

CDC/NPCR

2021 Data Quality Evaluation

Diagnosis year: 2018

2020 Data Evaluation Reports

Diagnosis Years 1995-2019

FCDS Annual Virtual Meeting

August 19, 2021

Megsys Herna, BA, CTR



NPCR

Data Quality Evaluation (DQE)

- Purpose
- Process
- Results
- Recommendations

Purpose of the NPCR DQE

- **Assess the quality of the data, including the accuracy and completeness of coding.**
 - Complete and accurate data are important for estimating variations in and changes among population subgroups over time.
- **These data are a critical because they are used for planning, operating, funding, and evaluating cancer surveillance systems.**

3

Purpose of the NPCR DQE

- **Data evaluated had emphasis on policies and procedures for the following areas:**
 1. Assessment of data quality
 2. Text documentation
 3. Data consolidation
 4. Accuracy of staging, treatment and treatment related information

4

Confidentiality and Security

- All electronic data was de-identified; all patient identifiers were removed.
- Staff signed all necessary confidentiality agreements before accessing confidential materials involved in the evaluation process.

5

Methodology

FCDS prepared two extract files:

1. A master abstract file of the consolidated cases of the five primary sites with two or more abstracts
 - Primary sites
 - Bladder
 - Pancreas
 - Kidney and renal pelvis
 - Ovary
 - Melanoma of the skin
 - 2018 Diagnosis year

6

Methodology

2. Extract file containing source abstract level data for the consolidated cancer cases.

- A random sample of 809 records were selected from the submitted data file.
- These 809 cases were reconsolidated and compared to the FCDS consolidated cases.
- 365 reconsolidated records reviewed

7

Data Elements Reviewed

Cancer Identification

Date of Diagnosis
 Primary Site
 Laterality
 Diagnostic Confirmation
 Histologic Type ICD-O-3
 Grade Clinical
 Grade Pathological
 Grade Post Therapy

Staging/Prognostic

Factors

Regional Nodes Positive
 Regional Nodes Examined
 Summary Stage 2018
 Tumor Size Summary

Treatment

Date of 1st Course RX—CoC RX Date—Radiation Flag
 Date of 1st Crs RX—CoC Flag
 Phase I Radiation Treatment Modality
 OR Reason No Radiation
 Date of Initial Rx—SEER RxSumm—Surg/Rad Seq
 Date of Initial Rx-SEER Flag RX Date—Chemo
 AND RX Date—Chemo Flag
 RX Summ--Treatment Status RX Summ—Chemo
 RX Date — Surgery RX Date—Hormone
 RX Date — Surgery Flag RX Date—Hormone Flag
 RX Date Mst Defn Srg RX Summ—Hormone
 RX Date Mst Defn Srg Flag RX Date — BRM
 RX Summ—Surg Prim Site RX Date — BRM Flag
 RX Summ—Scope Reg LN Sur RX Summ—BRM
 RX Summ—Surg Other Reg/Dis RxSumm—Systemic/Sur Seq
 Reason No Surgery RX Summ—Transplnt/Endocr
 RX Date — Other RX Date — Other
 RX Date — Other Flag RX Date — Other Flag

8

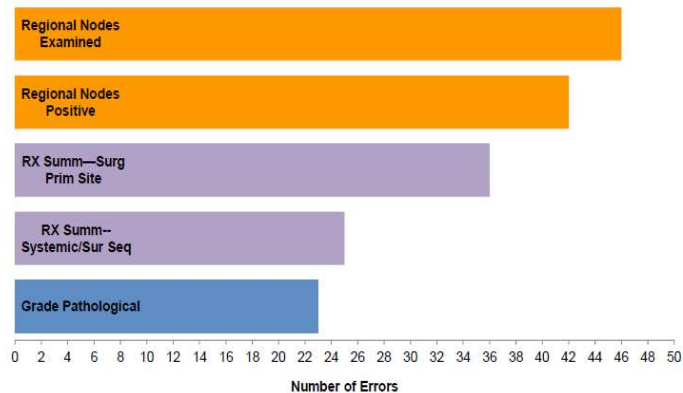
Number of Data Elements Reviewed by Site

Primary Site	Number of Data Elements Reviewed	Number of Cases Reviewed	Percentage of Cases Reviewed
Melanoma of the skin	43	73	20%
Bladder	43	73	20%
Pancreas	43	73	20%
Kidney and Renal Pelvis	43	73	20%
Ovary	43	73	20%
Total	215	365	100%

9

Top Data Elements with Major Errors across All Sites Combined

For Cancer Identification, Stage/Prognostic Factors & Treatment Data Elements



10

DQE Results Case Consolidation

- Of a total of 15,695 possible data elements that could have had errors, only 336 data elements (2.1%) were found to have errors.
- Data accuracy rate was **97.9%**.

11

NPCR Recommendations

- Continue conducting visual editing.
- Review basic abstracting principles with staff and data reporters.
- Continue to emphasize to all reporting facilities the importance of text documentation to support data elements.

12

NPCR 2020 Data Evaluation Reports (DER)

- Call for Data 1995-2019
- Data Submitted November 2020
 - 24 month file is 2018
 - 12 month file is 2019

Special Thank You to Brad Wohler!

13

NPCR 2020 Data Evaluation Reports (DER)

- 24 month data (2018) evaluated must meet data quality standards on the following five criteria:
 1. Percent Complete- 95%
 2. Unresolved Duplicate Rate- 1 per 1,000 or fewer unresolved duplicate rate
 3. Percent Death Certificate Only Cases- 3% or fewer death certificate only cases
 4. Percent Missing Critical Data Elements- maximum percent missing
 1. 2% AGE
 2. 2% SEX
 3. 3% RACE
 4. 2% COUNTY
 5. Percent Passing Edits- 99% pass on a set of standard edits

14

NPCR 2020 Data Evaluation Reports (DER)

- 12 month data (2019) evaluated must meet data quality standards on the following five criteria:
 1. Percent Complete- 90%
 2. Unresolved Duplicate Rate- 2 per 1,000 or fewer unresolved duplicate rate
 3. Percent Death Certificate Only Cases- 3% or fewer death certificate only cases
 4. Percent Missing Critical Data Elements- maximum percent missing
 1. 3% AGE
 2. 3% SEX
 3. 5% RACE
 4. 3% COUNTY
 5. Percent Passing Edits- 97% pass on a set of standard edits

15

NPCR 2020 Data Evaluation Reports (DER)

2018 DCO Rate = 1.75%

- DCO Rate for 2017 went down from 1.66% to 1.47%... (late cases)

Completeness met for 24 month data (2018); Missed target for 12 month data (2019)

- 24 Month Standard: 95.00% **FCDS 24 Month: 101.51%**
- 12 Month Standard: 90.00% **FCDS 12 Month: 82.72%**

24 month completeness met standard

12 month completeness is 7% below standard

Race Unknown near or over the 3% threshold for National Data Quality Standard

- 2018 % Race unknown (2.82%) near the National Data Quality Standard (<=3%)
- 2017 % Race unknown (2.74%) near the National Data Quality Standard (<=3%)
- 2016 % Race unknown (3.43%) over the National Data Quality Standard (<= 3%)

20 variable categories above NPCR median for 2018 data

- 1 variables was below last year and are now higher
 - RX Summ Radiation [1360] Blank

22 variable categories below NPCR median for 2018 data

16

20 variable categories above NPCR median for 2018 data

Variable Name	FCDS	NPCR Median
NHIA Derived Hisp Origin [191] Hispanic (1-8) and Spanish/Hispanic Origin [190] Unknown (9)	13.87	2.14
Birthplace Country [254] NOS (ZZN, ZZC, ZZS, ZZP, ZZE, ZZF, ZZA, ZZX)	0.38	0.04
Census Tract 2010 [135] Not Tracked (000000)	0.18	0.00
Census Tr Cert 2010 [367] Unknown (9)	0.18	0.01
Sequence Number-Central [380] Two or more (01 - 35)	29.23	22.50
Topography [400] Unknown Primary Site (C80.9)	1.78	1.51
Morphology [522] Non-specific Neoplasms (8000 - 8005)	5.40	3.32
Diagnostic Confirmation [490] excludes DCO Unknown (9) and Blank	1.58	0.49
Laterality [410] paired organs only - Only 1 side and side NOS (3)	0.10	0.05
Laterality [410] paired organs only - Unknown (9, blank)	1.99	1.74
RX Summ Surg Prim Site [1290] Surgery, NOS (90)	2.81	0.28
RX Summ Scope Reg LN Sur [1292] Unknown (9)	8.97	8.46
RX Summ Radiation [1360] Blank	100.00	22.88
RX Summ Surg/Rad Seq [1380] Unknown (9)	1.33	0.04
Merged Rad Regional RX Modality [1570] and RX Summ Radiation [1360] Both Blank or Both Unknown	100.00	23.05
RX Summ Surg/Rad Seq [1380] Unknown (9)	0.57	0.10
Over-Ride SeqNo/DxConf [2000] Overused	0.45	0.03
Over-Ride Site/Lat/SeqNo [2010] Overused	0.15	0.01
Over-Ride Histology [2040] 1 = Morphology--Type & Behavior (SEER MORPH)	0.55	0.01
Over-Ride Ill-defined Site [2060] Overused	0.16	0.02

17

22 variable categories below NPCR median for 2018 data

Variable Name	FCDS	NPCR Median
County at Diagnosis [90]	0.00	0.02
Birthplace Country [254] Blank or Unknown	54.26	59.54
Birthplace Country [254] Not born in USA (excluding NOS)	62.61	63.09
Census Tract 2010 [135] Blank	0.00	1.14
Census Tr Cert 2010 [367] Blank	0.00	0.32
Date of Diagnosis [390] Month (blank)	0.07	0.14
Diagnostic Confirmation [490] excludes DCO Not Microscopically Confirmed (5, 6, 7, 8)	3.12	4.01
RX Summ Surg Prim Site [1290] Blank and Unknown Combined	0.64	1.67
RX Summ Surg Oth Reg/Dis [1294] Blank and Unknown Combined	0.35	1.37
Reason for No Surgery [1340] Blank and Unknown Combined	0.59	2.61
RX Summ Chemo [1390] Blank and Unknown Combined	1.40	2.13
RX Summ Hormone [1400] Blank and Unknown Combined	0.86	1.83
RX Summ BRM [1410] Blank and Unknown Combined	0.75	1.72
RX Summ Other [1420] Blank and Unknown Combined	0.21	1.41
Rad Regional RX Modality [1570] Blank and Unknown Combined where RX Summ Radiation [1360] Indicates treatment conflict between recorded treatment	0.00	8.60
RX Summ Transplnt/Endocr [3250] Blank and Unknown Combined	0.42	1.20
Over-Ride Age/Site/Morph [1990] Overused	0.05	0.08
Over-Ride Site/Type [2030] Overused	0.20	0.22
Date of Last Contact [1750] Year within 18 months - alive vital status only	40.64	41.41
Follow-Up Source Central [1791] Unknown (99)	0.00	0.03
Cause of Death [1910] decedents only - Death Cert. Avail., But No COD (7797)	0.01	0.41
Cause of Death [1910] decedents only - Death Certificate Not Available (7777)	0.00	22.62

18

