Spring 5-3-2016

Thyroid Cancer and Tumor Collaborative Registry (TCCR).

Oleg Shats
University of Nebraska Medical Center, oshats@unmc.edu

Whitney Goldner
University of Nebraska Medical Center, wgoldner@unmc.edu

Jianmin Feng
University of Nebraska Medical Center, jmfeng@unmc.edu

Alexander Sherman
University of Nebraska Medical Center, asherman@unmc.edu

Russell B. Smith
University of Nebraska Medical Center

See next page for additional authors

Thyroid Cancer and Tumor Collaborative Registry (TCCR)

Oleg Shats[1,2], Whitney Goldner[3], Jianmin Feng[1], Alexander Sherman[1], Russell B. Smith[3,4] and Simon Sherman[1,2]

[1]Eppley Institute for Research in Cancer, University of Nebraska Medical Center, Omaha, NE, USA. [2]Progenomix, Inc., Omaha, NE, USA. [3]College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA. [4]Nebraska Methodist Hospital, Omaha, NE, USA.

ABSTRACT: A multicenter, web-based Thyroid Cancer and Tumor Collaborative Registry (TCCR, http://tccr.unmc.edu) allows for the collection and management of various data on thyroid cancer (TC) and thyroid nodule (TN) patients. The TCCR is coupled with OpenSpecimen, an open-source biobank management system, to annotate biospecimens obtained from the TCCR subjects. The demographic, lifestyle, physical activity, dietary habits, family history, medical history, and quality of life data are provided and may be entered into the registry by subjects. Information on diagnosis, treatment, and outcome is entered by the clinical personnel. The TCCR uses advanced technical and organizational practices, such as (i) metadata-driven software architecture (design); (ii) modern standards and best practices for data sharing and interoperability (standardization); (iii) Agile methodology (project management); (iv) Software as a Service (SaaS) as a software distribution model (operation); and (v) the confederation principle as a business model (governance). This allowed us to create a secure, reliable, user-friendly, and self-sustainable system for TC and TN data collection and management that is compatible with various end-user devices and easily adaptable to a rapidly changing environment. Currently, the TCCR contains data on 2,261 subjects and data on more than 28,000 biospecimens. Data and biological samples collected by the TCCR are used in developing diagnostic, prevention, treatment, and survivorship strategies against TC.

KEYWORDS: biomedical informatics, thyroid cancer, registry, software, standardization, metadata

CITATION: Shats et al. Thyroid Cancer and Tumor Collaborative Registry (TCCR). *Cancer Informatics* 2016:15 73–79 doi: 10.4137/CIN.S32470.

TYPE: Technical Advance

RECEIVED: December 29, 2015. **RESUBMITTED:** March 08, 2016. **ACCEPTED FOR PUBLICATION:** March 20, 2016.

ACADEMIC EDITOR: J. T. Efird, Editor in Chief

PEER REVIEW: Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 652 words, excluding any confidential comments to the academic editor.

FUNDING: The development of the TCCR was partially supported by a grant from the National Cancer Institute (1R03CA175668-01A1, WG and SS are the PIs). The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: ssherm@unmc.edu

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE). Provenance: the authors were invited to submit this paper.

Published by Libertas Academica. Learn more about this journal.

Introduction

Cancer is the second leading cause of death in the USA.[1] It is estimated that 1,658,370 new cases of cancer will be diagnosed in the USA and 589,430 individuals will die from cancer in 2015.[2] The number of people living beyond a cancer diagnosis will reach 19 million in 2024.[3] To determine the risk factors of cancer development and progression and to design the novel strategies for screening, early detection, and personalized treatment of cancer, a large amount of multidimensional, standardized data have to be collected and analyzed. These data should be longitudinal, have broad geographical coverage, and be collected from patients diagnosed with different types of cancer, high-risk individuals, and normal controls.

To achieve this goal, we have been developing collaborative cancer-specific registries for the standardized collection of comprehensive demographic, lifestyle, physical activity, dietary habits, family history, and quality of life (QoL) data, as well as data on medical history, diagnosis, treatment, and outcomes. The first two registries, namely, Pancreatic Cancer Collaborative Registry (PCCR) and the Breast Cancer Collaborative Registry (BCCR), were described previously.[4,5] In this work, we describe the Thyroid Cancer and Tumor Collaborative Registry (TCCR, http://tccr.unmc.edu) that utilizes *novel* metadata-driven software architecture, as well as *modern* standards and best practices for data collection, sharing, and interoperability.

Thyroid cancer (TC) is the most common endocrine-related malignancy. According to the National Cancer Institute (NCI), the estimated number of new TC cases in the United States is expected to be 62,450 in 2015,[2] with women diagnosed three times more often than men.[6] The incidence rate of TC has been increasing sharply since the mid-1990s, and it is the fastest increasing cancer in both men and women with a growth rate of about 6% a year.[7] It is projected that by 2030 TC will become the fourth (after breast, prostate, and lung cancers) leading cancer diagnosis.[8]

There are several known risk factors for TC, including gender, race, head, and neck irradiation, iodine deficiency, autoimmune thyroid disease, and genetic risks,[9–12] whereas the roles of other potential risk factors are still being investigated.[13–21] To find the possible trends and causes of TC, researchers aim to assess and analyze the significance of suspected risk factors. Standardized collection and comprehensive analysis of data on TC and thyroid nodule (TN) patients

is required to gain a better knowledge of TC's etiology and development to improve prevention, detection, and treatment of this disease.

These challenges have prompted us to develop the TCCR, a multicenter, web-based registry, which is open to any institution willing to adopt the TCCR standard questionnaire for data collection. The TCCR's mission is to provide a collaborative framework as well as standardized data and annotated biospecimens to support searches of risk factors for TC occurrence and progression and for the development of novel strategies for screening, early detection, and personalized treatment of TC patients.

Methods

In 2008, we developed the first version of the TCCR, which utilized traditional multitier web architecture. Since then, the best practices for data collection, management, and sharing have been significantly advanced. This prompted us to reengineer the TCCR. Our goal was to improve the maintainability and sustainability of the TCCR, provide a foundation for its interoperability with other data sources, and make the TCCR more attractive for collaborative research. To achieve this goal, we have (i) improved the questionnaire and workflow by incorporating feedback from clinicians, epidemiologists, and the end-users; (ii) annotated the TCCR's common data elements; (iii) harmonized these data elements with the NCI's cancer Data Standards Registry and Repository (caDSR)[22]; and (iv) reengineered the TCCR using novel metadata-driven software architecture.

Questionnaire and common data elements. The TCCR collects various data on adult individuals aged 19 years and older, who have a personal diagnosis of TC or TNs and are able to provide informed consent. Data are collected at the time of diagnosis, 3, 6, and 12 months after diagnosis, and then annually. The *personal, demographic, lifestyle, physical activity, dietary habits, medical history, family history*, and *QoL* data are provided and may be entered into the registry by the subject. Information on *diagnosis*, *treatment, surgery,* and *outcome* are entered by the clinical personnel involved in the subject's care. The vast majority of questions have a predefined list of permissible values to streamline data entry, reduce potential errors, and simplify data mining. The system also generates a unique *subject identification code* and automatically records administrative data elements, including the *date* when the case was submitted, *registering institution code*, and *person* completing the case.

To streamline the workflow, we divided the TCCR patient questionnaire into two parts: primary/core and extended/optional. Examples of core data elements are presented in Table 1, and examples of optional data elements are presented in Table 2.

To create the foundation for interoperability with other related data sources, we have annotated all TCCR data elements (using standard vocabularies, such as SNOMED CT[23]

Table 1. Examples of core data elements.

Demographic	**Family history of cancers and major diseases**
Date of birth	Relation
Date of enrollment	Cancer type/Major diseases
Age at diagnosis	Age diagnosed
Birth country/state	Smoking status
Current City/State/Zip	Current vital status
Gender	**Clinical data**
Race/ethnicity	First symptom
Marital status	Current height and weight
Education – highest level completed	Maximum weight ever
Household income	History of cancers and major diseases
Lifestyle	Surgical and hospitalization history
Tobacco/smoking habits (current and past)	Imaging studies done
Alcohol consumption (current and past)	Staging (pathologic and clinical)
Occupation and Environment	Histologic type
Current employment status	Site(s) of involvement
Jobs held and duration at each	Treatment (type, schedule, response)
Toxic exposures	Laboratory test results
Genetic testing	Functional/medical changes after treatment
Genes tested/mutations found	Treatment outcome
	Vital status

Table 2. Examples of optional/extended data elements.

Women's Health	**Quality of Life**
Age at menarche	SF-36
Age of menopause	**Lifestyle**
Number of pregnancies and live births	Dietary habits
Medical history	Vitamins/supplements intake
Extended history of other diseases and procedures	Coffee drinking habits
Medications	Caffeinated beverages drinking habits
	Physical activity
	Sleep patterns

and the NCI Thesaurus[24]) and began their harmonization with the NCI's caDSR. We have registered new data elements (that could not be mapped) with the caDSR to make them available to the broad health community. At present, we have harmonized/registered the core data elements; the harmonization of the TCCR's optional/expended data elements is underway. Upon completion of this metadata modeling and semantic mapping process, all TCCR data elements will be (i) explicitly numerated; (ii) described by an unambiguous, nonredundant definition; and (iii) controlled by and available from the caDSR. It will allow us to establish semantic relationships between concepts that will be necessary for data aggregation and integration and will provide a foundation for the TCCR interoperability with other relevant data sources.

Metadata-driven software architecture. The reengineered TCCR can be defined as a metadata-driven system "that relies on a detailed, structured description of a problem domain to facilitate automation, extensibility, and maintainability."[25] However, due to the fact that the TCCR uses a complex relational database as a back end, we could not use any existing metadata-driven architecture for the TCCR reengineering. For instance, the REDCap, the most widely used software engine for electronic data capture, is limited to building systems with simple, flat-table data models, static user interfaces, and uncomplicated business rules.[26] This is why we enhanced the conventional metadata-driven approach for the TCCR's reengineering.

In our new approach, we expanded the metadata by including the definitions of business rules (such as the relationship between questions, data validation rules, and user rights), hierarchy of the elements used for data collection (eg, question-group-section-form-questionnaire), navigation between forms, and descriptions of look-and-feel elements (such as form layout and object style). In collaboration with Progenomix, Inc., we have developed the Metadata-Driven Online Registry Generator (mdORG) that we utilized to reengineer the TCCR.[27]

The mdORG consists of the Metadata Loading Engine (MLE), the Automated Registry Builder (ARB), and the Runtime Data Manipulation Library (RDML). The MLE is used to load the metadata (that can be maintained in a spreadsheet) into an Oracle schema. This schema is used by the ARB to automatically build a registry's relational database. The ARB consists of Java servlets, packaged into the Java library, and a set of stored procedures and functions. The ARB supports the creation of fairly complex relational databases with multiple levels of relations. This allows a registry generated by the ARB to (i) accommodate multiple types of questionnaires (ie, *initial*, *six-month* or *yearly* follow-ups, etc.); (ii) split questionnaires into several groups (*must-have*, *should-have*, and *extended/optional*) to collect core data first; (iii) repeat blocks of questions to collect longitudinal or detailed data (ie, data on multiple treatments or procedures, detailed family history for individual relatives, etc.); and (iv) collect multiple values per question (such as more than one race and multiple medications taken).

Figure 1 shows the implementation of the mdORG for the TCCR development and function. The reengineered

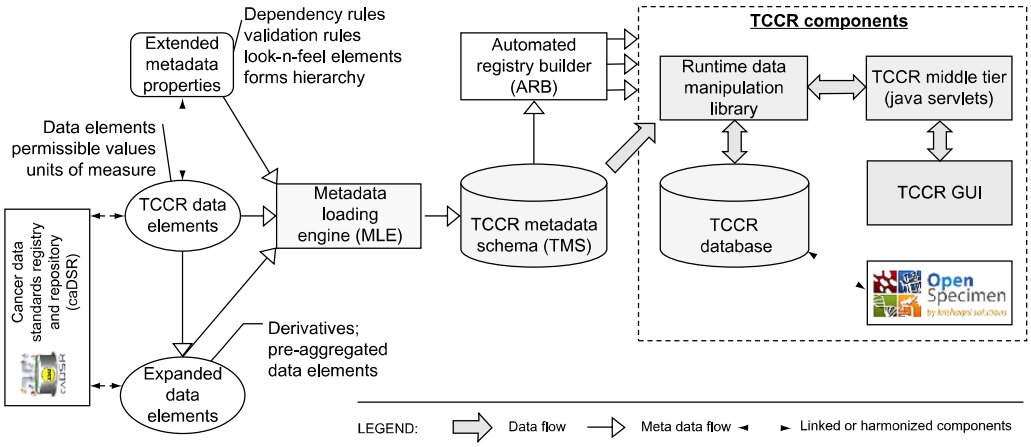

Figure 1. The TCCR system architecture diagram.

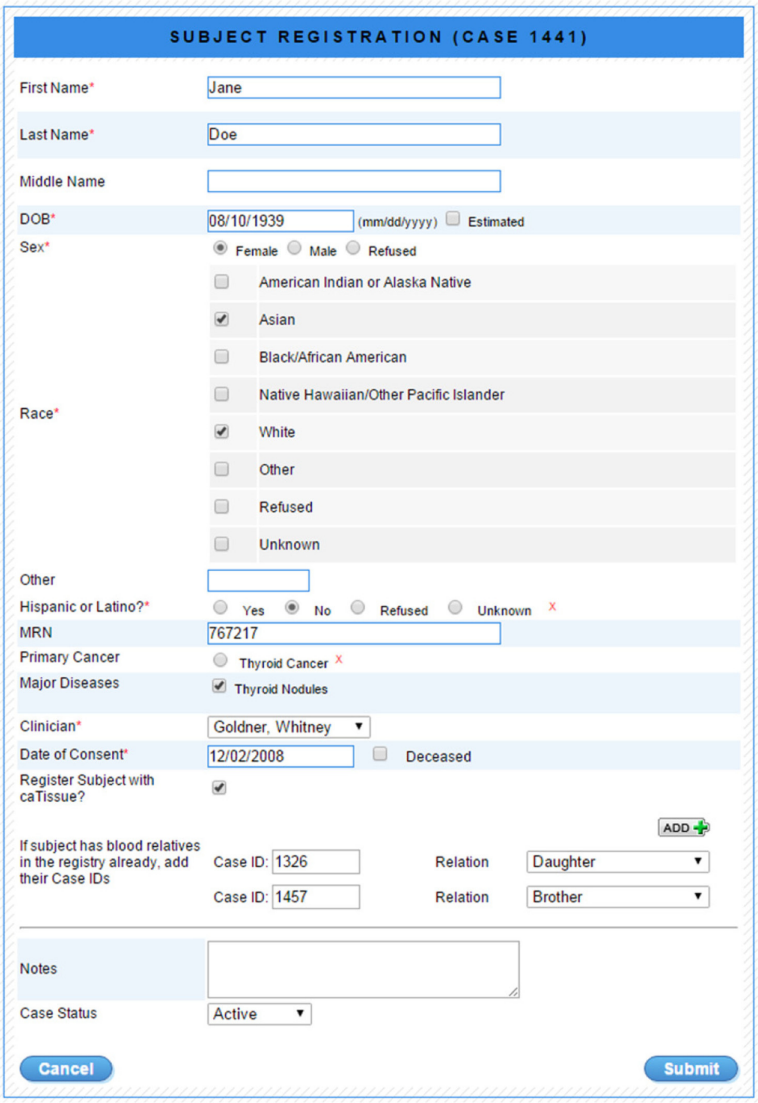

Figure 2. The TCCR's user interface: Subject Registration form.

TCCR is dynamically generated by the mdORG and consists of the backend (Oracle database), the middle tier (Java servlets), Javascript library and cascading style sheets (CSS) for runtime graphical user interface (GUI) generation, and the RDML. We also created a connector that couples the TCCR with the OpenSpecimen biospecimen management system.[28]

When an end-user logs in to the TCCR, the RDML extracts business rules and extended properties from the metadata, applies them to TCCR data, and sends preprocessed data to the TCCR middle tier. The middle tier of the TCCR consists of Java servlets that dynamically build the GUI (front-end) based on rules defined in the extended metadata and processes end-user input.

The use of the modern open-source technologies [such as HTML5 markup language, jQuery JavaScript library (https://jquery.com), AngularJS (https://angularjs.org) JavaScript framework, and CanvasJS HTML5 JavaScript charting library (http://canvasjs.com)] combined with the CSS tailored to PCs, Macs, iPads, or Android tablets, allowed us to make the TCCR GUI compatible with all major browsers and capable to automatically adjust to the end-user devices. The GUI consists of the TCCR online questionnaire, the reporting module, and the administrative module that allows user management. The online questionnaire was designed to assist in accuracy and ease of data collection by providing a predefined selection of choices whenever possible, a simple navigation between forms, and validation components that prevent users from entering erroneous information.

Figures 2 and 3 show examples of the TCCR GUI.

Operating procedures. To govern the TCCR and oversee all studies utilizing the TCCR data, the TCCR Steering Committee has been formed. The committee consists of

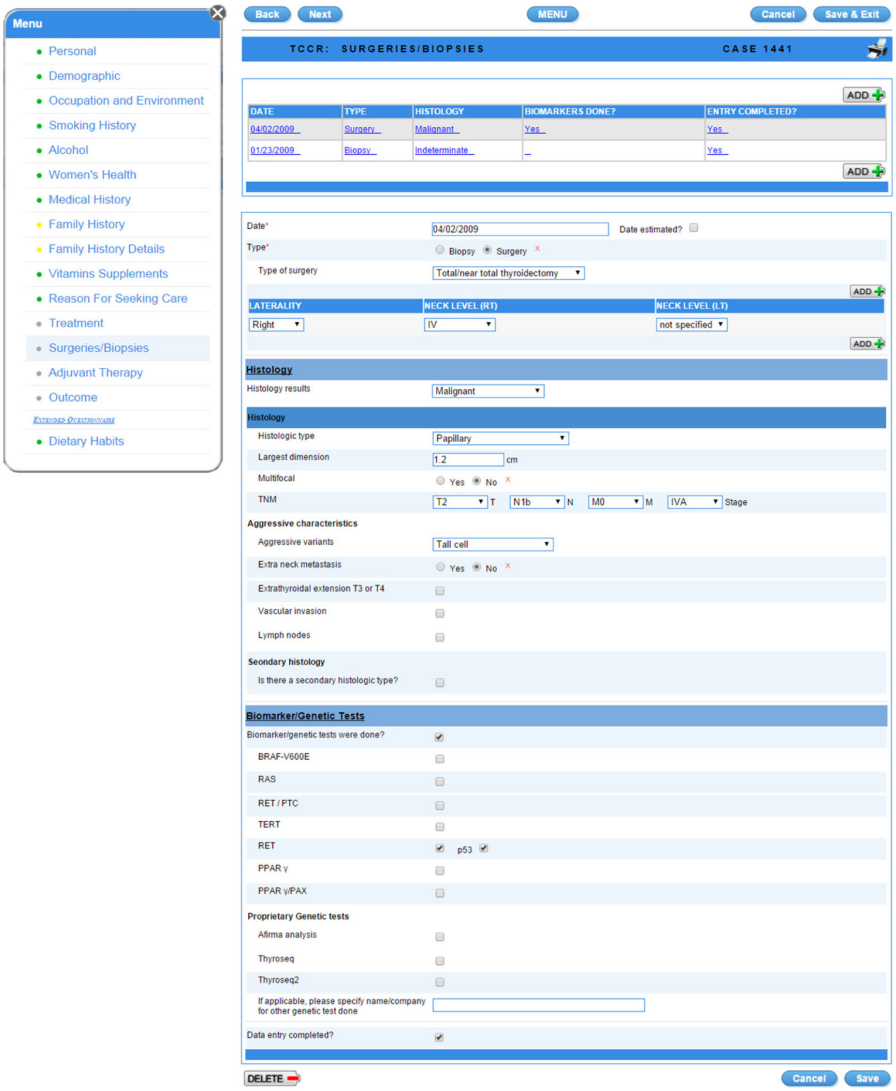

Figure 3. The TCCR's user interface: Biopsies/Surgeries form.
Note: This figure combines three screenshots to improve readability.

an appointed member from each participating institution. To guarantee an equal partnership to any institution, regardless of its size or location, we utilized the confederation model that was successfully implemented in the PCCR and BCCR.[4,5] This model provides each participating institution with an equal representation in the registry's steering committee and reassures that each center retains all rights to its own data. The data collected at any center can be used by other registry's users, but only after obtaining required permissions and upon agreement to provide appropriate references and acknowledgments.

The collaborators and representatives from the centers utilizing the TCCR have developed the standard organizational and operating procedures, standardized Institutional Review Board (IRB) applications, and common consent forms, as well as the Bylaws for institutions participating in multicenter collaborations. The TCCR Bylaws describe a framework for the TCCR's operation and management, including the responsibilities of individual centers and their designated administrators, the role of the steering committee and the TCCR coordinator, and the guidelines for collaborative studies and publications.

The security of the TCCR has been addressed using the electronic information security standards mandated by HIPAA. User authentication, access control, and data encryption issues were addressed during application development, while physical protection and network management are provided by the University of Nebraska Medical Center (UNMC) IT Services. The TCCR utilizes servers located in the state-of-the-art, secure UNMC data center. UNMC IT Services systematically conducts penetration testing of all servers. The

results of tests are analyzed, and all actions necessary to fix the security issues (if any) are taken.

The TCCR is accessible to registered users only. To ensure that users have the authority to proceed with data entry, authorized users are issued their own unique electronic signature – a combination of the user name and a password. Each user has an appropriate level of access to data (Table 1).[5] All subjects' Protected Health Information (PHI) collected in the TCCR is encrypted when entered into the database. These data can only be decrypted if a user has the corresponding level of access. The TCCR utilizes secure web server communication and supports the Secure Socket Layer (SSL; an Internet encryption method that provides two-way encryption along the entire route that data travels to and from a user's computer) and HTTPS authentication (the communications standard used to transfer pages on the Web). A 2048-bit SSL certificate has been purchased for the production application server. The proposed security procedures have been reviewed by the UNMC Information Security Team and found to be compliant with the HIPAA regulations.

To ensure that patients' PHI and the rights of data owners are protected, the data sharing strategy is implemented in the following way: (1) all data released to researchers is either de-identified or within the limited data set; (2) these data are used only if approved by the TCCR Steering Committee and IRB; (3) end-users are required to register and accept a license agreement during the registration; and (4) while de-identified data are available to all registered users, PHI within the limited data set is provided only to users who sign a data use agreement.

Before starting data collection, the participating centers are required to obtain approval from their respective IRB. All participating researchers and clinicians are required to complete the computer-based training course on the Protection of Human Research Subjects. All information gathered in the TCCR should be compliant with IRB approvals at participating sites that are monitored by each center's IRB. The TCCR coordinator opens new accounts and enables data entry into the TCCR only after receiving the documented proof of IRB protocol approval. In order to enter data, a copy of the consent form for each subject must be submitted to the TCCR coordinator.

Under the informed consent process, study participants have been asked to voluntarily participate in the TCCR. The potential participants are asked about their willingness to share the information they provided in the TCCR with research collaborators. The information the participants provide is collected for research purposes only. The subjects are informed in the consent that their PHI will be encrypted and that the web-based registry is accessible to authorized users only. Identifiers will never be released in order to protect participant confidentiality. Participants have also been informed that they may revoke the authorization to use and share their PHI at any time by contacting the principal investigator in writing. If they revoke the authorization, they may no longer participate in the research studies and the use or sharing of future PHI will be stopped, but the PHI which has already been shared may still be used.

Many subjects enrolled in the TCCR agree to donate blood and urine samples and release a portion of tissue from a previous biopsy or surgical excision. The TCCR utilizes the OpenSpecimen to track the collection, storage, and distribution of specimens that provides quality assurance for these activities.[28] The OpenSpecimen is an open-source tissue bank repository tool that is used to collect and manage the biospecimen data in a standard and efficient way. Biospecimen data collected by each participating center can be either submitted into the central repository or stored locally, if a center maintains its own installation of the OpenSpecimen. The TCCR subject ID is used to link the biospecimen data with the subject's other data, collected in the TCCR database. When a subject is registered in the TCCR, a corresponding record is automatically created in the OpenSpecimen to make it ready for biospecimen data collection.

Results

Currently, the University of Nebraska Medical Center (Omaha, NE), Methodist Health System (Omaha, NE), Avera Research Institute (Sioux Falls, SD), and Sanford Health System (Sioux Falls, SD) utilize the TCCR for data collection and management. Several regional hospitals from Nebraska, Iowa, and South Dakota are in the process of joining the TCCR.

As of March 1, 2016, there are 2,261 patients enrolled in the TCCR (1,396 subjects with TNs and 865 subjects with TC); data on more than 28,000 biospecimens (including derivatives and aliquots) for 1,266 subjects are annotated in the OpenSpecimen. Table 3 shows the detailed enrollment statistics.

Table 3. TCCR enrollment as of March 1, 2016.

BY GENDER	NUMBER OF CASES
Females	1885
Males	376
By Race	
American Indian or Alaska Native	6
Asian	12
Black or African American	88
Native Hawaiian or Other Pacific Island	3
White	1866
Multi-racial	18
Other	21
Unknown/Refused	247
By Ethnicity	
Hispanic	50
Non-Hispanic	1818
Unknown/Refused	393

To date, there were many collaborative projects utilizing data collected in the TCCR.[29–33]

The implementation of our innovative programming architecture greatly simplified support, maintenance, and future upgrades of the TCCR, and improved its interoperability with other related systems. Compared with the previous version of the TCCR, the value-added functionalities of the reengineered TCCR include the following: (i) compatibility with all major browsers; (ii) adaptability to end-user devices (including Apple iPads and Android tablets); and (iii) ability to modify the online questionnaire without recoding the underlying software. Integration with the OpenSpecimen significantly expanded the registry's capabilities, allowing researchers to manage and mine biospecimen data for subjects enrolled in the TCCR.

Recently, we integrated the reengineered TCCR with other registries, developed by the Biomedical Informatics Core Facility at the Fred and Pamela Buffett Cancer Center under the umbrella of the integrated Cancer Data Repository for Cancer research (iCaRe^2).[34] The TCCR core data elements have formed the Common Core Questionnaire shared among all iCaRe^2 registries.

Acknowledgments

The authors sincerely thank the TCCR's coordinators, Kelly Treude, Alice Kueh, and Shelby Pracht, for their help with finalizing the questionnaire and testing the registry.

Author Contributions

Conceived and designed the experiments: OS, WG, JF, AS, RBS, SS. Analyzed the data: OS, WG, SS. Wrote the first draft of the manuscript: OS, SS. Contributed to the writing of the manuscript: OS, WG, JF, AS, RBS, SS. Agree with manuscript results and conclusions: OS, WG, JF, AS, RBS, SS. Jointly developed the structure and arguments for the paper: OS, WG, SS. Made critical revisions and approved final version: OS, WG, JF, AS, RBS, SS. All authors reviewed and approved of the final manuscript.

REFERENCES

1. CDC/National Center for Health Statistics. *Leading Causes of Death*. Available at: http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm. Accessed December 21, 2015.
2. American Cancer Society. *Cancer Facts and Figures 2015*. Available at: http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2015. Accessed December 21, 2015.
3. National Cancer Institute. *Cancer Statistics*. Available at: http://www.cancer.gov/about-cancer/what-is-cancer/statistics. Accessed December 21, 2015.
4. Sherman S, Shats O, Ketcham MA, et al. PCCR: pancreatic cancer collaborative registry. *Cancer Inform*. 2011;10:83–91.
5. Sherman S, Shats O, Fleissner E, et al. Multicenter breast cancer collaborative registry. *Cancer Inform*. 2011;10:217–26.
6. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin*. 2013;63(1):11–30.
7. National Cancer Institute. *Surveillance, Epidemiology, and End Results Program (SEER) Stat Fact Sheets: Thyroid Cancer*. Available at: http://seer.cancer.gov/statfacts/html/thyro.html. Accessed December 21, 2015.
8. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res*. 2014;74(11):2913–21.
9. Meinhold CL, Ron E, Schonfeld SJ, et al. Nonradiation risk factors for thyroid cancer in the US Radiologic Technologists Study. *Am J Epidemiol*. 2010;171:242–52.
10. Dal Maso L, Bosetti C, La Vecchia C, Franceschi S. Risk factors for thyroid cancer: an epidemiological review focused on nutritional factors. *Cancer Causes Control*. 2009;20(1):75–86.
11. Horn-Ross PL, Morris JS, Lee M, et al. Iodine and thyroid cancer risk among women in a multiethnic population: the Bay Area Thyroid Cancer Study. *Cancer Epidemiol Biomarkers Prev*. 2001;10(9):979–85.
12. Balasubramaniam S, Ron E, Gridley G, Schneider AB, Brenner AV. Association between benign thyroid and endocrine disorders and subsequent risk of thyroid cancer among 4.5 million U.S. male veterans. *J Clin Endocrinol Metab*. 2012; 97(8):2661–9.
13. Mack WJ, Preston-Martin S, Dal Maso L, et al. A pooled analysis of case-control studies of thyroid cancer: cigarette smoking and consumption of alcohol, coffee, and tea. *Cancer Causes Control*. 2003;14:773–85.
14. Rossing MA, Cushing KL, Voigt LF, Wicklund KG, Daling JR. Risk of papillary thyroid cancer in women in relation to smoking and alcohol consumption. *Epidemiology*. 2000;11:49–54.
15. Engeland A, Tretli S, Akslen LA, Bjørge T. Body size and thyroid cancer in two million Norwegian men and women. *Br J Cancer*. 2006;95:366–70.
16. Chatenoud L, La Vecchia C, Franceschi S, et al. Refined-cereal intake and risk of selected cancers in Italy. *Am J Clin Nutr*. 1999;70(6):1107–10.
17. Randi G, Ferraroni M, Talamini R, et al. Glycemic index, glycemic load and thyroid cancer risk. *Ann Oncol*. 2008;19:380–3.
18. Guignard R, Truong T, Rougier Y, Baron-Dubourdieu D, Guénel P. Alcohol drinking, tobacco smoking, and anthropometric characteristics as risk factors for thyroid cancer: a countrywide case-control study in New Caledonia. *Am J Epidemiol*. 2007;166(10):1140–9.
19. Frentzel-Beyme R, Helmert U. Association between malignant tumors of the thyroid gland and exposure to environmental protective and risk factors. *Rev Environ Health*. 2000;15(3):337–58.
20. Peterson E, De P, Nuttall R. BMI, diet and female reproductive factors as risks for thyroid cancer: a systematic review. *PLoS One*. 2012;7(1):e29177.
21. Sakoda LC, Horn-Ross PL. Reproductive and menstrual history and papillary thyroid cancer risk: the San Francisco Bay Area thyroid cancer study. *Cancer Epidemiol Biomarkers Prev*. 2002;11(1):51–7.
22. Komatsoulis GA, Warzel DB, Hartel FW, et al. caCORE version 3: implementation of a model driven, service-oriented architecture for semantic interoperability. *J Biomed Inform*. 2008;41(1):106–23.
23. International Health Terminology Standards Development Organization. *SNOMED CT* (*Systematized Nomenclature of Medicine – Clinical Terms*). Available at: http://www.ihtsdo.org/snomed-ct/. Accessed December 21, 2015.
24. National Cancer Institute. *NCI Thesaurus* (*NCIt*). Available at: http://ncit.nci.nih.gov/. Accessed December 21, 2015.
25. Nadkarni PM. Metadata-Driven Software Systems in Biomedicine, Health Informatics. London: Springer-Verlag; 2011:227.
26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81.
27. PROGENOMIX. *The Metadata-Driven Online Registry Generator* (*mdORG*). Progenomix, Inc. Available at: http://progenomix.com/mdORG.html. Accessed December 21, 2015.
28. OpenSpecimen by Krishagni Solutions. *Revolutionary Biobanking Informatics*. Available at: http://www.openspecimen.org/. Accessed December 21, 2015.
29. Laney N, Meza J, Lyden E, Erickson J, Treude K, Goldner W. The prevalence of vitamin D deficiency is similar between thyroid nodule and thyroid cancer patients. *Int J Endocrinol*. 2010;2010:805716. doi: 10.1155/2010/805716.
30. Baehr KM, Lyden E, Treude K, Erickson J, Goldner W. Levothyroxine dose following thyroidectomy is affected by more than just body weight. *Laryngoscope*. 2012;122(4):834–8. doi: 10.1002/lary.23186.
31. Zahid M, Goldner W, Beseler CL, Rogan EG, Cavalieri EL. Unbalanced estrogen metabolism in thyroid cancer. *Int J Cancer*. 2013;133(11):2642–9. doi: 10.1002/ijc.28275.
32. Stansifer KJ, Guynan JF, Wachal BM, Smith RB. Modifiable risk factors and thyroid cancer. *Otolaryngol Head Neck Surg*. 2015;152(3):432–7. doi: 10.1177/0194599814564537.
33. Bennett RG, Wakeley SE, Hamel FG, High RR, Korch C, Goldner WS. Gene expression of vitamin D metabolic enzymes at baseline and in response to vitamin D treatment in thyroid cancer cell lines. *Oncology*. 2012;83(5):264–72.
34. *The Integrated Cancer Data Repository for Cancer Research* (*iCaRe^2*). Available at: http://icare2project.org. Accessed December 21, 2015.