original report

PROSPeCT: A Predictive Research Online System for Prostate Cancer Tasks

Maria Cutumisu, PhD[1]; Catalina Vasquez, MSc[1]; Maxwell Uhlich; Perrin H. Beatty, PhD[1]; Homeira Hamayeli-Mehrabani, PhD[1]; Rume Djebah, MBBS, MHA[1]; Albert Murtha, MD, FRCPC[1]; Russell Greiner, PhD[1]; and John D. Lewis, PhD[1]

abstract

PURPOSE An online clinical information system, called Predictive Research Online System Prostate Cancer Tasks (PROSPeCT), was developed to enable users to query the Alberta Prostate Cancer Registry database hosted by the Alberta Prostate Cancer Research Initiative. To deliver high-quality patient treatment, prostate cancer clinicians and researchers require a user-friendly system that offers an easy and efficient way to obtain relevant and accurate information about patients from a robust and expanding database.

METHODS PROSPeCT was designed and implemented to make it easy for users to query the prostate cancer patient database by creating, saving, and reusing simple and complex definitions. We describe its intuitive nature by exemplifying the creation and use of a complex definition to identify a "high-risk" patient cohort.

RESULTS PROSPeCT is designed to minimize user error and to maximize efficiency without requiring the user to have programming skills. Thus, it provides tools that allow both novice and expert users to easily identify patient cohorts, manage individual patient care, perform Kaplan Meier estimates, plot aggregate PSA views, compute PSA-doubling time, and visualize results.

CONCLUSION This report provides an overview of PROSPeCT, a system that helps clinicians to identify appropriate patient treatments and researchers to develop prostate cancer hypotheses, with the overarching goal of improving the quality of life of patients with prostate cancer. We have made available the code for the PROSPeCT implementation at https://github.com/max-uhlich/e-PROSPeCT.

INTRODUCTION

Globally, prostate cancer (PCa) is the second most commonly diagnosed cancer. In the developed world, it is the most common cancer in men,[1] with similar statistics for Canada[2] and the United States.[3] Although the PCa survival rate has increased overall, the incidence of PCa is increasing, triggering increased cost of treatment[4] with significant repercussions to the global economy. PCa is being detected earlier due to the use of the prostate-specific antigen (PSA) blood test, which, unfortunately, leads to overdiagnosis and excessive use of invasive treatment methods, even in low-risk patients with cancer. The most influential driver of patient survivability is the quality of health care, including the management of the disease,[5-7] informed by the results of research on patient cohorts. For clinicians, the use of an active surveillance strategy with low-risk patients has been proposed as an alternative to invasive, unnecessary therapies.[8] Being able to accurately distinguish low-risk from high-risk patients with PCa is critical for this approach. Consequently, it is important to develop tools that help clinicians effectively treat patients and help medical researchers identify relevant patient cohorts and answer questions about such groups. There are different types of digital query-based tools that clinicians and researchers can use to access patient information, clinical standards, best health care practices, literature reviews, and to analyze patient or cohort data for diagnostic or hypothesis-building purposes. These range from general spreadsheet-based programs to relational database management systems to highly specialized algorithm and clinical information systems (CISs),[9-12] where the user either constructs the query logic manually to perform database queries or uses query-building tools.[13]

For clinicians, the tools must support the collection and management of up-to-date patient data, because patients with PCa generally require regular testing after initial diagnosis to estimate individual risk. These tools are used to inform decisions regarding the appropriate course of treatment. For medical researchers, such tools should help identify and answer relevant research questions on the basis of regularly updated

ASSOCIATED CONTENT

Appendix

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on March 1, 2019 and published at ascopubs.org/journal/cci on XXXX, 2019: DOI https://doi.org/10.1200/CCI.18.00144

CONTEXT

Key Objective We introduce and describe the Predictive Research Online System Prostate Cancer Tasks (PROSPeCT), an online clinical information system.

Knowledge Generated PROSPeCT provides a visual interface aiding researchers and clinicians in generating hypotheses and constructing prostate cancer–related queries; identifies patient cohorts with specified characteristics; aids clinicians in managing individual patients by providing possible diagnosis, treatment plans, outcomes, and identifying possible complications; and interrogates patient information from the Alberta Prostate Cancer Registry hosted by the Alberta Prostate Cancer Research Initiative. We also compare PROSPeCT with related digital tools.

Relevance The PROSPeCT system can be applied in a clinical context to facilitate the work of researchers and clinicians in managing and improving prostate cancer outcomes for patients.

data. Clinicians and researchers need a query-building platform that is intuitive to use and does not require knowledge of query languages or of database models.[10,11,13] Currently, there are many general-purpose tools that allow clinicians and researchers to query databases of patients with PCa, such as cancer information systems; these tools vary greatly in terms of ease of use and analytical power.[14] The purpose of this report is to describe one such tool called Predictive Research Online System Prostate Cancer Tasks (PROSPeCT).[15]

Here, we introduce and describe PROSPeCT, an online CIS that (1) provides a visual interface aiding researchers and clinicians in generating hypotheses and constructing PCa-related queries; (2) identifies patient cohorts with specified characteristics; (3) aids clinicians in managing treatment for a single patient by providing possible diagnosis, treatment plans, outcomes, and identifying possible complications; and (4) interrogates patient information from the Alberta Prostate Cancer Registry hosted by the Alberta Prostate Cancer Research Initiative (APCaRI).[16] And we compare PROSPeCT with related digital tools.

METHODS

PROSPeCT is a web application powered by the Apache Tomcat web server. It is implemented using Java Google Web Toolkit(http://www.gwtproject.org/) modules and it integrates the Alberta Prostate Cancer Registry PostgreSQL database[17] and APCaRI Python import modules.[18] Many of its data visualization components also use the D3.js JavaScript library, as detailed in the Data Supplement.

Rationale for Creating the User-Friendly, Query-Building Tool PROSPeCT

PROSPeCT was designed to meet the needs and preferences of the target users: clinicians and researchers. The main features driving its creation are outlined as follows:

1. Query facilitation: Relevant subsets of patients should be easy to identify from the database. This may require complex queries but needs to be feasible without the use of complex text-based programming.
2. Incorporation of predefined features: Complicated characteristics or features, such PCa risk stratification and PCa recurrence or progression, should be predefined and, therefore, easy to incorporate into the queries. These features often require defining terms that correspond to a specific time interval for the query. Thus, the system must include predefined features (eg, Interval Query) to allow the user to easily define and use time intervals (eg, from birth to onset).
3. Cohort visualization: Important results about the identified patient subpopulations should be easily visualized with general dashboard summaries that display useful population characteristics from large data sets in a scalable way. It must also be easy to prepare commonly used graphical images (eg, Kaplan-Meier curves, aggregated PSA view).
4. Individual visualization: Important results about individual patients should be easily visualized with the "Single Patient Summary" that displays useful patient characteristics, including commonly used graphical images, such as PSA plots and the PSA doubling time (PSADT) between pairs of time points.

RESULTS

Dataflow

PROSPeCT allows users to query the database of patient information. The dataflow to produce and maintain that database is a stepwise process originating with the patient and progressing through the clinic, to the Research Electronic Data Capture (REDCap; https://www.project-redcap.org/)-managed Alberta Prostate Cancer Registry, and to PROSPeCT, which allows querying by the clinicians and researchers (Fig 1).

Once the patient provides consent to be part of the study, REDCap captures (1) patient-reported information, (2) PCa diagnosis-related data, (3) disease-specific information, (4) results from biomarker analysis, and (5) inventory of biosamples collected for the Alberta Prostate Cancer Registry and Biorepository. Note that some features have multiple time-linked values. This information includes data fields,

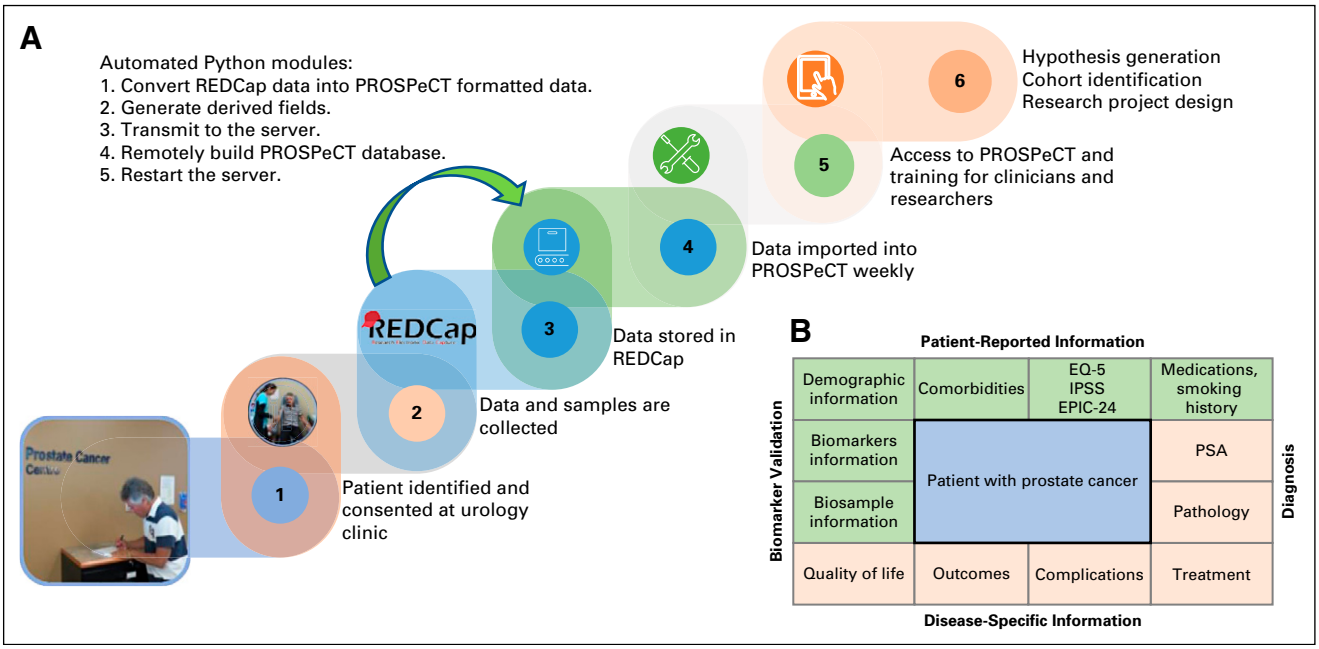

FIG 1. The data flow from Research Electronic Data Capture (REDCap) to Predictive Research Online System Prostate Cancer Tasks (PROSPeCT) and the types of data and analysis available with PROSPeCT. (A) The data file is manually exported from REDCap, then processed by automated modules that extract fields and update the database, and finally imported into the PROSPeCT web application. (B) The types of reported patient information and analysis available in PROSPeCT. EPIC-26, Expanded Prostate Cancer Index Composite; EQ-5D, EuroQol Group Standardized Instrument; IPSS, International Prostate Symptom Score; PSA, prostate-specific antigen.

such as demographic, comorbidities, use of medications, clinical, patient-reported outcomes, and biosample information. The data in REDCap are downloaded, processed, and uploaded once weekly into PROSPeCT, using five APCaRI Python modules.

Overview of the PROSPeCT Interface

Once the data are imported, users can probe it through the interface that contains five panes, shown in Figure 2: Fields (primary and PROSPeCT derived), Definition Builder, Defined Populations, Operations, and Results.

Fields. Primary fields contain unprocessed Alberta Prostate Cancer Registry patient data that are imported directly from the REDCap-managed database, which includes patient demographics; medical history; clinical results, including PSA levels, biopsy results, treatment information and outcomes, and biosample information; and quality-of-life surveys.

Incorporation of predefined features. PROSPeCT-derived fields are computed offline by the import modules during the data transfer from REDCap and include PCa recurrence or progression (defined here by biochemical recurrence or PSA failure) and risk stratification. PCa recurrence or progression in men treated with prostatectomy or radiation therapy is computed on the basis of their PSA levels over time, graphed as a PSA trajectory plot and visualized in the Single Patient Summary.

PROSPeCT also provides several browser-computed (ie, on-the-fly) data points, such as the patient's age as computed from their birth date and today's date, and the PSADT in the PSA trajectory graph from the Single Patient Summary.

Query facilitation: Definitions. One of the greatest strengths of PROSPeCT is the set of tools it provides to allow users to easily create, store, and reuse their own definitions by sequentially selecting a field (from the Fields Pane; Fig 2A) and dragging and dropping that field into the Definitions Builder Pane. For each field, the user then chooses an operator (eg, = , <, >, <= , >=, IS NULL, IS NOT NULL) and a value. Definitions are listed in the Definitions Builder Pane and they can be simple, using one field, such as "Clinical staging (T) = T1" or complex, using two or more fields, which involve arbitrary Boolean combinations.

Figure 3 depicts an example of a complex user-created definition based on four fields, designed to identify the patient cohort: "All patients whose total Gleason score (overall) is at least 8, or whose PSA level is over 20, or whose clinical staging is either T3 or T4." First, the user dragged and dropped the "Total Gleason score (overall)" field from the Fields Pane to the Definition Builder Pane and chose the mathematical operator ">=" from a drop-down menu, entering the value "8" in the adjacent field. Second, the user dragged and dropped the PSA field from the Fields Pane to the Definition Builder Pane and again chose the

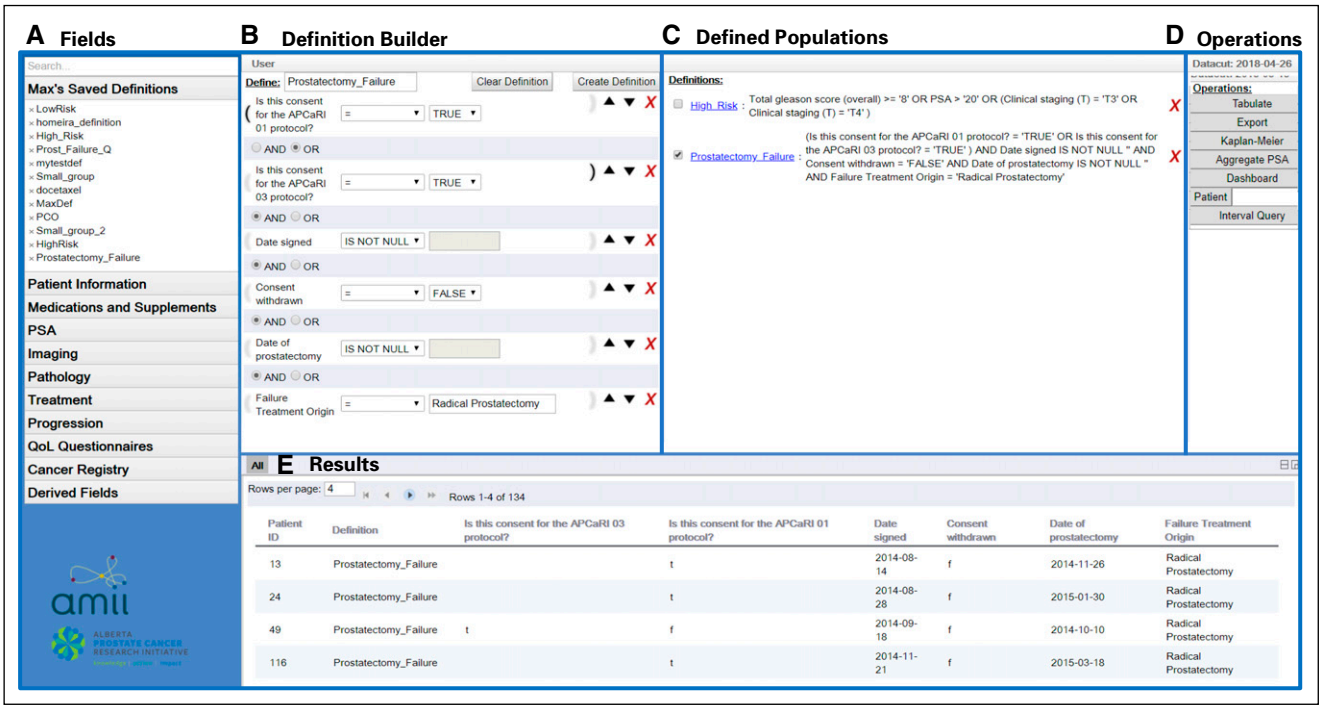

FIG 2. The Predictive Research Online System Prostate Cancer Tasks (PROSPeCT) interface used to query the database and visualize the results. Users can see the following panes on their monitors simultaneously: (A) Fields, (B) Definitions Builder, (C) Defined Populations, (D) Operations, and (E) Results.

correct operator, entering the value associated with the PSA level (ie, "> 20"). The user continued building the definition with the "Clinical staging" field, choosing the necessary operators, and inputting the required values. Because the user-created definition in Figure 3 contained four fields combined by "or" instead of "and," the user must choose that operator in the Definition Builder Pane. When the user then clicked on "Create Definition," PROSPeCT automatically identified those instances from the data set and displayed them in a table in the Results Pane. This definition is also summarized in the "Defined Populations Pane," where the user can provide a definition name, such as "High_Risk." If the definition is needed later, the user can right click on the blue label and click "Save this Definition" in the drop-down menu that appears.

Cohort visualization. Once the user has queried the database and identified an individual patient or a patient cohort, the user can visualize the results in the Results Pane, which contains tabs of types: Table, Kaplan-Meier Estimator, Aggregate PSA View, Single Patient Summary, or Cohort/Group Statistics Dashboard (Fig 4). The user can create a new tab by clicking the relevant button in the Operations Pane.

The bottom of Figure 2 shows the "Table" type (in the Results part of the display), which displays the result of performing the union of all the Defined Populations selected in the Defined Population Pane. The rows include all the instances selected by any of the definitions; the columns include all the features mentioned by any of these definitions. Notably, PROSPeCT also allows the user to easily add another column (field) to the current list of patients by selecting the new field from the Fields Pane and dragging and dropping it onto the Results (Table) pane to populate that column with the values of the associated feature, for each patient listed.

The Kaplan-Meier Estimator and the Aggregate PSA View are built-in applications in PROSPeCT that greatly aid PCa treatment decisions (Fig 4A and 4B).[19] The Aggregate PSA View produces a graph superimposing all PSA trajectories of the current defined population-patient cohort over time, centered at a user-selected date (eg, date of biopsy; Fig 4B).

Individual visualization. The Single Patient Summary allows the user to generate a page that summarizes important aspects associated with a patient (Fig 4D). This summary includes a graph of the PSA trajectory for the patient over time (Fig 4C). Figure 4D superimposes five events over that plot, and Figure 4E shows that the user can easily compute the PSADT (eg, "32 months" between January 2015 and June 2015). The Cohort/Group Statistics Dashboard application generates graphs comparing the data from fields across patient cohorts (Fig 4C).

Thus, PROSPeCT offers a convenient, easy, and fast method of querying the database to identify and examine the specified patient cohorts. It was designed to be used with very little training, owing to the drag-and-drop interface

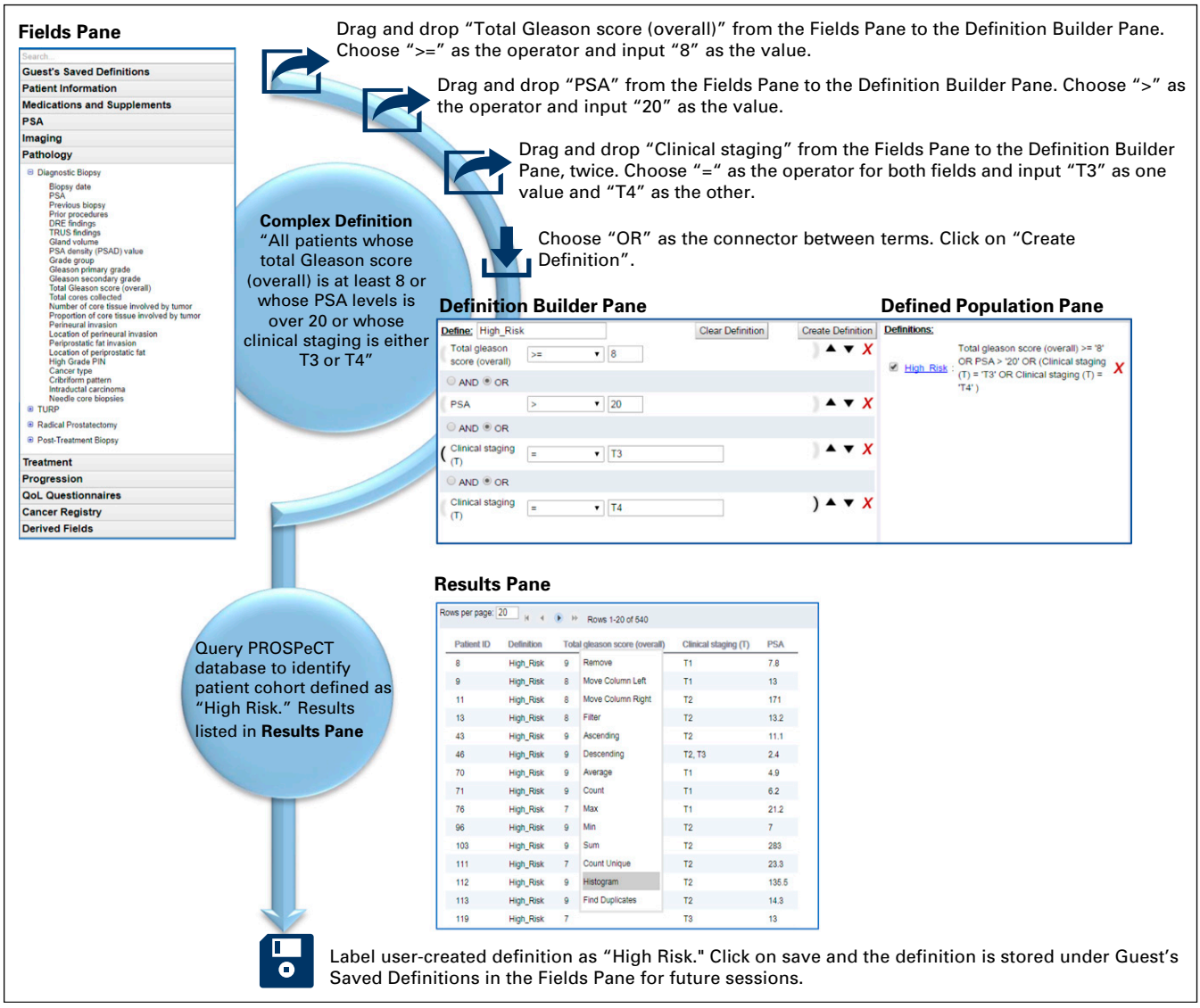

FIG 3. The use of the Predictive Research Online System Prostate Cancer Tasks (PROSPeCT) interface is illustrated by outlining the steps required to create the multifield definition called "High-Risk" into the Definition Builder Pane, using fields in the Fields Pane. The definition is then listed in the Defined Populations Pane. The query based on this definition generates the list of patients satisfying the High_Risk definition (the Defined Population) and displays their Gleason scores, PSA values, and clinical staging values in the Results Pane. The Fields Pane includes the primary and Research Electronic Data Capture (REDCap)–derived fields as well as the Derived Fields computed by PROSPeCT. The tab called "Guest's Saved Definitions" contains the user-created definitions stored by the user for future use. PSA, prostate-specific antigen.

and the preprogramming of many complex PCa-specific fields for the user.

Comparison of PROSPeCT With Other Digital Tools

There are many web-based tools that examine the generalized populations of patients with cancer (ie, general-population statistics) based on the SEER,[20] National Cancer Database,[21] or the University of California, San Francisco Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE)[22] databases. In contrast, PROSPeCT examines individuals within a specified cohort. Similar to PROSPeCT, even though other databases can be queried by many web-based tools that also offer survival reports, calculate survival by stage at diagnosis, and determine trends and incidence rates for various cancer sites over time, some databases store only PCa data (eg, CaPSURE), whereas others include data for other types of cancer as well (eg, SEER, National Cancer Database).

Although desktop-based programs such as Microsoft Excel or Microsoft Access (Redmond, CA) can help analyze patient data extracted from a small-scale database, this analysis can also take more computer time per query, be prone to user errors, restrict the input variables to predefined formats, and impose a flat data model that cannot easily capture relationships detectable by complex queries.[9,13] Automated extraction of patient cohorts from

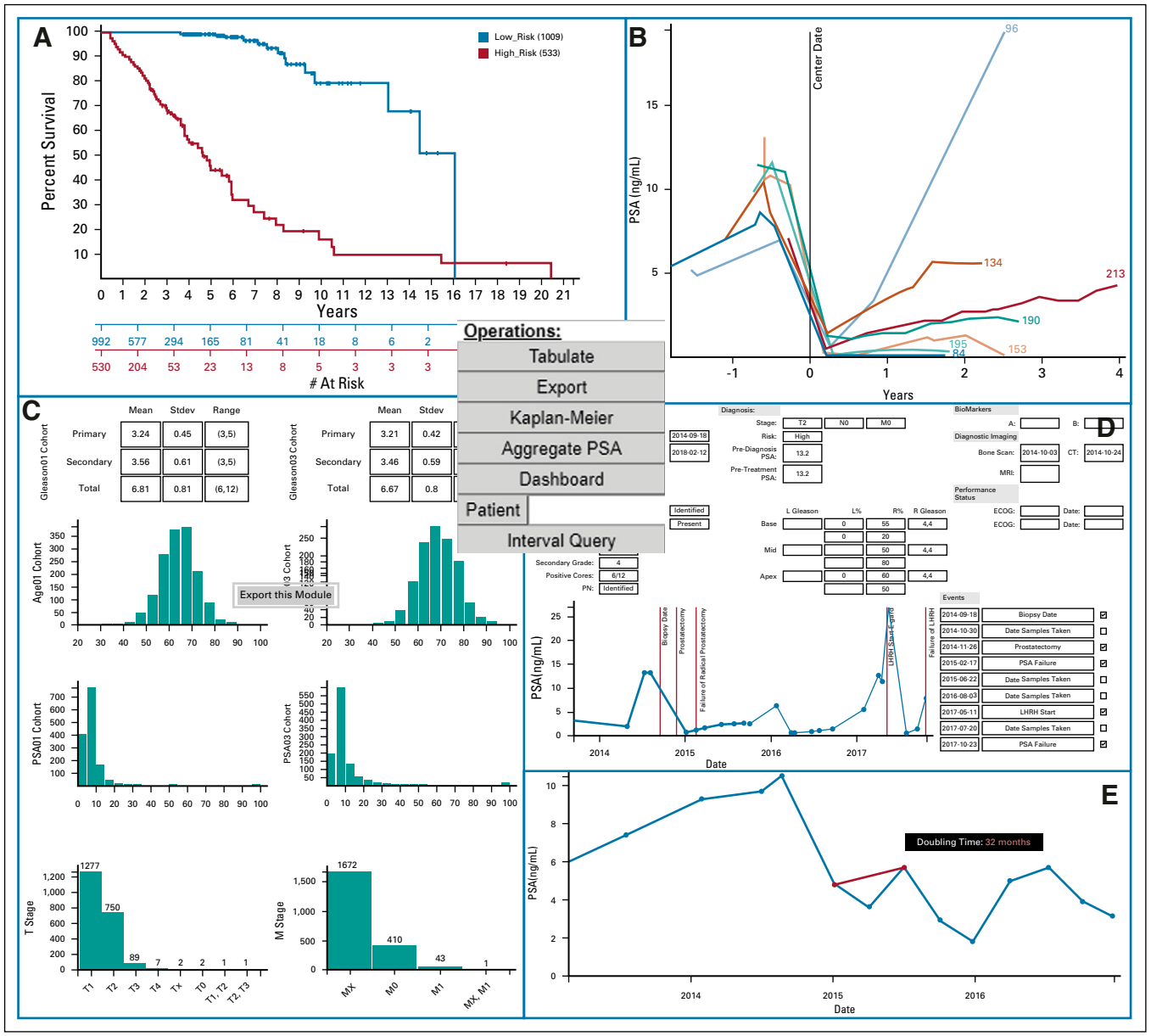

FIG 4. The Operations Pane with the built-in application options for listing, exporting, and analyzing query results from definitions and summarizing single patient data. (A) Kaplan-Meier estimator–generated Kaplan-Meier curve from a high-risk and a low-risk patient cohort. (B) PSA aggregate trajectory plot of eight patients. (C) Patient cohort summary dashboard. (D) Single patient summary with a PSA trajectory plot. (E) A single patient PSA trajectory plot showing the calculated PSA doubling time. PSA, prostate-specific antigen.

well-populated databases through user-defined queries would improve the accuracy of data retrieval, which, in turn, would greatly reduce the time spent by clinicians and researchers on data collection and analysis.

Our overarching goal is to create a general-purpose system that can help answer a range of questions. This differs from special-purpose systems that can answer only a single, predefined question. Gregg et al[23] recently developed a natural-language processing algorithm to extract PCa risk stratification data from existing electronic medical record databases. This allows clinicians to characterize PCa disease risk. Researchers hypothesized that PCa risk groups could be accurately determined from natural-language processing–extracted data in at least 90% of patients. Although extremely useful, this tool could only generate results for this specific question.

Table 1 compares PROSPeCT with several other digital query tools on the basis of ease of use, generation and visualization of results, data management, security, and cost. The tools can store a data set and perform queries on that data set, but differ in scale, ease of use, and flexibility in their programmed capabilities. Clinicians and researchers

TABLE 1. Comparison of PROSPeCT With Related Digital Tools

Feature	Spreadsheet Programs	Laboratory Information Management Systems	Expert Systems	EMR/EHR Systems	Natural Language Processing of EMR	Visual Data-Discovery Dashboards	Clinical Information System	
Example	Microsoft Excel	Clarity LIMS[24]	MYCIN[24]	GE Centricity[8]	Algorithm[23]	InSightive Analytics[25]	REDCap[26]	PROSPeCT
Description	Desktop-based GUI	Automate workflows, instruments, manage samples, data	Identify severely infectious bacteria, suggest effective therapies with a predefined model	Storage of patient data for records, billing, appointments, limited analysis	Automated PCa risk stratification data extraction method from EMRs by NLP	Custom-built, interactive dashboards; user examines data for patterns, trends	Server-based GUI for research databases	Secure web application designed to query and visualize data
Ease of use								
Data storage	Any data type	Laboratory or literature data	Data for model	Any data type	Data stored in EMR/EHR	Any data type	Any data type	Any data type
Data-handling scale	Not designed for large data files	Efficient for few features, large databases	Efficient for few features, large databases	Efficient for few features, large databases	Efficient for limited applications	Efficient for many features, large databases	Efficient for few features, large databases	Efficient for many features, large databases
Database updated	Manual	Variable	Variable	Point of care	Automated	Automated	Automated	Automated
Interface	Manual	Variable	Questions prompt user to input data	Questions prompt user to input data	Questions prompt user to input data	Wizard guided	Designed for data capture	Drag and drop, intuitive
Simple definition	Easy for user	Easy for user	Limited to built-in model	Variable, easy for user	Limited to built-in application	Easy for user	Easy for user	Easy for user
Complex definition	Difficult, code-writing needed	N/A	Limited to built-in model	N/A	Limited to built-in application	N/A	N/A	Easy, Boolean-logic capacity
Time spent: query	High: ≥ 20 minutes	N/A	Medium: 10-20 minutes	Medium: 10-20 minutes	Medium: 10-20 minutes	Low: ≤ 10 minutes	N/A	Low: ≤ 10 minutes
Generation and visualization of patient's results								
Single patient summary	Hard, write code	N/A	Variable	Easy, built-in application	Easy, built-in application	External analysis	N/A	Easy, built-in application
KME	Hard, write code	N/A	N/A	N/A	N/A	External analysis	N/A	Easy, built-in application
PSA trajectory plot	Hard, write code	N/A	N/A	N/A	N/A	External analysis	N/A	Easy, built-in application
PSA-doubling time	Hard, write code	N/A	N/A	N/A	Easy, built-in application	External analysis	N/A	Easy, built-in application
Security								

(Continued on following page)

TABLE 1. Comparison of PROSPeCT With Related Digital Tools (Continued)

Feature	Spreadsheet Programs	Laboratory Information Management Systems	Expert Systems	EMR/EHR Systems	Natural Language Processing of EMR	Visual Data-Discovery Dashboards	Clinical Information System
Permissions	File-level password protection	Variable	Variable	Multitiered access levels to specific sections	Variable	Multitiered access levels to specific sections	Multitiered access levels to specific sections

Abbreviations: EHR, electronic health record; EMR, electronic medical record; GUI, graphical user interface; KME, Kaplan-Meier estimator; LIMS, laboratory information management system; N/A, not applicable; NLP, natural language processing; PCa, prostate cancer; PSA, prostate-specific antigen; PROSPeCT, Predictive Research Online System Prostate Cancer Tasks; REDCap, Research Electronic Data Capture.

can greatly benefit from using PROSPeCT to query the Alberta Prostate Cancer Registry—or any other compatible PCa database—because it was designed with a user-friendly interface, it can generate query results quickly, and it can allow users to perform PCa-specific data calculations and visualizations via built-in applications.

We also compared the time spent by a user to query the Alberta Prostate Cancer Registry database with a complex, multifield definition using REDCap, Microsoft Excel, and PROSPeCT, asking the same user to run the same complex query with the three tools: "interrogate the whole population to extract features on specific patient cohorts with PCa disease-recurrence after receiving specific treatments." REDCap was not able to complete the query, because it is designed for retrieving records by field or table and it did not have any option to use multiple dates from different timelines and query forms. Microsoft Excel was able to perform this query, but the user had to write and implement extensive macro programming to generate the defined population list. Because PROSPeCT contains prederived PSA recurrence fields, the user could quickly build the definition used for this query and generate the defined population that fits this query, spending much less time using PROSPeCT to build this definition and produce the query report, compared with using Excel (Table 1).

DISCUSSION

The PROSPeCT interface is designed to make it easy for clinicians and researchers to interrogate the large Alberta Prostate Cancer Registry database of more than 3,600 individuals, with more than 1,500 features, and to save relevant definitions for use on updated versions of this database. There are many advantages to using PROSPeCT as an online database interface relative to other digital tools. First, it uses a drag-and-drop paradigm, so it is easy and quick for users with no programming expertise to build complex definitions to query the database. In contrast, as Microsoft Excel requires the user to write macros to run complex queries, Excel users are likely to spend much more time generating queries, in comparison with PROSPeCT users. Second, it contains several precomputed fields relevant to PCa (eg, PCa risk stratification and PCa recurrence or progression) that allow users to quickly identify relevant cohorts of patients and then important characteristics of these patients. Third, it contains several built-in applications relevant to PCa, such as the Kaplan-Meier estimator, Aggregated PSA View, PSADT, and Interval Query. Even though the PSADT is relatively simple to calculate, and clinicians and researchers find that the doubling-time information is useful for comparing treatment options and research studies, it is difficult to compute in Excel. This is due to the way data are organized in spreadsheets, because it requires users to extract multiple patient-specific data points from the database, then compute the doubling time for each chosen pair of data points. This complexity is why many clinicians do not use it

to compute PSADT.[3] However, the PSADT facility incorporated in "Single Patient Summary" makes this extrapolated result easily obtainable when using the PROSPeCT interface. Finally, it allows users to easily visualize single patient summaries, including a plot of the patient's PSA values over time with overlays showing relevant events, and to compare defined populations using the cohort/group dashboard feature.

Limitations

First, the current system is limited by the synchronization with the REDCap data management tool, which is not automated, because REDCap does not provide the facilities needed to make this process seamless. Second, although, in general, a user does not require computer programming skills to use PROSPeCT, some tasks need to be defined by a programmer (eg, PSA Failure). Also, although the PROSPeCT system does have an adjustable template for its dashboard (Appendix Fig. A1), this action currently requires a programmer to adjust. Third, although PROSPeCT was designed to maximize expressiveness while minimizing complexity, there is a boundary where certain queries are not expressible by the system. This stems from PROSPeCT trying to strike a balance between complex and expressive query languages like Standard Query Language (SQL), which require much training to use, and simpler, visual interfaces that can be used with little training. Fourth, this research is limited by the nature of the data stored in the PCa database, because the queries depend on definitions that draw on basic database fields. Finally, although informal tests of accuracy and efficiency have been conducted, this report focuses on descriptive and not empirical research.

Future Work

The current PROSPeCT system gets patient values on a scheduled basis. The near-term goal for PROSPeCT is to have a direct, real-time link to electronic medical records, meaning it could be used to generate high-value data analysis to aid treatment decision-making for a current patient. The long-term goal involves connecting PROSPeCT with databases for other types of cancer or diseases, to allow cross-referencing of patient data that could dramatically increase the effectiveness of PROSPeCT as both a diagnostic and research tool for PCa and other diseases. Overall, PROSPeCT enables users to easily and quickly create elaborate, error-free queries, especially those related to PCa.

AFFILIATION

[1]University of Alberta, Edmonton, Alberta, Canada

CORRESPONDING AUTHOR

John D. Lewis, PhD, Department of Experimental Oncology, University of Alberta, Edmonton, AB T6G 2E1 Canada; e-mail: jdlewis@ualberta.ca

SUPPORT

Supported by Alberta Cancer Foundation Program Grant No. 26491. J.D.L. holds the Frank and Carla Sojonky Chair in Prostate Cancer Research funded by the Alberta Cancer Foundation.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST AND DATA AVAILABILITY STATEMENT

Disclosures provided by the authors and data availability statement (if applicable) are available with this article at DOI https://doi.org/10.1200/CCI.18.00144.

AUTHOR CONTRIBUTIONS

Conception and design: Maria Cutumisu, Catalina Vasquez, Rume Djebah, Albert Murtha, Russell Greiner, John D. Lewis

Financial support: John D. Lewis

Administrative support: Catalina Vasquez, Rume Djebah, Russell Greiner, John D. Lewis

Provision of study material or patients: Catalina Vasquez, Rume Djebah, John D. Lewis

Collection and assembly of data: Maria Cutumisu, Catalina Vasquez, Maxwell Uhlich, Homeira Hamayeli-Mehrabani, Albert Murtha, John D. Lewis

Data analysis and interpretation: Maria Cutumisu, Catalina Vasquez, Perrin H. Beatty, Albert Murtha, Russell Greiner, John D. Lewis

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Catalina Vasquez

Employment: Nanostics

Stock and Other Ownership Interests: Nanostics

Travel, Accommodations, Expenses: Nanostics

Perrin H. Beatty

Consulting or Advisory Role: Nanostics

John D. Lewis

Employment: Nanostics

Leadership: Nanostics

Stock and Other Ownership Interests: Nanostics, Entos Pharmaceuticals

Patents, Royalties, Other Intellectual Property: Authorship on patents and patent applications in the fields of diagnostics, drug delivery, and oncology therapeutics.

No other potential conflicts of interest were reported.

Acknowledgment

We thank Nawaid Usmani, MD, for expert guidance with clinical features.

REFERENCES

1. Stewart BW, Wild CP, eds: World Cancer Report 2014. Geneva, Switzerland, World Health Organization, 2015
2. Canadian Cancer Statistics Advisory Committee. Canadian Cancer Statistics. A 2018 Special Report on Cancer Incidence by Stage. Toronto, ON, Canada, Canadian Cancer Society, 2018
3. Shariat SF, Karakiewicz PI, Roehrborn CG, et al: An updated catalog of prostate cancer predictive tools. Cancer 113:3075-3099, 2008
4. Evans SM, Millar JL, Wood JM, et al: The Prostate Cancer Registry: monitoring patterns and quality of care for men diagnosed with prostate cancer. BJU Int 111:E158-E166, 2013 (4 Pt B)
5. Burnett AL: Racial disparities in sexual dysfunction outcomes after prostate cancer treatment: Myth or reality? J Racial Ethn Health Disparities 3:154-159, 2016
6. Jayadevappa R, Chhatre S, Johnson JC, et al: Variation in quality of care among older men with localized prostate cancer. Cancer 117:2520-2529, 2011
7. Schroeck FR, Kaufman SR, Jacobs BL, et al: Regional variation in quality of prostate cancer care. J Urol 191:957-962, 2014
8. Knighton AJ, Belnap T, Brunisholz K, et al: Using electronic health record data to identify prostate cancer patients that may qualify for active surveillance. EGEMS (Wash DC) 4:1220, 2016
9. Courtwright AM, Gabriel PE: Clinical databases for chest physicians. Chest 153:1016-1022, 2018
10. Horvath MM, Rusincovitch SA, Brinson S, et al: Modular design, application architecture, and usage of a self-service model for enterprise data delivery: The Duke Enterprise Data Unified Content Explorer (DEDUCE). J Biomed Inform 52:231-242, 2014
11. Horvath MM, Winfield S, Evans S, et al: The DEDUCE Guided Query tool: Providing simplified access to clinical data for research and quality improvement. J Biomed Inform 44:266-276, 2011
12. Nigrin DJ, Kohane IS: Data mining by clinicians. Proceedings/AMIA: Annual Symposium Proceedings/AMIA Symposium:957-961, 1998
13. Huser V, Narus SP, Rocha RA: Evaluation of a flowchart-based EHR query system: A case study of RetroGuide. J Biomed Inform 43:41-50, 2010
14. Kim C-S, Lee JY, Chung BH, et al: Report of the Second Asian Prostate Cancer (A-CaP) Study meeting. Prostate Int 5:95-103, 2017
15. e-PROSPeCT. https://github.com/max-uhlich/e-PROSPeCT.
16. Alberta Prostate Cancer Research Initiative (APCaRI): Alberta Prostate Cancer Research Initiative homepage. https://apcari.ca.
17. The PostgreSQL Global Development Group: PostgreSQL: the world's most advanced an open source relational database. https://www.postgresql.org/.
18. van Rossum G: Python tutorial. Technical Report CS-R9526. Amsterdam, the Netherlands, Centrum voor Wiskunde en Informatica, 1995
19. Goel MK, Khanna P, Kishore J: Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res 1:274-278, 2010
20. National Cancer Institute: SEER data & software. https://seer.cancer.gov/data-software.
21. American College of Surgeons, American Cancer Society: National Cancer Database. https://www.facs.org/quality-programs/cancer/ncdb.
22. University of California, San Francisco, Department of Urology: CaPSURE. https://urology.ucsf.edu/research/cancer/capsure.
23. Gregg JR, Lang M, Wang LL, et al: Automating the determination of prostate cancer risk strata from electronic medical records. JCO Clin Cancer Inform [epub ahead of print on June 8, 2017]
24. Febbo PG, Mulligan MG, Slonina DA, et al: Literature Lab: A method of automated literature interrogation to infer biology from microarray analysis. BMC Genomics 8:461, 2007
25. Varian Medical Systems: InSightive analytics. https://www.varian.com/oncology/products/software/information-systems/insightive-analytics.
26. Harris PA, Taylor R, Thielke R, et al: Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 42:377-381, 2009

APPENDIX

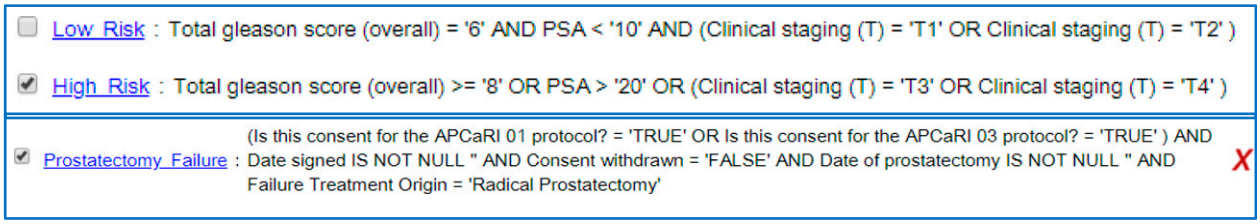

FIG A1. Adjustable template for Predictive Research Online System Prostate Cancer Tasks (PROSPeCT) system dashboard.