

I. 総括研究報告

厚生労働科学研究費補助金（がん対策推進総合研究事業）
（総括）研究報告書

全国がん登録を基盤とした長期記述疫学研究用特定匿名化情報の整備に関する研究
研究代表者 柴田 亜希子 国立がん研究センターがん対策情報センターがん登録センター 室長

研究要旨

本研究は、平成 31 年 1 月予定の全国がん登録情報の提供開始の前に、米国の SEER をモデルとして、がん登録推進法その他関連法令等と照らし、我が国の実情に即して長期持続可能な、全国がん登録を基盤とする「特定匿名化情報」の提供並びに「がん診療情報」の収集の仕組み（日本版 SEER）の提案を目的とする。初年度に SEER の全体像を理解し、最終年度に日本の実情に即した仕組みを提案する。初年度の平成 29 年度は、SEER の制度や組織の仕組みを理解するために、ウェブサイト等の公開情報の調査及び公開情報で不明な点の現地聞き取り調査を行い、SEER Data の内容や利用手続きを理解するために SEER Data を用いた記述疫学研究を開始した。SEER Data の実利用、ユタ州がん登録と SEER 事務局の調査分析を通して、SEER 事業が長期にわたり継続可能な背景要因について、米国の医療制度、研究体制及び法制度、文化との関連づけを行いながら整理した。これを踏まえ、がん登録推進法その他の関連法令等及び自らの医療情報の提供に関する我が国の国民意識と照らして、我が国に導入可能な点の整理を開始した。特に、SEER のように長期に渡る公的研究事業が我が国では維持されにくいこと、研究と個人情報保護のバランスに関する米国と我が国の法体系及び国民意識の違いを焦点に、平成 30 年度は、引き続き、我が国において、がん登録データの発生源である医療機関から「がん診療情報等」を収集し、がん登録推進法下で比較的活用しやすいと想定される情報である「特定匿名化情報」として提供可能なあり方を検討する予定である。

研究分担者

伊藤秀美
愛知県がんセンター・研究所
遺伝子医療研究部 室長
井上真奈美
（国研）国立がん研究センター
社会と健康研究センター
コホート連携研究部 部長
大木いずみ
（地独）栃木県立がんセンター
がん予防情報相談部 部長
金村政輝
宮城県立がんセンター研究所
がん疫学・予防研究部 部長
西野善一
金沢医科大学
医学部公衆衛生学 教授
林櫻松
愛知医科大学
医学部公衆衛生学 准教授

東尚弘
（国研）国立がん研究センター
がん対策情報センター
がん登録センター センター長
片野田耕太
（国研）国立がん研究センター
がん対策情報センター
がん統計・総合解析研究部 部長
堀芽久美
（国研）国立がん研究センター
がん対策情報センター
がん統計・総合解析研究部 研究員
松田智大
国立がん研究センター
がん対策情報センター
がん登録センター 全国がん登録室長

研究協力者

平田公一
札幌医科大学消化器・総合、乳腺・内分泌
外科 客員研究員

A. 研究目的

がん登録推進法では、登録情報の十分な活用と国民への結果の還元が求められている。この基本理念の推進を目的として、提供要望頻度が高いと見込まれる情報について、あらかじめ、全国がん登録情報の匿名化情報（特定匿名化情報）を整備し、全国がん登録データベースに記録できる、とされている。また、基本理念には、がん対策の充実のためには、全国がん登録の実施のほか、がんの診療の状況を適確に把握することが必要であることに鑑み、院内がん登録により得られる情報その他のがんの診療に関する詳細な情報（がん診療情報）の収集が図られなければならない、とされている。

米国では、National Cancer Institute (NCI) が、協力州から基本的な州がん登録情報その他のがん診療情報の提供を受けて SEER (The Surveillance, Epidemiology, and End Results) Data を約 40 年以上にわたって整備維持している。そのデータは簡単な申請で研究者に提供されているため、定型の公表集計値以外の部位や組織型の罹患率や生存率の記述疫学研究等に幅広く活用されている。SEER Data の特徴は、最大で協力 18 州のデータに限られるが信頼性の高い罹患率と生存率を算出できること、時代時代で関心の高い情報を追加で収集し、整備されていることである。我が国では 10 数年前から米国を参考に院内がん登録及び地域がん登録の仕組みの標準化と登録精度向上に努め、2013 年にがん登録推進法が成立した。同法に基づき保存される 2016 年罹患の全国がん登録情報の提供は 2019 年 1 月に開始予定である。研究代表者らは、平成 27 年度から、全国がん登録情報の第三者提

供に関する国民意識の調査並びにがん診療情報の収集方法の一つとして既存のがん診療情報との連携可能性に関する研究を続けてきた。その結果を踏まえ、我が国で SEER と同等の情報提供が可能であれば、調査研究への速やかな積極的な活用を促し、日米等共同研究に対して比較可能な基本的な情報を提供できると考える。

本研究は、平成 31 年 1 月予定の全国がん登録情報の提供開始の前に、SEER をモデルとして、がん登録推進法その他関連法令等と照らし、我が国の実情に即して長期持続可能な、全国がん登録を基盤とする「特定匿名化情報」の提供並びに「がん診療情報」の収集の仕組み（日本版 SEER）の提案を目的とする。

B. 研究方法

SEER の制度や組織の仕組みを理解するために、ウェブサイト等の公開情報の調査及び公開情報で不明な点の現地聞き取り調査を、SEER Data の内容や利用手続きを理解するために SEER Data を用いた記述疫学研究を 2 年計画で実施する計画とした(図 1)。

研究班は、前者を担当する仕組み検討班と後者を担当する記述疫学研究班の 2 つの作業部会で構成した。作業部会は、研究分担者の実績に応じて担当した。

仕組み検討班 (柴田、松田、東、西野、大木、金村、伊藤)

1 年目の調査によって以下の点を明らかにすることとした。

- ・ 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：柴田、松田、西野）。

- ・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)。
- ・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、柴田)。

本分野を担当する研究分担者は、将来的に、SEER 協力州に相当する都道府県の候補となり得る都道府県の、がん登録に見識のある研究者とした。

訪問先は、SEER の最も古い 1973 年からの 9 つの協力州の一つである Utah Cancer Registry (ユタ州がん登録) と SEER 事務局に決定し、2017 年 8 月 29 日、31 日にそれぞれを訪問した。ユタ州がん登録には柴田、井上、伊藤、大木、金村が、SEER 事務局には柴田、井上、東、堀、林が訪れた。林は、自身の記述疫学研究の相談のためにハワイ州がん登録を訪問した。

記述疫学研究班 (伊藤、井上、林、片野田、堀)

研究を通じて、以下の点を明らかにする

ことにした。

- ・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。

2つの作業部会は、それぞれの調査で知り得たことを班会議その他で共有し、調査の効率化を図った。

(倫理面への配慮)

本研究においては人体から採取された試料は用いない。仕組み検討班においては、個々のがん登録情報を用いず、運営や仕組みに関する検討を中心に行うため、個人情報保護上、特に問題は発生しない。記述疫学研究班においては、研究用に整備された匿名化された情報を用いる予定であるが、「人を対象とする医学系研究に関する倫理指針」に準拠し、必要に応じて国立がん研究センター倫理審査委員会の審査を受けるものとする。他の公的統計資料を利用する場合は、それぞれの利用手続きに則った。

年度	平成29 (1年目)				平成30 (2年目)						
年次	2017				2018				2019		
月	1-3	4-6	7-9	10-12	1-3	4-6	7-9	10-12	1-3	4-6	7-9
全国がん登録予定				全国がん登録情報提供指針		全国がん登録情報提供説明会	厚生科学審議会がん登録部会による提供審議		情報提供開始予定		
研究班予定	仕組み作りWG 記述疫学WG	SEER公開情報による知見収集 記述疫学研究計画利用手続	班会議 SEER開取調査 提供データ分析	班会議 調査分析知見統合と日本版SEERの検討		日本版SEER仕組み案まとめ 記述疫学研究まとめ	班会議 日本版SEER仕組み案まとめ 記述疫学研究まとめ	班会議	研究総括		

図1 平成29年度からの2年の研究計画

C. 研究結果

SEER 事業について

研究分担者の伊藤、片野田、林、堀が、疫学研究を計画し、SEER Data の利用手続きを行った。SEER linked database を用いた研究も検討されたが、このデータベースの国外からの利用は受け付けていないことが分かった。一方、SEER Data については、SEER のウェブサイトからオンラインで利用申請し、利用規約に対する誓約書を電子メールで送付することで、国外からの利用申請が可能で、データが提供された。提供されたデータから判断する限り、氏名は含まないものの、年齢は各歳階級、地理情報は日本では郡に相当する County レベル、がんの組織型については国際腫瘍学分類の組織型約 2,000 分類を含む詳細なデータであった。これらの情報を組み合わせると、集計結果が容易に 1 例になる場合があると考えられた。利用規約において、個人を特定できるデータを公表しない、少数例の集計値の公表を避ける、等の少数集計値による個人識別性への公表時の対応を求め、提供に当たり遵守を誓約させていた。利用手続きの実際の詳細は、研究分担者の堀らが分担報告書にまとめた。

公開情報や実際に提供されたデータからは、前述以上の匿名加工処理が行われているか分からなかったため、現地聞き取り調査時の確認事項とした。聞き取り調査では、SEER Data は、氏名除去、生年月日の年齢表記と生年処理、診断日の年月処理、生存期間の生存期間表記、地理情報の County までの処理の匿名加工処理がされていることが分かった。氏名は収集段階で別途管理されており、SEER の担当部署では取り扱わない。

法的背景には、Health Insurance Portability and Accountability Act (HIPAA) は、その他の公衆衛生目的の情報収集と同様に、がんサーベイランス事業には適用されないこと、また、研究に関する米国連邦法により、個人情報には除かれているが、原情報に連結可能なデータ (de-identified data) を取り扱う研究は、原情報に連結不可能な非特定化データ (anonymous data) ではなくても「ヒトを対象とする研究」とは見なされないこと、があることが分かった。さらに、HIPAA が非特定化手法を例示しており、SEER 事業ではそのセーフハーバールールを参考にしていただけたことだった (付録 1)。データの提供を受けることの叶わなかった SEER linked database のように、がん登録データベースを他のデータベースとリンクすることについては、前述のとおり、個人情報の管理組織を別にする配慮はしているものの、それに関して個人情報の取扱い上問題だという議論は特になかったことだった。詳細は、分担研究者の東が付録 2 にまとめた。

SEER Data には、各がんの詳細な病期の情報が含まれていることが最大の特徴である。その他、期間限定ではあるが、腫瘍マーカー、治療手技、ホルモンレセプターの情報も含まれている (付録 3)。このような詳細な診療情報を、住民ベースのがん登録で正確に収集するのは、一般的に容易ではない。我々は、SEER に協力している州がん登録がどのような手法で詳細な診療情報を収集しているのかを知るために、ユタ州がん登録を調査した。前もって、我々の研究上の関心事項を先方の担当者に連絡したところ、訪問時に以下の 7 項目の議事に分けて詳細な

説明を受けた（付録 4）。

- 1) 州がん登録の概要と環境
- 2) データの収集について
- 3) データの編集、集約について
- 4) データの機密保持について
- 5) 追跡について
- 6) データ照合
- 7) データシェア

ユタ州では州法でユタ州健康局が疾病の実態把握をすることと定められており、ユタ州がん登録は、州の行政規則によって、ユタ州健康局の代理でがんの情報を管理する組織と定められている。現在は、ユタ大学の一部門である。ユタ州がん登録のデータは、ユタ州健康局に帰属し、情報公開される。その他の情報公開については、年 2 回、ユタ州健康局、病院、研究者の代表から構成される「研究諮問委員会」で検討される。現在、ユタ州がん登録データは、ユタ州健康局、SEER 事業、CDC（Centers for Disease Control and Prevention）の事業である NPCR（National Program of Cancer Registries、ユタ住民データベース、研究申請、非研究申請に対して提供されている。ユタ州がん登録と SEER 事業の関係は、単純には「契約」とのことだった。ユタ州がん登録は、契約に従って、SEER から資金その他の支援を得て、SEER の求める診療情報も含めて医療機関から情報収集し、SEER に提出する。その他の複数の研究事業とも同様の協力契約を交わしている。州の行政規則は、がん登録について、収集するデータの形式や変数等の具体的な内容に言及していないため、州がん登録は、SEER 事業、NPCR、その他の研究事業が必要とする項目を収集することができるとしている。ユタ州がん登

録では、SEER 事業財源で、複数名の実務者が雇用されている。

さらに、ユタ州の行政規則において、すべての病院その他のがんの診断・治療を行う施設はがん登録に報告しなければならないとされている。これによって、病院等は、届出のための情報の抽出にかかる費用を自ら払わなければならない。病院等は、届出のために、腫瘍登録士を雇用したり、同様のサービスを提供する会社と契約している。ユタ州の病院等は概ね協力的であり、ユタ州がん登録は病院等に適切な時期の報告について勧奨する連絡は絶やさないが、ユタ州がん登録や州が報告に関する行政規則を行使したことはないとのことであった。

さらに、米国では、がん登録のデータの標準化とその適用が進んでいる。North American Association of Central Cancer Registries（NAACCR）が情報の集約のための標準的なレイアウト形式を提供している。SEER や NPCR は標準方式の決定機関であり、NAACCR は標準方式の普及や標準方式の決定機関の協力機関である。ユタ州には、ユタ州がん登録と院内がん登録を支援するユタ州がん登録協議会という地方組織があり、年に一度の会議の開催の他、日頃から電子メールや電話で連絡をとって、助け合っている。ここまでの詳細は、研究分担者の金村が分担報告書にまとめた。

データの収集について、ユタ州がん登録の特筆すべき点は、独立したがんの情報源として病理報告を含めており、その 88% が e-Path という SEER が開発した電子的病理報告情報収集システムで収集されている。本システムはがん患者のケースファインディング用であり、がん登録の報告用ではな

い。ユタ州では、病院や検査機関に本システムの導入を進め、がん以外の全病理報告を収集し、ユタ州がん登録において一部機械的判定も使いながら新規がん患者の見つけ出しを行っていた。病理報告があつて、がん登録の届出がない場合には、病院に届出を促すさかのぼり調査を行うので、最終的に病理報告のみでがん登録されているケースは約5%程度とのことであつた。

医療機関からがん登録への届出の方法は、電子的な届出については安全性の確立されたファイル送達システムや暗号化されたウェブポータルサイトも提供されているが、基本的には届出電子情報が暗号化されていれば、一般的なメール添付も可能とのことであつた。手書きの届出も一部残っており、郵送のみならずファックスの届出も受け付けていた

その他の住民ベースのがん登録の基本的な業務であるがん届出情報の確認、編集、既登録腫瘍か否かの判定と情報の集約には、SEER や NPCR が開発して提供しているツールその他の独自開発のツールを使って効率的に問題のあるデータの抽出と編集を行っていた。米国独自の認定資格である「腫瘍登録士」が、データの編集、集約を行い、品質管理に責任を持っていた。ここまでの詳細は、研究分担者の大木が分担報告書にまとめた。

同一人物の情報の連結(リンケージ)は、がんの届出情報同士のみならず、死亡その他の追跡情報との連結の作業とまとめ、独立した業務とされていた。個人識別指標は、姓、名、生年月日、社会保障番号で、住所は照合時の確認に用いるとのことであつた。照合ソフトは、SEER の開発した SEER Link

Plus を利用していた。登録患者の生死状況の追跡には、様々な資料源を用いられていた。SEER 事業に参加することで、SEER の支援の元、国の死亡情報と社会保険庁データとの照合を行うことができる。その他、院内がん登録、公的保険加入名簿、有権者登録名簿、担当医師及び患者への直接連絡で、追跡情報を得ていた。ユタ州全体の追跡率は99%とのことであつた。

ユタ州がん登録情報の SEER や NPCR へのデータ提供は、データの機密保護及び被験者保護の連邦ガイドラインとがん研究促進のための要件を含む契約にて行われていた。一般的な随時の研究への提供は、研究に対する諮問委員会で審議して決定されることであつた。ここまでの詳細は、研究分担者の伊藤が分担報告書にまとめた。

SEER Data を用いた記述疫学研究について

研究分担者の林は、日米ともに増加傾向にある膵がんについて、罹患率・死亡率の年次推移、診断時ステージ、組織型分布、ステージ別5年生存率の日米比較を計画した。今年度は、SEER Data の利用申請を行い、SEER が提供している統計解析ソフト SEER*Stat の Client-Server Mode を用いて、米国分の解析を行った。また、日系アメリカ人のがん罹患率や死亡率の情報は、SEER Data には含まれていなかったが、別の Asian Pacific Islander (API) Specialized Population Dataset に含まれており、別途の仕様申請で提供を受けられた。2000年以降、米国では膵がんの罹患率、死亡率ともに上昇傾向が続いていること、診断時ステージ分及び5年生存率からみて、依然として早期発見が困難で、予後が

大変不良であることが分かった。詳細は、分担報告書にまとめた。

研究分担者の堀と片野田は、卵巣がんの年齢別、組織型別罹患率の日米比較を計画した。今年度は、SEER Data の利用申請を行い、SEER*Stat の Client-Server Mode を用いて、米国分の解析を行った。罹患数は1歳刻みの年齢区分が利用可能であるが、人口データは1歳年齢階級に対応していないことが分かった。また、地理情報は County まで利用可能で、さらに County に対応した American Community Survey データの利用が可能なので、地域別の社会・経済指標とがん罹患・死亡との関連分析が可能であることが分かった。組織型については、ICD-O-3 分類に基づく組織型/性状情報が含まれ、診断年その他の患者情報(性別、年齢階級、人種等)を研究者自らの任意の組み合わせる集計が可能であった。SEER*Stat から取得したデータを用いて、卵巣がんの年齢階級・人種・組織型別罹患率を計算した。White, Black, American Indian/Alaska Native, API 別では、API とその他の人種の卵巣がん罹患率に、年齢別、組織型別に違いが見られたため、日米比較を行う意義が確認された。詳細は、分担報告書にまとめた。

D. 考察

記述疫学研究班の研究計画と結果概要は、日本版 SEER によってもたらされるだろう我が国のがんによる負荷に関する記述統計の充実と、それによる公平、効率的ながん対策企画立案への貢献の可能性を示している。SEER Data の素晴らしさは、住民ベースのがん登録データであること、診療情報を含むこと、高品質であることである。Data で

示すとおり、これらの3点を長期にわたり満たすのは容易ではない。一般的に、より正確に数(がん罹患者)を把握しようとするればその他の項目の正確な取得が、個々の患者に関する診療情報をより多く把握しようとするれば全数登録は難しい。また、SEER Data 提供の仕組みの素晴らしさは、原則として国内外問わず、簡単な手続きで誰でも利用できることである。日本では、2015年の個人情報保護法改正によって匿名加工情報が規定され、匿名加工情報の定義を満たす情報を、事業者は本人の同意なく第三者に提供できるようになった。がん登録推進法における全国がん登録情報の匿名化された情報の定義は、改正個人情報保護法の匿名加工情報とは独立しており、本研究班の発足時に厳密な定義が未定であったが、個人情報保護法の考え方は尊重されるであろうことから、「匿名化」が易利用性において重要になってくる。我々は、SEER 事業が、住民ベースで診療情報を含む高品質のデータを整備し続ける仕組み及び彼らの匿名化の考え方について調査した。

まず、匿名化の考え方について、米国では、非特定化データと匿名化データの区別があり、非特定化データを用いる研究は「人を対象とする研究」と見なさないとする法規制が前提となる。日本の場合、個人情報保護法における「匿名化」の定義は、米国の匿名化データの定義(個人情報除かれて、原情報に連結不可能)に近い。一方で、HIPAA の非特定化データに関するセーフハーバー規則は、日本の個人情報保護法ガイドライン(匿名加工情報編)による匿名加工情報の適正な加工に近い。日本の個人情報保護法ガイドラインでは、情報を相互に連結する

符号もがあれば必ず削除するか置き換えをしなければならないこと、個人情報データベース等の性質に応じて復元可能性について個別具体的に判断する必要があることがHIPAAのセーフハーバー規則とは異なるが、氏名と個人識別符号の削除又は置き換え、生年月日等の年月までの加工、地理情報の最小基準までの加工、特異な記述の削除や置き換えについては、ほぼ同様の基準である。

SEER Dataは、氏名除去、個人識別符号除去、生年月日の年齢表記と生年処理、診断日の年月処理、生存期間の生存期間表記、地理情報のCountyまでの処理の匿名加工済みの非特定化データであり、原情報への復元可能性は否定されない。その点は、利用者との利用規約において、提供された情報から万が一個人を識別できてしまった場合は事務局に報告すること、その情報は削除すること、少数集計値では公表しないこと、を規定することで配慮している。

がん登録推進法上の匿名化は、「他の情報との照合による識別を含め、当該がん罹患した者の識別ができないように加工すること」と定義されている。個人情報保護法ガイドライン（匿名加工情報編）では、個人情報保護法における「特定の個人を識別することができる」とは、情報単体又は複数の情報を組み合わせて保存されているものから社会通念上そのように判断できるものをいい、一般人の判断力又は理解力をもって生存する具体的な人物と情報の間に同一性を認めるに至ることができるかどうかによるものである。匿名加工情報に求められる「特定の個人を識別することができない」という要件は、あらゆる手法によって特定する

ことができないよう技術的側面から全ての可能性を排除することまでを求めるものではなく、少なくとも、一般人及び一般的な事業者の能力、手法等を基準として当該情報を個人情報取扱事業者又は匿名加工情報取扱事業者が通常の方法により特定できないような状態にすることを求めるものである、とされていることから、がん登録推進法上の匿名化についても、ほぼ同様に考えてよいであろう。

以上より、本研究班は、全国がん登録情報の匿名化の加工とは、全国がん登録の原情報から、氏名除去、個人識別符号除去、情報を連結する符号の除去、日付項目の置き換え（生年月日の年齢表記、一定年齢以上の高齢年齢の置き換え、死亡日等の生存期間表記）、診断時住所の市区町村より詳細部分の削除を最低限の加工とし、加工後の情報に特定の個人を識別可能な特異な記述が含まれるか否かは、全国がん登録情報等の提供に係る審議会等の意見を聴くことを提案する。

SEER事業のもう一つの素晴らしい特質である、住民ベースで診療情報を含む高品質のデータを整備し続ける仕組みについては、SEER事業が、資金提供、教育及び登録システムの環境整備支援を条件とした、データの提供元である州がん登録との共同研究事業であることが要であることが分かった。各州のがん登録の存在根拠や基本的な財源は各州の法律と財源によるが、それを越える活動についてはSEER事業その他の研究事業との契約によって、自らの可能な範囲で行っている。また、SEER事業を継続するために、NCIのSEERを管理する部門は、常に事業改善計画を考え、新しい事業を付加

し、事業価値を維持、高める工夫をしている。

日本の場合、がん登録推進法施行前は、米国の州がん登録のように都道府県がん登録が存在し、SEER 事業のように厚生労働省の研究班が日本全体のがん罹患の実態を把握する役割を担ってきた。しかしながら、日本の一つの公的研究事業が SEER 事業のように 40 年も持続することはまれであり、各都道府県がん登録を必要十分に支援できる仕組みでもなかった。現在は、日本の住民ベースのがん登録は、研究事業ではなく、がん登録推進法に基づく国の事業であり、各都道府県には法定受託事務として必要な経費の 2 分の 1 が国から補助されている。この体制において可能な、全国がん登録届出項目を越えるがん診療情報を収集できる日本版 SEER の仕組みを提案する必要がある。

研究班は、検討開始当初、全国がん登録届出項目にがん登録推進法に基づく標準的な院内がん登録項目を加えて、院内がん登録実施病院かつデータ提供に同意する病院からのデータからなる日本版 SEER の在り方を検討した。しかし、この方法で収集されるデータはがん診療を専門とする病院情報というバイアスがかかり、日本全体の診療実態の把握は難しいこと、分母となる集団を想定できないため罹患率を計測できないこと、別途行われている院内がん登録全国集計事業との違いは病院間の重複症例の有無だけであることから、本研究班で目指す日本版 SEER は住民ベースを目指すことにした。この問題については、研究分担者の井上が、「疫学研究として活用するためのデータベースの条件」をテーマとして分担報告書にまとめた。また、全国がん登録項目の他

のがん診療情報を住民ベースに収集するにあたっての課題を、ユタ州がん登録の運営を参考に研究分担者の金村が抽出した（分担報告書）。これらの課題については、期間限定の短期の公的研究事業で着手可能なこと、施行から 5 年、つまり 2020 年過ぎを目処に予定されているがん登録推進法の改正が必要なことの視点で、平成 30 年度に引き続き検討する。

平成 29 年度は、年度末に「全国がん登録情報の提供マニュアル」が厚生労働省らによって公開された他、改正個人情報保護法が施行されたことを受けて、個人情報を含む情報の活用のための第三者への提供に関連する分野の規則や指針の新設や見直しが進められた。平成 30 年度は、このような我が国固有の法的背景、環境を考慮して、本研究班で引き続き検討の必要な課題が何点か残っている。

一点目は、「全国がん登録 情報の提供マニュアル」において全国がん登録情報の提供及び利用の在り方が明確化され、その枠組みでは、たとえ我が国の定義による匿名化された情報が整備できたとしても、その情報を SEER Data のように国内外を問わず、簡単な手続きで利用できる仕組みからは遠いようにみえることである。「全国がん登録情報の提供マニュアル」の別添として示される予定の利用規約及び提供を受けた全国がん登録情報の利用者が行うべき安全管理措置の内容の十分な検討が必要である。全国がん登録情報利用者に求められる安全管理措置の現状と課題について、研究分担者の西野が分担研究報告書にまとめた。また、井上が分担研究報告書で述べているように、単なる集計結果の比較ではなく、非特定化

データであっても SEER Data のように個票単位のデータを使えることで記述疫学の利用価値として高まるが、我が国の現環境で可能な形と将来のあるべき形の提案に分けて、どこまで SEER Data の提供の仕組みに近づけるか考えたい。

二点目は、がん登録推進法の施行によって、我が国のがんの長期記述疫学研究の用に資するデータの利用の仕組みが、2015 年以前と 2016 年以後に分断されてしまったことに対する対応策である。井上が分担研究報告書に記載しているように、対策のためのエビデンスが求められている現在、長期情報は当面解析できないのは問題である。2015 年以前の都道府県事業によって収集されていたがん罹患情報を、2016 年以後の匿名化された情報と一体として、我が国のがんの長期記述疫学研究の用に資するために利用できるように整備される必要がある。

E. 結論

平成 29 年度は、概ね、研究計画とおりに進捗した。SEER Data の実利用、ユタ州がん登録と SEER 事務局の公開情報調査分析及び訪問を通して、SEER 事業が長期にわたり継続可能な背景要因について、米国の医療制度、研究体制及び法制度、文化との関連づけを行いながら整理した。これを踏まえ、がん登録推進法その他の関連法令等及び自らの医療情報の提供に関する我が国の国民意識と照らして、我が国に導入可能な点の整理を開始した。特に、SEER のように長期に渡る公的研究事業が我が国では維持されにくいこと、研究と個人情報保護のバランスに関する米国と我が国の法体系及び国民意識の違いを焦点に、平成 30 年度は、引き続

き、我が国において、がん登録データの発生源である医療機関から「がん診療情報等」を収集し、がん登録推進法下で比較的活用しやすいと想定される情報である「特定匿名化情報」として提供可能なあり方を検討する予定である。

F. 健康危険情報

特になし

G. 研究発表

1. 論文発表

1) Katanoda K, Shibata A, Matsuda T, Hori M, Nakata K, Narita Y, Ogawa C, Munakata W, Kawai A, Nishimoto H. Childhood, adolescent and young adult cancer incidence in Japan in 2009–2011. *Jpn J Clin Oncol.* 2017; 47: 762–771

2. 学会発表

- 1) 4. 柴田亜希子. 全国がん登録からみた臓器がん登録. 第 25 回日本乳癌学会学術集会 (シンポジウム) : 2017. 7: 福岡県.
- 2) 柴田亜希子, 松田智大. 全国がん登録の匿名化情報の研究利用に関する一般意識調査. 第 76 回日本癌学会学術集会 (ポスター, 英語) : 2017. 10: 神奈川県.
- 3) 柴田亜希子. 全国がん登録の現状. 第 55 回日本癌治療学会学術集会 (シンポジウム) : 2017. 10: 神奈川県.
- 4) 柴田亜希子. がん登録推進法の 5 つの基本理念 : 基本理念の実現に向けて. 第 30 回日本放射線腫瘍学会学術大会 (シンポジウム, 英語) : 2017. 11: 大阪府.

H. 知的財産権の出願・登録状況（予定を含む。）

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

付録1 米国における非特定化データと匿名化データの考え方（米国 NCI 提供資料）

付録2 2017年8月31日 NCI DCCP 訪問
まとめ（研究分担者 東尚広）

付録3 SEER RESEARCH DATA RECORD
DESCRIPTION CASE DIAGNOSED in 1973-
2014*（SEER Data 利用申請者提供資料）

付録4 ユタ州がん登録提供資料

Anonymous data; de-identified data

Valentina Petkov, MD, MPH
NCI Surveillance Research Program

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



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



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>

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 とその他のがん登録は (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
 - SEER研究事業参加登録のこと
 - 地域がん登録のこと
 - セーフ・ハーバー・ルール
 - 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
 - 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>

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

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・顔写真に限らず、
- その他の個人を特定できる特性すべて

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

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。

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.

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。

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

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
Total PHI	26740	206	26946	0.992	100	0.975
Path Numbers	1678	1310	2988	0.562	810	0.798
Institutions	1355	1673	3028	0.447	825	0.794
Total de-id info	29773	3189	32962	0.903	1735	0.566

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



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.

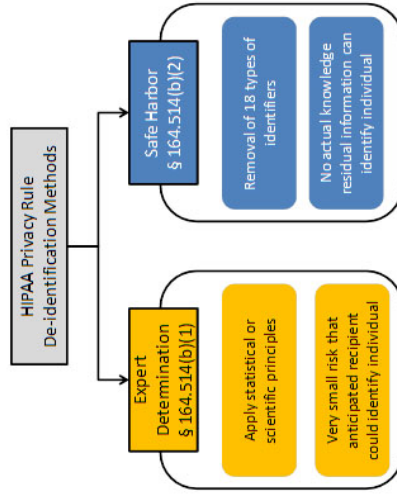


Figure 1. Two methods to achieve de-identification in accordance with the HIPAA Privacy Rule.

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

データの活用について

- ・データの活用はNCIの性格として研究を推進することを中心としていることから、活用が容易にできるようになるのは自然な流れという雰囲気。時代的（1973年当時から）も一つの要素ある
- ・County levelで希少がんだと、データがユニークになってしまうこともあるが、その人をもともと知っていないと、それが当該人物だとわからないので、誓約書一つでレコードレベルのデータを出すことに違和感はない。ただ、議論はいつもあるのも事実。特にどの程度のユニークなレコードがあったらいけない等の基準はない。
- ・過去に漏洩事故というのは起きたことが無いので、それも信用の一つにはなっている。
- ・特に誓約書を守っているかは調べていないが、守っていなかった事例については、「データを商業的に売っている」という事例であり、同僚などの通報で発覚した。
- ・HIPPAはがん登録には適用されないが、18の個人情報定義されているので参考にはされる。またFederal ruleにより、De-identifiedな情報は(個人情報が削除されたもの)はanonymous（連結不可能匿名化）ではなくても、Human Subject research とは見なされない。

データの質について

- ・年に2回データを集めていて、4月に報告書が出る。
- スケジュールは、
- Nov: 18registry から収集 (IMS で処理)
 - Dec: SEER では Outlier を見ることに、過去のデータと比べてチェック
 - Jan: SEER Stat file をまず作ってみる
 - Feb: 再度 Quality Check
 - Mar: データ準備
 - Apr: リリース
- ・死亡データは、NDI、SSA（年金）、CMS（保険）の全てからチェック
これらの相違があったときのアルゴリズムがある。
 - ・名寄せなどは、Link-plus, Big Match などのソフトが使われる。
 - ・過去のデータは、毎回集め直している。過去データの訂正はそのときに反映される。

データリンクについて

- ・リンクを行うことは必然である。それに関して特に個人情報の問題だという議論はない。
- ・有用性とプライバシーリスクを常に考えるべき、という意識のもとでリンクが妨害されることはない。
- ・どこでリンクをしているのか、は不明。例えば、SEER-MEDICARE は、Web の説明では、SEER に Personal Identifier が送られてきて CMS でリンクしていると書いてあるが、実際はどこでやっ

ているのかは不明。作業は Information Management System (IMS) という。(薬の卸データをあつ
えめている IMS とは別物)

- ・ NCI 内でそれをやっているのは Healthcare Delivery Program であり、別部署

e-Path

- ・ De-ID というソフトも有り、個人情報らしい物が出てきたら削除するという機能がある。(性能については Valentina のスライド参照)
- ・ CDC、NCI 両方作っているが、病理検査会社にインストールして、情報を送っていく機能
- ・ NCI が配っているソフトは、AIM というカナダの会社が作ったもの。Case-finding を行ってデータをリアルタイムに送ってくるため。Rapid Case Ascertainment という、臨床試験へのリクルートなどに役に立つ。(未)
- ・ CDC が配っている E-mark は別物で、データを抽出する機能が主となっている。(未)

その他

- ・ 様々な活用ソフトを用意している。特にランキングは重要だが、不安定なランキングであることを示すために、ランキングの 95%CI 等も出している。
- ・ CIS-NET は USPSTF と密接に連携している。特に RCT で効果が出た検診について、シミュレーションで Replicate したあとに、その上で間隔や対象などのパラメータを動かしてみても最適な方法を検討する事などに使われている。
- ・ SEER multiple primary のルールは、solid tumor という名前に変わる、2018 年からの適用を目指して作業しているが、ICD-O-3 が 3.2 という形になってくるので、遅れるか、部分的になるかもしれない。(後から浮かんだ疑問：ICD が関係するのは、Histology rule だけでは?)
- ・ ICD-O-3 よりも進んで、WHO 分類が新しいコードを作ってしまうことがある。このときにはそれはとくに排除しない。
- ・ 未確認情報だが、がん登録の情報を集める仕組みができてはいるものの、細かくどの項目をあつめる、という範囲は、それほど明確に決まっているわけではなさそう。そのために病理譲歩を集めるなどの事が可能になったり、CVS Pharmacy から情報を集めることが可能になっている?
- ・ 州レベルでのリンクは非常に良く行われている。ルイジアナなどは好例。
- ・ residual tumor registry というものがある。E-path で病理レポートがあるため、それをたどれば組織検体を収集することも可能になる。
- ・ Virtual Pooled Registry

様々な Pool をする。Hash 関数を使って匿名化した上での Registry 間の名寄せをする。一つのファイウォールの後ろで処理をするので安全

Centralized IRB のサービス(?)も行っている。多くの Registry のデータを使う際に全部の registry から IRB を受けなくても良くなる。

SEER RESEARCH DATA RECORD DESCRIPTION

CASES DIAGNOSED IN 1973-2014*

Submission: *November 2016*

Follow-up Cutoff Date: *December 31, 2014*

Documentation Version: *April 2017*

Diagnosis Years: *1973-2014*

** This documentation describes the data files in the incidence/yr1973_2014.seer9, yr1992_2014.sj_la_rg_ak, yr2000_2014.ca_ky_lo_nj_ga , and yr2005.lo_2nd_half directories. Refer to individual variable definitions to determine the differences between the directory files.*

Cervix in situ cases after 1995 are not included.

COMPUTER RECORD FORMAT

NAACCR Name	NAACCR Item #	SAS Variable Name	Applicable Years	Position	Length
Patient ID number	20	PUBCSNUM		1-8	8
Registry ID	40	REG		9-18	10
Marital Status at DX	150	MAR_STAT		19	1
Race/Ethnicity	160	RACEIV		20-21	2
NHIA Derived Hispanic Origin	191	NHIADE		23	1
Sex	220	SEX		24	1
Age at diagnosis	230	AGE_DX		25-27	3
Year of Birth	240	YR_BRTH		28-31	4
Sequence Number—Central	380	SEQ_NUM		35-36	2
Month of diagnosis	390	MDXRECOMP		37-38	2
Year of diagnosis	390	YEAR_DX		39-42	4
Primary Site	400	PRIMSITE		43-46	4
Laterality	410	LATERAL		47	1
Histology (92-00) ICD-O-2	420	HISTO2V		48-51	4
Behavior (92-00) ICD-O-2	430	BEHO2V		52	1
Histologic Type ICD-O-3	522	HISTO3V		53-56	4
Behavior Code ICD-O-3	523	BEHO3V		57	1
Grade	440	GRADE		58	1
Diagnostic Confirmation	490	DX_CONF		59	1
Type of Reporting Source	500	REPT_SRC		60	1
EOD—Tumor Size	780	EOD10_SZ	1988-2003	61-63	3
EOD—Extension	790	EOD10_EX	1988-2003	64-65	2
EOD—Extension Prost Path	800	EOD10_PE	1985-2003	66-67	2
EOD—Lymph Node Involv	810	EOD10_ND	1988-2003	68	1
Regional Nodes Positive	820	EOD10_PN	1988+	69-70	2
Regional Nodes Examined	830	EOD10_NE	1988+	71-72	2
EOD—Old 13 Digit	840	EOD13	1973-1982	73-85	13
EOD—Old 2 Digit	850	EOD2	1973-1982	86-87	2
EOD—Old 4 Digit	860	EOD4	1983-1987	88-91	4
Coding System for EOD	870	EOD_CODE	1973-2003	92	1
Tumor Marker 1	1150	TUMOR_1V	1990-2003	93	1
Tumor Marker 2	1160	TUMOR_2V	1990-2003	94	1
Tumor Marker 3	1170	TUMOR_3V	1998-2003	95	1
CS Tumor Size	2800	CSTUMSIZ	2004+	96-98	3
CS Extension	2810	CSEXTEN	2004+	99-101	3
CS Lymph Nodes	2830	CSLYMPHN	2004+	102-104	3
CS Mets at Dx	2850	CSMETS DX	2004+	105-106	2
CS Site-Specific Factor 1	2880	CS1SITE	2004+*	107-109	3
CS Site-Specific Factor 2	2890	CS2SITE	2004+*	110-112	3
CS Site-Specific Factor 3	2900	CS3SITE	2004+*	113-115	3
CS Site-Specific Factor 4	2910	CS4SITE	2004+*	116-118	3
CS Site-Specific Factor 5	2920	CS5SITE	2004+*	119-121	3

COMPUTER RECORD FORMAT

NAACCR Name	NAACCR Item #	SAS Variable Name	Applicable Years	Position	Length
CS Site-Specific Factor 6	2930	CS6SITE	2004+*	122-124	3
CS Site-Specific Factor 25	2879	CS25SITE	2004+*	125-127	3
Derived AJCC T	2940	DAJCC T	2004+	128-129	2
Derived AJCC N	2960	DAJCC N	2004+	130-131	2
Derived AJCC M	2980	DAJCC M	2004+	132-133	2
Derived AJCC Stage Group	3000	DAJCCSTG	2004+	134-135	2
Derived SS1977	3010	DSS1977S	2004+	136	1
Derived SS2000	3020	DSS2000S	2004+	137	1
Derived AJCC—Flag	3030	DAJCCFL	2004+	138	1
CS Version Input Original	2935	CSVFIRST	2004+	141-146	6
CS Version Derived	2936	CSVLATES	2004+	147-152	6
CS Version Input Current	2937	CSVCURRENT	2004+	153-158	6
RX Summ—Surg Prim Site	1290	SURGPRIF	1998+	159-160	2
RX Summ—Scope Reg LN Sur	1292	SURGSCOF	2003+	161	1
RX Summ—Surg Oth Reg/Dis	1294	SURGSITF	2003+	162	1
RX Summ—Reg LN Examined	1296	NUMNODES	1998-2002	163-164	2
Reason for no surgery	1340	NO_SURG		166	1
RX Summ—Surgery Type	1640	SS_SURG	1973-1997	170-171	2
RX Summ—Scope Reg 98-02	1647	SURGSCOP	1998-2002	174	1
RX Summ—Surg Oth 98-02	1648	SURGSITE	1998-2002	175	1
SEER Record Number	2190	REC_NO		176-177	2
SEER Type of Follow-up	2180	TYPE_FU		191	1
Age Recode <1 Year olds	N/A	AGE_1REC		192-193	2
Site Recode ICD-O-3/WHO 2008	N/A	SITERWHO		199-203	5
Recode ICD-O-2 to 9	N/A	ICDOTO9V		204-207	4
Recode ICD-O-2 to 10	N/A	ICDOT10V		208-211	4
ICCC site recode ICD-O-3/WHO 2008	N/A	ICCC3WHO		218-220	3
ICCC site rec extended ICD-O-3/WHO 2008	N/A	ICCC3XWHO		221-223	3
Behavior Recode for Analysis	N/A	BEHTREND		224	1
Histology Recode—Broad Groupings	N/A	HISTREC		226-227	2
Histology Recode—Brain Groupings	N/A	HISTRECB		228-229	2
CS Schema v0204+	N/A	cs0204schema		230-232	3
Race recode (White, Black, Other)	N/A	RAC_RECA		233	1
Race recode (W, B, AI, API)	N/A	RAC_RECY		234	1
Origin recode NHIA (Hispanic, Non-Hisp)	N/A	ORIGRECB		235	1
SEER historic stage A	N/A	HST_STGA		236	1
AJCC stage 3 rd edition (1988-2003)	N/A	AJCC_STG		237-238	2
SEER modified AJCC Stage 3 rd ed (1988-2003)	N/A	AJ_3SEER		239-240	2
SEER Summary Stage 1977 (1995-2000)	N/A	SSS77VZ	1995-2000	241	1
SEER Summary Stage 2000 (2001-2003)	N/A	SSSM2KPZ	2001-2003	242	1

COMPUTER RECORD FORMAT

NAACCR Name	NAACCR Item #	SAS Variable Name	Applicable Years	Position	Length
First malignant primary indicator	N/A	FIRSTPRM		245	1
State-county recode	N/A	ST_CNTY		246-250	5
Cause of Death to SEER site recode	N/A	CODPUB		255-259	5
COD to site rec KM	N/A	CODPUBKM		260-264	5
Vital Status recode	N/A	STAT_REC		265	1
IHS Link	192	IHSLINK		266	1
Summary stage 2000 (1998+)	N/A	SUMM2K	1998+	267	1
AYA site recode/WHO 2008	N/A	AYASITERWHO		268-269	2
Lymphoma subtype recode/WHO 2008	N/A	LYMSUBRWHO		270-271	2
SEER Cause-Specific Death Classification	N/A	VSRTSADX		272	1
SEER Other Cause of Death Classification	N/A	ODTHCLASS		273	1
CS Tumor Size/Ext Eval	2820	CSTSEVAL	2004+	274	1
CS Lymph Nodes Eval	2840	CSRGEVAL	2004+	275	1
CS Mets Eval	2860	CSMTEVAL	2004+	276	1
Primary by international rules	N/A	intprim		277	1
ER Status Recode Breast Cancer (1990+)	N/A	erstatus	1990+	278	1
PR Status Recode Breast Cancer (1990+)	N/A	prstatus	1990+	279	1
CS Schema -AJCC 6th ed (previously called v1)	N/A	csschema		280-281	2
CS Site-Specific Factor 8	2862	CS8SITE	2004+*	282-284	3
CS Site-Specific Factor 10	2864	CS10SITE	2004+*	285-287	3
CS Site-Specific Factor 11	2865	CS11SITE	2004+*	288-290	3
CS Site-Specific Factor 13	2867	CS13SITE	2004+*	291-293	3
CS Site-Specific Factor 15	2869	CS15SITE	2004+*	294-296	3
CS Site-Specific Factor 16	2870	CS16SITE	2004+*	297-299	3
Lymph vascular invasion	1182	VASINV	2004+*	300	1
Survival months	N/A	srv_time_mon		301-304	4
Survival months flag	N/A	srv_time_mon_flag		305	1
Insurance recode (2007+)	N/A	INSREC_PUB	2007+	311	1
Derived AJCC-7 T	3400	DAJCC7T	2010+	312-314	3
Derived AJCC-7 N	3410	DAJCC7N	2010+	315-317	3
Derived AJCC-7 M	3420	DAJCC7M	2010+	318-320	3
Derived AJCC-7 Stage Grp	3430	DAJCC7STG	2010+	321-323	3
Breast Adjusted AJCC 6 th T (1988+)	N/A	ADJTM_6VALUE	1988+	324-325	2
Breast Adjusted AJCC 6 th N (1988+)	N/A	ADJNM_6VALUE	1988+	326-327	2
Breast Adjusted AJCC 6 th M (1988+)	N/A	ADJM_6VALUE	1988+	328-329	2
Breast Adjusted AJCC 6 th Stage (1988+)	N/A	ADJAJCCSTG	1988+	330-331	2
CS Site-Specific Factor 7	2861	CS7SITE	2004+*	332-334	3
CS Site-Specific Factor 9	2863	CS9SITE	2004+*	335-337	3
CS Site-Specific Factor 12	2866	CS12SITE	2004+*	338-340	3
Derived HER2 Recode (2010+)	N/A	her2	2010+	341	1

COMPUTER RECORD FORMAT

NAACCR Name	NAACCR Item #	SAS Variable Name	Applicable Years	Position	Length
Breast Subtype (2010+)	N/A	brst_sub	2010+	342	1
Lymphomas: Ann Arbor Staging (1983+)	N/A	ANNARBOR	1983+	348	1
CS Mets at Dx-Bone	2851	CSMETSDXB_PUB	2010+	349	1
CS Mets at Dx-Brain	2852	CSMETSDXBR_PUB	2010+	350	1
CS Mets at Dx-Liver	2853	CSMETSDXLIV_PUB	2010+	351	1
CS Mets at Dx-Lung	2854	CSMETSDXLUNG_PUB	2010+	352	1
T value - based on AJCC 3rd (1988-2003)	N/A	T_VALUE	1988-2003	353	2
N value - based on AJCC 3rd (1988-2003)	N/A	N_VALUE	1988-2003	355	2
M value - based on AJCC 3rd (1988-2003)	N/A	M_VALUE	1988-2003	357	2
Total Number of In Situ/malignant Tumors for Patient	N/A	MALIGCOUNT		359-360	2
Total Number of Benign/Borderline Tumors for Patient	N/A	BENBORDCOUNT		361-362	2

*2004+ varying by schema

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

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

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

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

Other U.S. Cancer Registry Organizations

In addition to federal government agencies NCI and CDC, several professional organizations in the U.S. are involved in defining Cancer Registry data content, data formats, and standards for cancer registrar training and certification.

- North American Association of Central Cancer Registries (NAACCR)
- American College of Surgeons Commission on Cancer (CoC)
- National Cancer Registrars Association (NCRA)
- American Joint Commission on Cancer (AJCC)

Utah Cancer Registry (UCR)

- Utah's state-wide, population-based cancer registry
- Operating at the University of Utah since 1966
- Member of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program since 1973
- Participating in CDC National Program of Cancer Registries (NPCR) effective 2017



State Laws Mandating Cancer Registries

- In the U.S., each state has its own laws and regulations that govern reporting of health data for surveillance purposes.

Authorization for Cancer Registry Reporting in Utah

In Utah, the relevant law and rules include:

- Utah Code Title 26 Utah Health Code, Chapter 5 Chronic Disease Control, Section 3, System for detecting and monitoring diseases established by department
- Utah Administrative Code Rule R384-100, Cancer Reporting Rule
(<http://www.rules.utah.gov/publicat/code/r384/r384-100.htm>)

Features of State Laws Governing Cancer Registries

These laws and regulations typically address the following:

- Hospitals and other health care providers are required to report cancer
- Central cancer registries receive identifiable cancer data
- Central cancer registries are required to protect privacy and confidentiality

Utah Code Title 26-5-3
System for Detecting and Monitoring Diseases
Established by Department

"The department (*of health*) shall develop and maintain a system for detecting and monitoring chronic diseases within the state and shall investigate and determine the epidemiology of those conditions which contributed to preventable and premature sickness, or both, and to death and disability."

Key language from Utah Administrative Code Rule R384-100, Cancer Reporting Rule:

Reporting

- "All hospitals, radiation therapy centers, pathology laboratories licensed to provide services in the state, nursing homes, and other facilities and health care providers involved in the diagnosis or treatment of cancer patients shall report or provide information related to a cancer or reportable benign tumor to the Registry."

A comment on the impact of the Utah Cancer Reporting Rule on hospitals in the state:

In order to compile details of each cancer case for reporting to UCR, hospitals employ cancer registrars or contract with companies for this service. The hospital must pay the cost of this "abstracting".

Utah hospitals and health care providers are generally very cooperative with reporting. The registry actively contacts hospitals and providers to encourage timely reporting. The UCR and state have never had to use legal enforcement of the cancer reporting rule.

Key language from Utah Administrative Code Rule R384-100, Cancer Reporting Rule:

Responsibilities

- "Cancer records are managed by the Utah Cancer Registry (Registry) on behalf of the Utah Department of Health."
- "The Utah Department of Health retains ownership and all rights to the records."
- "All reports required by this rule are confidential..."

UCR Data Confidentiality and Data Release, Utah Health Code

Title 26, Utah Health Code

- 26-3-7. *Disclosure of health data -- Limitations.*

(<http://le.utah.gov/xcode/Title26/Chapter3/26-3-57.html>)

- Disclosures permitted include:
 - with consent of the individual or NOK,
 - to a government entity if consistent with purpose for which data were collected, or
 - "for a specified period, solely for bona fide research and statistical purposes".

A comment on UCR's experience with the Utah Cancer Reporting Rule:

The rule does not go into specifics such as format of data or variables. Because the language is not specific, the rule does not need to be revised when SEER, NPCR, NAACCR etc. request changes in registry variables or data formats.


Oversight of Disclosure of UCR Data

Utah Department of Health (UDOH)

- The cancer registry data are owned by UDOH
- UCR sends an annual data file to UDOH, provides data summary reports, and performs data linkages for UDOH public health purposes
- Memorandum of Agreement, Utah Department of Health and University of Utah, defines other authorized data disclosures
- UCR leadership meet monthly with UDOH

67

NCI



**DATA STANDARDS, STANDARD SETTERS
& SUPPORTING ORGANIZATIONS**

SUANN MCFADDEN, CTR
OPERATIONS MANAGER
UTAH CANCER REGISTRY

UNIVERSITY OF UTAH HEALTH, 2017


DATA STANDARDS

- Provides consistency among groups
- Evolved over time
- Determines:
 - What is collected
 - How data are processed
 - How data are used
- Primary authority to address questions on data & data management

 UNIVERSITY OF UTAH HEALTH, 2017


DATA STANDARDS

- Shared standards:
 - Ensures clarity of communication
 - Protects integrity of pooled or compared data
 - Focuses attention on key aspects of cancer care & cancer control

 UNIVERSITY OF UTAH HEALTH, 2017


DATA STANDARDS

- Computerized registry data systems have facilitated ability to share data
 - NAACCR layout format for abstracts

 UNIVERSITY OF UTAH HEALTH, 2017


STANDARD SETTERS

- Standard Codes for topography, morphology & extent of tumor spread:
 - The World Health Organization
 - ICD-O-3 site & histology codes
 - Rules for assigning codes
 - American Cancer Society
 - Early coding system; forerunner to ICD-O
 - American Joint Committee on Cancer
 - Tumor-Node-Metastasis (TNM) prognostic system
 - In cooperation with International Union against Cancer (UICC)

 UNIVERSITY OF UTAH HEALTH, 2017

STANDARD SETTERS

- Standards for facility & population-based registries
 - Commission on Cancer of the American College of Surgeons (CoC)
 - Defined role of facilities in cancer management
 - National Cancer Data Base (NCDB)
 - The National Cancer Institute's Surveillance Epidemiology & End Results Program (SEER)
 - Central registry procedures for cancer monitoring
 - The National Program of Cancer Registries (NPCR) – Centers for Disease Control & Prevention
 - Guidelines for population-based registries

 UNIVERSITY OF UTAH HEALTH, 2017

SUPPORTING ORGANIZATIONS


- The National Cancer Registrars Association (NCRA)
 - Support for cancer registrars
- The North American Association of Central Cancer Registries (NAACCR)
 - Development of standards
 - Support of standard setters
- Utah Cancer Registrars Association
 - Local organization to support central & hospital cancer registrars



REPORTABILITY REQUIREMENTS

- Based on governing agency of facility
- Overlap of required cases
 - Utah Department of Health/Utah Cancer Registry
 - Utah CoC approved facilities
 - Hospital Cancer Committees
 - 'Reportable by Agreement' cases
 - Hospitals must meet UCR requirements
 - UCR will accept hospital cases not required by SEER




HEALTH
UNIVERSITY OF UTAH


**DATA COLLECTION, TRAINING &
HOSPITAL COLLABORATION: WHAT WE
COLLECT**

SUANN MCFADDEN, CTR
OPERATIONS MANAGER

©UNIVERSITY OF UTAH HEALTH, 2017


DATA COLLECTION

- Data collection area includes all residents of Utah
- Out-of-state cases included in database if submitted by hospital registries
- Extensive case finding activities performed to ensure complete case capture
 - Early stage prostate, melanoma of skin & chronic myelocytic leukemia difficult to identify

 ©UNIVERSITY OF UTAH HEALTH, 2017


SEER REPORTABLE NEOPLASMS

- Invasive & in situ neoplasms
 - Cervix in situ not reportable beginning 1/1/1996
- Non-malignant brain and central nervous system tumors diagnosed beginning 1/1/2004
- Excludes basal & squamous cell carcinomas of skin

 ©UNIVERSITY OF UTAH HEALTH, 2017


SOURCE OF CANCER CASES

- 48 Hospitals in Utah
 - Hospitals with cancer registries submit cancer reports in "NAACCR Abstract" format
 - Field hospitals
- 7 Free-standing radiation centers
- Physician Offices & clinics
- Pathology Labs (hospital & free-standing)
- Death certificates
- Nursing Homes
- Data exchange with other state registries

 ©UNIVERSITY OF UTAH HEALTH, 2017

PATHOLOGY REPORT COLLECTION

- Electronic pathology
 - 88% of path received through e-Path
 - 36 of 48 hospitals
 - 5 free-standing labs
 - Working with other labs to onboard to AIM
 - Artificial Intelligence in Medicine (AIM)
- Non-electronic pathology
 - Received by fax, mail, secure FTP

 ©UNIVERSITY OF UTAH HEALTH, 2017

PATHOLOGY REPORT COLLECTION

- Hospital pathology labs
 - 28 hospitals with pathology labs
 - Some read path for multiple hospitals
- Free-standing pathology labs
 - 45 Utah & out-of-state labs
 - 21 submit cases regularly
 - Others sporadically
- Pathology lab completeness
 - Survey sent to dermatologists, urologists & oncologists to identify labs being used

 ©UNIVERSITY OF UTAH HEALTH, 2017

PATHOLOGY REVIEW

- Each pathology report is reviewed by a Utah Cancer Registry staff member to determine whether it is a reportable case



UNIVERSITY OF UTAH HEALTH, 2017

PATHOLOGY FOLLOW-BACK PROCEDURE

- If a pathology report is received and can not be linked to a hospital abstract, a "follow-back" letter is sent to the physician listed on path report
- Request detailed information
 - Confirm if hospitalized for cancer indicated on path report
 - Diagnosing information
 - Treatment



UNIVERSITY OF UTAH HEALTH, 2017

HOSPITAL ABSTRACT COLLECTION

- All hospital abstracts are submitted electronically
 - Submitted through our secure FTP site
 - Transferred from SEER*DMS abstracting tool
- 33 hospitals report through on-site registries or contract abstracting companies
- UCR does abstracting for 15 hospitals
- Cases expected to be submitted within six months of diagnosis



UNIVERSITY OF UTAH HEALTH, 2017

CASES RECEIVED BY RECORD TYPE

Type of Record	Average Number of Records Received per Year (2012-2015)
HL7 E-Path	44466
Casefinding (non-ePath)	4035
NAACCR Abstract	15083
Death Cert	17293



UNIVERSITY OF UTAH HEALTH, 2017

DEATH CLEARANCE PROCESS

- Yearly death files submitted from Utah State Vital Records Department
- Matched against registry database
 - Name, social security number, date of birth
- Matched cases updated
 - Date of death, cause of death
 - Place of birth, race, name
- Unmatched cases designated as possible 'death certificate only' (DCO) cases until follow-back provides additional information



UNIVERSITY OF UTAH HEALTH, 2017

FOLLOW-BACK PROCESS FOR DEATH CERTIFICATE ONLY CASES

- Hospital deaths
 - Hospitals contacted to determine if abstract should be submitted
- Nursing home deaths
 - Contact made for information
- Other locations of deaths
 - Letter sent to physician who signed death certificate
 - Letter sent to informant listed on death certificate
 - Limited information can be used per death clearance guidelines
 - Request timeline, physician who diagnosed



UNIVERSITY OF UTAH HEALTH, 2017

FOLLOW-BACK PROCESS FOR DEATH CERTIFICATES

- Reportability Determined
 - Utah residents at diagnosis
 - Reportable diagnosis
- Report Source Assigned
 - Hospital, physician office, death certificate only (DCO)
- NAACCR Death Clearance Manual used to correctly apply DCO rules
- SEER allows >0% & <1.5% DCO cases
- 2014 – 0.37%; 44 DCO cases



©UNIVERSITY OF UTAH HEALTH, 2017

2014 DEATH CLEARANCE REPORT

- Non-matched death certificates - 816
- Cases added to UCR Database - 285
 - Hospital - 8
 - Radiation & Oncology Center - 4
 - Physician Office - 261
 - Nursing Home/Hospice - 5
 - Autopsy Only - 7
- Death Certificate Only Cases - 44



©UNIVERSITY OF UTAH HEALTH, 2017

2014 DEATH CLEARANCE REPORT

- Death certificates followed but did not result in a reportable case - 487
 - Linked to an existing case - 286
 - Out of area - 117
 - Cause of death metastatic site - 30
 - not a new primary
 - Cause of death not reportable - 54
 - Non-reportable skin malignancies



©UNIVERSITY OF UTAH HEALTH, 2017

UTAH CANCER REGISTRY CASES BY REPORT SOURCE – 2014

Report Source	Number of Cases
1 – Hospital Inpatient	10,293
2 - Radiation Treatment Center or Medical Oncology Center	15
3 – Pathology Only	454
4 – Physician’s Office	1,252
5 - Nursing Home	5
6 - Autopsy Only	8
7 – Death Certificate Only	44
8 – Other Hospital Outpatient Units/Surgery Center	4
TOTAL	12,075



©UNIVERSITY OF UTAH HEALTH, 2017

TRAINING/QUALITY CONTROL PROCEDURES

- Experienced and certified (Certified Tumor Registrar, CTR) Registry Editors & Quality Improvement Specialists
- Cases evaluated for completeness, quality & correct application of coding rules
- Abstracts not meeting standards are returned to originating facilities to be corrected



©UNIVERSITY OF UTAH HEALTH, 2017


TRAINING/QUALITY CONTROL PROCEDURES

- Good working relationships with Utah hospital registrars
- Technical assistance provided
- Training provided through Utah Cancer Registrar’s Association meetings, NAACCR & NCRA webinars, SEER*Educate



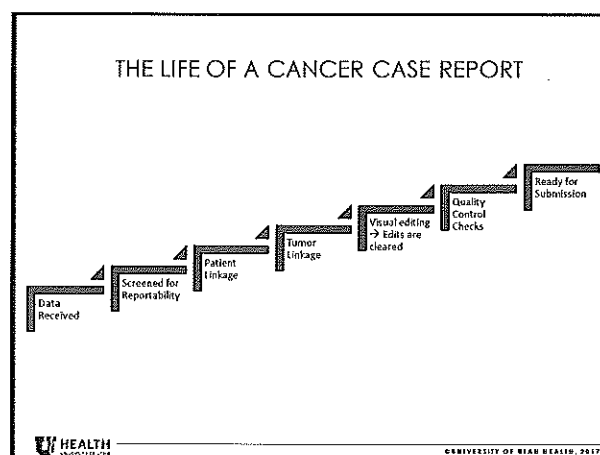
©UNIVERSITY OF UTAH HEALTH, 2017

Questions?


**CASE CONSOLIDATION
EDITING
SUBMISSION PREPARATION**


KACEY WIGREN, RHIT, CTR
 LEAD CERTIFIED TUMOR REGISTRAR
 UTAH CANCER REGISTRY

© UNIVERSITY OF UTAH HEALTH, 2017




DATA RECEIVED

- Data is received from multiple reporting sources
 - Types of data received:
 - Hospital facility abstracts
 - Pathology reports
 - Physician office reporting forms
 - Death certificates
 - Other state central cancer registry abstracts
 - Treatment and diagnosing information from freestanding cancer clinics and treatment centers
- Data is converted into system records


© UNIVERSITY OF UTAH HEALTH, 2017


SCREENING FOR REPORTABILITY

- All data records received are screened to ensure reportability
- Reportable cancer data records move through workflow to case consolidation


© UNIVERSITY OF UTAH HEALTH, 2017


CONSOLIDATION: PATIENT LEVEL LINKAGE

- Multiple reportable records for the same patient are combined/consolidated into one comprehensive patient set
 - Patient demographics are verified
- A single reportable record is consolidated into a patient set


© UNIVERSITY OF UTAH HEALTH, 2017

CONSOLIDATION: TUMOR LEVEL LINKAGE

- Data records in a single patient set are consolidated into one or more cancer cases based on:
 - Primary site
 - Morphology
 - Multiple Primary/Histology rules
- A sequence number is assigned to each reportable cancer case within a patient set


© UNIVERSITY OF UTAH HEALTH, 2017

VISUAL EDITING

- All cancer cases are visually edited:
 - Review text fields in all records
 - Verify accuracy of all coded data items
 - Assess completeness
 - Request further information from reporting sources if necessary
 - Provide necessary feedback to hospital registrars

DATA EDITS

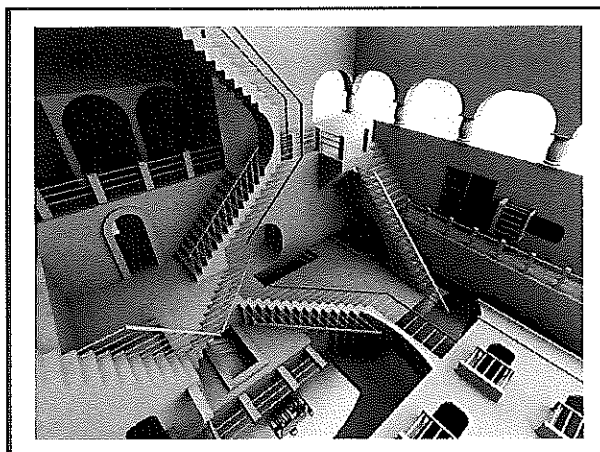
- Data edits test the logical effects of coding rules or natural relationships
 - For example: Squamous cell carcinoma is common in the lung, but is not expected to arise in the pleura. An automated edit would question a case coded as squamous cell carcinoma of the pleura.
- Through the leadership of NAACCR, edit standards have been compiled for data elements across standard data sets

DATA EDITS AT THE UTAH CANCER REGISTRY

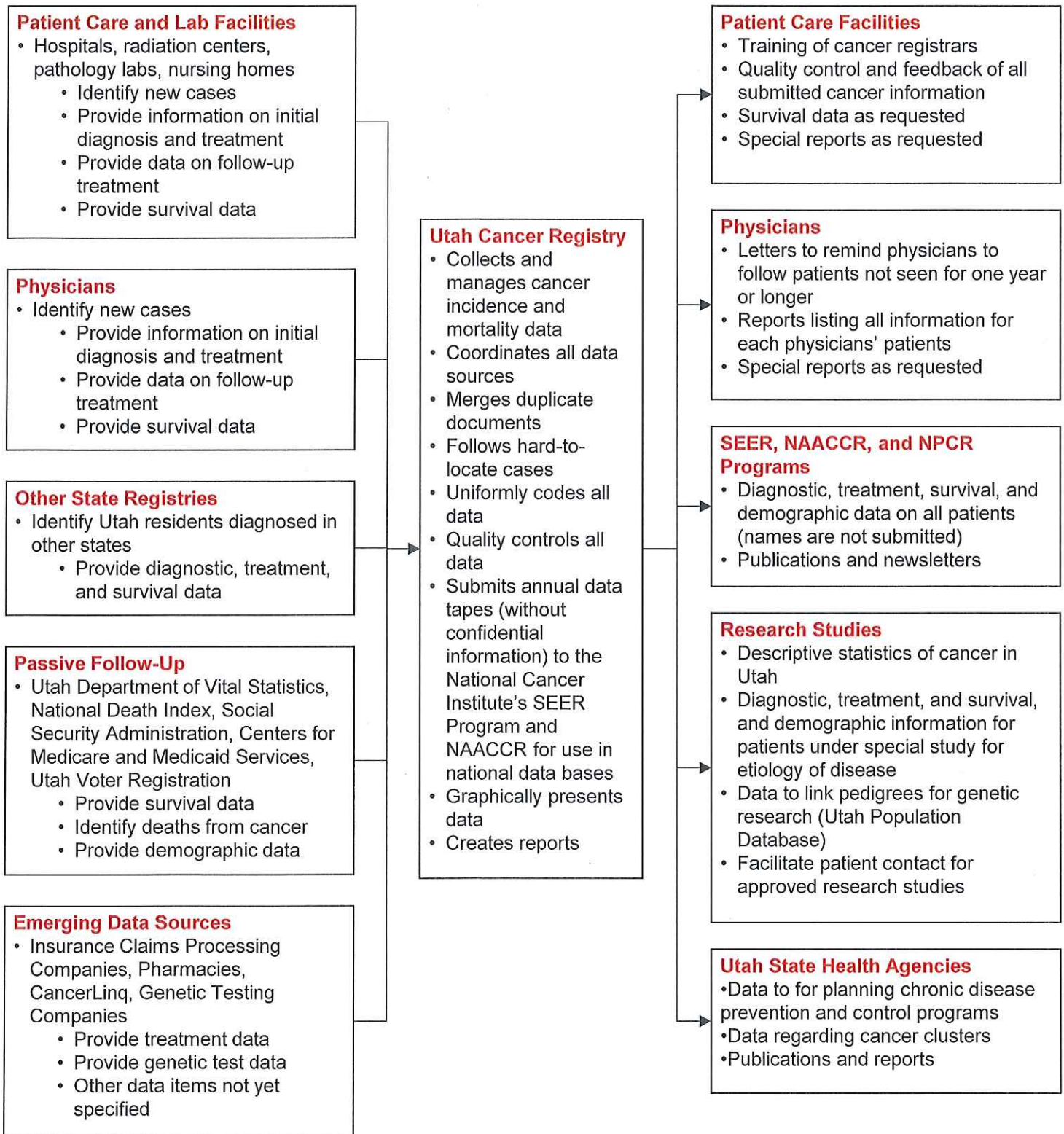
- SEER, NPCR and local edits are automated and executed in every patient set
- Failing edits are clearly listed with a detailed description
- Edits are refreshed continually throughout the visually editing process
 - Failures are often resolved while visually editing
- All edits are cleared prior to each submission


SUBMISSION PREPARATION

- Request and finalize any missing abstracts from facilities
- Complete death certificate follow-back
- Ensure all cases have been visually edited and edits are clear
- Ensure follow-up rates are above acceptable threshold
- Perform quality control reviews
 - For example: unknown laterality, unknown/ill-defined sites, run duplicate case reports



Utah Cancer Registry Data Flow





**DATA SECURITY, DATA STRUCTURE,
AND DATA TRANSMISSION**

CARRIE BATEMAN BS
SENIOR INFORMATICS SPECIALIST
UTAH CANCER REGISTRY

©UNIVERSITY OF UTAH HEALTH, 2017

Utah Cancer Registry Data Security

©UNIVERSITY OF UTAH HEALTH, 2017

PURPOSE OF DATA SECURITY

- To protect identifiable data from intentional or inadvertent unauthorized disclosure.
 - Identifiable data means any item, collection, or grouping of data that, either from the data disclosed or in combination with other data, identifies the individual described in it or the individual or organization who supplied the data.

U HEALTH UNIVERSITY OF UTAH ©UNIVERSITY OF UTAH HEALTH, 2017

APPLICABLE POLICIES, LAWS, RULES, AND AGREEMENTS

- University of Utah Policy 4-004
 - Information Security Policy
- Utah Code Title 26 Chapter 3 Section 10
 - Department Measures to Protect security of health data
- Utah Administrative Code Rule R384-100
 - Cancer Reporting Rule
- Utah Department of Health and University of Utah Memorandum of Agreement
- Federal Information Security Management Act (FISMA)
- Privacy Act of 1974

U HEALTH UNIVERSITY OF UTAH ©UNIVERSITY OF UTAH HEALTH, 2017

DATA SECURITY REQUIREMENTS

- Document data security procedures
- Restrict access to authorized individuals
- Maintain physical and electronic device security
- Limit individual access to and authorized release of identifiable data to the minimum necessary to complete the task
- Encrypt identifiable data in flight and at rest
- Provide regular data security training to staff
- Report unauthorized disclosure of identifiable data to all appropriate organizations

U HEALTH UNIVERSITY OF UTAH ©UNIVERSITY OF UTAH HEALTH, 2017

DOCUMENT DATA SECURITY PROCEDURES

- Procedures are in place that describe how staff must ensure the security of identifiable data
- Procedures are reviewed as new policies, laws, agreements, and rules are implemented to ensure compliance

U HEALTH UNIVERSITY OF UTAH ©UNIVERSITY OF UTAH HEALTH, 2017

RESTRICT ACCESS

- Restrict access to identifiable data to staff who:
 - Need to access the information to complete their assigned tasks
 - Have completed all required data security training
 - Have signed all applicable confidentiality and or non-disclosure agreements
- Review access regularly



UNIVERSITY OF UTAH HEALTH, 2017

MAINTAIN PHYSICAL AND ELECTRONIC DEVICE SECURITY

- Physical Security Standards
 - Office space is restricted to authorized personnel
 - Paper containing identifiable data
 - Stored behind two locks
 - Transported in a locked container
 - Discarded in the approved shredding bin
 - Electronic devices containing identifiable data
 - Encrypted
 - Password protected
 - Anti-virus and anti-malware software installed
 - All software set to automatically update



UNIVERSITY OF UTAH HEALTH, 2017

MINIMUM NECESSARY

- Limit individual access to and authorized release of identifiable data to the minimum necessary to complete the task
- Example
 - A researcher would like to calculate survival rates of all Utah residents diagnosed with late stage breast cancer and requests that we release full date of diagnosis and full date of last contact. To meet the minimum necessary requirement we instead would release the calculated date interval because the researcher does not require the full dates to complete the aims of the project.



UNIVERSITY OF UTAH HEALTH, 2017

ENCRYPT IDENTIFIABLE DATA IN FLIGHT AND AT REST

- All electronic identifiable data must be encrypted
- Encryption must meet federal standards
- In flight encryption is accomplished by using encrypted email, secure file transfer protocol (SFTP), or encrypting the information being sent
- At rest encryption is accomplished by encrypting the hard drive where the data are stored



UNIVERSITY OF UTAH HEALTH, 2017

PROVIDE REGULAR DATA SECURITY TRAINING TO STAFF

- Online data security training module twice a year
- Quarterly presentations about data security in staff meeting
- Monthly data security reminders by email
- Specific data security training provided in the event of an unauthorized disclosure of identifiable data



UNIVERSITY OF UTAH HEALTH, 2017

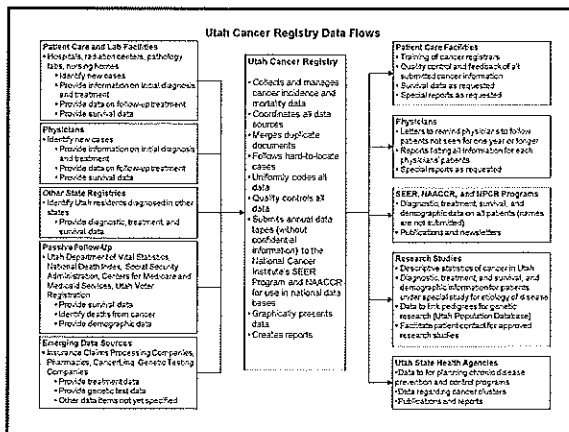
REPORT UNAUTHORIZED DISCLOSURE OF IDENTIFIABLE DATA

- Report unauthorized disclosure of identifiable data
 - University of Utah Information Security and Privacy Office
 - Utah Department of Health
 - National Cancer Institute Surveillance, Epidemiology, and End Results Program
 - Centers for Disease Control National Program of Cancer Registries
- Report Contents
 - How the unauthorized disclosure happened
 - What steps were taken to mitigate the impact of the disclosure
 - Corrective actions to be taken to prevent future disclosures



UNIVERSITY OF UTAH HEALTH, 2017

Utah Cancer Registry Data Structure and Transmission




- ## CANCER DATA SOURCES AND STRUCTURE
- Patient Care and Laboratory Facilities
 - NAACCR Abstracts
 - HL7
 - Excel or CSV lists
 - PDF files
 - Mailed or faxed paper reports
 - Physicians
 - HL7 CDA
 - PDF files
 - Mailed or faxed paper reports

- ## CANCER DATA SOURCES AND STRUCTURE - CONTINUED
- Other State Registries
 - NAACCR Abstracts
 - Excel or CSV lists
 - Passive Follow-up
 - Excel or CSV lists
 - Emerging Data Sources
 - Claims data
 - ANSI 837 Format
 - Other data sources
 - To be determined

- ## CANCER DATA TRANSMISSION METHODS
- Secure file transfer protocol (SFTP)
 - Encrypted storage devices
 - Encrypted web portal
 - Encrypted email

- ## RESOURCES
- University of Utah Policy 4-004
 - Information Security Policy
 - <http://regulations.utah.edu/f/4-004.php>
 - Utah Code Title 26 Chapter 3 Section 10
 - Department Measures to Protect security of health data
 - https://le.utah.gov/xcode/Titles/26/Chapter/3/26-3-S10.html?v=C26-3-S10_160001010001001
 - Utah Administrative Code Rule 8384-100
 - Cancer Reporting Rule
 - Federal Information Security Management Act (FISMA)
 - <https://www.dhs.gov/fisma>
 - Privacy Act of 1974
 - <https://www.archives.gov/about/laws/privacy-act-1974.html>
 - NAACCR Abstract Data Dictionary
 - <https://www.naacr.org/data-standards-data-dictionary/>
 - NAACCR HL7 Data Dictionary
 - <https://www.naacr.org/pathology-laboratory-electronic-reporting/>
 - ANSI 837 Format
 - <https://www.cms.gov/Outreach-and-Education/Medical-Learning-Network-MALN/Products/Downloads/837P-CMS-1500.pdf>


**LINKAGES & DATA EXCHANGE
PROCESS**

KIM HERGET, MSTAT
 BIOSTATISTICIAN III

UNIVERSITY OF UTAH HEALTH, 2017

DATA LINKAGE PURPOSES

- Link multiple incoming records that belong to the same patient
- Combine new records to existing patients in database
- Conduct 'passive follow up' to update date of last contact
- Assist researchers in identifying cancer patients in their study cohort

UNIVERSITY OF UTAH HEALTH, 2017

PROBABILISTIC LINKAGES

- Uses approximate matching criteria to score likelihood of match

File 1				File 2			
Last Name	First Name	Date of Birth	Social Security Number	Last Name	First Name	Date of Birth	Social Security Number
Doe	John	1/1/1960	999-45-6789	Doe	John	1/1/1960	999-45-6789
Jones	Kathy	2/5/1956	999-67-8912	Jones	Kathryn	2/5/1956	999-67-8912
Smith	William	1/2/1973	000-95-6723	Smith	William	1/2/1973	999-95-6723
Johnson	Sarah	8/5/1942	000-55-3232	Johnson-Brown	Sarah	8/5/1942	000-55-3232
Davis	Robert	9/6/1962	000-45-1232	Robert	Davis	6/9/1962	000-45-1222

UNIVERSITY OF UTAH HEALTH, 2017

RECORD CONSOLIDATION

- Link incoming records from hospitals to existing records within database
 - Pathology reports linked to NAACCR records
- Link incoming records to existing patient sets
 - Pathology report to an existing patient in our database
- De-duplicate multiple patient records within our database
 - Compare two separate patient sets to verify they are different people

UNIVERSITY OF UTAH HEALTH, 2017

PASSIVE FOLLOW UP

- Link to state and national databases to determine vital status
 - Social Security Administration
 - Centers for Medicare and Medicaid Services
 - National Death Index
 - Utah Voter Records Database
 - Utah Department of Health's Death Certificate Records
- Link to other resources for improved data
 - Indian Health Services for data on race
 - Oncotype linkage for better tumor marker data
 - All Payer's Claims Database for treatment data

UNIVERSITY OF UTAH HEALTH, 2017

RESEARCH STUDIES


- UCR will link to study cohorts for research projects
 - American Cancer Society's Cancer Prevention Study
 - Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial
 - Adventist Health Study
 - Center for Disease Control's Breast and Cervical Cancer Screening Program
 - Forfeo Patient Registry

UNIVERSITY OF UTAH HEALTH, 2017

DATA EXCHANGE PROCESS

- Utah residents seek treatment in other states, and non-Utah residents come to Utah for treatment
- Utah needs to report all resident diagnoses, despite location of diagnosis or treatment
- Utah belongs to NAACCR's Interstate Data Exchange Agreement, which allows Utah to exchange data with all other member states
- Twice annual exchange of data, with more frequent exchanges between neighboring states of Idaho, Wyoming, and Colorado

Questions?




FOLLOW-UP PROCESS FOR REPORTABLE CANCER CASES

SUANN MCFADDEN, CTR
OPERATIONS MANAGER
UTAH CANCER REGISTRY

UNIVERSITY OF UTAH HEALTH, 2017

WHY IS FOLLOW-UP IMPORTANT?


- Improve Cancer Patient Care
 - American College of Surgeons Recommendation
 - 1920's
 - Health care providers to see patients at least once a year
- Survival Analysis
 - Outcomes Research
 - Predictors of Outcomes



UNIVERSITY OF UTAH HEALTH, 2017

TYPES OF FOLLOW-UP


- **Active:** Contact with a primary source (e.g. hospital, physician, family, patient)
 - Direct Contact: phone call
 - Indirect Contact: letter, obituaries
- **Passive:** Information updated from linkages to external databases (e.g. vital statistics, voter registration, federal compensation agencies)



UNIVERSITY OF UTAH HEALTH, 2017

ACTIVE FOLLOW-UP SOURCES


- Hospitals: Readmissions, registry FU, Outpatient visits
 - Provides approximately 15% of annual FU
- Physicians: Phone calls or letters
- Patients & Next of Kin: Phone calls or letters



UNIVERSITY OF UTAH HEALTH, 2017

ACTIVE FOLLOW-UP


- Active sources used only when no other source provides needed information
- Can be labor intensive/expensive
- Physician records may be unavailable/sent to storage
- Electronic medical record information is generally available



UNIVERSITY OF UTAH HEALTH, 2017

PASSIVE FOLLOW-UP SOURCES

- Vital Statistics Information from State Department of Health
 - Death Information (e.g. death date, cause of death, place of death etc.)
- National Death Index
 - Cause of death source for patients who die outside of Utah
 - Obtained through yearly application process



UNIVERSITY OF UTAH HEALTH, 2017

PASSIVE FOLLOW-UP SOURCES

- Social Security Administration (SSA)
 - Persons contributing for future benefits
 - "Working Population"
 - Provides highest percentage of updates
 - When done first eliminates need for other sources
 - Provides approximately 69% of annual FU



© UNIVERSITY OF UTAH HEALTH, 2017

PASSIVE FOLLOW-UP SOURCES

- Centers for Medicare & Medicaid Services (CMS)
 - People over age 65 & those receiving benefits due to special circumstances
 - Provides approximately 6% of annual FU
- Election (Voter) Registration Records
 - Records date of registration & date last voted
 - Processed during major election years



© UNIVERSITY OF UTAH HEALTH, 2017

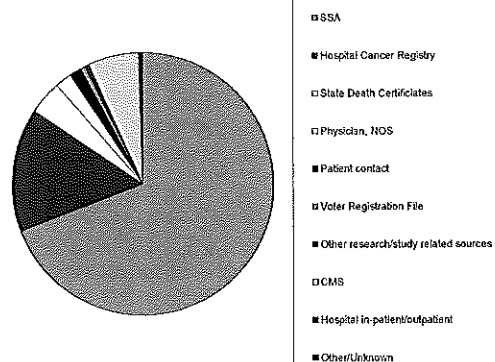
PASSIVE FOLLOW-UP

- Relatively inexpensive/limited staff time needed (some manual review needed)
- FU received for large volume of patients
- Social Security Number needed for linkage
- Accurate data needed for linkage
- May exclude some groups
 - Pediatric cases
 - Name changes



© UNIVERSITY OF UTAH HEALTH, 2017

Number of Records by Follow-Up Source: Utah



SUMMARY

- Electronic linkages provide largest percentage of follow-up
 - SSA, CMS, Voter Registration
- Follow-up from registry hospitals helps
- Other sources used as needed/available



© UNIVERSITY OF UTAH HEALTH, 2017


FOLLOW-UP STANDARDS

- American College of Surgeons (ACoS)
 - Lost to FU at 12 months from diagnosis or at last FU date
 - 90% (all ages, types)
- SEER
 - Age 0-64: 90% (not lower than 80%)
 - Age 65+: 95% (not lower than 90%)
 - In Situ Cases: 90% (not lower than 80%)



© UNIVERSITY OF UTAH HEALTH, 2017

reluctant


**RESEARCH USING UTAH CANCER
REGISTRY DATA**
 MARJORIE CARTER, MSPH
 RESEARCH MANAGER
 UTAH CANCER REGISTRY
 AUGUST 29, 2017

UNIVERSITY OF UTAH HEALTH, 2017

DISCLOSING CANCER CASE INFORMATION

- State cancer surveillance
- National cancer surveillance
- Patient care
- Approved research

UNIVERSITY OF UTAH HEALTH, 2017

DATA DISCLOSURE REQUIREMENTS

- Utah Department of Health
 - Memorandum of Agreement
 - Data security and confidentiality
 - Circumstances in which data can be disclosed
- SEER
 - Contract
 - Federal guidelines on data security, protection of human subjects
 - Requirement to facilitate cancer research
- University of Utah
 - Data security and confidentiality policies

UNIVERSITY OF UTAH HEALTH, 2017

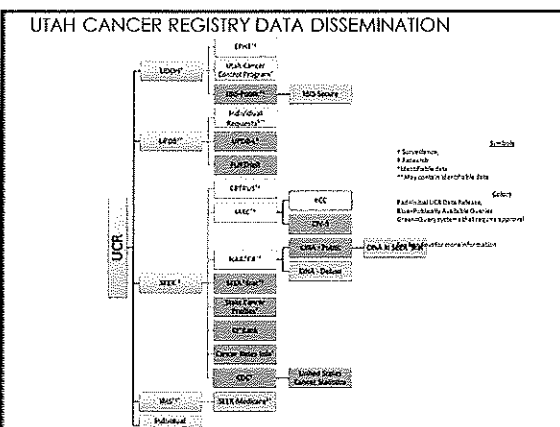
ADVISORY RESEARCH COMMITTEE

- Oversees Registry policies and procedures for data disclosures
- Reviews requests submitted to Registry for data
 - Assesses whether protocol complies with UCR policies and procedures for data security and protection of human subjects

Includes representatives from

- Utah Department of Health
- University of Utah
- Huntsman Cancer Institute
- Intermountain Healthcare
- Community-based health care providers and health care facilities

UNIVERSITY OF UTAH HEALTH, 2017



APPROVAL FOR DATA DISCLOSURES TO RESEARCHERS

- Requests to use Registry data in state & national research datasets
 - Application submitted to the agency
 - Limited variables, deidentified
- Requests submitted to Registry
 - Application
 - Consultation with Registry representatives
 - Review by Advisory Research Committee
 - Investigator signs data use agreement, confidentiality agreement
 - Data set may contain identifiers (if approved)

UNIVERSITY OF UTAH HEALTH, 2017

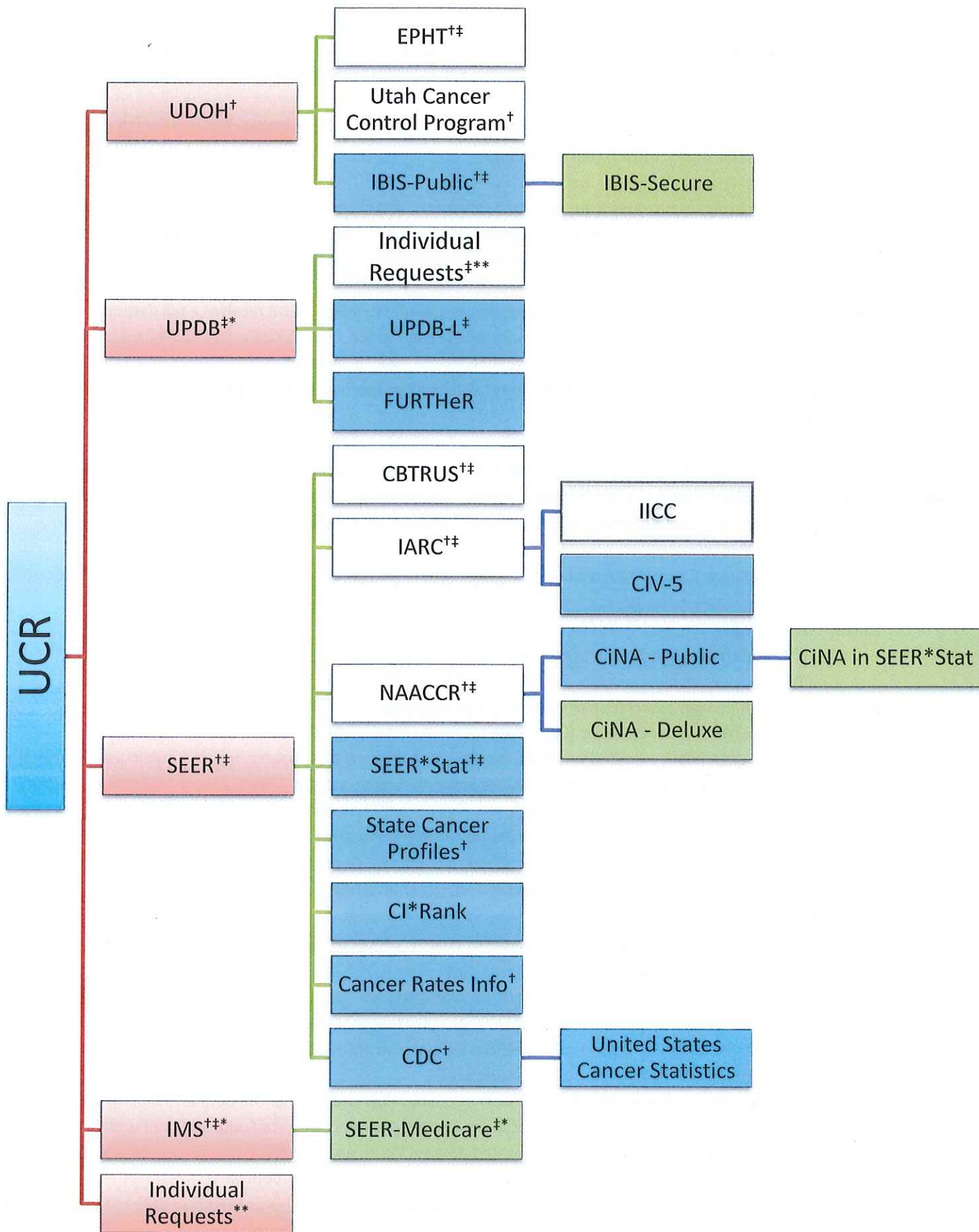
SUMMARY OF UTAH CANCER REGISTRY
SUPPORT OF RESEARCH, 2011-2016

- Projects using Registry data: 167
 - 152 external,
 - 15 internal
- Utah Cancer Registry Role
 - Special datasets
 - Medical record abstraction
 - Patient contact
 - Research resources
- Peer-reviewed publications using Registry data: 287

TABLE 1. UTAH CANCER REGISTRY SUPPORT OF CANCER
RESEARCH,
JULY 2016-JULY 2017

Utah Cancer Registry Role	Number of Projects
Participant contact	8
Abstraction of medical records	5
Pre-research data requests	8
Data linkages or queries	21
Projects with current approval to use UCR data available in Utah Population Database	75
Peer-reviewed publications resulting from projects using Registry data	35

Utah Cancer Registry Data Dissemination Chart



† Surveillance, ‡ Research, * Identifiable data, **May contain identifiable data

Red=Initial UCR Data Release, Blue=Publically Available Queries, Green=Query systems that require approval

UDOH – Utah Department of Health. Governed by a memorandum of agreement, UDOH receives a full dataset, which is used for surveillance and research. (Identifiability: full dates and addresses, but no names.)

- EPHT – Environmental Public Health Tracking. This UDOH program uses UCR data for cancer cluster investigations and to provide reports to various CDC programs, including the national EPHT program. (<http://epht.health.utah.gov/epht-view/>)
- Utah Cancer Control Program – information on cancer screenings performed in Utah are linked to UCR data for public health surveillance purposes. (<http://cancerutah.org/>)
- IBIS-PH – Indicator-Based Information System for Public Health (<http://ibis.health.utah.gov/>). This UDOH Bureau of Health Promotion program provides a query system for cancer, and other health topics, available to the public.
 - IBIS Public: de-identified, queryable by cancer site, diagnosis year, demographic variables
 - IBIS Secure: Includes zip code and census tract. Users must go through a review process similar to an IRB to gain access

UPDB – Utah Population Database. Governed by a memorandum of understanding, UPDB receives a full dataset

<https://healthcare.utah.edu/huntsmancancerinstitute/research/updb/>

(Identifiability: They receive full identifying information (Name, DOB, SSN, etc.))

- Individual requests – Data released to individual researchers after approval by IRB and RGE. May or may not contain identifiable data.
- UPDB-L – this data is public, de-identified, and queryable (<https://hci-updblapp.hci.utah.edu/updbl/>)
- FURTHeR – U of U Query Tool, same data as UPDB-L, limited to U of U patients (<http://www.further.utah.edu/>)

SEER – The Surveillance Epidemiology and End Results program. Governed by a contract, primary source of UCR funding, part of National Cancer Institute (Identifiability: They receive full dates for DOB/DOD/DODx, but no other identifiers.)

<https://seer.cancer.gov/>

- CBTRUS – Central Brain Tumor Registry of the US – They receive benign brain tumor cases (<http://www.cbtrus.org/>)
- IARC – International Agency for Research on Cancer - this data is used for surveillance and research (<https://www.iarc.fr/>)
 - IICC – International Incidence of Childhood Cancer, Publication (<http://iicc.iarc.fr/>)
 - CIV-5 – Cancer Incidence in Five Continents, Public and queryable (<http://ci5.iarc.fr/>)
- NAACCR – North American Association of Central Cancer Registries
 - Cancer in North America Public (CiNA) data sets
 - CiNA Public – limited variables, requires DUA (<https://www.naaccr.org/cina-public-use-data-set/>)
 - CiNA in SEER*Stat – needs NAACCR approval to access (<https://www.naaccr.org/cina-in-seerstat/>)
 - CiNA Deluxe –used for research, requires full application/review for access (<https://www.naaccr.org/cina-deluxe-for-researchers/>)
- SEER*Stat –data set is public and queryable after signing a confidentiality form (<https://seer.cancer.gov/seerstat/>)
- State Cancer Profiles –data set is public and queryable, very limited query options (<https://seer.cancer.gov/statistics/scp.html>)
- CI*Rank – Cancer incidence and mortality, queryable by county, cancer site, diagnosis year, demographics (<https://surveillance.cancer.gov/cirank/>)
- Cancer Rates Info –data set is public and queryable. Data through 2009 only (<http://www.cancer-rates.info/>)
- Centers for Disease Control and Prevention – cancer data collected by SEER registries and CDC’s National Program of Cancer Registries are combined to produce the official federal statistics on cancer incidence. Deidentified data sets can be downloaded for analysis. (<https://nccd.cdc.gov/uscs/>)

IMS – Information Management Systems. IMS contracts with NCI to develop and manage the Database Management System (SEER*DMS) for SEER registry data. IMS hosts Utah Cancer Registry data. UCR staff add, edit, and update cancer case information in SEER*DMS, working on the IMS-hosted data set via an encrypted virtual private network (VPN) in real time (Identifiability: UCR sends identifiers to IMS) <https://www.imsweb.com/products/>

- SEER Medicare – this data is used for research and is restricted. IMS links and sends Medicare data to SEER, who can link using a SEER Medicare ID (<https://healthcaredelivery.cancer.gov/seermedicare/>)

Individual Requests - Data released to individual researchers as requested and approved by UCR, IRB, RGE, and/or ARC committees (committee review depends on nature of request). May or may not contain identifiable data.