How We Know What We Know About Cancer Data

A Conversation with Dr. Nadia Howlader, NCI Surveillance Research Program

LBCA Advocate Chat Series 2022
How We Know What We Know About Cancer Data

Dr. Nadia Howlader
National Cancer Institute
Surveillance Research Program
Cancer Surveillance, SEER Registries & Resources for Research, and Future Directions

Nadia Howlader, PhD
Surveillance Research Program, DCCPS

Lobular Breast Cancer Alliance
September 21, 2022
Outline

• Introduction to SEER
• Clinically Relevant Data Presentation and Analysis
• Challenges and Opportunities with Cancer Registries
• Selected New Initiatives to Enhance Cancer Registries
Surveillance, Epidemiology, and End Results (SEER)

• SEER mission to support research
  o Diagnosis, treatment, and outcomes of cancer since 1973
  o Provide baseline data on U.S. cancer incidence and survival trends

• Population-based registries representing 48% of the U.S. population

• Over 850,000 incident cases reported annually
SEER Data: Quality and Timeliness

• Data quality is an important part of keeping data consistent and a reliable source for cancer statistics

• North American Association of Central Cancer Registries (NAACCR) issues certification for data quality including:
  o Case ascertainment 95% or higher completeness
  o Fewer than 3% of cases based on death certificates only
  o Less than 2% of cases have missing age, sex, and county

• Three year lag between diagnosis and reporting of cancer
  o November 1st 2021 data submission -> cases diagnosed through 2019 -> Data release on April 15, 2022
• Dark Blue represents Core Registries (Reporting data)

• Light Blue represents Research Support Registries (participate in special projects)
Evolution of SEER Over Time

- SEER 9 covering years 1975+
  - San Francisco-Oakland, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta
  - Covers 9.4% of the US population

- SEER 13 covering years 1992+
  - SEER 9 plus San Jose-Monterey, Los Angeles, Rural Georgia, Alaska Natives
  - Covers 13.4% of the US population

- SEER 18 covering years 2000+
  - SEER 13 plus California (excluding SF/SJM/LA), Kentucky, Louisiana, New Jersey, Georgia (excluding Atlanta and Rural Georgia)
  - Covers 27.8% of the US population
Evolution of SEER Over Time (Cont.)

• SEER 21 covering years 2000+
  - SEER 18 plus Idaho, New York, and Massachusetts
  - Covers about 35% of US population
  - As of June 2021, SEER now covers 48% of the US population, represents >850,000 incident cancers reported annually
Population Coverage by Race/Ethnicity

PERCENT OF THE POPULATION COVERED BY RACE

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>SEER 9</th>
<th>SEER 13</th>
<th>SEER 18</th>
<th>SEER 21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>9.4</td>
<td>13.4</td>
<td>27.8</td>
<td>34.6</td>
</tr>
<tr>
<td>White</td>
<td>8.7</td>
<td>11.5</td>
<td>24.9</td>
<td>31.9</td>
</tr>
<tr>
<td>Black</td>
<td>8.8</td>
<td>11.3</td>
<td>25.6</td>
<td>30.7</td>
</tr>
<tr>
<td>American Indian/Alaska</td>
<td>12.5</td>
<td>19.3</td>
<td>30.6</td>
<td>49.3</td>
</tr>
<tr>
<td>Asian</td>
<td>17.7</td>
<td>31.0</td>
<td>50.4</td>
<td>57.5</td>
</tr>
<tr>
<td>Native Hawaiian/Pacific</td>
<td>43.3</td>
<td>49.9</td>
<td>66.5</td>
<td>68.5</td>
</tr>
</tbody>
</table>

National Cancer Institute
How Does SEER compare to US?

- Foreign Born: SEER 17.9%, US 13.2%
- Less than H.S. Diploma (person aged 25+): SEER 14.2%, US 13.0%
- Below Poverty Level: SEER 15.3%, US 15.1%
Data in SEER Registries

• **Demographic**: age, gender, area of residence, race and ethnicity, insurance status, marital status.

• **Tumor (cancer)**: primary cancer site, histology, morphology, stage, lab values and tumor markers.

• **Treatment**: treatment in hospital more complete than outpatient (e.g., first course chemotherapy, surgery, radiation).

• **Outcome** (follow-up for vital status): living or deceased, month and year of death and cause of death.

• SEER data is linked at the county level with Census data and provide socio-economic variables at the county of residency.

• **Data is consolidated and available for analyses in SEER*Stat**
Clinically Relevant Tumor Markers Collected by SEER

- **Breast** (ER/PR, HER2, Multigene assays)
- **CRC** (CEA, KRAS)
- **Testis** (hCG, AFP, LDH)
- **Pharynx** (HPV)
- **Liver/ biliary tract** (AFP)
- **Oropharyngeal** (HPV)

- **Ovary** (CA-125)
- **Neuroendocrine** (Serum Chromogranin, urinary 5-HIAA)
- **Prostate** (PSA)
- **Hematologic Malignancies** (JAK2)
- **Melanoma** (LDH)
Cancer Statistics Review – Most recent cancer incidence, mortality, survival, prevalence and lifetime risk statistics

Fact Sheets – Plain language summaries of key statistics by cancer site
http://seer.cancer.gov/statfacts/

SEER*Explorer
https://seer.cancer.gov/explorer/

State Cancer Profiles (Interactive Maps)
http://statecancerprofiles.cancer.gov/

Annual Report to the Nation - provides an annual update of cancer incidence, mortality, and trends in the United States.
How SEER Data Can be Accessed?

• Two ways to obtain the SEER research data: both require access through an Internet connection.

• 1. SEER*Stat's Client-Server Mode

• 2. Download Compressed Files

https://seer.cancer.gov/data/options.html
Standard Statistics Reported Using SEER Data
Standard Statistics Reported Using SEER Data

- Incidence: Rate per 100,000
- Trends in incidence, annual percent change in rates, or average annual percent change over a specified time frame
- Prevalence of people alive with a previous diagnosis of cancer
- Cancer Survival Statistics (Relative or Cause Specific Survival)
- Probability of developing or dying of cancer over a lifetime
All Cancer Sites Combined

Long-Term Trends in SEER Age-Adjusted Incidence Rates, 1975-2019
By Sex, Delay-adjusted SEER Incidence Rate, All Races, All Ages

Legend (Sex)
- Female
- Male

Rate per 100,000

Year of Diagnosis

National Cancer Institute
All Cancer Sites Combined
Long-Term Trends in SEER Age-Adjusted Incidence Rates, 1975-2019
By Sex, Delay-adjusted SEER Incidence Rate, All Races, All Ages

2015-2019: 5-year trend
Males - 0.2
Females 0.0
Increasing or Decreasing Cancer Incidence Trends in the US, Among Men and Women

**MEN**

- Liver & Intrahepatic Bile Duct: 2.7%
- Melanoma of the Skin: 2.3%
- Thyroid: 1.9%
- Kidney & Renal Pelvis: 1.5%
- Myeloma: 1.3%
- Oral Cavity & Pharynx: 1.2%
- Pancreas: 1.0%
- Leukemia: 0.7%
- Non-Hodgkin Lymphoma: -0.3%
- Esophagus: -0.5%
- Brain & Other Nervous System: -0.6%
- Stomach: -0.6%
- Bladder: -0.9%
- Colon & Rectum: -1.5%
- All Sites: -2.1%
- Larynx: -2.6%
- Lung & Bronchus: -2.3%
- Prostate: -6.1%

**WOMEN**

- Liver & Intrahepatic Bile Duct: 3.8%
- Melanoma of the Skin: 1.7%
- Myeloma: 1.4%
- Corpus & Uterus: 1.2%
- Thyroid: 1.0%
- Leukemia: 0.9%
- Oral Cavity & Pharynx: 0.7%
- Cervix: 0.5%
- Kidney & Renal Pelvis: 0.5%
- Breast: 0.4%
- Stomach: 0%
- All Sites: 0%

**AVERAGE ANNUAL PERCENT CHANGE (AAPC) 2011-2015**

*AAPC is significantly different from zero (p<.05). Rates were adjusted for reporting delay in the registry.

Source: Annual Report to the Nation

[seer.cancer.gov]
Increasing or Decreasing Cancer Mortality Trends in the US, Among Men and Women

Average Annual Percent Change (AAPC) 2012-2016

*AAPC is significantly different from zero (p<.05).

Source: Annual Report to the Nation

National Cancer Institute
Prevalence of people alive with a previous diagnosis of cancer
As of January 2016, it is estimated that there are xx 15.3 million cancer survivors in the United States.

Diagram: Estimated Number of Cancer Survivors in the US

Projections

Year


Millions


National Cancer Institute
Breast and Prostate Cancer Are The Most Common Sites Among Cancer Survivors

Estimated Number of Persons Alive in the U.S. Who Were Diagnosed With Cancer, by Site (as of January 1, 2014)
Total Cancer Survivors, N=14.5M

- Female breast: 22%
- Prostate: 21%
- Colorectal: 13%
- Gynecologic: 9%
- Hematologic: 7%
- Urinary Bladder, Kidney, Renal Pelvis: 7%
- Other: 4%

Disparity in Cancer Incidence and Mortality Trends
Liver & Intrahepatic Bile Duct

- Incidence data from SEER 13 1992–2010
- Mortality data from NCHS

*Hispanic is not mutually exclusive from other groups

Incidence data from SEER 13 1992–2011, Mortality data from NCHS
Female Breast Cancer

Delay-Adjusted Incidence

Rate per 100,000

Delay-Adjusted Mortality

Year of Diagnosis


White

Black
Clinically Relevant Data Presentation and Analysis
Breast Cancer
Female Invasive Breast Cancer, SEER

Delay-Adjusted Incidence

Mortality
Breast Cancer Molecular Subtypes

• The major subtypes of breast cancer are approximated by whether the tumor over expresses any or all of three markers
  - Human Epidermal Growth Factor Receptor 2 (HER2), Estrogen Receptor (ER), and Progesterone Receptor status (PR)
  - ER and PR jointly defined as Hormone receptor status (HR)

• Four major breast cancer molecular subtypes
  - HR+/HER2- (approximates Luminal A*)
  - HR+/HER2+ (approximates Luminal B*)
  - HR-/HER2+ (HER2 enriched)
  - HR-/HER2- (Triple Negative)

*Approximation as Ki67 defines Luminal cancers and is not available in SEER/NAACCR
Breast Cancer Molecular Subtypes

• For 2010 cases, all registries from the SEER program for the first time collected Human Epidermal Growth Factor 2 (HER2) receptor status

• Estrogen Receptor (ER) and Progesterone Receptor (PR) status were collected since the 1990s

• With the availability of HER2/ER/PR, demographic and clinical assessment of major breast cancer subtypes are now possible
  o Covering ~28% of the US population
Breast Cancer Incidence by Molecular Subtypes 2010, SEER

HR⁺/HER2⁻  Triple-negative  HR⁺/HER2⁺  HR⁻/HER2⁺

HR+ in white women

Triple negative in black women
US Incidence of Breast Cancer Subtypes Defined by Joint Hormone Receptor and HER2 Status

Nadia Howlader, Sean F. Altekruse, Christopher I. Li, Vivien W. Chen, Christina A. Clarke,
Lynn A. G. Ries and Kathleen A. Cronin

Author Affiliations

Correspondence to: Nadia Howlader, MS, Mathematical Statistician, Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, 9609 Medical Center Dr, Bethesda, MD 20892 (e-mail: howladern@mail.nih.gov)
Tracking and Evaluating Molecular Tumor Markers With Cancer Registry Data: IGER2 and Breast Cancer

William G. Anderson, Philip S. Rosenberg and Hormuzd A. Katki

+ Author Affiliations

Correspondence to: Hormuzd A. Katki, PhD, National Cancer Institute, NIH, Room 7E506, 9609 Medical Center Dr, Bethesda, MD 20892-9780 (e-mail: katki.h@mail.nih.gov).

First noted in 1948 by Johannes Clemmesen from the Danish Cancer Registry (1), patterns of female breast cancers in developed countries are consistent with a "mixture model" with at least two main parts (2-5). The
Breast Cancer Survival By Molecular Subtypes

Howlader et al. CEBP 2018
Breast cancer-specific survival by molecular subtypes, SEER-18 (After imputation)
Breast cancer-specific survival by stage

2A: Stage I

<table>
<thead>
<tr>
<th>HR+/HER2-</th>
<th>HR+/HER2+</th>
<th>HR-/HER2+</th>
<th>Triple Negative</th>
</tr>
</thead>
</table>

2B: Stage II

<table>
<thead>
<tr>
<th>HR+/HER2-</th>
<th>HR+/HER2+</th>
<th>HR-/HER2+</th>
<th>Triple Negative</th>
</tr>
</thead>
</table>

2C: Stage III

<table>
<thead>
<tr>
<th>HR+/HER2-</th>
<th>HR+/HER2+</th>
<th>HR-/HER2+</th>
<th>Triple Negative</th>
</tr>
</thead>
</table>

2D: Stage IV

<table>
<thead>
<tr>
<th>HR+/HER2-</th>
<th>HR+/HER2+</th>
<th>HR-/HER2+</th>
<th>Triple Negative</th>
</tr>
</thead>
</table>
Differences in Breast Cancer Survival by Molecular Subtypes in the United States

Nadia Howlader¹, Kathleen A. Cronin¹, Allison W. Kurian², and Rebecca Andridge³
Challenges and Opportunities with Cancer Registries
Current Registry Data Items: Challenges & Opportunities

Fill in clinical gaps

Current SEER — Diagnosis
First round of treatment/surgery
Recurrence
Subsequent treatments/surgery
Death (Survival)

Challenges:
1. Biomarkers
2. Detailed treatment (agent, dose)
3. Outcome (recurrence, patient generated health outcome)

Source: California Health Care Foundation Report 2014
Current Registry Data Collection Process: Challenges & Opportunities

- Overall data capture process is mostly manual.
- 65%–80% patient-specific data items come from unstructured texts\(^1\)
  - E-path reports or clinical notes
  - Natural Language Processing (NLP) is mature enough for text mining to obtain rich clinical information
- Due to aging of population, cancer cases expected to grow in the future
  - Information needs to be abstracted with flat resources

\(^1\) Hiatt et al. JNCI 2015
Conclusions
Conclusions

• Many new strategic initiatives underway to enhance the surveillance infrastructure

• Result in collection of more clinically relevant data elements
  o Leading up to diagnosis (e.g., risk factors)
  o Between diagnosis and death (e.g., recurrence, detailed treatment, biomarkers)

• Position cancer registries to support next generation cancer research
Thank you!
Questions

Nadia Howlader
E-mail: howladerm@mail.nih.gov