

## 2017年10月27日会議資料 日本版SEER研究班 平成29年度研究進捗確認

仕組み検討班（伊藤、大木、金村、西野、東、松田）

記述疫学研究班（伊藤、井上、林、堀、片野田）

### A資料内容

- 2ページ 平成28年度 SEERの実態詳細確認点
- 2ページ スケジュール
- 3ページ ユタ州がん登録訪問 2017年8月29日
  - ・ 他関連資料
- 7ページ SEER訪問 2017年8月31日
  - ・ 他関連資料

## 平成28年度 SEERの実態詳細確認点

### 仕組み検討班

- 「がん診療情報」の収集の仕組みの提案に関し、SEERのために収集されたデータを第三者の研究者に提供可能な理由。特に病院等での情報取得時の本人通知及び本人同意の有無。  
(主担当：西野、松田)
- 「がん診療情報」の収集の仕組みの提案に関し、SEER特別項目を州がん登録及び各州の病院から届出協力してもらう仕組み。必要経費や教育支援の有無と内容。  
(主担当：大木、金村、伊藤)
- 「特定匿名化情報」の提供に関し、希少がん等の個人識別可能性のあるデータの匿名化処理の有無。匿名化処理を行っている場合はその手法。  
(主担当：東、柴田)

### 記述疫学研究班 (担当：伊藤、井上、林、堀、片野田)

- 利用手続きの実際。
- 通常の利用申請手続きで利用可能なデータの内容や粒度。
- 通常の利用申請手続きで利用可能なデータを用いて、どの程度の粒度の記述疫学研究が可能か。

3

## スケジュール

### 仕組み検討班

- 班会議 (6/27) まで
  - 担当分野に関して不明点、疑問について、日本語又は英語の課題リストを作成する。
- 7月末まで
  - 担当分野に関して不明点、疑問について、英語の課題リスト作成の継続
  - 不明点、疑問について、サイト情報等から答えを得られることは解決する。
- 8月20日頃まで
  - 柴田：課題リストレイアウトをそろえる。
- 8月28日～9月1日 現地訪問 (伊藤、大木、金村、東)
- 9月末頃まで
  - 現地訪問で理解、解決したことを課題リストにまとめる。
- 10月～11月 第2回班会議

### 記述疫学研究班

- 班会議 (6/27)
  - SEER site : Accessing the 1973-2014 SEER Dataの説明 (堀)
  - 解析内容の案提示 (伊藤、井上、林、堀、片野田)
- 8月20日頃まで
  - 研究計画案の完成、班員共有
  - 各種利用申請手続き (必要に応じて、MCIJ利用申請、研究倫理審査)
  - 利用できる情報による研究計画の修正
- 8月28日～9月1日 現地訪問 (井上、林、堀)
- 10月～
  - 解析、暫定結果報告
- 10月～11月 第2回班会議
  - 進捗報告

## ユタ州がん登録訪問 2017年8月29日

伊藤、大木、金村、井上、柴田

- 研究計画上の確認事項
  - SEER特別項目を州がん登録及び各州の病院から出してもらおう仕組み
  - 必要経費や教育支援の有無と内容

1966年開始、SEER9州の一つ  
2013年 人口290万人、がん罹患数 約9,500

- 事前質問
  - Topics: Data collection; additional data items for clinical researches, basis laws (acts, regulations), public awareness for personal information protection, incentives for data providers

5

## 事前質問 続き

- We are interested in NPCR-AERRO Project and Comparative Effectiveness Research Data Collection Enhancement Project in relation to collecting additional data items for clinical research. I suppose that your registry has not participated in these projects. If you don't mind me asking, I would like to know whether there are any reasons not be able to join in those projects or not.
- In addition, I know that the Utah cancer registry is to achieve a high quality of data, not only NPCR data but also SEER data. It would be very helpful if we can learn how to collect data from hospitals in your registry.

6

# AGENDA

## Center for Cancer Registries, National Cancer Center, Japan Visit with Utah Cancer Registry

August 29, 2017

0900 – 1700

**Japan National Cancer Center Attendees:** Akiko Shibata, MD – Center for Cancer Registries; National Cancer Center, Japan  
 Hidemi Ito, MD – Aichi Cancer Registry, Nagoy  
 Izumi Oki, MD – Tochigi Cancer Registry; Tochigi Cancer Center Hospital Registry  
 Seiki Kanemura, MD – Miyagi Cancer Registry; Miyagi Cancer Center Hospital Cancer Registry  
 Manami Inoue, MD – Center for Public Science; National Cancer Center, Japan  
 Daisuke Kawakita, MD – Otolaryngologist; Nogoya City University Hospital, Aichi, Japan

**Utah Cancer Registry Attendees:** Carol Sweeney, PhD - Professor, Internal Medicine and Director, Utah Cancer Registry  
 Mia Hashibe, PhD - Associate Professor, Family and Preventive Medicine, Huntsman Cancer Institute, and Utah Cancer Registry  
 SuAnn McFadden, CTR - Operations Manager, Utah Cancer Registry  
 Kacey Wigren, RHIT, CTR - Lead Certified Tumor Registrar, Utah Cancer Registry  
 Carrie Bateman, BS - Senior Informatics Specialist, Utah Cancer Registry  
 Kim Herget, MStat - Biostatistician, Utah Cancer Registry  
 Marjorie Carter, MSPH - Research Manager, Utah Cancer Registry

<b>0900 – 0915</b>	<b>Welcome, Introduction &amp; Schedule Overview</b>	Carol Sweeney, SuAnn McFadden
<b>0915 – 0930</b>	<b>Overview of Center for Cancer Registries, National Cancer Center, Japan</b>	Japan National Cancer Center Representative
<b>0930 – 1030</b>	<b>Registry Oversight and Relationship to US State and Federal Agencies</b>	Carol Sweeney
<b>1030 – 1045</b>	<b>Break</b>	
<b>1045 – 1100</b>	<b>Data Standards, Standard Setters &amp; Supporting Organizations</b>	SuAnn McFadden
<b>1100 – 1130</b>	<b>Data Collection &amp; Hospital Collaboration</b>	SuAnn McFadden
<b>1130 – 1300</b>	<b>Lunch The Point at Huntsman Cancer Institute</b>	
<b>1300 – 1330</b>	<b>Editing, Case Consolidation, Data Preparation for Submission</b>	Kacey Wigren

<b>1330 – 1430</b>	<b>Data Security, Data Structure &amp; Data Transmission</b>	Carrie Bateman
<b>1430 – 1445</b>	<b>Break</b>	
<b>1445 - 1515</b>	<b>Follow-up Process</b>	SuAnn McFadden
<b>1515 – 1530</b>	<b>Linkages and Data Exchange Process</b>	Kim Herget
<b>1530 - 1545</b>	<b>Data Sharing and Participant Recruitment for Research</b>	Marjorie Carter
<b>1545 – 1700</b>	<b>Open Discussion, Questions, Wrap-up</b>	
<b>1700</b>	<b>Return to hotel</b>	

**Additional Information:**

August 28, 2017 1634, Arrive in Salt Lake City, Utah

Hotel: Little America Hotel

500 South Main Street

Salt Lake City, Utah 84101

Telephone: 801-596-5700

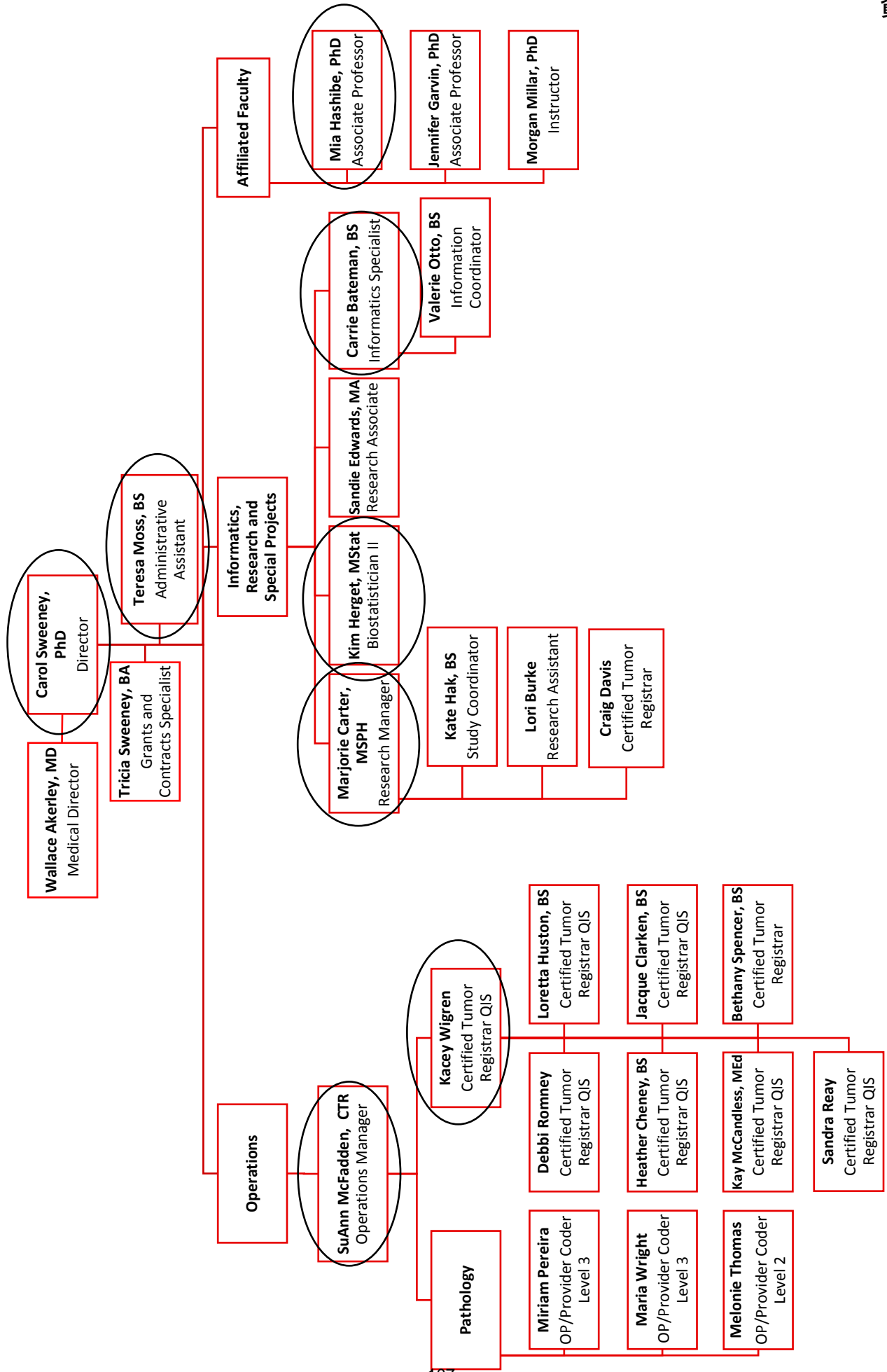
August 29, 2017: 0830, Take Taxi to Utah Cancer Registry (650 Komas Drive, Suite 106B) entrance is on the west side of building; office is in southeast corner.

August 30, 2017: 0735, Flight from Salt Lake City, Utah to Washington, D.C.

Utah Cancer Registry Contact Numbers while in Salt Lake City:

Teresa Moss, Administrative Assistant 801-581-8407

SuAnn McFadden, Operations Manager 801-213-3247 or 801-755-0494



# SEER訪問 2017年8月31日

堀、林、東、井上、柴田

- 研究計画上の確認事項
  - SEERの制度や組織の仕組みの理解
- 事前質問
  - Topics  
Public use data; SEER\*Stat Databases, anonymous data, de-identified data, basis Laws (Acts, Regulations), public awareness for personal information protection
- 1. How the SEER solve the privacy issues in providing the de-identified dataset to researchers. More specifically,
  - What criteria of anonymity do you apply, and how do you test the dataset satisfy the criteria?

7

## 事前質問 続き

- Who decides the criteria of anonymity on what processes?
  - Why the level of security protocol much stricter when dealing with the SEER-Medicare data? I understand that the level of identifiability is much higher than standalone SEER data, but is there any other rules or factors that contribute to the restriction, such as the different laws that apply?
  - How does SEER ensure that the researchers implement the security protocol properly?
  - Was there any incidence that the security protocols or other rules for data usage were violated?
- 
- 2. When the new idea of data linkage, such as linkage with CAHPS or MHOS, who decides the rules?
- 3. How do you incorporate e-path systems into the SEER systems.

8

**August 31, 2017**

**National Cancer Center Japan Visit to National Cancer Institute US**

**Location**

National Cancer Institute  
9609 Medical Center Drive  
Rockville, MD 20850  
Room 2W912 (West tower, second floor)

**Logistics**

Enter building 9609 Medical Center Drive and proceed to Security Desk. Call Peggy 240-276-6867 from Security Desk. Peggy will escort visitors from security to meeting room.

**Agenda**

- 9:00 Arrival
- 9:30 Introductions
- 10:00 SEER program overview – Dr. Kosary
- 11:00 Anonymous data, de-identified data – Dr. Petkov
- 12:00 Lunch
- 1:00 Public use data – Dr. Yu
- 2:00 SEER\*Stat Databases – Dr. Yu, Dr. Howlader
- 3:00 Basis Laws (Acts, Regulations) – Dr. Kosary
- 4:00 Wrap-up
- 5:00 Adjourn

*Dr. Feuer.*  
 < *GISNET and Analytic Tools*

**Participants**

**National Cancer Center Japan**

Akiko Shibata, MD, PhD  
Center for Cancer Registries  
National Cancer Center, Japan  
Takahiro Higashi, MD, PhD  
Center for Cancer Registries  
National Cancer Center, Japan  
Manami Inoue, MD, PhD  
Center for Public Science  
National Cancer Center, Japan



Megumi Hori, PhD  
 Center for Cancer Control and Information Services  
 National Cancer Center, Japan

Yingsong Lin, MD, PhD  
 Department of Public Health, School of Medicine  
 Aichi Medical University

**National Cancer Institute US**

Peggy Adamo, BS, RHIT, CTR  
 Data Quality Team Lead  
 Data Quality, Analysis, and Interpretation Branch  
 Surveillance Research Program  
 Division of Cancer Control and Population Sciences

*25 yrs*

Carol Kosary, PhD  
 Program Manager, Surveillance, Epidemiology and End Results (SEER)  
 Program  
 Office of the Associate Director  
 Surveillance Research Program  
 Division of Cancer Control and Population Sciences

*25 yrs*

Valentina I Petkov, MD, MPH  
 Health Scientist/Program Officer  
 Data Quality, Analysis, and Interpretation Branch  
 Surveillance Research Program  
 Division of Cancer Control and Population Sciences

Mandi Yu, Ph.D.  
 Statistician  
 Statistical Research and Applications Branch  
 Surveillance Research Program  
 Division of Cancer Control and Population Sciences

Nadia Howlader, PhD, MS  
 Mathematical Statistician  
 Data Analysis Branch  
 Surveillance Research Program  
 Division of Cancer Control and Population Science

Serban Negoita, MD, DrPH, CTR  
 Branch Chief  
 Data Quality, Analysis, and Interpretation Branch  
 Surveillance Research Program  
 Division of Cancer Control and Population Sciences

Edward (Ted) L. Trimble, MD, MPH  
 Director, Center for Global Health  
 National Cancer Institute

# DCCPS Organization and Leadership



**Healthcare Delivery Research Program**  
Dr. Paul Jacobsen, ASSOCIATE DIRECTOR

**Healthcare Delivery Research Program**  
Dr. Ann Geiger, DEPUTY ASSOCIATE DIRECTOR

**Healthcare Assessment Research**  
Dr. Paul Doria-Rose, ACTING BRANCH CHIEF

**Health Systems and Interventions Research**  
Dr. Sarah Kobrin, ACTING BRANCH CHIEF

**Outcomes Research**  
Dr. Ashley Wilder Smith, BRANCH CHIEF

**Surveillance Research Program**  
Dr. Lynne Penberthy, ASSOCIATE DIRECTOR

**Surveillance Research Program**  
Dr. Kathleen Cronin, DEPUTY ASSOCIATE DIRECTOR

**Data Quality, Analysis, and Interpretation**  
Dr. Serban Negoita, BRANCH CHIEF

**Surveillance Informatics**  
Dr. Paul Fearn, BRANCH CHIEF

**Data Analysis**  
Dr. Angela Mariotto, BRANCH CHIEF

**Statistical Research and Applications**  
Dr. Eric Feuer, BRANCH CHIEF

**Epidemiology and Genomics Research Program**  
Dr. Kathy Helzlsouer, ASSOCIATE DIRECTOR

**Epidemiology and Genomics Research Program**  
Dr. Emily Harris, DEPUTY ASSOCIATE DIRECTOR

**Methods and Technologies**  
Dr. Mukesh Verma, BRANCH CHIEF

**Environmental Epidemiology**  
Dr. Gary Ellison, BRANCH CHIEF

**Genomic Epidemiology**  
Dr. Elizabeth Gillanders, BRANCH CHIEF

**Clinical and Translational Epidemiology**  
Dr. Andrew Freedman, BRANCH CHIEF

**Risk Factor Assessment**  
Dr. Susan Krebs-Smith, BRANCH CHIEF

**Behavioral Research Program**  
Dr. William Klein, ASSOCIATE DIRECTOR

**Behavioral Research Program**  
Dr. Linda Nebeling, DEPUTY ASSOCIATE DIRECTOR

**Basic Biobehavioral and Psychological Sciences**  
Dr. Paige Green, BRANCH CHIEF

**Health Behaviors Research**  
Dr. Susan Czajkowski, BRANCH CHIEF

**Health Communication and Informatics Research**  
Dr. Bradford Hesse, BRANCH CHIEF

**Tobacco Control Research**  
Dr. Michele Bloch, BRANCH CHIEF

**CONNECT WITH US** (<https://twitter.com/NCICancerCtrl>)



## Surveillance Research Program (/)

☰ Menu

🔍 Search

📄 Section Navigation



### Surveillance Research Program Staff



#### SRP Main Numbers

Phone: 240-276-6740 / Fax: 240-276-7908

#### Office of the Associate Director (OAD)

Photos	Name	Areas of Expertise	Contact Info
 <a href="/about/bios/penberthyl.html">(/about/bios/penberthyl.html)</a>	<b>Penberthy, Lynne - MD, MPH</b> <a href="/about/bios/penberthyl.html">(/about/bios/penberthyl.html)</a> Associate Director	Medical Informatics Clinical Trials and Informatics Cancer Surveillance Automation Secondary Data (claims, electronic medical records)	<a href="mailto:schumacherpentl@mail.nih.gov">schumacherpentl@mail.nih.gov</a> <a href="mailto:schumacherpentl@mail.nih.gov">mailto:schumacherpentl@mail.nih.gov</a> Phone: 240-276-6864
 <a href="/about/bios/cronink.html">(/about/bios/cronink.html)</a>	<b>Cronin, Kathy - PhD</b> <a href="/about/bios/cronink.html">(/about/bios/cronink.html)</a> Deputy Associate Director	Developing Population Survival Measures Modeling the Impact of Risk Factors, Screening & Treatment on Breast Cancer Incidence & Mortality	<a href="mailto:cronink@mail.nih.gov">cronink@mail.nih.gov</a> <a href="mailto:cronink@mail.nih.gov">mailto:cronink@mail.nih.gov</a> Phone: 240-276-6836





[\(/about/bios/friedmans.html\)](/about/bios/friedmans.html)

Friedman, Steve - MHSA  
[\(/about/bios/friedmans.html\)](/about/bios/friedmans.html)  
Senior Advisor for  
Operations

Program Operations  
Clinical Trials and  
Informatics  
Data Quality

[friedmast@mail.nih.gov](mailto:friedmast@mail.nih.gov)  
<mailto:friedmast@mail.nih.gov>

Phone: 240-276-6138



[\(/about/bios/garsona.html\)](/about/bios/garsona.html)

Garson, Amy - BS  
[\(/about/bios/garsona.html\)](/about/bios/garsona.html)  
Program Analyst

In-House &  
Extramural Budget  
Administration  
SEER Program  
Budget  
Administration  
Meeting Support  
Contract  
Management

[garsona@mail.nih.gov](mailto:garsona@mail.nih.gov)  
<mailto:garsona@mail.nih.gov>

Phone: 240-276-6694



[\(/about/bios/husseys.html\)](/about/bios/husseys.html)

Hussey Sarah - BA  
[\(/about/bios/husseys.html\)](/about/bios/husseys.html)  
Fellow

Communications  
Social Media

[Sarah.hussey@nih.gov](mailto:Sarah.hussey@nih.gov)  
<mailto:Sarah.hussey@nih.gov>

Phone: 240-276-6926



[\(/about/bios/kosaryc.html\)](/about/bios/kosaryc.html)

Kosary, Carol - DMgt  
[\(/about/bios/kosaryc.html\)](/about/bios/kosaryc.html)  
Program Manager, SEER

SEER Program  
Management  
SEER Data  
Management  
System  
Cancer Rates &  
Trends  
Electronic Data  
Capture

[kosaryc@mail.nih.gov](mailto:kosaryc@mail.nih.gov)  
<mailto:kosaryc@mail.nih.gov>

Phone: 240-276-6732



[\(/about/bios/mebane.html\)](/about/bios/mebane.html)

Mebane, Ana – MA  
[\(/about/bios/mebane.html\)](/about/bios/mebane.html)  
Fellow

Health  
Communications  
Health Writing  
Social Media

[ana.mebane@nih.gov](mailto:ana.mebane@nih.gov)  
<mailto:ana.mebane@nih.gov>

Phone: 240-276-6590



[\(/about/bios/mintzr.html\)](/about/bios/mintzr.html)

Mintz, Rebecca Ilene – MPH  
[\(/about/bios/mintzr.html\)](/about/bios/mintzr.html)  
Fellow

Health Education  
Community  
Outreach  
Communications

[rebeccailene.mintz@nih.gov](mailto:rebeccailene.mintz@nih.gov)  
<mailto:rebeccailene.mintz@nih.gov>

Phone: 240-276-6740



[\(/about/bios/mossj.html\)](/about/bios/mossj.html)

Moss, Jennifer - PhD,  
MSPH  
[\(/about/bios/mossj.html\)](/about/bios/mossj.html)  
Cancer Prevention Fellow

Adolescent Vaccines  
Prevention of  
Reproductive/Genital  
Cancers  
Geographic  
Disparities in Cancer  
Prevention  
Behaviors

[jennifer.moss@nih.gov](mailto:jennifer.moss@nih.gov)  
<mailto:jennifer.moss@nih.gov>

Phone: 240-276-5048



[\(/about/bios/murphyt.html\)](/about/bios/murphyt.html)

Murphy, Trish - MS  
[\(/about/bios/murphyt.html\)](/about/bios/murphyt.html)  
Program Analyst

Communications  
Program Support  
Program Operations  
Personnel

[patricia.murphy@nih.gov](mailto:patricia.murphy@nih.gov)  
<mailto:patricia.murphy@nih.gov>

Phone: 240-276-6903



[\(/about/bios/scottss.html\)](/about/bios/scottss.html)









Scott, Susan - MPH  
[\(/about/bios/scottss.html\)](/about/bios/scottss.html)  
Public Health Advisor

Scientific Editing  
Communication  
Planning  
Grants Management

[scottss2@mail.nih.gov](mailto:scottss2@mail.nih.gov)  
<mailto:scottss2@mail.nih.gov>

Phone: 240-276-6951

## Data Analytics Branch (DAB)







Photos	Name	Areas of Expertise	Contact Info
 (/about/bios/mariottoa.html)	<u>Mariotto, Angela - PhD</u> (/about/bios/mariottoa.html) Chief, DAB	Development, Evaluation & Dissemination of Cancer Progress Measures Development and Application of Models to Predict Cancer Incidence and Prevalence Models for Survival Projections	<u>mariotta@mail.nih.gov</u> (mailto:mariotta@mail.nih.gov)  Phone: 240-276-6698
 (/about/bios/coughland.html)	<u>Coughlan, Diarmuid - PhD,</u> <u>MPharm, MSc</u> (/about/bios/coughland.html) Cancer Prevention Fellow	Cost of Cancer Care Pharmacy	<u>diarmuid.coughlan@nih.gov</u> (mailto:diarmuid.coughlan@nih.gov)  Phone: 240-276-7647
 (/about/bios/forjaz.html)	<u>Forjaz de Lacerda, Gonçalo</u> (/about/bios/forjaz.html) Guest Researcher	Cancer Registration CanReg-5 Descriptive Epidemiology	<u>goncalo.forjaz@nih.gov</u> (mailto:goncalo.forjaz@nih.gov)  Phone: 240-276-7149
 (/about/bios/freduar.html)	<u>Fredua, Rose - BS</u> (/about/bios/freduar.html) Program Analyst	Grants Policy & Management Program Evaluation Contracts & Budget Management Interagency Agreements	<u>freduar@mail.nih.gov</u> (mailto:freduar@mail.nih.gov)  Phone: 240-276-6923
★  (/about/bios/howlader.html)	<u>Howlader, Nadia - PhD, MS</u> (/about/bios/howlader.html) Mathematical Statistician	Statistical Methods for the Estimation of Population-Based Cancer Statistics Survival Analysis Missing Data	<u>howlader@mail.nih.gov</u> (mailto:howlader@mail.nih.gov)  Phone: 240-276-6891
 (/about/bios/lamc.html)	<u>Lam, Clara - PhD</u> (/about/bios/lamc.html) Mathematical Statistician	Development and Improvement of Linkage Claims Treatment Data Population-based Cancer Statistics Cancer Trends	<u>kimcj2@mail.nih.gov</u> (mailto:kimcj2@mail.nih.gov)  Phone: 240-276-7388
 (/about/bios/lewisd.html)	<u>Lewis, Denise - PhD, MPH</u> (/about/bios/lewisd.html) Epidemiologist	Cancer Rates and Trends Adolescent and Young Adult (AYA) Cancers Geospatial Analysis Determinants of Health Disparities	<u>lewisde@mail.nih.gov</u> (mailto:lewisde@mail.nih.gov)  Phone: 240-276-6757
 (/about/bios/noonea.html)	<u>Noone, Anne-Michelle - MS</u> (/about/bios/noonea.html) Mathematical Statistician	Longitudinal Data Analysis Population-Based Cancer Statistics Cancer Trends	<u>noonea@mail.nih.gov</u> (mailto:noonea@mail.nih.gov)  Phone: 240-276-6705

[Return to top]

## Data Quality , Analysis, and Interpretation Branch (DQAIB)

Photos	Name	Areas of Expertise	E-mail
 (/about/bios/negoita.html)	<b>Negoita, Serban – MD., DrPH</b> (/about/bios/negoita.html) Chief, DQAIB	Cancer registry operations Cancer data quality assessments Racial/ethnic disparities in cancer treatment	<a href="mailto:serban.negoita@nih.gov">serban.negoita@nih.gov</a> (mailto:serban.negoita@nih.gov)  Phone: 240-276-5409
★  (/about/bios/adamop.html)	<b>Adamo, Peggy - BS, RHIT, CTR</b> (/about/bios/adamop.html) Team Lead, Quality Improvement Team	Cancer Data Collection, Classification, Quality, Completeness, Timeliness Cancer Registrar Training	<a href="mailto:adamom@mail.nih.gov">adamom@mail.nih.gov</a> (mailto:adamom@mail.nih.gov)  Phone: 240-276-6867
 (/about/bios/croppm.html)	<b>Cropp-Puckett, Monique</b> (/about/bios/croppm.html) Program Analyst	Budget Administration Acquisition Support Program Support	<a href="mailto:croppma@od.nih.gov">croppma@od.nih.gov</a> (mailto:croppma@od.nih.gov) 240-276-6109
 (/about/bios/dickiel.html)	<b>Dickie, Lois - CTR</b> (/about/bios/dickiel.html) Public Health Analyst	Cancer Registry Operations QC/QI American Indian/Alaskan Native Data Collection	<a href="mailto:dickielo@mail.nih.gov">dickielo@mail.nih.gov</a> (mailto:dickielo@mail.nih.gov)  Phone: 240-276-6862
 (/about/bios/ehrenkranzr.html)	<b>Ehrenkranz, Rebecca - MPH</b> (/about/bios/ehrenkranzr.html) Fellow	Clinical Trials Surveillance Data Quality	<a href="mailto:rebecca.ehrenkranz@nih.gov">rebecca.ehrenkranz@nih.gov</a> (mailto:rebecca.ehrenkranz@nih.gov)  Phone: 240-276-7810
★  (/about/bios/petkovv.html)	<b>Petkov, Valentina - MD, MPH</b> (/about/bios/petkovv.html) Health Sciences Administrator/Program Officer	Data Assessment (Scientific and clinical relevance; data quality) Automation of data collection (Data linkages, Natural Language Processing) Population-based cancer statistics Precision Medicine Clinical Trials	<a href="mailto:Valentina.petkov@nih.gov">Valentina.petkov@nih.gov</a> (mailto:Valentina.petkov@nih.gov)  Phone: 240-276-5387
 (/about/bios/riesl.html)	<b>Ries, Lynn - MS</b> (/about/bios/riesl.html) Health Statistician	Childhood, Adolescent, & Young Adult Cancers Cancer Survival Analyses Cancer Incidence Trends by Stage and Cancer Type Cancer Survival for Older Aged Americans	<a href="mailto:rieslynn@mail.nih.gov">rieslynn@mail.nih.gov</a> (mailto:rieslynn@mail.nih.gov)  Phone: 240-276-6863

## Statistical Research &amp; Applications Branch (SRAB)

Photos	Name	Areas of Expertise	E-mail
★  <a href="/about/bios/feuerr.html">(/about/bios/feuerr.html)</a>	<b>Feuer, Rocky - PhD</b> <a href="/about/bios/feuerr.html">(/about/bios/feuerr.html)</a> Chief, SRAB	Cancer Control Modeling Statistical Methods for the Analysis, Interpretation, and Presentation of Population-based Cancer Statistics Development of New Cancer Progress Measures Survival Analysis	<a href="mailto:feuerr@mail.nih.gov">feuerr@mail.nih.gov</a> <a href="mailto:feuerr@mail.nih.gov">mailto:feuerr@mail.nih.gov</a>  Phone: 240-276-6772
 <a href="/about/bios/chenh.html">(/about/bios/chenh.html)</a>	<b>Chen, Huann-Sheng - PhD</b> <a href="/about/bios/chenh.html">(/about/bios/chenh.html)</a> Mathematical Statistician	Statistical Genetics Genetic Epidemiology Biostatistics	<a href="mailto:chenh6@mail.nih.gov">chenh6@mail.nih.gov</a> <a href="mailto:chenh6@mail.nih.gov">mailto:chenh6@mail.nih.gov</a>  Phone: 240-276-6797
 <a href="/about/bios/liub.html">(/about/bios/liub.html)</a>	<b>Liu, Benmei - PhD</b> <a href="/about/bios/liub.html">(/about/bios/liub.html)</a> Mathematical Statistician	Sample Survey Design & Analysis Small Area Estimation Imputation for Missing Data	<a href="mailto:liub2@mail.nih.gov">liub2@mail.nih.gov</a> <a href="mailto:liub2@mail.nih.gov">mailto:liub2@mail.nih.gov</a>  Phone: 240-276-6718
 <a href="/about/bios/tatalovichz.html">(/about/bios/tatalovichz.html)</a>	<b>Tatalovich, Zaria - PhD</b> <a href="/about/bios/tatalovichz.html">(/about/bios/tatalovichz.html)</a> Geospatial Scientist	Geographic Information Systems, Science, and Technology Geospatial analysis, modeling, and visualization of population-based cancer data	<a href="mailto:tatalovichzp@mail.nih.gov">tatalovichzp@mail.nih.gov</a> <a href="mailto:tatalovichzp@mail.nih.gov">mailto:tatalovichzp@mail.nih.gov</a>  Phone: 240-276-6976
★  <a href="/about/bios/yum.html">(/about/bios/yum.html)</a>	<b>Yu, Mandi- PhD</b> <a href="/about/bios/yum.html">(/about/bios/yum.html)</a> Team Lead, Population and Mortality Data	Statistical Disclosure Limitation Methods for Missing Data Statistical Estimation of Health Disparities Statistical Matching and Record Linkage Methods	<a href="mailto:yum3@mail.nih.gov">yum3@mail.nih.gov</a> <a href="mailto:yum3@mail.nih.gov">mailto:yum3@mail.nih.gov</a>  Phone: 240-276-6866
 <a href="/about/bios/zhul.html">(/about/bios/zhul.html)</a>	<b>Zhu, Li - PhD</b> <a href="/about/bios/zhul.html">(/about/bios/zhul.html)</a> Mathematical Statistician	Spatial Statistics, Bayesian Statistics, Hierarchical Modeling Geographic Information Systems Disease Mapping	<a href="mailto:li.zhu@nih.gov">li.zhu@nih.gov</a> <a href="mailto:li.zhu@nih.gov">mailto:li.zhu@nih.gov</a>  Phone: 240-276-6851

[\[Return to top\]](#)

**Anonymous data; de-identified data**  
*Valentina Petkov, MD, MPH*  
*NCI Surveillance Research Program*



October 16, 2017

www.cancer.gov      www.cancer.gov/espanol

**General points**

- Some terminology
  - De-identified data - personal identifiers are removed but can be linked back to original data
  - Anonymous data – de-identified data that cannot be linked back to the original data
  - Protected Health Information (PHI): patient’s health records that can be linked to that patient
- De-identification depends on data source
  - Discrete data
  - Text data
  - Images
  - Genomic data



NIH NATIONAL CANCER INSTITUTE

3

**総論**

- 定義
  - 非特定化データ- 個人情報では除かれているが、原情報に連結可能なデータ
  - 匿名化データ - 原情報に連結不可能な非特定化データ
  - Protected Health Information (PHI): 特定の患者にひも付く医療記録
- 非特定化は資料源によって異なる
  - 離散データ
  - 文字データ
  - 画像
  - ゲノムデータ



NIH NATIONAL CANCER INSTITUTE

4

**De-identification in the context of laws and regulations**

- Protected Health Information – HIPAA law (Health Insurance Portability and Accountability Act)
  - Cancer surveillance is exempt along with other public health related reporting activities
    - SEER registries and other cancer registries (central/state registries and hospital registries) collect and maintain personal identifiers
      - Identifiers needed for long-term follow-up of cancer patients and for data linkages
- HIPAA specifies 2 methods of de-identification
  - Statistical certification of de-identification
  - Safe Harbor rule
- Summary of the law at US HHS
  - <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>

NIH NATIONAL CANCER INSTITUTE

5

**法律や規制上の「非特定化」**

- Protected Health Information – HIPAA law (Health Insurance Portability and Accountability Act)
  - がんサーベイランスは、他の公衆衛生関連の報告活動同様に適用除外
    - SEER 登録 と 其他のがん登録は (中央/州がん登録 and 院内がん登録) 個人特定指標を集めて、保有している。
      - 個人特定指標は、がん患者の長期追跡やデータ連結に必要である。
  - HIPAA は2つの非特定化手法を述べている。
    - Statistical certification of de-identification
    - セーフ・ハーバー・ルール
  - 米国保健福祉省によるHIPAA法の要約
    - <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>

SEER研究事業参加登録のこと

地域がん登録のこと

別添付けました。

予め定められた一定のルールのもとで行動する限り、違法ないし違反にならないとされる範囲

NIH NATIONAL CANCER INSTITUTE

6



## Data handling at NCI/SEER

- SEER registries submit de-identified data to NCI/SEER program twice/year
  - Data are submitted to the Information Management Systems, Inc, a contractor of NCI that manages NCI/SEER data
- Data are submitted with registry ID number, which is unique in combination with registry but does not have any meaning outside the registry
- This number is never released to the research community
- Each year a random generated number is assigned to each subject listed in the submitted data
- The de-identified data contain partial dates (month and year of diagnosis and first course of treatment) and county of residence at the time of diagnosis

13

## NCI/SEERにおけるデータの取扱い

- SEER登録は、年に1-2回、NCI/SEER 研究事業に、非特定化データを提出する。
  - データは、NCI / SEERデータを管理するNCIの請負業者であるIMS社に提出される。
- データは登録ID番号付きで提出される。これは登録室と組み合わせると一意だが、登録室外では意味を持たない。
- この番号は、研究の場には決して公開されない。
- 毎年、ランダムに生成された番号が、提出されたデータにリストされた各対象に割り当てられる。
- 非特定データには、部分的な日付(診断日と初回治療開始日の診断の月と年)と診断時の居住地の郡が含まれる。

14

## Release of SEER data

- Researchers submit a request for access to the data and SEER\*Stat software.
- A signed [SEER Research Data Agreement](#) is required
  - New Research Data Agreement is required with each data release
  - Each member of the research team needs to sign the agreement
- Main stipulations of the agreement
  - Use only for research
  - No linking
  - No publication of individual data or small group stats
  - No attempts to re-identify the data
  - If identity discovered: no use of it, no disclosure, notify SEER
  - No re-release of data

15

## SEER dataの公開・提供

- 研究者が、データとSEER \* Statソフトウェアへのアクセス申請を提出。
- SEER研究データ契約への署名が求められる。
  - データ公開の度に新しい研究データ契約が必要
  - 研究メンバーの全員の署名が必要
  - 契約書の主な規程は以下のとおり
    - 研究目的だけに使う
    - リンクしない
    - 個人データ及び少数例の公表をしない
    - データの再特定をしようとするしない
    - もし、特定された場合、それを使わない、開示しない、SEERに知らせる
    - 提供されたデータの二次提供をしない

16

## Data release: additional safeguards

- County of residence and Census Tract Socioeconomic Indexes are not release in the same data set
- Census Tract is a geographical area with 4000-8000 individuals living in it
- Information on socioeconomic variables is available at census tract level
- Those variables are used to create SES index which is released in quintiles

17

## データ提供: その他の安全管理対策

- 居住郡および国勢調査のTract Socio Economic Indexesは同じデータセットでは提供しない
- Census Tractとは4000-8000人が住む地理的エリア
- 社会経済的変数に関する情報はCensus Tractのレベルで利用可能
- これらの変数は、五分位数で公表されるSESインデックスを作成するために使用される。

18

## 18 HIPAA 'Safe Harbor' Identifiers

- Name
- Address (all geographic subdivisions smaller than state, including street address, city county, and zip code)
- All elements (except years) of dates related to an individual (including birthdate, admission date, discharge date, date of death, and exact age if over 89)
- Telephone numbers
- Fax number
- Email address
- Social Security Number
- Medical record number
- Health plan beneficiary number
- Account number
- Certificate or license number
- Any vehicle or other device serial number
- Web URL
- Internet Protocol (IP) Address
- Finger or voice print
- Photographic image - Photographic images are not limited to images of the face.
- Any other characteristic that could uniquely identify the individual

NATIONAL CANCER INSTITUTE

7

## HIPPA法による18の「セーフ・ハーバー」指標

- 氏名
- 住所 (州より小さいすべての地理区画、道路名、市郡、郵便番号)
- 個人に関連する年を除くすべての日付要素 (生年月日、受診日、退院日、死亡日、90歳以上の正確な年齢)
- 電話番号
- 電子メールアドレス
- 社会保障番号
- 診療録番号
- Health plan beneficiary number
- 口座番号
- 免許証番号
- 車両番号
- ウェブ URL
- インターネット (IP) Address
- 指紋、声紋
- 写真・顔写真の限らず、
- その他の個人を特定できる特性すべて

NATIONAL CANCER INSTITUTE

8

## De-identification in the context of laws and regulations (cont.)

- Cancer surveillance/cancer registries establishment, function and authorities are mandated by state specific laws
  - Data used for public health reporting
  - Research
- U.S. Code of Federal Regulations (CFR) governing research
  - The Common Rule: Title 45, part 46
  - Title 21, parts 50, 54, 59
    - Using de-identified data is not considered human subjects research
    - CFR does not provide specific guidelines on de-identification

NATIONAL CANCER INSTITUTE

9

## 法律や規制上の「非特定化」(cont.)

- がんサーベイランス及びがん登録の設立、機能、典拠は、州個別法による。
  - 公衆衛生目的のデータ利用
  - 研究
- 研究を統括するCFR 米合衆国・連邦規則集
  - The Common Rule: Title 45, part 46
  - Title 21, parts 50, 54, 59
    - 非特定化データの利用は、ヒトに関する研究とはみなされない。
    - CFR は非特定化について、特定のガイドラインを提供していない。

アメリカ合衆国の連邦政府により連邦官報の中で公布される、一般的かつ永続的な規則・規定を集成した法典。アメリカ合衆国の行政法として位置づけられることもある。

NATIONAL CANCER INSTITUTE

10

## Data handling at individual SEER registries

- Data are collected with patient identifiers (name, date of birth, address, telephone, alphanumeric identifiers)
- Each registry has its own policy and procedure governing release of their data for research
- Registries use the services of IMS for data storage and management
- IMS serves as an Honest Broker between the registries and NCI
- IMS security systems meet each registry requirement and FISMA
  - The Federal Information Security Management Act (FISMA) is United States legislation that defines a comprehensive framework to protect government information, operations and assets against natural or man-made threats.

NATIONAL CANCER INSTITUTE

11

## 個々のSEER登録におけるデータの取扱い

- データは個人情報付きで収集される (氏名、生年月日、住所、電話番号、alphanumeric identifiers)
- 各登録が、研究へのデータ提供を管理する指針や手法を持っている。
- SEER登録では、データの蓄積・管理に[IMSサービス]を利用している。
- IMSが、SEER登録とNCIの間の[公正な仲介者]の役割を果たしている。
- IMSの安全管理システムは各登録の要件とFISMAに適合している。
  - FISMAとは、政府の情報、業務、資産を自然災害や人為的脅威から保護するための包括的な枠組みを定める米国の法律。

IMS, Inc. 2016年10月3日 (米国東部標準時)、Quintiles Transnational Holdings Inc. (米国クインタイルズ社)とIMS Health Holdings, Inc. (米国IMS Health社)は合併してQuintilesIMSとなりました。

NATIONAL CANCER INSTITUTE

12

## De-identification of clinical text documents

- SEER registries collect pathology reports and other clinical documents
- Currently these are not submitted to NCI/SEER due to patient identifiers
  - Registries do release these for research purposes with relevant IRB approvals and Data Use Agreements
- SEER is actively searching for a reliable text de-identification tool
- Conducted evaluation of 2 de-identification tools
- Suboptimal results
- Plans for testing more de-identification tools

## 診療情報の非特定化

- SEER登録では病理報告書及びその他の診療書類を集めている。
- 現在、これらは、患者の特定指標のため、NCI/SEERには集められていない。
- 各登録室は、関連するIRBの承認およびデータ使用許諾契約書とともに研究目的でこれらを提供している。
  - SEERは、信頼できる文字情報非特定化ツールを積極的に探している。
- 2種類の非特定化ツールの評価を行った。
  - 最適ではない結果
  - 他の非特定化ツールのテストを計画中

## Performance of De-ID™ in five SEER registry

PHI type	De-Id phrases N	Missed phrases N	All PHI phrases	PII phrase DelD rate	N pts w/ missed PII	Pt level DelD rate
Names	13030	88	13118	0.993	19	0.995
Dates	8717	31	8748	0.996	23	0.994
Phone Numbers	909	0	909	1.000	0	1.000
Places	1532	0	1532	1.000	0	1.000
Street Addresses	350	10	360	0.972	7	0.998
Zip Codes	844	0	844	1.000	0	1.000
ID Numbers	1358	77	1435	0.946	51	0.987
<b>Total PHI</b>	<b>26740</b>	<b>206</b>	<b>26946</b>	<b>0.992</b>	<b>100</b>	<b>0.975</b>
Path Numbers	1678	1310	2988	0.562	810	0.798
Institutions	1355	1673	3028	0.447	825	0.794
<b>Total de-id info</b>	<b>29773</b>	<b>3189</b>	<b>32962</b>	<b>0.903</b>	<b>1735</b>	<b>0.566</b>

## Resources



- NISTIR 8053: De-Identification of Personal Information (Oct. 2015)
  - <http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf>
- NIST Special Publications 800-188: De-Identifying Government Datasets (second draft, Dec. 2016)
  - [http://csrc.nist.gov/publications/drafts/800-188/sp800\\_188\\_draft2.pdf](http://csrc.nist.gov/publications/drafts/800-188/sp800_188_draft2.pdf)

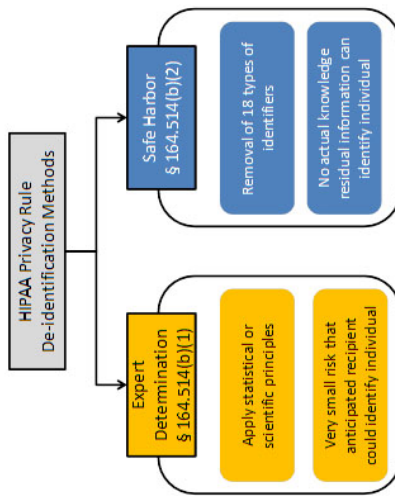


## The De-identification Standard

Section 164.514(a) of the HIPAA Privacy Rule provides the standard for de-identification of protected health information. Under this standard, health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual.

**§ 164.514 Other requirements relating to uses and disclosures of protected health information.**  
 (a) *Standard: de-identification of protected health information.* Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information.

Sections 164.514(b) and (c) of the Privacy Rule contain the implementation specifications that a covered entity must follow to meet the de-identification standard. As summarized in Figure 1, the Privacy Rule provides two methods by which health information can be designated as de-identified.



**Figure 1. Two methods to achieve de-identification in accordance with the HIPAA Privacy Rule.**

**The first is the "Expert Determination" method:**

- (b) *Implementation specifications: requirements for de-identification of protected health information.* A covered entity may determine that health information is not individually identifiable health information only if:
- (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable;
  - (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
  - (ii) Documents the methods and results of the analysis that justify such determination; or

Cancer Surveillance

# E-Path Monitoring Console

## Cancer Data Now

Reviewing electronic cancer reporting



# E-Path Monitoring Console

The E-Path Monitoring Console provides submitting sites with several important functions to assist in the management of an automated E-Path system. These include:

- Process Monitoring
- Pathology Report Review
- HIPAA Disclosure Reporting

The following descriptions of these functions assume some familiarity with the E-Path technology provided by AIM. The E-Path Monitoring Console incorporates the functionality of AIM's ISIS Registrar, Coding Edition product. Information about E-Path can be made available to you at your request.

## Process monitoring

E-Path is an automated process, designed to operate in an unattended manner. The process monitoring function provides, on demand, process reports on the operation of the system along with summary information about reports that have been processed or failed to process.

Summaries may be obtained for daily, monthly and annual periods.

**Summaries tabulate the following:**

- Total number of reports processed
- Total number of reports selected (as being reportable to the registry)
- Total number of reports that could not be processed
- Breakdown of the reports processed by report type (e.g. Pathology, Gynecological Cytology, Bone Marrow, Autopsy, etc.)

The count of pathology reports on summaries is not unique. Thus, if a report is sent to the system twice, for example, because of an amendment, it will be counted each time.

All pathology reports that could not be processed are placed in an error log. The E-Path Monitoring Console does not provide any function to retrieve these from the error log (they may be viewed using any text editor).

Processing summaries may be viewed on screen, printed or saved to text files.

## Pathology report review and forwarding

The functionality of ISIS Registrar, Coding Edition, has been incorporated into the Console to provide it with the means to search for, view, forward and review the coding of pathology reports that have been selected by the E-Path filter.

Selected pathology reports are loaded into the Console's database. Individual reports may then be viewed by selecting them from a list of available reports.

Alternatively a search function is provided to find reports by various criteria, including diagnosis. Pathology reports are displayed in their entirety and in a standard format. They may not be edited. Pathology reports may be coded with the system's AutoCode function. The assigned topography and morphology codes are displayed and may be edited. The user may also select the codes, in the event there are several, under which the report is to be classified.

The coding system normally used by AutoCode is ICD-O-3. Other coding systems may be used, but this will require the system to be reconfigured by AIM.

Although E-Path is generally configured to automatically forward selected pathology reports to the designated central cancer registry, they may also be forwarded to other destinations determined by the users. In many cases, other destinations include the local (e.g. hospital or laboratory) cancer registry system and researchers within and external to the institution.

**Pathology Reports may be exported in the following formats:**

- NAACCR Abstract v 9.1 or 10.1\*\*
- NAACCR HL7
- NAACCR Flat File

**Exported pathology reports may be:**

- Forwarded in patient identified or de-identified format
- Directed to a file
- Emailed

The selected pathology reports are stored in the Console's database and are available until the data are purged from the system. The design of the system assumes that pathology reports from several years will be on file at any one time.

## HIPAA Disclosure Reports

The HIPAA Disclosure Report lists those pathology reports that have been automatically selected and forwarded from the institution to the central registry by the E-Path system. It will not include any pathology reports that are sent using the ISIS Registrar function to forward reports.

HIPAA Disclosure Reports may be viewed and printed for daily, monthly or annual periods. The user may select any or all of the following data elements to include in the report, provided they are available in the source records:

- MRN
- Accession Number
- Patient Name
- Patient Sex
- Patient Date of Birth
- Patient Account Number

### E-Path HIPAA Patient Information Disclosure Report

Pleasant Valley Hospital

Report for the Period 2004-May-01 to 2004-May-31

MRN	Accession No.	Patient Name	Sex	Date of Birth	Date Disclosed	Disclosed To
MRN1234456788	GP-10336	Jones, Kathryn	F	1970-12-21	2004-05-01 13:37:01	State Cancer Registry
MRN1234456789	GP-10334	Kirk, Cameron	M	1963-08-10	2004-05-01 16:58:24	State Cancer Registry
MRN012234567	GP-11155	Armstrong, Michael	M	1955-02-15	2004-05-02 16:58:30	State Cancer Registry
MRN990123456	GP-90123	McDonald, Ailey	F	1975		
MRN1234456789	GP-10334	Frasier, Derek	M	1969		
MRN012234567	GP-11155	Baird, John	M	1970		
MRN990123456	GP-90123	Smart, Susan	F	1972		
MRN1234456789	GP-10334	Croft, Teresa	F	1966		
MRN012234567	GP-11155	Appleton, William	M	1959		
MRN990123456	GP-90123	Whitley, Anthony	M	1958		
MRN1234456788	GP-10336	Sean, Cynthia	F	1976		
MRN1234456789	GP-10334	Brian, David	M	1977		
MRN012234567	GP-11155	Chow, Melanie	F	1980		
MRN990123456	GP-90123	DeCruz, Grace	F	1976		

### E-Path Processing Summary

Pleasant Valley Hospital  
Report for the Period of 2004-March-01 to 2004-March-31

#### Process Summary

Category	Total
Number of Reports Received from LIS:	702
Number of Reports Processed:	702
Number of Warnings:	0
Number of Reports Transmitted:	121
Number of Reports not Transmitted:	581
Number of Reports not Processed:	0

#### Report Type Summary

Report Type	Number of Reports
Pathology	311
Cytology	91
Gynecological Cytology (Pap)	270
Bone Marrow	18
Autopsy	4
Other	8
Unknown	0
Total processed:	702

- Social Security Number
- Date Disclosed

The recipient and reason for disclosure may also be specified and included in the report

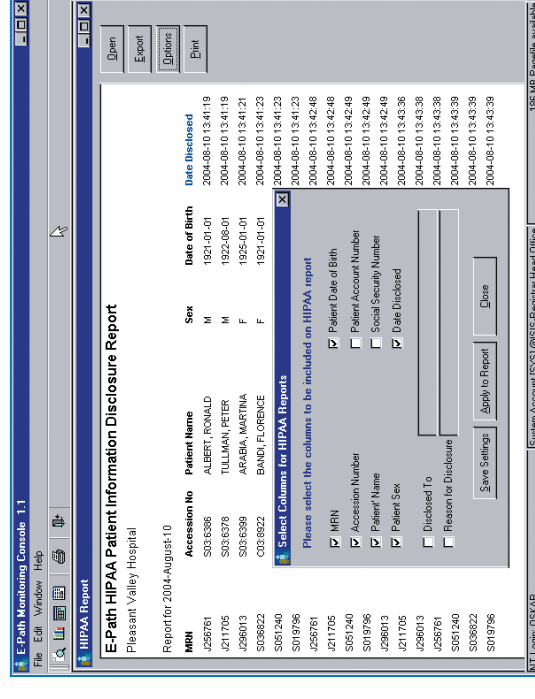
#### Hardware Requirements and Installation

It is recommended that the E-Path Monitoring Console be installed on a different computer from the one on which E-Path itself is deployed. The two computers need to be connected to the same network.

The system is intended to run on a standard desktop PC running any recent version of Microsoft Windows.

**Recommended specifications include:**

- Intel Pentium II Processor
- 300 MHz or greater
- 64 MB RAM
- 2 GB available disc space



*Note: The abstract record will contain only the components available in the pathology report.*