Childhood Cancer Data Collection: A Trend Analysis of Completeness and Data Quality from NPCR-ECC (October 2012-October 2016 Submissions)

K Zhang[1], O Galin[1], J Stanger[1], Y Ren[1], S Ranasinghe[1], R Wilson[2], T Williams[2], L Douglas[2]

[1] ICF, Fairfax, VA; [2] Centers for Disease Control and Prevention, Atlanta, GA

INTRODUCTION

Cancer incidence in children is less common than in adults and thus it is problematic to obtain timely, meaningful and sufficiently large incident data sets to support scientific studies and treatment evaluation research. The Early Case Capture (ECC) of Pediatric and Young Adult Cancers (PYAC) program was created to address this issue. Built on the existing National Program of Cancer Registries – Cancer Surveillance System (NPCR-CSS), the ECC project captures state surveillance data on childhood cancers from the latest available year within 30 days of diagnosis.

This study examines trends in the case completeness percentage and data quality indicators of childhood cancer incidence data collected from the ECC system. This system collects data every six months, beginning with cases diagnosed in January of 2012 from a selected subset of NPCR states. Only five states that provided data in all of the submission periods were used in the analysis.

METHODS

The data selected for analysis were drawn from states that continuously participated in the data collection (between October 2012 and October 2016). The data quality indicators (DQI) were chosen from a set of demographic variables that were collected in every submission and consistently populated. For each indicator, the state aggregate was used to calculate the statistics presented in the analysis. The case completeness percentage was also presented as a state aggregate. This measure is based on the NAACCR methodology, but omits the adjustment for background mortality due to the limitation of mortality data. However, it still compares each state's incident count with an expected incident count that is derived from national data. To minimize the issues inherent in using small case counts, cases were pooled across race, sex, and site and grouped cases into five age groups.

RESULTS AND DISCUSSION

Part I. To examine state level trends, the completeness percentage for cases diagnosed in 2012 was followed across all of the 6 month data submission periods.

Figure 1. State Level Case Completeness Percentages

Figure 2. Median Case Completeness Percentage (all states)
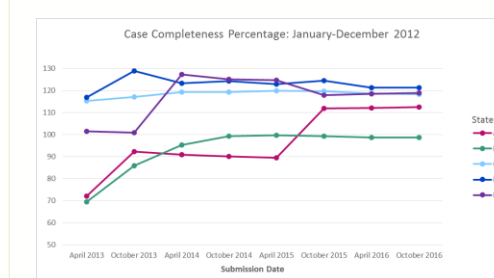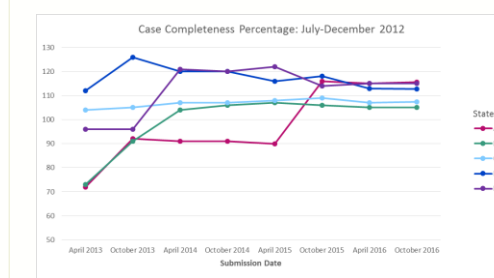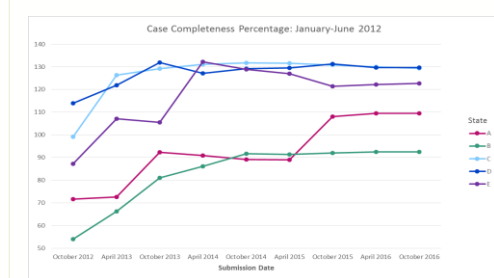
We observed a notable improvement in the case completeness percentage with later submissions having higher percentages than earlier submissions. For most of the states the case completeness percentage stabilized after the third data submission.
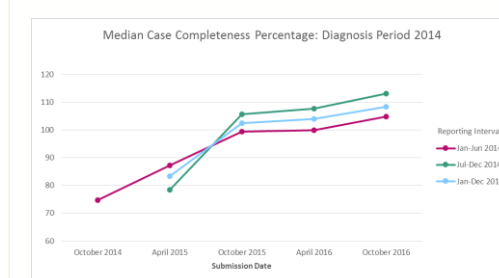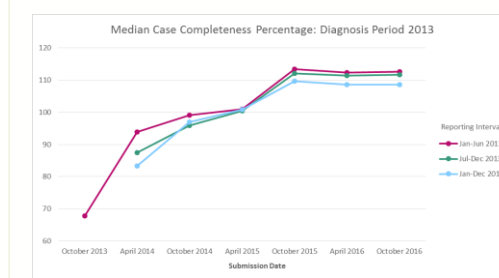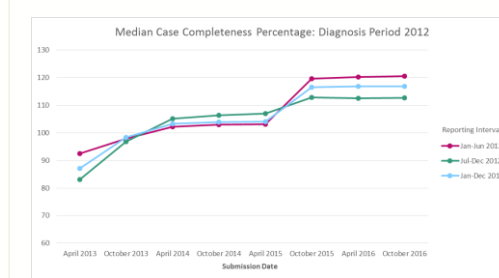
To compare the 6 and 12 month reporting intervals, the median case completeness percentage of all states was shown for cases diagnosed in 2012, 2013, and 2014.

A comparison between the projected case completeness percentage at 6 month intervals and the corresponding observed percentage at 12 months shows a strong correlation between the two measures.

Part II. Trends in data quality were assessed by examining the percentage change of invalid and/or blank values for selected variables, or data quality indicators (DQI), across data submissions. Only DQIs with an error percentage of 5 or greater were considered for the analysis. These were: Spanish/Hispanic Origin; Diagnostic Confirmation, County at Diagnosis; Laterality; and Race. The figures below present the results as all states combined.

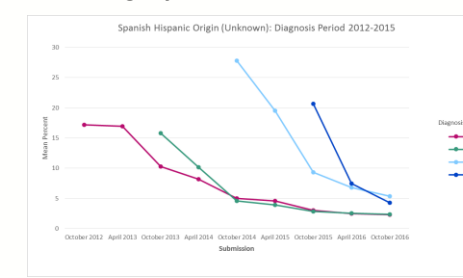*Figure 3. Spanish/Hispanic Origin: Mean Percentage of Unknown Cases*

Figure 4. Spanish/Hispanic Origin: Mean, Maximum and Minimum Percentage for Unknown Cases

Figure 5. County at Diagnosis: Mean Percentage of Unknown Cases

Figure 6. County at Diagnosis: Mean, Maximum and Minimum Percentage for Unknown Cases

Figure 7. Race: Mean Percentage of Missing and Unknown Cases

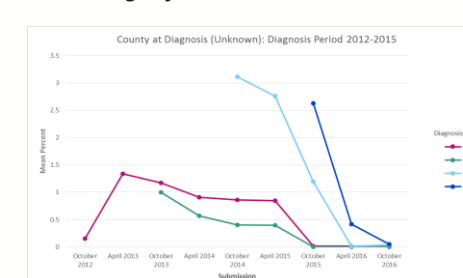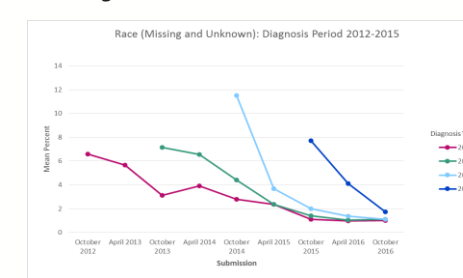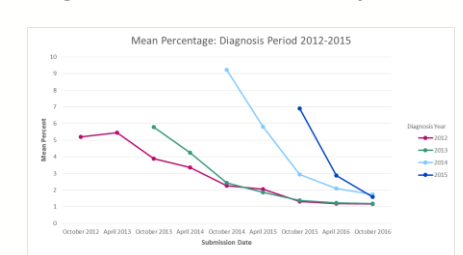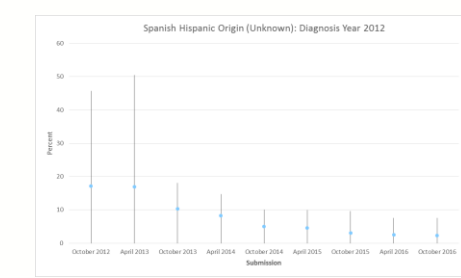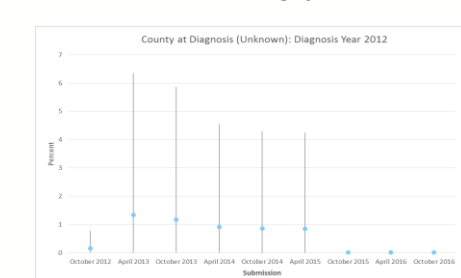*Figure 8. Race: Mean, Maximum and Minimum Percentage for Missing and Unknown Cases*

Figure 9. Combined Data Quality Indicators

The average percentage of invalid and/or blank values found in the Data Quality Indicators generally improved and stabilized after several data submissions. This trend is similar to the pattern of improvement observed in the case completeness percentage.
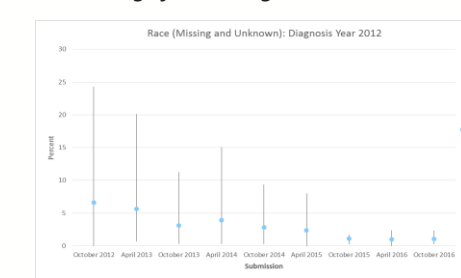
CONCLUSIONS

We observed that both the case completeness percentage and data quality indicators gradually improved over the first few submission periods and then stabilized. We also found that case completeness percentage of data submitted at 6 month intervals tracked closely with the same data submitted at the 12 month intervals.

ACKNOWLEDGEMENTS

CONTACT INFORMATION

Kevin Zhang kevin.zhang@icf.com

Division of Cancer Prevention and Control

RELIABLE   TRUSTED   SCIENTIFIC

National Center for Chronic Disease Prevention and Health Promotion
Division of Cancer Prevention and Control