Libertas Academica
FREEDOM TO RESEARCH

ORIGINAL RESEARCH

PCCR: Pancreatic Cancer Collaborative Registry

Simon Sherman[1,2], Oleg Shats[1,2], Marsha A. Ketcham[1], Michelle A. Anderson[3], David C. Whitcomb[4], Henry T. Lynch[5], Paola Ghiorzo[6], Wendy S. Rubinstein[7], Aaron R. Sasson[8], William E. Grizzle[9], Gleb Haynatzki[10], Jianmin Feng[1], Alexander Sherman[1], Leo Kinarsky[1,2] and Randall E. Brand[4]

[1]Eppley Institute for Research in Cancer, University of Nebraska Medical Center, Omaha, NE, USA. [2]Progenomix, Inc., Omaha, NE, USA. [3]Division of Gastroenterology, University of Michigan Medical Center, Ann Arbor, MI, USA. [4]Division of Gastroenterology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA. [5]Department of Preventive Medicine, Creighton University School of Medicine, Omaha, NE, USA. [6]Department of Oncology, Biology and Genetics, University of Genoa, Genoa, Italy. [7]Department of Medicine, North Shore University Health System, Evanston, IL and University of Chicago Pritzker School of Medicine, Chicago, IL, USA. [8]Department of Surgery, University of Nebraska Medical Center, Omaha, NE, USA. [9]Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA. [10]Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE, USA.
Corresponding author email: ssherm@unmc.edu

Abstract: The Pancreatic Cancer Collaborative Registry (PCCR) is a multi-institutional web-based system aimed to collect a variety of data on pancreatic cancer patients and high-risk subjects in a standard and efficient way. The PCCR was initiated by a group of experts in medical oncology, gastroenterology, genetics, pathology, epidemiology, nutrition, and computer science with the goal of facilitating rapid and uniform collection of critical information and biological samples to be used in developing diagnostic, prevention and treatment strategies against pancreatic cancer. The PCCR is a multi-tier web application that utilizes Java/JSP technology and has Oracle 10 g database as a back-end. The PCCR uses a *"confederation model"* that encourages participation of any interested center, irrespective of its size or location. The PCCR utilizes a standardized approach to data collection and reporting, and uses extensive validation procedures to prevent entering erroneous data. The PCCR controlled vocabulary is harmonized with the NCI Thesaurus (NCIt) or Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT). The PCCR questionnaire has accommodated standards accepted in cancer research and healthcare. Currently, seven cancer centers in the USA, as well as one center in Italy are participating in the PCCR. At present, the PCCR database contains data on more than 2,700 subjects (PC patients and individuals at high risk of getting this disease). The PCCR has been certified by the NCI Center for Biomedical Informatics and Information Technology as a cancer Biomedical Informatics Grid (caBIG®) Bronze Compatible product. The PCCR provides a foundation for collaborative PC research. It has all the necessary prerequisites for subsequent evolution of the developed infrastructure from simply gathering PC-related data into a biomedical computing platform vital for successful PC studies, care and treatment. Studies utilizing data collected in the PCCR may engender new approaches to disease prognosis, risk factor assessment, and therapeutic interventions.

Keywords: biomedical informatics, pancreatic cancer, registry, caBIG® bronze compatible system

This article is available from http://www.la-press.com.

Introduction

Pancreatic Cancer (PC) is the fourth leading cause of cancer death in both men and women.[1,2] The anatomic location of the pancreas, deep in the abdomen, has hampered early detection. Despite progress in imaging techniques, most cancers smaller than 1 cm remain undetected. Since symptoms occur late, patients with PC usually present at an advanced disease stage. About 80% of patients have metastatic disease by the time of their first clinical visit. Despite intense research in this area, the prognosis of patients with PC has remained dismal. The etiology of the disease is obscure and the epidemiological data is, in part, controversial. Some suggested factors that seem to affect the risk of PC development include age, race, gender, genetics, and lifestyle habits (such as physical activity, smoking, drinking, and diet); however, inconsistent results hamper general understanding of their role in PC induction and/or promotion.[3–11]

Some PC cases exhibit a definite familial clustering and about 10% of cases are due to the hereditary factors.[12–14] The role of hereditary factors and genetic predisposition to this disease are also poorly understood and appear to be complex. The reported data are difficult to compare and risk factors are not yet well defined. The registration and monitoring of PC-prone families are very useful resources for the identification of susceptibility genes and the development of early detection, surveillance and prevention strategies. The obviously higher incidence of PC in this high-risk population offers an avenue to prospectively collect detailed clinical, biochemical, genetic, nutritional, habitual, molecular biological and environmental information that can shed light on risk factors, prevention and therapy, as well as an elucidation of early symptoms of this disease. A comparison of the information gained from this group and from sporadic cases is of considerable clinical and epidemiological importance.

In most cases PC development appears sporadic, with most well established risk factors having only a small effect. Resolving a multifactor and complex condition requires careful ascertainment and comparison of large numbers of subjects. Since even the top centers see only a few hundred cases per year, adequately powered PC studies will require recruitment of PC patients across the country and around the world. Furthermore, it will be necessary to establish a close information partnership between multiple centers with expertise in PC epidemiology, genetics, biology, pathology, early detection and patient care. This can be accomplished only by design and implementation of an advanced biomedical computing platform thereby allowing the researchers to collect and process PC-related data in a unified and convenient way. The development of a standardized manner of data collection, together with a comprehensive plan for registering all PC cases and collecting data on individuals at high risk of developing PC, as well as on PC-free individuals for the case control studies, are fundamental needs for achieving a better knowledge of this disease. Standardization of the data and terminology models is essential in order to develop methods and databases that will allow clinicians and researchers to make computerized validation, mine clinical and biological data, and define more reliable environmental and behavioral risk factors, which may contribute to the development of this disease.

To address all of these challenges, the Pancreatic Cancer Collaborative Registry (PCCR) was initiated in 2000 as a pilot project of the NCI Specialized Programs of Research Excellence (SPORE) grant awarded to the UNMC Eppley Cancer Center. In 2001, representatives of several universities and the NCI agreed to participate in the development of the expanded PCCR infrastructure that would provide a standardized collection and sharing of clinical and demographic information on PC patients and individuals at high risk of PC development, as well as information on biospecimens collected from these subjects.

The goal of this paper is to describe the biomedical informatics aspects of the PCCR development, organizational procedures supporting the PCCR activities, and to present overall statistics of the PCCR enrollment.

Methods
PCCR architecture

The PCCR team elected to use a *"confederation model",* as opposed to traditional registry or network models that assume control of an individual center's data. "Confederation" assumes that each participating site (institution) retains all rights to the acquired data that can be used by other PCCR participants only after obtaining required permissions and by providing

corresponding references and acknowledgements. It was also recognized that for this model to be successful, it is essential to have a standardized approach to data collection and reporting. A confederation would also encourage participation of any interested center, regardless of its size or location.

The PCCR has been implemented using a three-tier, Web-based architecture: i) a client that interacts with the user; ii) an application server that contains the business logic of the application; and iii) a resource manager that stores the data. The PCCR utilizes Java Servlet/JSP technology and has Oracle 10g database as a back-end.

Standardization of data

To create a potential for integration with other related data sources, the PCCR team defined and established the criteria for standardization of collection forms and identified research questions that had to be addressed. Questionnaires well recognized in the cancer research community (for example, the NCI Quick Food Scan Percent Fat Screener questionnaire[15] and the NCCN/FACT hepatobiliary/pancreas symptom index[16]) have been implemented into the PCCR. When creating the physical activity form, the American Cancer Society's (ACS) examples of moderate versus vigorous physical activity guidelines for cancer prevention were used.[17]

Consistent with the "confederation" concept, the PCCR team has agreed upon a Core Data Set of information that all participating centers must provide and share with the PCCR community. The required (core) data elements include the most common questions used in clinical, epidemiological, nutritional, pharmaceutical, and quality of life studies. The core data set also includes the following administrative data elements: date when the case was submitted, current status of the case, registering institution code, person completing the case, and subject identification code.

Optional data elements cover a variety of questions that provide additional information for PC researchers. It is the individual participating center's decision to collect additional data and detailed information that make up the Complete Data Set.

The main areas covered by the PCCR questionnaires are presented in Table 1.

According to rules accepted by the PCCR team, a subject may provide information on personal, demographic, lifestyle, physical activity, dietary

Table 1. Core data elements.

Category	Data elements
Demographic	Race, Ethnicity, Gender, Year of birth, State/Country of residence, Education
Lifestyle	Smoking habits and Alcohol consumption
Family history	Data on relatives with history of cancer: Year of birth, Cancer type, Smoking status, Survival
Medical data	Weight, Height, BMI, Cancer type, Year of diagnosis, Stage, Major diseases
Diagnostic	Dates and types of diagnostic procedures
Treatment	Type of treatment, Agent, Schedule, Dosage, Response
Biospecimens	Types of collected specimens, Amount, Date, Pre/post-operative, Storage
Outcome	Date of last contact, Survival

habits, family history, women's health, genetics data, symptoms, quality of life, and medical history; whereas information on diagnostic studies, pathology/staging, treatment, surgeries, biospecimens, and survival can be provided only by the clinical personnel.

All terms of the PCCR controlled vocabulary have an explicit numeration and are described by an unambiguous, non-redundant definition drawn from the NCI Thesaurus or SNOMED-CT. The data elements have been defined based on the NCICB Cancer Data Standards Registry and Repository (caDSR) convention. When possible, the PCCR data elements were mapped to the caDSR.

PCCR web interface

The PCCR public website (see Fig. 1) can be accessed at http://pccrproject.org. The PCCR user interface is designed to eliminate ambiguity in the data collected and assist users with accuracy and ease of completion. The collection forms have predetermined selection choices whenever possible. As an example, Figure 2 presents a fragment of the dietary collection form. The validation components in the PCCR interface prevent the users from entering erroneous information into the system.

Integration with the caTissue

To ensure that the biospecimen data are collected in a standard and efficient way, the caTissue Suite[18] has been adopted and integrated into the framework to manage biospecimen data. The PCCR's case number

Figure 1. The PCCR website.

is used to link the biospecimen data with data collected in a registry.

PCCR security

The patients' personal information collected in the PCCR is protected using recommendations of the Healthcare Information and Management Systems Society (HIMSS) Privacy and Security Toolkit and the electronic information security standards mandated by the Federal Health Insurance Portability and Accountability Act (HIPAA). According to the HIMSS, a complete security solution that maximizes the benefits of networked data communications must contain the following elements: User Authentication, Access Control, Encryption, Physical Protection, and Management.[19] User authentication, access control, and data encryption issues have been addressed during the application development, while physical protection and network management have been provided by the UNMC Information Technology Services.

The PCCR system is accessible to registered users at participating sites 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 a user name and a password.

Each user has an appropriate level of access to data. There are six user roles, each of which has a particular type of authority (see Table 2).

All subjects' protected health information (PHI) collected in the PCCR is encrypted when entered into the database. It can only be decrypted if the user has the corresponding level of access. To decrypt the PHI, the private key is sent to the database function from the application. The private key is only sent if the

Figure 2. Example of the PCCR user interface.

user is authorized to access this data. The fact that the database and application servers are located in different network zones minimizes the risk of unauthorized access to PHI, even if the database is compromised.

The system utilizes secure web server communication and supports 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).

IRB and subject recruitment

To utilize the PCCR, each participating center is required to obtain approval from its Institutional Review Board (IRB) or Ethics Committee. Standard protocol templates and privacy assurances in procedures of informed consent were formulated detailing the use of Web-based tools. The template also includes: a) methods and procedures applied to human subjects; b) data storage and confidentiality; c) potential risk assessment for human subjects; d) risk classification; e) protection against the potential risks for human subjects; f) potential benefit assessment for human subjects; g) potential benefits to society; and h) alternatives to participation. In addition to the elements of informed consent required by 45CFR46.116(a), a model consent form includes the following HIPAA-mandated information: i) a specific description of the information to be used or disclosed; ii) the person or entity to whom disclosure will be made; iii) the purpose of the use or disclosure; iv) an expiration date or event

Table 2. PCCR user roles and their authority.

Role	Authority
Subject/patient	Can enter/update personal, demographic, lifestyle, symptoms, QOL, family and medical history data
Lab technician	Can enter/update biospecimen data only
Clinician	All of the above + Enter medical data, retrieve and edit existing cases of his/her patients
Coordinator	All of the above for the assigned clinicians
Center manager	All of the above + Retrieve and edit cases of the patients of the center/institution
System coordinator	All of the above + Retrieve and edit all cases, activate/suspend users, assign the user authorities

for use of the information; v) an explanation of how authorization may be revoked; and vi) any restrictions placed on the subject's access to the information with access granted upon completion of the research. All participating investigators are able to use these standardized statements to assist with IRB applications.

All investigators and research staff are required to complete the computer-based training course on the Protection of Human Research Subjects, approved by the Office for Human Research Protections of the US Department of Health and Services. All information gathered in the PCCR must be in accordance with IRB approval and monitored with an annual review by each center's IRB. A copy of the IRB approval letter for each site has been submitted to the PCCR Coordinator at UNMC.

In order to enter data into the PCCR, a consent form for each subject must be obtained. Under the informed consent process, study participants are asked to voluntarily participate in the PCCR. The potential participants, a PC patient or unaffected subject, are asked about their willingness to share the information they provided in the PCCR with research collaborators. The subjects are informed in the consent that: i) the information they provide is collected for research purposes only; ii) their PHI will be encrypted prior to the submission of the data into the PCCR; iii) they may revoke the authorization to use and share their PHI at any time by contacting the principal investigator in writing; and iv) if they revoke this

authorization, the use or sharing of future PHI will be stopped, and that the PHI, which has already been collected, may still be used. The instructions for completing the Web-based or the paper questionnaires are given to the participant in person while they attend the clinic or mailed to their home address. Available staff gives a verbal explanation of the forms or assistance with the Website. The subject may also contact the PCCR Coordinator for assistance.

Descriptions of human subjects' rights regarding privacy and consent issues in epidemiological research were addressed according to the HIPAA regulations 45CFR46.102(d) to ensure that patient confidentiality is appropriately protected and that research can be accomplished with minimum disclosure of the patient's PHI. For the purposes of data mining, the subject's PHI is de-identified by deleting specific identifiers; such as names, all geographic subdivisions smaller than a state or a zip code, all elements of dates (except year) for dates directly related to an individual (birth date, etc.), telephone numbers, fax numbers, electronic mail addresses, social security numbers, medical record numbers, and any other unique identifying number, characteristic, or code. Researchers with an IRB approval and Data Use Agreement (see below) can utilize and disclose PHI that contains a limited data set without HIPAA authorization or a waiver of consent granted by the IRB. The limited data set is PHI from which the 16 direct identifiers listed at 45CFR164.514(e)(2) of the HIPAA Privacy Rule have been removed.

Importantly, a multi-center Certificate of Confidentiality (COC) from the National Cancer Institute was obtained on behalf of the Secretary of the Department of Health and Human Services to prevent a court of law from being able to subpoena the PCCR records. The COC applies to all institutions participating in the registry. The certificate protects the researchers from revealing the personal identity of a participant in this registry in any federal, state, local, civil, administrative, legislative, or other proceedings. The Certificate of Confidentiality is intended to protect only the confidentiality of the medical history and genetic test result data collected as part of this research study. It does not protect information from this registry required for audit or program evaluation by the US Department of Health and Human Services, requested by the Food and Drug Administration, placed in a

health care provider's medical records, or disclosed voluntarily by the subject or someone with whom the subject has shared this information.

Data use agreements

An Advisory and Prioritization Committee (Steering Committee) has been formed to oversee the development of studies and projects that utilize information collected in the PCCR. The committee consists of an appointed member from each participating institution, as well as four external advisors, including a patient advocate, and the PCCR Coordinator. A major responsibility of this committee is to review and prioritize proposed projects that utilize the registry and its resources.

Quality control

Quality and completeness of the collected data are ensured by multi-layer control (clinician, center manager, system coordinator). Authorized users are provided with ad-hoc audit reports that highlight missing, incomplete, or potentially erroneous records. The comprehensive training materials, defining a set of procedures and lines of responsibility, were distributed to the centers. To ensure consistency and reliability of data collection across the participating centers, the PCCR system coordinator continuously provides educational training sessions and audits submitted data, whereas individual center managers review the data submitted from their respective centers.

Results and Discussions

The PCCR is implemented as a bank of information with the ability to reference the institution and investigator where additional files/samples are located. The PCCR maintains an audit trail of all data entries to protect the authenticity, integrity and confidentiality of all data entries. The collaborators and representatives from the centers utilizing the PCCR system have developed organizational and operational procedures, standardized IRB applications and common consent forms, as well as bylaws for stockholders participating in multi-center collaborations.

Until recently, the PCCR was the only cancer registry that has been certified by the NCI Center for Biomedical Informatics and Information Technology as a caBIG® Bronze Compatible product. In January 2011, a similar system developed and implemented at the UNMC, the Breast Cancer Collaborative Registry (BCCR, http://bccr.unmc.edu/), has also been certified as a caBIG® Bronze compatible product. Data collected in the PCCR are being used as a platform for several research projects, which should lead to a better understanding and treatment of PC. Participants have being recruited through the PCCR website as well as advertising at professional conferences and meetings. At present, seven cancer centers in the USA (University of Nebraska Medical Center, Creighton University, North Shore University Health System, University of Pittsburgh, University of Alabama at Birmingham, University of Michigan Medical Center, and the Memorial Sloan Kettering Cancer Center), as well as one center in Italy (University of Genoa) are participating in the PCCR. Currently, the PCCR database contains data on more than 2,700 subjects (PC patients and high-risk individuals). The statistics of the current enrollment (as of 1/5/2011) are presented in Table 3.

The PCCR's success is based on the collaborative efforts of multiple institutions with expertise in medical oncology, gastroenterology, genetics, pathology, pancreatic cancer epidemiology, nutrition, and computer science. One of the PCCR's advantages is that all clinical and epidemiological information is collected while a patient is still in clinic, which is imperative for the aggressive cancers such as PC. The PCCR offers a number of benefits to its users, including: i) standardized data elements, vocabulary and forms for data collection; ii) collecting data by effective, secure, and easy to use web-based system; iii) computerized audit and control of data quality; iv) access to the bigger data set for data analysis (separate agreement with collaborating centers is required); and v) ability to collaborate with other centers participating in the PCCR activities.

The PCCR has all the necessary prerequisites for subsequent evolvement of the developed infrastructure from simply gathering PC-related data into a biomedical computing platform vital for successful PC studies, care and treatment.

In the future developments, the caBIG® tools will be utilized to develop application programming interfaces (APIs) to virtually integrate data collected in the PCCR with data from "external" registries, such as Surveillance Epidemiology and End Results (SEER) and Behavioral Risk Factor

Table 3. Descriptive statistics of the PCCR enrollment (as of 12/23/2010).

Total enrollment report: number of subjects enrolled to date (cumulative) by ethnicity and race				
Ethnic category	Females	Males	Unknown or not reported	Total
Hispanic or Latino	22	22	0	**44**
Not Hispanic or Latino	907	798	1	**1706**
Unknown (Individuals not reporting ethnicity)	451	479	21	**951**
Ethnic category: Total of all subjects*	**1380**	**1299**	**22**	**2701**
Racial categories				
American Indian or Alaska Native	0	2	1	**3**
Asian	11	1	0	**12**
Native Hawaiian or other Pacific Islander	1	1	0	**2**
Black or African American	44	30	0	**74**
White	1172	1112	0	**2284**
More than one race	6	5	0	**11**
Unknown or not reported	146	148	21	**315**
Racial categories: Total of all subjects*	**1379**	**1299**	**22**	**2701**

Surveillance System (BRFSS). It will allow researchers to: i) assess cancer risk factors; ii) study behavioral changes among cancer survivors; iii) assess outcomes among racially and ethnically diverse communities; iv) perform comparative analyses of effectiveness of different types of cancer treatment; v) develop preclinical diagnostic and prognostic methods for early detection of cancer; and vi) develop predictive mathematical models for treatment outcomes.

Disclosures

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer Statistics, 2009. *CA Cancer J Clin*. 2009;59:225–49.
2. Jemal A, Siegel R, Xu J, Ward E. Cancer Statistics, 2010. *CA Cancer J Clin*. 2010;60:277–300.
3. Lowenfels AB, Maisonneuve P. Pancreatic cancer: development of a unifying etiologic concept. *Ann NY Acad Sci*. 1999;880:191–200.
4. Wong HH, Lemoine NR. Pancreatic cancer: molecular pathogenesis and new therapeutic targets. *Nature Reviews Gastroenterology and Hepatology*. 2009;6:412–22.
5. Stathis A, Moore MJ. Advanced pancreatic carcinoma: current treatment and future challenges. *Nature Reviews Clinical Oncology*. 2010;7:163–72.
6. Li D-H, Day RS, Bondy ML, et al. Dietary Mutagen Exposure and Risk of pancreatic Cancer. *Cancer Epidemiol Biomarkers Prev*. 2007;16:655–61.
7. Nöthlings U, Wilkens LR, Murphy SP, Hankin JH, Henderson BE, Kolonel LN. Meat and Fat Intake as Risk Factors for Pancreatic Cancer: The Multiethnic Cohort Study. *J Natl Cancer Inst*. 2005;97:1458–65.
8. Gupta S, Wang F, Holly EA, Bracci PM. Risk of pancreatic cancer by alcohol dose, duration, and pattern of consumption, including binge drinking: a population-based study. *Cancer Causes Control*. 2010;21:1047–59.
9. Yadav D, Whitcomb DC. The role of alcohol and smoking in pancreatitis. *Nat Rev Gastroenterol Hepatol*. 2010;7:131–45.
10. Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic cancer: an update. *Dig Dis*. 2010;28:645–56.
11. Anderson LN, Cotterchio M, Gallinger S. Lifestyle, dietary, and medical history factors associated with pancreatic cancer risk in Ontario, Canada. *Cancer Causes Control*. 2009;20:825–34.
12. Greer JB, Lynch HT, Brand RE. Hereditary pancreatic cancer: a clinical perspective. *Best Pract Res Clin Gastroenterol*. 2009;23:159–70.
13. Chu D, Kohlmann W, Adler DG. Identification and screening of individuals at increased risk for pancreatic cancer with emphasis on known environmental and genetic factors and hereditary syndromes. *JOP*. 2010;11:203–12.
14. Rebours V, Boutron-Ruault MC, Jooste V, et al. Mortality rate and risk factors in patients with hereditary pancreatitis: uni- and multidimensional analyses. *Am J Gastroenterol*. 2009;104:2312–7.

15. http://riskfactor.cancer.gov/diet/screeners/fat/percent_energy.pdf.
16. http://www.facit.org/LiteratureRetrieve.aspx?ID=45180.
17. Kushi LH, Byers T, Doyle C, et al. American Cancer Society Guidelines on Nutrition and Physical Activity for Cancer Prevention: Reducing the Risk of Cancer With Healthy Food Choices and Physical Activity. *CA Cancer J Clin*. 2006;56:254–81.
18. https://cabig.nci.nih.gov/tools/catissuesuite.
19. HIMSS Privacy and Security Toolkit. Managing Information Privacy and Security in Healthcare—The HIPAA Final Security Rule in Plain English. Available at: http://www.himss.org/content/files/CPRIToolkit/version6/v6%20pdf/D15_HIPAA_Final_Standard_for_Data_Security_in_Plain_English.pdf. Accessed October 8, 2010.