1. THE STORY

I was one of five co-authors for the article “Epidemiological Mapping of Florida Childhood Cancer Clusters” in 2010, which received wide media attention in Florida due to the coincidental announcement, at nearly the same time, by the Centers for Disease Control and Prevention (CDC) that there was a pediatric cancer cluster in Palm Beach County, often referred to as “the Acreage Cluster.” When we had completed the initial cluster analysis in 2009, and had identified what we believed to be statistically significant cancer clusters, I contacted the Florida Department of Health (FDOH) to inform them about our findings. We had a conference call with several key personnel at the FDOH and the Florida Cancer Data System (FCDS), the official cancer registry in Florida. Because of the CDC announcement, and perhaps partly because of our article, there was public concern and discussion, especially among parents of children with brain tumors in Palm Beach County. As a result, shortly after the publication of Amin et al. (2010), I and my co-authors were asked to meet with the Surgeon General of Florida and her staff to discuss our findings.

That discussion did not lead to major policy changes. The Surgeon General (wisely) wanted our results to be replicated by independent analysts. That request led to this new project, in which several researchers used the same Florida Association of Pediatric Tumor Programs (FAPTP) data, but different epidemiological methods, to determine whether statistically significant pediatric cancer clusters exist in specific areas of Florida. These analyses used the updated dataset for the period 2000–2010. In parallel, the methods used in Amin et al. (2010) have been applied to the updated data, and the results are published as part of this special issue of Statistics and Public Policy.

2. THE DIFFICULTIES IN GETTING DETAILED PEDIATRIC CANCER DATA

The most challenging part of this kind of epidemiological study can be securing an accurate and detailed dataset for the cancer incidence cases. The most direct approach to requesting a “confidential dataset” in Florida is to contact the FCDS at the University of Miami. For a fee, the application is reviewed by a committee whose membership is kept anonymous throughout the data request period. If the data request is eventually approved, then the data request is forwarded to another committee at the FDOH to obtain Institutional Review Board (IRB) approval (for another fee). I tried only once to go through this process, and the first committee at the FCDS rejected my request for confidential data after many months of back-and-forth questions and answers. The main reason that was given to me was “there is insufficient depth in the epidemiological staff” on my research team. I am a statistics professor, and this seems to have been the primary barrier to being able to access the pediatric cancer data.

For the Amin et al. (2010) article, we were able to directly obtain a dataset from the Florida Association of Pediatric Tumor Programs (FAPTP) on all children who were treated at Florida hospitals for cancer during the period 2000–2007. Given the obstacles to procuring detailed cancer data from the state cancer registry, it was fortunate for us that Florida has a parallel system, the FAPTP, which has statewide pediatric cancer data, at a reasonable level of geographic specificity, going back to 1980. This updated dataset is the one used by the researchers in this special issue, and I would like to express my appreciation to FAPTP for allowing us all to use their dataset for this project.

3. TECHNICAL INFORMATION ABOUT THE FAPTP DATA

3.1. ZIP Codes and ZIP Code Tabulation Areas

Amin et al. (2010) used cancer incidence data for the years 2000–2007. The dataset lists the ZIP Codes of the address for each cancer case, as it is a standard practice to record information on the residence of cancer patients by hospitals. While the FCDS can provide the cancer data at the census tract level or even as geo-coded data, we did not have access to such fine-grained data. The FAPTP provides cancer case records that include age, gender, race, time of diagnosis, and ZIP code of the patient. The United States Postal Service ZIP Codes are a collection of mail delivery routes. ZIP codes change over time, and they are not readily suitable for a geospatial study. Instead, we used ZIP Code Tabulation Areas (ZCTAs), which are generalized areal representations of ZIP codes that the U.S. Census Bureau uses. ZIP codes and ZCTAs are generally similar but not identical. To match each ZIP code to the corresponding ZCTA (when this is possible), we would have to compare individual geo-coded addresses to the ZCTAs. Lacking the geo-coded addresses from the FCDS, this was not possible. Instead, we used an alternative methodology to convert ZIP codes to ZCTAs.
We replaced each ZIP Code of Residence in the FAPTP dataset with ZCTA2010, because the population size is provided only at the ZCTA Level by the U.S. Census Bureau. There are three cases:

1. The ZIP Code of Residence exists in the ZCTA for the year 2010. We keep this code.
2. The ZIP Code of Residence does not exist in the ZCTA for the year 2010. In that case, we manually assign a ZCTA on the basis of geographical software (ArcGIS) and a significant matching based on the other available covariates.
3. The ZCTA does not exist in the ZIP Code of Residence. These 37 cases were removed.

### 3.2. Data Cleaning

Amin et al. (2010) used the FAPTP dataset for the study period 2000–2007 without additional data cleaning, which may have been a limitation of study. We worked with the cancer cases in the dataset based on instructions that were given to us by one of the co-authors who was a childhood cancer oncologist. The updated FAPTP dataset for 2000–2010 was checked and cleaned to remove any obvious errors (typos, inconsistent records, impossible values, and so forth). The initial dataset for 2000–2010 had 7595 cases of cancer in its file. We deleted cases as follows:

- Duplicate cases.
- Cases with addresses that were outside Florida.
- Cases that had ages that were outside the study age period 0–19.
- Cases for which the ZIP codes did not have a corresponding ZCTA in the year 2010 (37 cases).

The data cleaning resulted in a data file with 6736 cases. This reduced dataset was shared with all of the authors whose analyses are appearing in the special issue of *Statistics and Public Policy*.

### 3.3. Estimation of Population Counts

Amin et al. (2010) covered the period 2000–2007, which included population data from the U.S. census in 2000. The Census Bureau does not provide population estimates at the ZCTA level for the years 2001–2009, which made it challenging to estimate the target population and thereby obtain the corresponding cancer rates. We chose a simple method to check for possible population change during this period in the southern part of Florida (where our work suggests that excess childhood cancers are located). After the most likely cluster was identified with the software SaTScan, the population growth was calculated from Census estimates from 2000 to 2007 for the cluster area and then for the rest of Florida. A test showed that the difference between the two growth rates was not significant. This was a limitation of that earlier study. But in the new study with the extended FAPTP dataset, we have data for the period 2000–2010, which allowed us to use the census for the years 2000 and 2010, and then to interpolate between 2000 and 2010 with SaTScan. The interpolation in SaTScan is linear, which is identical to the interpolation that is used by the U.S. Census Bureau. The cancer rates in the new study should be better estimates than those that were available in the previous article.

### Reference

Amin, R., Bohnert, A., Holmes, L., Rajasekaran, A., and Assanasen, C. (2010). “Epidemiologic Mapping of Florida Childhood Cancer Clusters,” *Pediatric Blood and Cancer*, 54, 511–518. [97,98]