

The Quality of Cancer Registry Data: Preliminary Results from CDC-NPCR's Patterns of Care Study

Robert R. German, DrPH, MPH; Tim Byers, MD, MPH;
Holly Wolf, PhD, MSPH; Jen Wike, MBA, MPH;
Linda Alley, PhD; and PoC Study Group

June 2006 NAACCR annual conference



SAFER • HEALTHIER • PEOPLE™



Acknowledgement of Authors

Robert R. German, DrPH, MPH*

Tim Byers, MD, MPH†

Holly Wolf, PhD†

Jen Wike, MBA, MPH*‡

Linda Alley, PhD*

* CDC's National Program of Cancer Registries, Atlanta, GA

† University of Colorado Health Services, Denver, CO

‡ Northrop Grumman Corporation, Atlanta, GA

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of CDC.

Agenda

- Background on CDC-NPCR's Breast, Colon, and Prostate Cancer Data Quality and Treatment Patterns of Care (PoC) Study
 - Data quality component
- Results of data quality analysis
- Discussion and concluding remarks

PoC Study, #1

- In 1999, the Institute of Medicine (Hewitt and Simone, 1999) recommended the use of existing data, such as CDC-NPCR's,
 - To measure variations in the use of appropriate standards of care
 - To assess the quality of cancer care in the US.
- A year later, the IOM (Hewitt and Simone, 2000) noted that some cancer patients may not be receiving the care known to be most effective for their condition

PoC Study, #2

- Collaborative study (McDavid et al., 2004)
 - CDC's National Program of Cancer Registries and statewide cancer registries in CA, CO, IL, LA, NY, RI, SC, and DC
- Main objectives
 - Assess quality and completeness of cancer data collected by registries
 - Evaluate the extent to which patients are receiving stage-specific, guidelines-based cancer treatments

PoC Study, #3

- Simple random samples
 - Localized female breast cancer (C50.0-C50.9)
 - Regional 3 or 4 colon cancer (C18.0-C18.9)
 - Localized prostate cancer (C61.9)
- Re-abstraction of data on patient, tumor, treatment
 - Hospital charts
 - Physician offices
 - Ambulatory surgery centers
 - Radiation therapy centers

Data Quality Analysis in PoC, #1

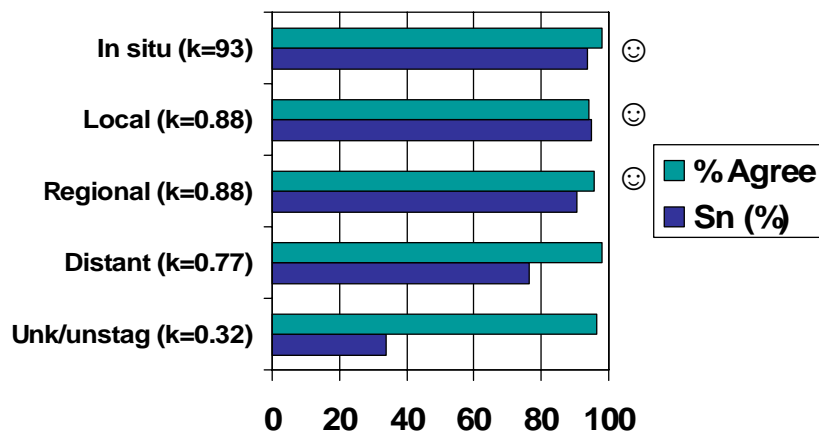
- Data from all states in PoC except DC
- Comparison of re-abstracted data with those originally coded in the registry
 - Original data archived or “frozen”
 - Merged by state and case identification number
 - Single primaries only

Data Quality Analysis in PoC, #2

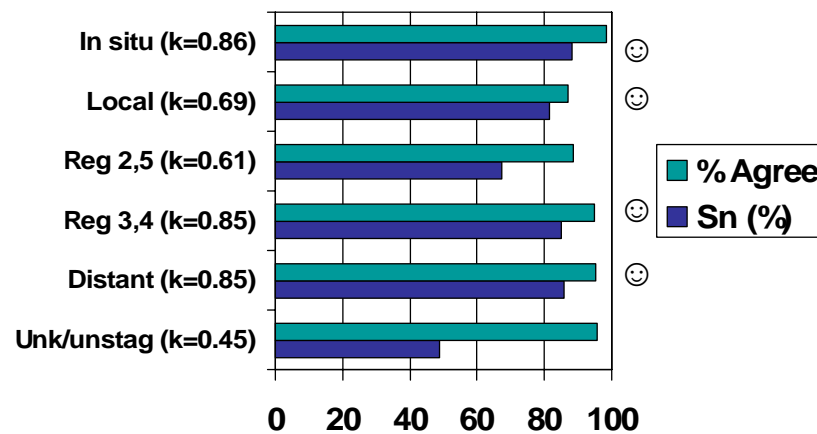
- All three stage-specific cancers in PoC
 - Stage at diagnosis with CONCORD data in PoC
- Measurements of agreement: kappa statistic, % agreement, sensitivity (Szklo and Nieto, 2000)
 - Re-abstracted data as ‘Gold Standard’
 - Good quality or better: $k \geq 0.60$ and % agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001)
- SEER summary stage 1977, race/ethnicity, breast and prostate grade, breast histology, surgery, radiation, chemotherapy, hormone therapy, and biologic response modifier

SEER Summary Stage Agreement, CONCORD data in PoC Study, DY 1997-1998

Female breast cancer (n=3,653)



Colon cancer (n=2,566)

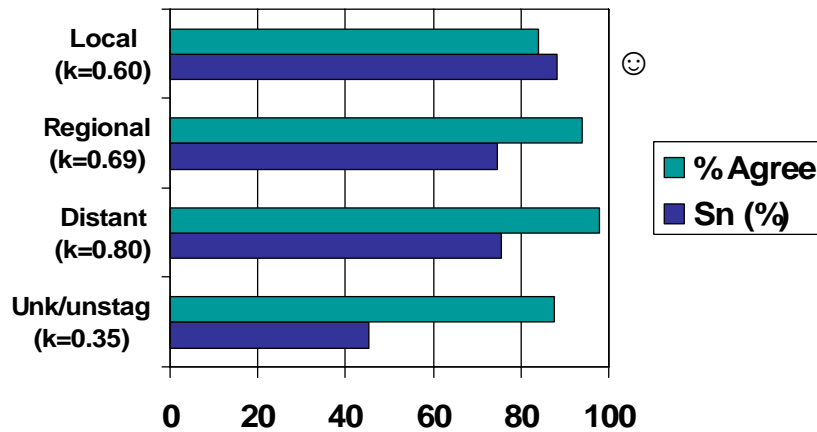


k = kappa statistic
% Agree = percent agreement
Sn = sensitivity

☺ Good quality or better based on kappa ≥ 0.60 and each of percent agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001).

SEER Summary Stage Agreement, #2

Prostate cancer (n=3,503)



k = kappa statistic
 % Agree = percent agreement
 Sn = sensitivity

☺ Good quality or better based on kappa ≥ 0.60 and each of percent agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001).

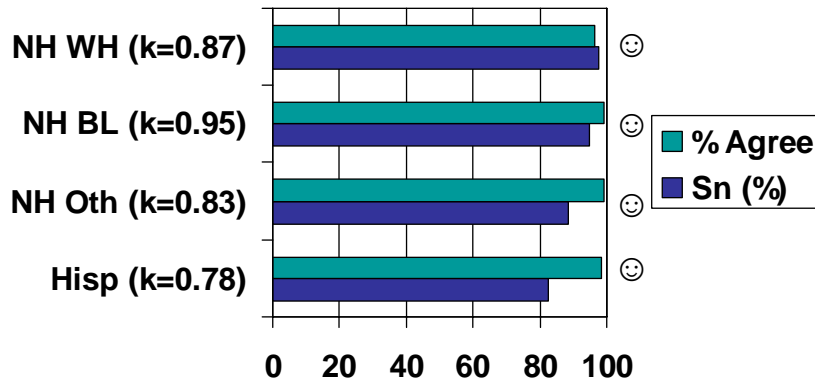
Summary of n-by-n tables

Cancer	% unknown/ unstaged re-abstracted to stage	% staged re-abstracted to unknown/ unstaged
Female breast	66%	2%
Colon	55%	2%
Prostate	61%	6%

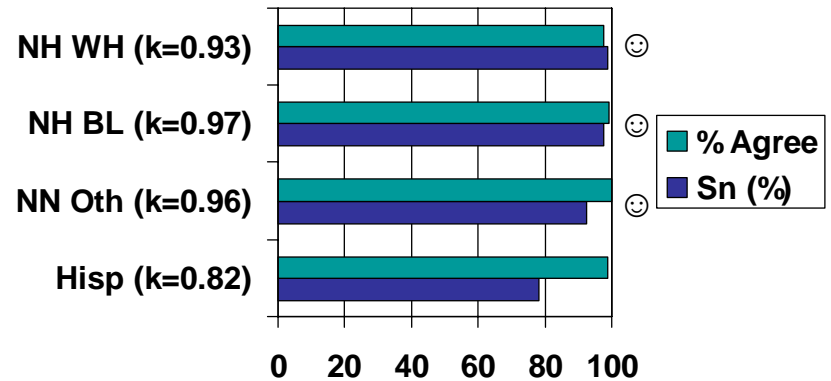
► More unknown/unstaged female breast cancers in re-abstracted data (n=101) than in original data (n=99)

Race/Ethnicity Agreement, PoC Study, DY 1997-1998

In situ or localized female breast cancer (n=3,551)



Regional 3 or 4 colon cancer (n=1,328)



Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

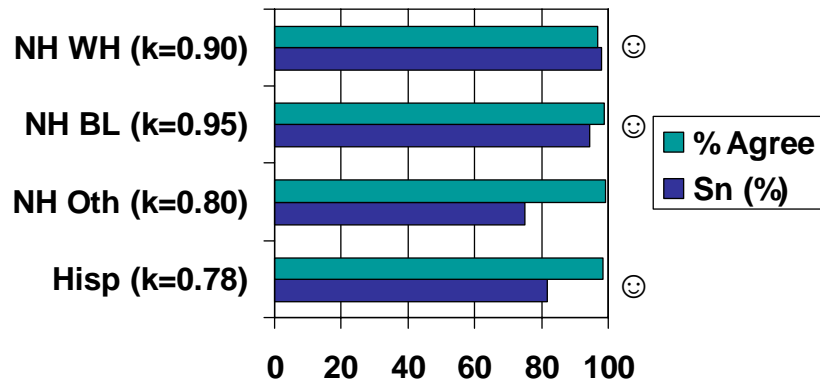
NH WH = non-Hispanic white
 NH BL = non-Hispanic black
 NH Oth = non-Hispanic other
 Hisp = Hispanic, all races

k = kappa statistic
 % Agree = percent agreement
 Sn = sensitivity

☺ Good quality or better based on kappa ≥ 0.60 and each of percent agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001).

Race/Ethnicity Agreement, #2

Localized prostate cancer (n=3,281)



Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

NH WH = non-Hispanic white
 NH BL = non-Hispanic black
 NH Oth = non-Hispanic other
 Hisp = Hispanic, all races

k = kappa statistic
 % Agree = percent agreement
 Sn = sensitivity

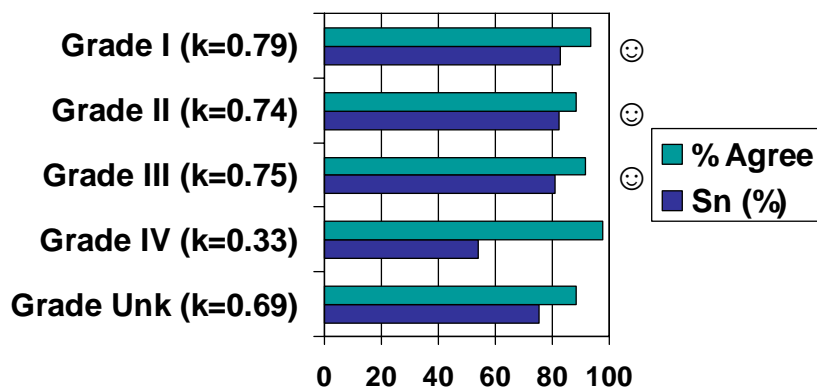
☺ Good quality or better based on kappa ≥ 0.60 and each of percent agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001).

Summary of n-by-n tables

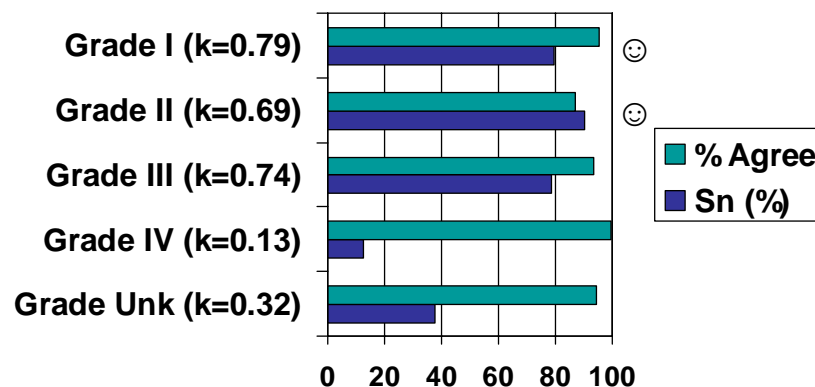
Cancer	% unknown race/ethnicity re-abstracted to race/ethnicity	% race/ethnicity re-abstracted to unknown race/ethnicity
In situ or localized female breast	76%	0.4%
Regional 3,4 colon	80%	0.2%
Localized prostate	74%	0.5%

Grade Agreement, PoC Study, DY 1997-1998

**In situ or localized
female breast cancer (n=3,549)**



**Localized
prostate cancer (n=3,281)**



Missing responses are excluded.

Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

k = kappa statistic.

% Agree = percent agreement

Sn = sensitivity

☺ Good quality or better based on kappa ≥ 0.60 and each of percent agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001).

Grade agreement, #2

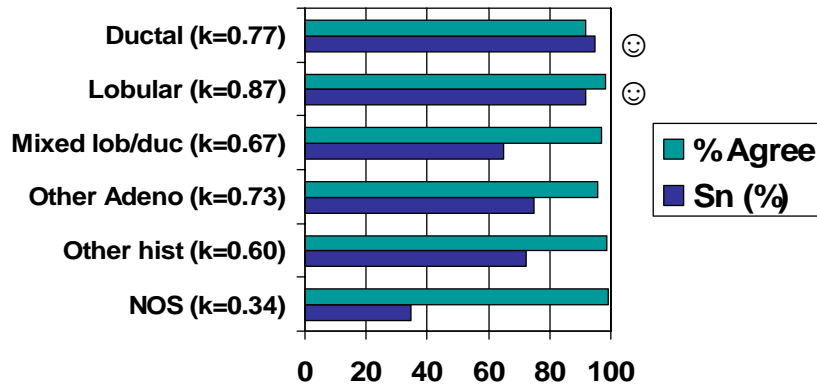
Summary of n-by-n tables

Cancer	% unknown grade re-abstracted to grade	% grade re-abstracted to unknown grade
In situ or localized female breast	21%	9%
Localized prostate	67%	3%

- More unknown grade for female breast cancers in re-abstracted data (n=919) than in original data (n=877)

Histology Agreement, PoC Study, DY 1997-1998

**In situ or localized
female breast cancer (n=3,551)**



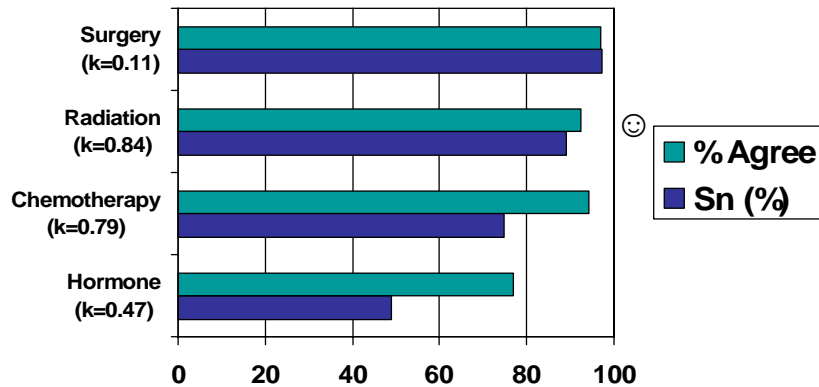
Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

k = kappa statistic
% Agree = percent agreement
Sn = sensitivity

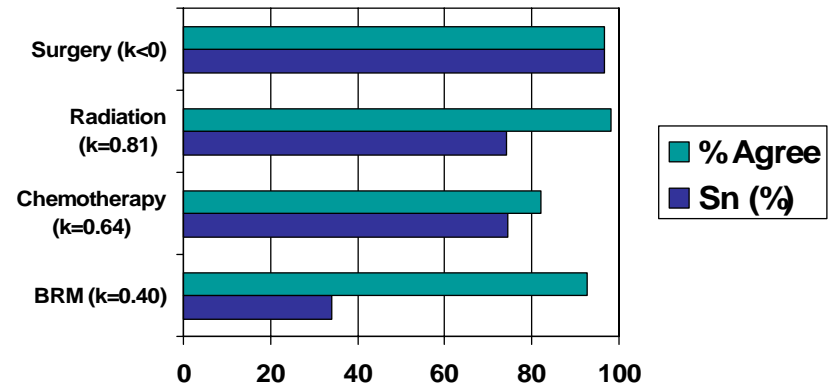
☺ Good quality or better based on kappa ≥ 0.60 and each of percent agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001).

Treatment Agreement, PoC Study, DY 1997

**In situ or localized
female breast cancer (n=3,019)**



**Regional 3 or 4
colon cancer (n=1,059)**



Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

Chemotherapy for colon cancer excludes missing responses.

k = kappa statistic; k<0 is not significant.

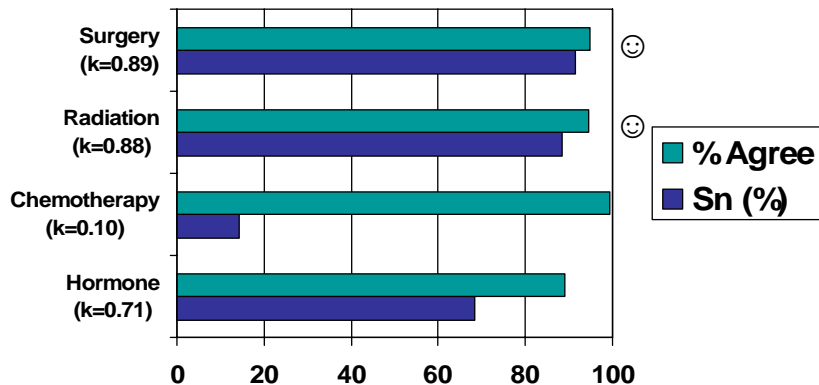
% Agree = percent agreement

Sn = sensitivity

☺ Good quality or better based on kappa ≥ 0.60 and each of percent agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001).

Treatment Agreement, #2

Localized prostate cancer (n=2,830)



Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction

k = kappa statistic

% Agree = percent agreement

Sn = sensitivity

☺ Good quality or better based on kappa ≥ 0.60 and each of percent agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001).

Summary of n-by-n tables

- Across these 12 treatment variables, there were no more than 15 unknowns for any variable in the original data
- Except for colon chemotherapy, all of the unknowns in the original were re-coded in the re-abstraction
- For 9 of the 12 variables (75%), there were more unknowns in the re-abstracted data ($n \leq 108$) than in the original data ($n \leq 15$)

Discussion, #1

- Results provide a check on quality of data recorded in population-based central cancer registries
- In this analysis, e.g.,
 - Localized stage achieved good quality or better for all three cancers, but other stages did not
 - Information on non-Hispanics whites and blacks was good quality for all three cancers, but not other groups
 - Very few treatment variables met the data quality criteria

Discussion, #2

- The analysis
 - Indicates the quality of variables routinely collected by cancer registries can vary by cancer site and across variables
 - Appears to support the observation that the use of supplemental sources of data (e.g., non-hospital) may be necessary to achieve good quality information on cancer treatment

Discussion, #3

- Limitations
 - Wide variation in state-specific measurements
 - Analysis is one measure of quality in cancer registry data
 - Changes in reporting requirements over time
 - “Gold standard” may not as pure or complete as preferred
 - Analyses are preliminary

Concluding Remarks

- Changes in cancer treatment regimens require assessing quality of cancer registry data
- The data quality analysis in the PoC studies guide efforts to improve the quality of data disseminated by a cancer registry
- Future PoC studies will apply lessons learned from the current analyses

References

- Cicchetti DV. The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology* 2001; 23(5): 695-700.
- Hewitt M, Simone JV. *Enhancing data systems to improve the quality of cancer care*. National Cancer Policy Board, Institute of Medicine, and Commission on Life Sciences, National Research Council. Washington, DC: National Academy Press, 2000.
- Hewitt M, Simone JV. *Ensuring Quality Cancer Care*. National Cancer Policy Board, Institute of Medicine, and Commission on Life Sciences, National Research Council. Washington, DC: National Academy Press, 1999.
- McDavid K, et al. Rationale and design of the National Program of Cancer Registries' breast, colon, and prostate cancer patterns of care study. *Cancer Causes and Control* 2004; 15: 1057-66.
- Szklo M, Nieto FJ. *Epidemiology: Beyond the basics*. Gaithersburg, MD: Aspen Publishers, Inc., 2000, pages 343-404.