A comparison of methods for assessing completeness of case ascertainment in data from the National Program of Cancer Registries

A. Blythe Ryerson, PhD, MPH

Cancer Surveillance Branch

2014 NAACCR Annual Conference — Ottawa, Ontario, Canada — June 21-26, 2014

BACKGROUND

- Research and policies meant to reduce the breast cancer burden depend on cancer ascertainment data.
- Data must meet high standards of quality and reliability.
- Completeness of incident case ascertainment is an essential component of quality reliability.
- Incomplete data results when registries are unable to collect accurate information on all incident cancer cases across specific geographic area within the given timeframe.
- Some registries have not implemented systems for collecting all incident cases.
- Some made incomplete reporting decisions.
- Some data collected incompletely.
- Methods of Completeness - Quantifies the percentage of actual incident cancer cases that could not be collected by a cancer registry.
- Estimating the truth observed vs. expected.
- The actual reported number of incident cases is unknown.
- The expected cases must be estimated from available data.

METHODS

I/M Ratio (SEER)

- Assumes the rate of age-adjusted incidence to mortality rates is constant across geographic areas for a given cancer site (I/M ratio, sex below & black only), and gender and combined
- Completeness indices weighted by race (white and black only) and gender and combined
- Uses national incidence rates from SEER only
- Estimated incidence rates (observed vs. expected)
- Easy to implement
- No variance estimates
- Includes linear relationship
- Assumes time period.
- Used national incidence rates from SEER only
- Makes use of only white and black race groups
- No variance estimates

I/M Ratio (NPCR)

- Can we improve estimates of expected incidence in NPCR registries by utilizing national NPCR incidence rates?
- Completeness indices weighted by race (white and black only), and gender and combined
- Uses national incidence rates from NPCR only
- Expected
- Simple method
- Easy to implement
- No variance estimates

Methods of Least Squares (Simple Linear Regression)

- SLR (Incidence) Completeness
- SLR (Mortality) Completeness

- Uses national incidence rates from SEER only
- Expected
- Simple method
- Easy to implement
- No variance estimates

OTHER METHODS

- Modifications of I/M ratio
  - Simple linear regression by other race and/or ethnicity
  - Example: Restrictions of I/M ratio calculations to Hispanic only in the Puerto Rico registry
- Multiple Linear Regression and/or Linear Transformation
  - Allows for adjustment for confounding
  - Allows non-linear relationships
  - Estimation methods
  - Weighted data for more recent years
  - Limitations
    - Would not identify ‘chronic under ascertainment’
    - May not identify confounders/misclassification
    - Doesn’t take into account correlated data
- Data factor/Dynamic Panel Models
  - Allows for trend over time
- Spatial Prediction Models
  - Includes mortality rates — can be seen as an extension of I/M ratio method
  - Incorporates information on socioeconomic status, geographically.
  - Health
  - Lifestyle factors
- Additional methods not yet explored:
  - Includes spatial/autoregressive models for areas with low case counts

FUTURE DIRECTIONS

- Find the best model
  - Accurate prediction
  - Easy to implement at registry level
  - Collaborate with partners

CONTACT INFORMATION

A. Blythe Ryerson, PhD, MPH
Aryerson@cdc.gov 770.488.2426

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