CanRisk Tool – A web interface for the prediction of breast and ovarian cancer risk and the likelihood of carrying genetic pathogenic variants

Tim Carver[1,*], **Simon Hartley**[2], **Andrew Lee**[1], **Alex P Cunningham**[1], **Stephanie Archer**[3], **Chantal Babb de Villiers**[3], **Jonathan Roberts**[4], **Rod Ruston**[5], **Fiona M Walter**[3,6], **Marc Tischkowitz**[4,7], **Douglas F Easton**[1,8], **Antonis C Antoniou**[1]

[1]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, UK

[2]Centre for Computational Biology, University of Birmingham, UK

[3]Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, UK

[4]East Anglian Medical Genetics Service, Cambridge Biomedical Campus, Addenbrooke's Hospital, Cambridge, UK

[5]Priory Analysts, PO Box 4030, Milton Keynes, UK

[6]Centre for Cancer Research and Department of General Practice, University of Melbourne, Australia

[7]Academic Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK

[8]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, UK

Abstract

Background—The CanRisk Tool (https://canrisk.org) is the next generation web interface for the latest version of the BOADICEA (Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm) state-of-the-art risk model and a forthcoming ovarian cancer risk model.

Methods—The tool captures information on family history, rare pathogenic variants in cancer susceptibility genes, polygenic risk scores, lifestyle/hormonal/clinical features and imaging risk factors to predict breast and ovarian cancer risks and estimate the probabilities of carrying pathogenic variants in certain genes. It was implemented using modern web frameworks, technologies and web services to make it extensible and increase accessibility to researchers and third-party applications. The design of the graphical user interface was informed by feedback from healthcare professionals and a formal evaluation.

*Corresponding author address: tjc29@medschl.cam.ac.uk.

Results—This freely accessible tool was designed to be user-friendly for clinicians and to boost acceptability in clinical settings. The tool incorporates a novel graphical pedigree builder to facilitate collection of the family history data required by risk calculations.

Conclusions—The CanRisk Tool provides healthcare professionals and researchers with a user-friendly interface to carry out multifactorial breast and ovarian cancer risk predictions. It is the first freely accessible cancer risk prediction program to carry the CE marking.

Impact—There have been over 3100 account registrations, and 98000 breast and ovarian cancer risk calculations have been run within the first 9 months of the CanRisk Tool launch.

Keywords

Breast cancer; ovarian cancer; risk prediction; carrier probability; PRS

Introduction

Cancer risk prediction is widely used to identify individuals at high risk who may benefit from increased surveillance or risk reducing interventions. The BOADICEA model, first developed in 2002, models the risks of breast and ovarian cancer based on family history and genotypes for variants in *BRCA1* and *BRCA2* (1). The model has since been extensively updated to incorporate the effects of pathogenic variants in other genes, common genetic variants (summarised as polygenic risk scores, PRS), lifestyle, hormonal and clinical features, breast density and disease pathology (2). The model has been extensively validated, both for the prediction of carrier probabilities and prediction of subsequent cancer risk (3, 4).

BOADICEA was first made available online in the BOADICEA Web Application (BWA) in 2007 (5). There have been several releases since then to accommodate extensions to the BOADICEA model and to implement improvements to the web interface (6,7). The BWA uses a series of web forms to capture cancer family history, *BRCA1/2* genetic test results and tumour pathology data. The BWA has been widely used in clinician settings, particularly clinical genetic centres, with more than 15000 user registrations from over 100 countries.

BOADICEA is currently recommended by several national bodies and organisations for determining eligibility for high-risk breast cancer screening, eligibility for *BRCA1/2* mutation screening and to inform breast cancer risk management. These include the UK NICE guidelines (https://www.nice.org.uk/guidance/cg164), the American Cancer Society (8), the Ontario Breast Screening program (https://www.cancercareontario.ca/en/guidelines-advice/cancer-continuum/screening/breast-cancer-high-risk-women) and the eviQ Australian guidelines for health professionals (https://www.eviq.org.au/cancer-genetics/adult/risk-management).

The motivation to develop a new web tool arose firstly from the need to accommodate advances in the underlying risk prediction algorithms. In addition, the new interface has been designed to enhance usage in multiple levels of clinical care, and to conform to more recent regulations on software for clinical application, specifically EU medical device regulations (9). In addition, there was a need to resolve limitations resulting from older web

technologies. For example, the BWA used multiple HTML forms (each form capturing details for a single pedigree member) to capture pedigrees of arbitrary size and structure because there were significant barriers to the implementation of graphical pedigree building tools using open standard web technologies (e.g. Scalable Vector Graphics) and Dynamic HTML when it was first developed. This was a novel and powerful approach at the time, but multiple form-based pedigree data entry can be complicated (requiring navigation backward/ forward through multiple pages) and time consuming to complete in practice. Finally, there was a need to present the outputs from the risk calculations in a wider range of formats.

To address these issues and enhance usability, increase accessibility, and enable rapid software development and maintainability, the web tool has now been wholly redeveloped from the ground-up using state-of-the-art software engineering techniques and frameworks. This article describes the resulting CanRisk Tool, which provides access to the latest version of the BOADICEA model (2) and related ovarian cancer prediction model (10).

Materials And Methods

The design of the CanRisk Tool was motivated by the desire to maximise usability for the healthcare professional (HCP) and to facilitate the patient consultation process. This was done by providing a user-friendly interface and by presenting results in a range of popular output formats to help HCPs convey their meanings to the patient. The initial design and subsequent development of the tool was informed by frequent input from the target HCPs including clinical geneticists, genetic counsellors, surgeons, oncologists, practice nurses and general practitioners, and included a formal evaluation (11). This study demonstrated that: *"the prototype CanRisk tool was generally acceptable to most participants due to its intuitive design."*

User-friendly data entry was implemented using standard HTML form elements (text/ number input fields) on a single web page that reacts dynamically and intelligently to HCP inputs. HCPs are presented with a simple and intuitive user interface. PRSs can also be incorporated in risk calculations either by uploading a variant call format (VCF) file containing the genotypes of common variants included in the PRSs (12), or by entering the PRSs directly.

A novel web based graphical pedigree builder was developed that allows HCPs to graphically build the family structure (13) and capture this information rapidly in a visual and intuitive manner.

A web services architecture has been employed to provide programmatic access to the risk models across the Internet. This means that the risk models can be accessed by both HCPs using the CanRisk Tool, and by researchers using their own bespoke research software. Web services are used to calculate PRSs from a VCF file for use in the risk calculations.

Data security was also of paramount importance in the design of the CanRisk Tool. The tool uses in-built security features provided by the Django web development framework, https:// www.djangoproject.com (e.g. cross site request forgery protection), as well as data encryption (via HTTPS) during transmission between the HCP's web browser and the

CanRisk server. To increase the security of the web pages, Open Web Application Security Project (https://owasp.org) security headers are implemented. The tool also removes patient identifiable information from the input dataset before it is submitted to the server for processing. Additionally, once the calculations are complete, the input data are deleted immediately from the server before the results are sent back to the web browser.

The tool was developed using state-of-the-art software engineering practices e.g. automated continuous integration, testing, version control, web application security, coding standards and web development frameworks (e.g. Django). These techniques were adopted to make the tool easier to maintain, shorten software release cycles, and facilitate extensibility. The software architecture (Figure 1) is based on the distributed system client-server model (14). Software development was carried out iteratively and prototypes were evaluated by HCPs in the UK, France and Germany (11). The web interface was refined using feedback from successive rounds of evaluation.

Test suites have been developed for the web services and web tool. These make the tool more robust, verify performance and increase development productivity by automatically testing the software at each stage of the development cycle. Web browser automation tests, using Selenium (https://www.selenium.dev/), validate the CanRisk Tool functionality and check compatibility with current web browser versions.

Results

The CanRisk Tool provides HCPs and researchers with a user-friendly interface to carry out multifactorial breast and ovarian cancer risk predictions. The tool can be customised by setting default preferences that are saved to the HCP's browser, such as the units of measurement and country of residence. To provide more specific risk predictions, HCPs are encouraged to input as much accurate information as possible and are led through the data input process via a series of sections in a vertical accordion style interface (Supplementary Video S1). Completed section headers are then marked as green (Figure 2). A novel feature is the inclusion of PRS input to the risk predictions. On uploading a VCF file or directly entering the PRSs, graphical representations of the PRSs are displayed.

The family history section uses a graphical pedigree builder (13), Supplementary Video S2. It uses standard pedigree nomenclature to represent a family tree (15). To enable rapid building of family structures in various clinical care settings, an initial family structure (e.g. nuclear, primary or secondary degree relatives) or the number of each type of relative (e.g. number of brothers, daughters) can be selected. Clicking on family members in the graphic reveals widgets (i.e. clickable icons) that are used to add family members (e.g. parents, siblings, partners). This approach provides the flexibility to allow for arbitrary family sizes and structures of varying complexity (including loops). On selecting an individual, their details (e.g. age, year of birth, cancer history, pathology, genetic test results) can be set.

While the CanRisk Tool implements the standard version of the risk models, several parameters in the models can be tailored as necessary by the HCP: these include the cancer incidences in the population, the frequencies of pathogenic variants, mutation testing

sensitivity and the PRS distribution in the population. These enable HCPs across the world to tailor risk modelling parameters to their region and setting.

The computed breast and ovarian cancer risks are presented in different output formats to help HCPs convey complex risk information (Supplementary Video S3). The remaining lifetime breast and ovarian cancer risks for the next 5 years, 10 years and to the age of 80 are presented both in text and icon array formats. Additionally, cancer risks are presented in a graph from the current age to the age of 80. For each representation, the personal cancer risks are given in the context of the general population risks. The breast cancer risk classification (i.e. near population, moderate or high risk), as defined by UK NICE clinical management guidelines, is presented in text, table and graphical formats. The probabilities of carrying pathogenic variants in the cancer susceptibility genes used in the models are also presented.

The CanRisk Tool can also generate a customised report that summarises the input dataset and results. Reports can be printed or downloaded in PDF format. For data protection reasons, the reports are generated within the browser so that patient identifiable data are not transmitted to the server. The input dataset can be saved to a CanRisk format data file that can later be loaded into the tool to recreate the session. Reports and data files can be attached to patient records in the HCP's patient management system.

The web services are used by the CanRisk Tool to run the risk calculations and calculate PRSs. However, the main advantage of using the web services layer is that it provides programmatic access to the underlying risk models over the Internet, thus widening access to third party software. For example, researchers have developed software to submit pedigrees to the CanRisk web service to obtain risk estimates in a non-graphical manner. This approach is particularly useful when researchers wish to run BOADICEA model risk calculations within their own software tools as part of an epidemiological analysis.

It is important for HCPs to be able to compute risks quickly. Calculation times are dependent on the size and complexity of pedigrees. The cancer risk models have been greatly optimised so that the CanRisk Tool is very responsive, with 50% of breast/ovarian cancer risks being completed within ~10s and 98% within ~60s.

Online documentation has been provided (https://canrisk.atlassian.net/wiki/spaces/FAQS/overview) and will be expanded to create a CanRisk knowledge base for users of the tool. In addition, the new tool has a dedicated helpdesk (https://canrisk.atlassian.net/servicedesk) to provide support. As a result of requests to the helpdesk for features and improvements, several updates to the tool have already been implemented (https://canrisk.org/releases/).

Discussion

The CanRisk Tool was developed to broaden access to the underlying cancer risk models and to facilitate data entry by HCPs. Within the first 9 months of its release, it has had over 3100 account registrations, and 98000 breast and ovarian cancer risk calculations have been run. In addition, the CanRisk Tool implementation of BOADICEA has now been incorporated into the NCCN guidelines for familial breast/ovarian cancer [16].

The CanRisk Tool provides a user-friendly web interface for HCPs across clinical settings. The breast and ovarian cancer risk predictions and pathogenic variant carrier probabilities can be invaluable to HCPs and patients in deciding the patient's best clinical management. The interface brings together multifactorial elements including, for the first time, PRSs and a novel graphical pedigree builder that enables HCPs to build and visualise large and complex pedigrees as they are building them. This approach makes data entry much quicker and easier than the form-based approach previously used by the BWA. Data sent to and from the web services are encrypted and pedigree data are anonymised. The CanRisk web services are a powerful addition, providing programmatic access to the cancer risk prediction models. They enable researchers, clinical users and third-party clinical applications to submit single/ batch data files and compute risks without human interaction, i.e. data entry via the CanRisk Tool web interface.

The CanRisk Tool is classified as a medical device under EU regulations. The tool carries the CE marking to indicate compliance with applicable safety and performance requirements for use by HCPs within the European Economic Area (EEA) (9). In particular, the CE marking: (i) shows that the manufacturer has checked that the product meets EU safety, health or environmental requirements; (ii) is an indicator of the product's compliance with EU legislation; and (iii) allows the free movement of products within the EEA. The CE marking process involves activities such as risk management, software verification and performance evaluation. These activities help to ensure that the software is implemented in accordance with software engineering best practice and that it fulfils stated user requirements. The CanRisk Tool obtained CE marking in November 2019: to our knowledge, this is the first breast and ovarian cancer risk prediction tool to achieve this.

Future development of the CanRisk Tool will accommodate further risk model extensions. Additional evaluations of the tool will determine how it can be further enhanced to meet the requirements of specific clinical settings, for example in GP consultations by tailoring the risk communication and management components. A patient facing interface for data capture is also planned. Future studies aim to assess the feasibility and acceptability among HCPs and patients of using the CanRisk Tool across different clinical settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Antoniou AC, Pharoah PDP, McMullan G, Day NE, Stratton MR, Peto J, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer. 2002; 86:76–83. [PubMed: 11857015]

2. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med. 2019; 21:1708–18. [PubMed: 30643217]

3. Choudhury PP, Brook MN, Wilcox AN, Lee A, Mulder CV, Coulson P, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort. medRxiv. 2020 2020.04.27.20081265

4. Lakeman IMM, Rodríguez-Girondo M, Lee A, Ruiter R, Stricker BH, Wijnant SRA, et al. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. Genetics in Medicine. 2020:1–9.

5. Cunningham AP, Antoniou AC, Easton DF. Clinical software development for the Web: lessons learned from the BOADICEA project. BMC Medical Informatics and Decision Making. 2012; 12:30. [PubMed: 22490389]

6. Lee AJ, Cunningham AP, Kuchenbaecker KB, Mavaddat N, Easton DF, Antoniou AC, et al. BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. Br J Cancer. 2014; 110:535–45. [PubMed: 24346285]

7. Lee AJ, Cunningham AP, Tischkowitz M, Simard J, Pharoah PD, Easton DF, et al. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. Genet Med. 2016; 18:1190–8. [PubMed: 27464310]

8. Smith RA, Manassaram-Baptiste D, Brooks D, Doroshenk M, Fedewa S, Saslow D, et al. Cancer screening in the United States, 2015: a review of current American cancer society guidelines and current issues in cancer screening. CA Cancer J Clin. 2015; 65:30–54. [PubMed: 25581023]

9. Department for Business EIS. CE marking 2012 [23/8/2019]. Available from: https://www.gov.uk/guidance/ce-marking

10. Jervis S, Song H, Lee A, Dicks E, Harrington P, Baynes C, et al. A risk prediction algorithm for ovarian cancer incorporating BRCA1, BRCA2, common alleles and other familial effects. Journal of Medical Genetics. 2015; 52:465–75. [PubMed: 26025000]

11. Archer S, Babb de Villiers C, Scheibl F, Carver T, Hartley S, Lee A, et al. Evaluating clinician acceptability of the prototype CanRisk tool for predicting risk of breast and ovarian cancer: A multi-methods study. PLoS ONE. 2020; 15 e0229999 [PubMed: 32142536]

12. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. The American Journal of Human Genetics. 2019; 104:21–34. [PubMed: 30554720]

13. Carver T, Cunningham AP, Babb de Villiers C, Lee A, Hartley S, Tischkowitz M, et al. pedigreejs: a web-based graphical pedigree editor. Bioinformatics. 2018; 34:1069–71. [PubMed: 29095980]

14. Sommerville, I. Software engineering. Tenth. Pearson; Boston: 2016.

15. Bennett RL, Steinhaus KA, Uhrich SB, O'Sullivan CK, Resta RG, Lochner-Doyle D, et al. Recommendations for standardized human pedigree nomenclature. Pedigree Standardization Task Force of the National Society of Genetic Counselors. Am J Hum Genet. 1995; 56:745–52. [PubMed: 7887430]

16. NCCN Clinical Practice Guidelines in Oncology. [September 8, 2020] Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic Version 1.2021. https://www.genomeweb.com/sites/default/files/nccn__genetic__cancer__risk__assessment.pdf
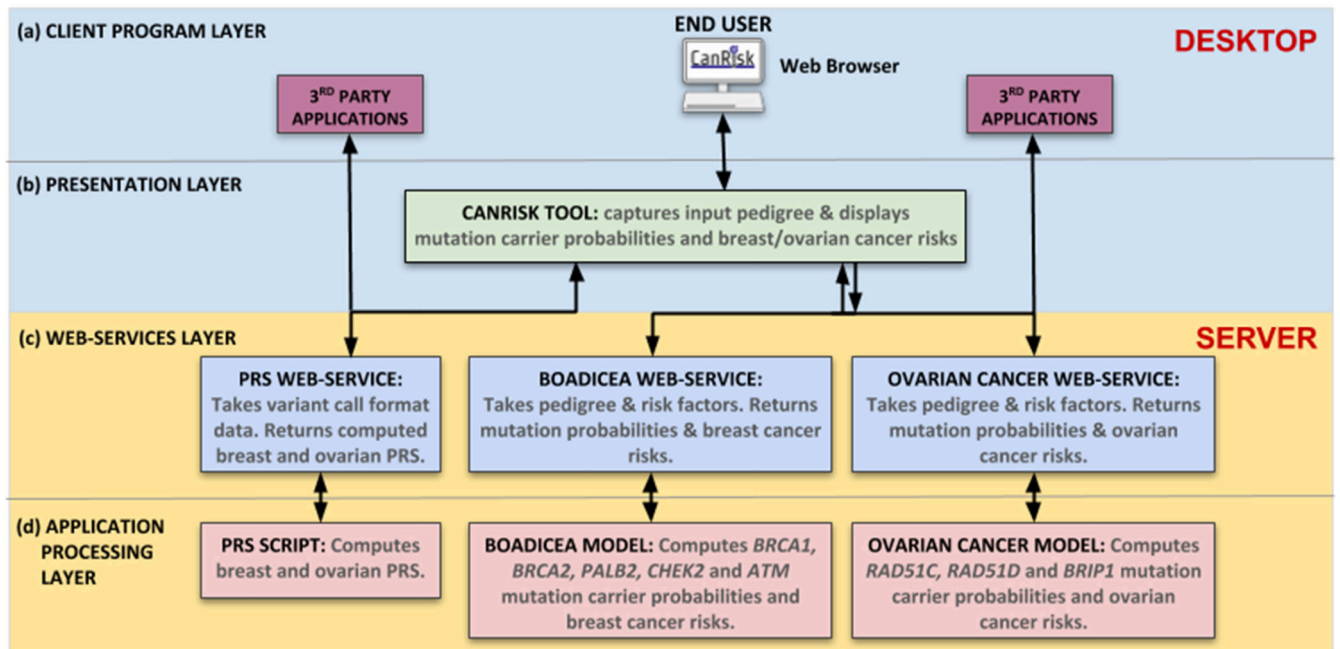
Figure 1. CanRisk Tool software architecture.
(a) Client Program Layer. The HCP accesses the CanRisk Tool via their web browser.
Similarly, researchers, clinical users and third-party clinical application software accesses
the services provided by the web services layer directly. (b) Presentation Layer. The tool
generates the web pages that are presented in the HCP's web browser. It captures the input
dataset (including pedigree information) and displays the output dataset (pathogenic variant
carrier probabilities and breast/ovarian cancer risks) in text, tables and graph formats. (c)
Web Services Layer. The PRS, BOADICEA and Ovarian Cancer web services are
application processing interfaces (APIs) that make the programs in the Application
Processing Layer accessible to other programs over the Internet. (d) Application Processing
Layer. The PRS Script computes breast and ovarian PRSs. The BOADICEA Model
computes the probabilities of carrying a pathogenic variant in BRCA1, BRCA2, PALB2,
CHEK2 and ATM and breast cancer risks. The Ovarian Cancer Model computes the
probabilities of carrying a pathogenic variant in RAD51C, RAD51D and BRIP1 and ovarian
cancer risks. Communications between the desktop and server are via secure HTTPS.

Figure 2. CanRisk Tool data input is separated into sections.
Each section heading turns green when data entry for the section is complete. CanRisk datafiles containing the pedigree and risk factors can be saved and later loaded to recreate a previous session, eliminating the need for the HCP to re-enter data.