# 2022 NPCR DISTRICT OF COLUMBIA SUCCESS STORY

District of Columbia Cancer Registry; Ian Quan and Maria Leuchert, RHIA, MHA, CTR

## Data Quality Improvement for Address at Diagnosis and Correction of Plastic Land Parcel Boundaries

## National Program of Cancer Registries
# SUCCESS STORY

## SUMMARY

In November of 2019, the DC Cancer Registry (DCCR) reviewed the 1996-2017 cases for completeness and correctness of geographic data. Visual review of DCCR's exported data identified gaps in blocks of older records which were not geocoded while other blocks displayed incorrect ward and census tract information for the date of diagnosis (ward boundaries are adjusted on the 2nd year of every decade, e.g., 2002, 2012, 2022 etc., while census tracts are adjusted at the start of every decade, e.g., 2000, 2010, 2020 etc.) While many of the records appeared to have correct coordinates, wards, and census tracts, to ensure the highest degree of completeness AND correctness, it was determined that a full re-geocode of the cancer registry database (1996-2017) was the most prudent approach.

The in-house geocoder, Master Address Repository (MAR), is a tool developed and maintained by the Office of the Chief Technology Officer of DC (OCTO). It accepts only DC addresses and returns coordinates in two geographic coordinate systems (latitude/longitude and UTM), current ward and census tract, matched address and zip code, and match score. Roughly 100,000 addresses can be geocoded in 90 minutes.

During the MAR linkage to the Rocky Mountain Cancer Database Systems (RMCDS), 71,100 records (with Dx dates between 1996 - 2017) were identified with 10,400 records failing to geocode. Over a 28-month period, the failed addresses were manually inspected and corrected, and the project was finally completed in April 2022.

## CHALLENGES

- DCCR's biggest challenge was due to attrition and repurposing of staff in response to COVID 19.

- Limitations with the MAR geocoder - it only returns [current] ward and census tract data, no historical delineations. This meant older records required alignment with their proper wards and census tracts through other means, i.e., GIS. Figures 1 and 2 show the differences between the current ward boundaries and previous iterations from different decades. If a patient from a previous decade received a cancer diagnosis AND reported a place of residence that fell within the orange area, the geocoder would return a false ward of residence. Only a spatial join within GIS software can correct these errors en masse.

**FIGURE 1.**
DIFFERENCES BETWEEN 2002-2011 AND 2022-2031 WARD BOUNDARIES
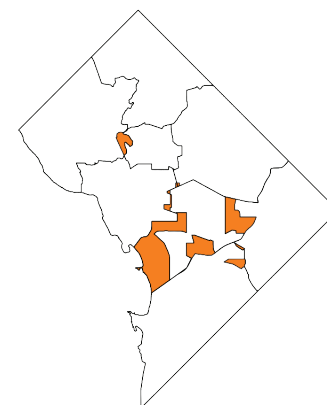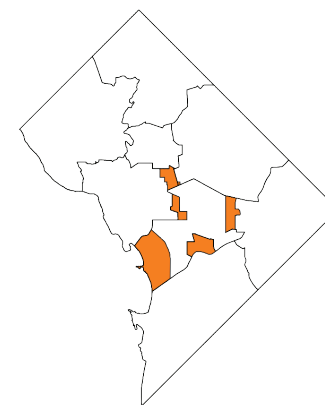


**FIGURE 2.**
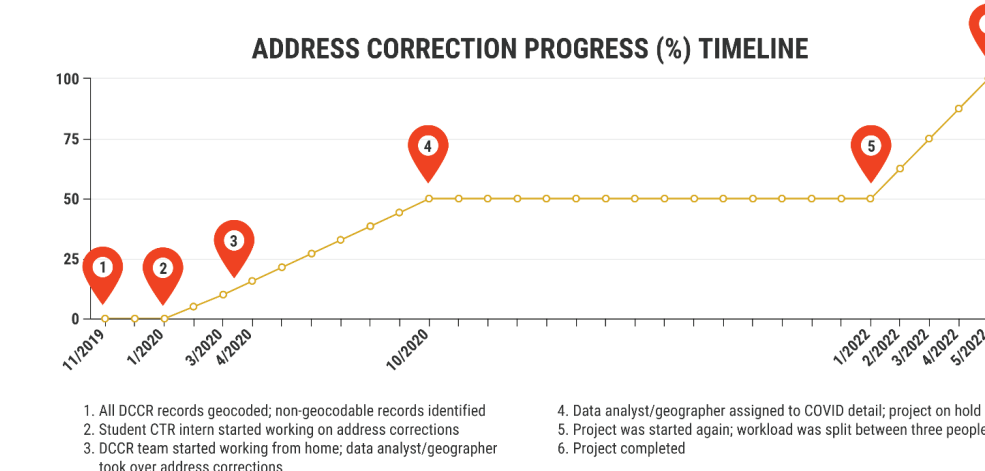DIFFERENCES BETWEEN 2012-2021 AND 2022-2031 WARD BOUNDARIES



## SOLUTION

To assist in better data collection, DCCR provided training to registry staff on the requirements of documenting address at diagnosis. DCCR requires all reporting facilities to collect only actual street address not including apartment, unit, or suite in this field.

DCCR staff reviewed 10, 400 records for incomplete addresses by reviewing the patient text fields (pathology and physical examination and treatment) and utilizing Lexis Nexis to verify residency. The DCCR registry intern completed 10% of the revisions prior to being released due to COVID-19. In an enormous team effort, DCCR staff reviewed and revised the remaining records. Expected project completion was estimated for June 2022, but project was completed in April 2022, two months ahead of schedule (See Figure 3).

## FIGURE 3



ADDRESS CORRECTION PROGRESS (%) TIMELINE

1. All DCCR records geocoded; non-geocodable records identified
2. Student CTR intern started working on address corrections
3. DCCR team started working from home; data analyst/geographer took over address corrections
4. Data analyst/geographer assigned to COVID detail; project on hold
5. Project was started again; workload was split between three people
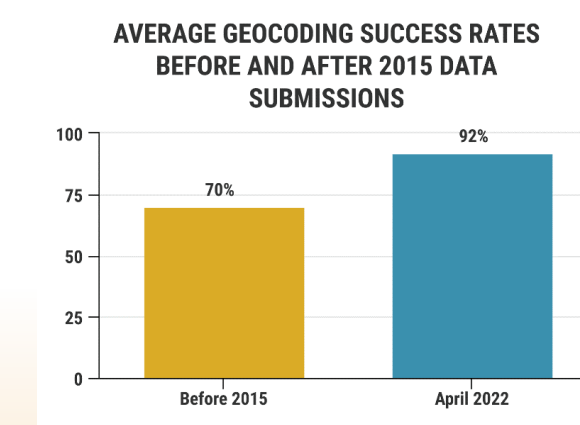6. Project completed

Upon completion of address corrections, the 10,400 records were linked to the MAR geocoder then corrected forward and census tract at time of diagnosis using QGIS.

Address errors that result in geocoding failures are expected but the number of such errors can be reduced with awareness of DC's unique address syntax. In January of 2020, DCCR implemented a new QA visual review process which includes the review of 25 randomly selected cancer cases from all DC reporting hospitals for every data submission. The address at diagnosis data items were one of the most common errors identified during the reviews.

## RESULTS

As of April 2022, DCCR's geocoding success rate increased from 70% in 2015 to 92% in 2022. (See Figure 4)

## FIGURE 4



AVERAGE GEOCODING SUCCESS RATES BEFORE AND AFTER 2015 DATA SUBMISSIONS

In November of 2019, a total of 10,400 un-geocoded addresses began to be reviewed by DCCR's student/intern who completed approximately 10% of the workload. The project was continued by DCCR's data analyst/geographer who was able to complete around 50% of the assignment.

However, in January of 2022, during the team approached review phase of all remaining 5,415 un-geocoded cases, various discoveries were made regarding patients' place of residence on original date of diagnosis. 672 cases (12.4%) of the total remaining 5,415 count were not residents of Washington DC at time of original diagnosis (See Figures 5 and 6).

- 240 cases belonged to neighboring states Maryland and Virginia.
- 159 cases were assigned P.O Boxes (unable to geocode) or addresses from DC military facilities (unable to geocode)
- 111 cases belonged to other US states.
- 56 cases belonged to other countries.
- 52 cases were determined to have unknown addresses.
- 19 cases were US cases of unknown states.
- 11 cases were military addresses abroad.

## FIGURE 5



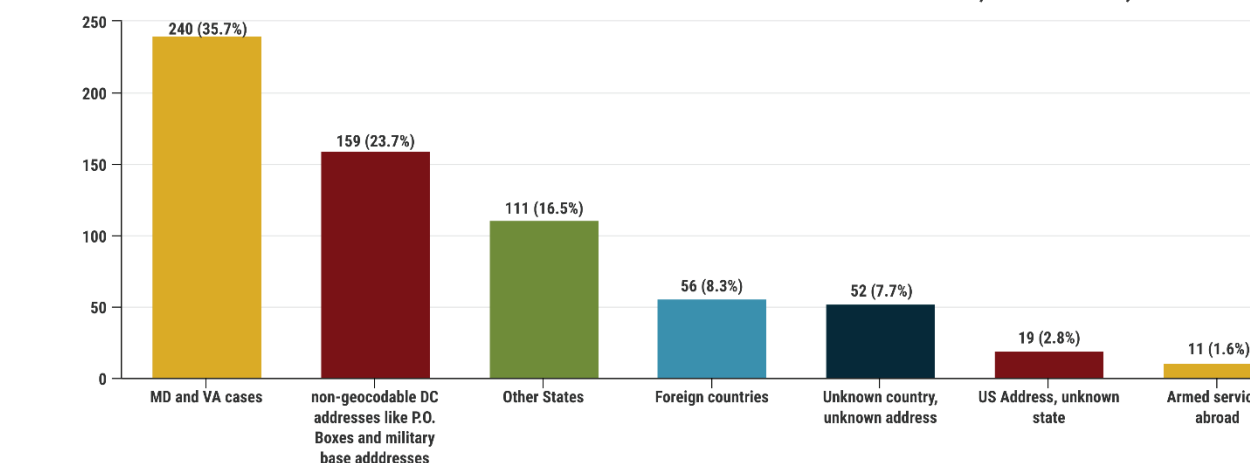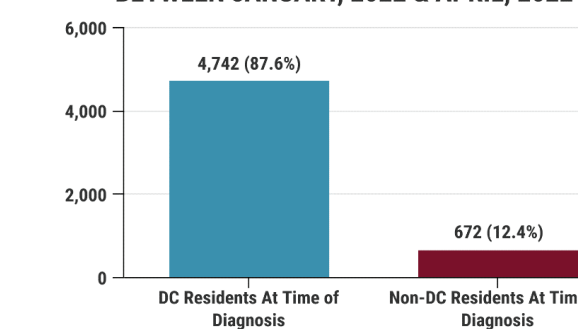BREAKDOWN OF NON-GEOCODABLE RECORDS CORRECTED BETWEEN JANUARY, 2022 & APRIL, 2022

## FIGURE 6



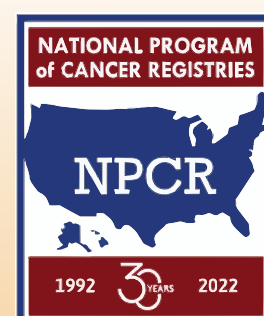RESULTS OF ADDRESS CORRECTION BETWEEN JANUARY, 2022 & APRIL, 2022

## SUSTAINING SUCCESS

DCCR will continue bi-annual, review of the "address at diagnosis" field focusing only on newly added records to the DCCR. A policy for this process will be included in the DCCR standard operating procedure manual containing step-by step instructions for identifying suspect addresses at diagnosis recorded at specific locations, such as U.S. embassies and DC hospitals.

## REGISTRY CONTACT INFORMATION

202-442-5955
DC Cancer Registry Website