	.0049	1325	0.1128	0.0064
11	1156	0.1273	0.0073	1497	0.1368	.0059	2378	0.1092	.0058	1302	0.1308	0.0080
12	1180	0.1253	0.0057	1583	0.1219	.0029	2190	0.1183	.0042	1506	0.1089	0.0049
13	1177	0.0965	0.0084	1375	0.1068	.0063	2515	0.1145	.0044	1394	0.1047	0.0033
14	1259	0.0829	0.0077	1518	0.0939	.0029	2349	0.0990	.0054	1476	0.1005	0.0052
15	1234	0.0813	0.0072	1554	0.0992	.0053	2460	0.1011	.0041	1414	0.1028	0.0047
16	1312	0.0812	0.0026	1507	0.0836	.0044	2157	0.0949	.0040	1571	0.1028	0.0040
17	1075	0.0989	0.0047	1608	0.1209	.0047	2350	0.1068	.0051	1494	0.0944	0.0061
18	1394	0.1067	0.0021	1645	0.1406	.0044	2411	0.1345	.0076	1363	0.1225	0.0069
19	1324	0.1505	0.0068	1655	0.1582	.0084	2252	0.1587	.0056	1452	0.1152	0.0046
20	1384	0.1964	0.0075	1538	0.1803	.0074	2378	0.1636	.0073	1397	0.1710	0.0108
21	1326	0.1917	0.0079	1416	0.1945	.0062	2374	0.1620	.0051	1485	0.1748	0.0052
22	1274	0.1722	0.0060	1622	0.1942	.0071	2217	0.1729	.0054	1429	0.1762	0.0075
23	1355	0.1951	0.0062	1634	0.2008	.0093	2386	0.2076	.0067	1267	0.1984	0.0104
24	1180	0.1913	0.0077	1624	0.2278	.0047	2348	0.1847	.0068	1335	0.1858	0.0070
25	1241	0.1896	0.0075	1408	0.2596	.0083	2360	0.2096	.0059	1362	0.2397	0.0104
26	1129	0.2790	0.0137	1457	0.2718	.0088	2160	0.2379	.0069	1331	0.2830	0.0108

Appendix D: Estimated Proportions for Children by Region for Selected two-week periods