



New Treatment Data Sources

Kathy Cronin

Surveillance Research Program

NCI Working Group to Think About Two Questions:

- How will this data be used within DMS (in-house by registries, NCI and IMS)?
 - Evaluate the unique contributions of a particular data source
 - Evaluate the representativeness and completeness of each data source
 - Use the data to develop algorithms for automation, including consolidation rules for multiple data sources
 - Other?
- How will data be released to researchers?
 - Present the data in a way that allows researchers to assess the potential biases associated with augmented data in their study
 - Communicate which data elements have been supplemented through linkages with additional data sources

Defining Requirements

After a few discussion we thought it was best to start with how the data will be release and work our way back to what will be needed from DMS.

- Initial question - what will researchers need to know when analyzing treatment data?

Other than the treatment data itself, researchers would need information to assess biases either introduced or reduced by new data sources

Population Representativeness

Did supplementing a portion of the cases introduce bias?

- Example - comparing treatment by age when over 65 was linked to Medicare claims
- Data need - **Identify cases that were and were not supplemented**
 - Allow researchers to limit the analysis to cases with more complete information
 - Assess representativeness of the cases with additional treatment information by comparing cases with and without supplemented information
 - Investigate geographical differences in treatment rates

Identifying the Appropriate Denominator

Identifying the appropriate inclusion criterion for an analysis?

- Example: estimating the percent of cases that received a treatment in SEER-MEDICARE the analysis would be limited to only cases that had MEDICARE coverage
- Data Need – **Identify cases that have similar opportunity to obtain treatment information**
 - In some cases the appropriate denominator for an analysis would be cases that linked to supplemental data sources even if there was not additional treatment information

Distinguishing Between “No” and “Unknown”

Identifying when the absence of treatment information implies none was given

- Example: When linking with CVS data, does the absence of a claim imply a cases did not receive a treatment when establishing a comparison group that did not receive a treatment?
- Data Need - **Provide guidance on the completeness of different data sources**
 - Allow researchers to decide if a case should be classified as no, unknown, or no/unknown

Additional Variables

- Variables to identify:

	Cases linked to additional data sources	Cases not linked to additional data sources
Adjuvant treatment information found		
Adjuvant treatment information not found		

- Not necessary for every potential data source, maybe by type of data source
 - Claims IV Chemo, Claims Oral Chemo, Data from Radiation Facilities, EMR information ...

Multiple Levels of Data Sets Containing Treatment Information

- Received chemotherapy/radiation, yes or no
 - Publicly available with treatment data use agreement
- Specific chemotherapy agent or type of radiation
 - Special data request with additional cautions related to used of the data and potential biases
- Detailed longitudinal information including agents, dates, dosage
 - Require IRB approval and appropriate tracking of data

Understanding How New Treatment Sources Will Incorporated In DMS

Very Generally - As new treatment information comes into the registry, a treatment page is created (TX page) in DMS which contains information on the source of the information and a summarization function is triggered which automatically fills in certain fields.

- This information can be used to create variables or flags would be helpful for evaluation and eventual data release

Questions For Discussion

- What additional variables or flags would be helpful for either evaluating or release new data?
- Are there procedures that can be put in place as part of the work flow that support the release of data later on?
- Are there specific functions that you would like to see added to DMS to evaluate data sources as they come into DMS?
- How will the summarization function be used to evaluate and combine new data sources?