# Which County is IT? When reported county does not match geocoded County

Recinda Sherman, MPH CTR
University of Miami Miller School of Medicine, Florida Cancer Data Systems

## I Introduction:

Florida Cancer Data Systems (FCDS) currently reports county level cancer rates based on the county denoted by the reporting facility. However, the process of geocoding cancer cases can often result in a change of "improvement" from one county to another—most often to a contiguous county or sometimes one in close proximity. This represents a problem with publishing cancer rates by county—which county is it?

We postulate a "move" to a contiguous county was often based on zip codes crossing county lines. And a move to a county in close proximity is often the result of the facility's county being reported instead of the patients. And, although this scenario is less common, we hypothesized that a cancer case reported in a county quite far from the geocoded county resulted from either data entry error or geocoding error.

FCDS is considering publishing cancer rates based on geocoded county. But as we consider this change, we need to understand the characteristics of cancer cases that are reported and geocoded to different counties.

This poster 1) describes the scope of the difference between rates using the reported county and rates using the geocoded county; and 2) describes the characteristics of cases that "move" across county lines to guide publication decisions.

## II Methods:

Invasive cases from 1981-2009 for Florida residents with a reported county at diagnosis were selected from the FCDS Commercial file—a static snapshot, annually appended with most recent and complete year of data, of the FCDS dataset that is used to produce age-adjusted rates for the Florida Annual Cancer Report.

All geocoding for FCDS is done by a proprietary vendor, Claritas/Nielson.

The reported county was compared to the county obtained through the geocoding process and delineated as missing, different, or same. Geocoded counties that were different from geocoded counties were further delineated as a county contiguous or non-contiguous to the reported county.

County movement (geocoding county does not equal reported county) was evaluated by demographic features (age, sex, race) as well as year of diagnosis, class of case, and report source. County movement whether the address at diagnosis was modified during the geocoding process and the NAACCR geocoding quality code, census tract certainty. County movement was also evaluated by county characteristics, the population size and number of hospital facilities

## III Results:

### Rate Changes:

Moving from using reported county to geocoded county for county level rates results in a net loss in age-adjusted rates for the majority of counties due to the lack of full geocoding coverage. Four low population counties increased rates using the geocoded county (see Figure 1). Using a composite county (based on geocoded county supplemented by reported county when geocoded county is missing) resulted in a net loss in about a third of the counties, the greatest being 14%, and a net gain in another third, the greatest being 40% (see Figure 2).

The changes in rates vary individually by county over time with the biggest impact on counties with medium size populations.

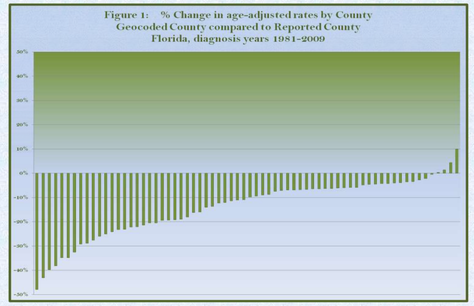
### County Movement:

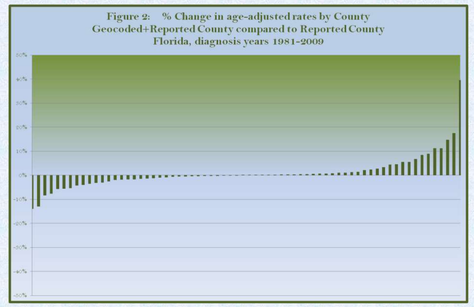
Overall, about 91% of the cases were geocoded to the same county that was reported. Slightly over 7% were not geocodable, 1.5% were moved to a contiguous county, and <1% were moved to a non-contiguous county. The changes in the percent of counties moved to non-contiguous counties decreased dramatically over time from 20% in 1981 to 0% in 2009 for an average of 4% of the counties that did not correlated between report and geocode. This trend parallels the decrease in the % un-geocoded over time, from 20% un-geocoded in 1981 to 3% in 2009, and likely reflects improvements in geocoding over time.

County movement (different reported county from geocoded count) did not differ by class, sex, race (although a greater percent of blacks were un-geocoded), or age. Death clearance only cases had a much lower percentage of county movement, but DCO's represent less than 3% of the cases. County movement also differed by match type. Cases geocoded to the zip code, had a higher percentage of cases moving to different counties)—5% of cases geocoded to zip code based on street address and 3% of cases geocoded to zip code based on a PO Box  moved counties.

Of the cases with same reported address as gecoded address (addresses that were not modified or "improved" during the geocoding process), 1.5% of  the cases with same moved to a contiguous county versus 2.1% with a different zip code and 1.4 % versus 2.4 % with a modified street address. The percent of county movement did not change based on modification of city during geocoding.

Counties that gained the most cases had fewer hospitals in the county; specifically fewer ACOS approved hospitals. The counties that gained the most cases were, in general, proximal to counties that lost the most cases.




## IV Conclusions:

It appears that the geocoded county is a more accurate placement than the reported county for a cancer patient's county at diagnosis.

The composite county, using the reported county only when the geocoded county is unavailable, appears to be the most complete and most accurate county variable to use for calculating county level cancer rates.

However, using the composite county will result in county level age-adjusted cancer rate changes that are artifacts, not real. Publishing the composite county rateswould require a coordinated public message to explain the difference.