

Mapping Accuracy: Addressing Data Errors in Virginia's Localities

AUTHORS: *Nikkia Ray, MPH; Larry Kirkland, MA; Shuhui Wang, MS*

SUMMARY

The Virginia Cancer Registry (VCR) faced challenges with data errors in the cancer incidence rates for localities with similar names or geographical proximity, such as Richmond City and Richmond County and Fairfax City and Fairfax County. These errors, caused by cancer registrars' unfamiliarity with Virginia's complex localities, led to inaccurate incidence rates that skewed public health data. In response, the VCR developed strategies to identify, correct, and prevent similar issues, focusing on data accuracy through improved quality assurance (QA) processes and new technology.

CHALLENGE

Inaccurate cancer incidence data was found for several Virginia cities and counties with similar or bordering geographical locations, including Richmond City and Richmond County and Fairfax City and Fairfax County. The errors were due to facility cancer registrars misclassifying counties, stemming from unfamiliarity with Virginia's localities. As a result, cancer incidence rates were inflated or skewed, affecting the reliability of data dashboards and reports. The inaccurate rates could undermine program transparency and public trust in data accuracy and public health decisions. The VCR took swift action to identify the root cause and implement corrective measures to resolve the issue and prevent future occurrences.

SOLUTION

A thorough review of cancer incidence data for 2016 to 2020 and 2017 to 2021 was conducted to identify discrepancies. The VCR used a combination of epidemiological analysis, data from the VCR website, and advanced geocoding tools like Texas A&M Geocoding System and Geocodio. These methods allowed for more accurate classification of counties and cities, ensuring correct data entry. The epidemiological review focused on error rates and trends in misreporting by locality and diagnosis year. A key component of the solution was to develop plans to implement tailored training seminars for cancer registrars at facilities where data entry errors were identified. Training depended on funding to support the education and training coordinator as the lead.

RESULTS

The key successes were identifying the problem and developing a robust strategy to prevent recurrences, although full resolution is an ongoing process. The strategy involved collaboration between data analysts, epidemiologists, and cancer registrars, and the use of new software tools like Geocodio. The new geocoding software corrected 98% of cases with improperly coded localities.

Lessons learned included the importance of tailored training for registrars on Virginia's localities and the value of technology in reducing manual errors. A key takeaway was the realization that automation—particularly in geocoding and data submission—could significantly reduce the need for manual oversight, leading to faster and more accurate data handling.

CONCLUDING REMARKS

The VCR will continue to improve its QA processes, with a focus on annual data submissions and further development of automated systems to streamline data collection and review. Efforts are ongoing to secure additional funding for data modernization and software development. The ultimate goal is to create a sustainable, automated process that minimizes manual intervention and reduces errors in future submissions. This ongoing work not only addresses the immediate issue but also positions the VCR for long-term success in maintaining high-quality, reliable cancer data for public health use.