University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Terry, MB;Phillips, KA;Daly, MB;John, EM;Andrulis, IL;Buys, SS;Goldgar, DE;Knight, JA;Whittemore, AS;Chung, WK;Apicella, C;Hopper, JL

Title:
Cohort profile: The breast cancer prospective family study cohort (ProF-SC)

Date:
2016-06-01

Citation:
Terry, M. B., Phillips, K. A., Daly, M. B., John, E. M., Andrulis, I. L., Buys, S. S., Goldgar, D. E., Knight, J. A., Whittemore, A. S., Chung, W. K., Apicella, C. & Hopper, J. L. (2016). Cohort profile: The breast cancer prospective family study cohort (ProF-SC). International Journal of Epidemiology, 45 (3), pp.683-692. https://doi.org/10.1093/ije/dyv118.

Persistent Link:
https://hdl.handle.net/11343/261487

License:
CC BY-NC

International Journal of Epidemiology, 2016, 683–692
doi: 10.1093/ije/dyv118
Advance Access Publication Date: 13 July 2015
Cohort Profile

Cohort Profile

Cohort Profile: The Breast Cancer Prospective Family Study Cohort (ProF-SC)

Mary Beth Terry,[1,2]* Kelly-Anne Phillips,[3,4,5] Mary B Daly,[6] Esther M John,[7,8] Irene L Andrulis,[9] Saundra S Buys,[10] David E Goldgar,[11] Julia A Knight,[12,13] Alice S Whittemore,[8] Wendy K Chung,[14] Carmel Apicella[4] and John L Hopper[4]*

[1]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA, [2]Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY, USA, [3]Peter MacCallum Cancer Centre and Sir Peter MacCallum Department of Oncology, [4]School of Population and Global Health, [5]St Vincent's Hospital, University of Melbourne, Melbourne, VIC, Australia, [6]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA, [7]Cancer Prevention Institute of California, Fremont, CA, USA, [8]Department of Health Research & Policy, Stanford University School of Medicine, Stanford, CA, USA, [9]Lunenfeld-Tanenbaum Research Institute, University of Toronto, Toronto, ON, Canada, [10]Department of Medicine, [11]Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA, [12]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, [13]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada and [14]Departments of Pediatrics and Medicine, Columbia University, New York, NY, USA

*Corresponding authors. Mary Beth Terry, Columbia University Mailman School of Public Health, Department of Epidemiology, 722 West 168th St, New York, NY 10032, USA. E-mail: mt146@columbia.edu; John L Hopper, The University of Melbourne, School of Population Health, Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, Level 3, 207 Bouverie Street, Carlton, Vic 3053, Australia. E-mail: j.hopper@unimelb.edu.au

Why was the cohort established?

For breast cancer, like most other common diseases, having a family history of the disease is associated with an increased risk. There is large heterogeneity in absolute and relative breast cancer risks associated with family history, depending on the age of the woman, the age(s) at diagnosis of her affected relative(s) and the genetic relationship(s). Women with one affected first-degree relative are on average at 2-fold increased risk of breast cancer relative to women with no first-degree family history, and this increases to 4-fold for women with three or more affected first-degree relatives[1].

Under a multiplicative risk model, the underlying risk distribution must be highly skewed, and most women in the population are well below average risk; see Figure 1, which illustrates the difference between women unselected for familial or genetic risk (blue line) and women enriched for familial/genetic risk (dotted red line). Given that epidemiological studies make inference about risk for the controls, almost all existing knowledge about risk factors is not relevant to 'women at average risk', but to women at lower than average risk. It is not known if this knowledge applies to women at increased, if not high, risk. To find evidence relevant to women across the full continuum of risk, with the potential for targeted risk modification and prevention, we have established and genetically-characterised a large prospective family-based cohort enriched for familial risk.

Over the past two decades, an increasing number of genetic risk factors have been identified.[2,3] Nevertheless, the majority of women with a family history of breast

© The Author 2015; all rights reserved. Published by Oxford University Press on behalf of the International Epidemiological Association

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

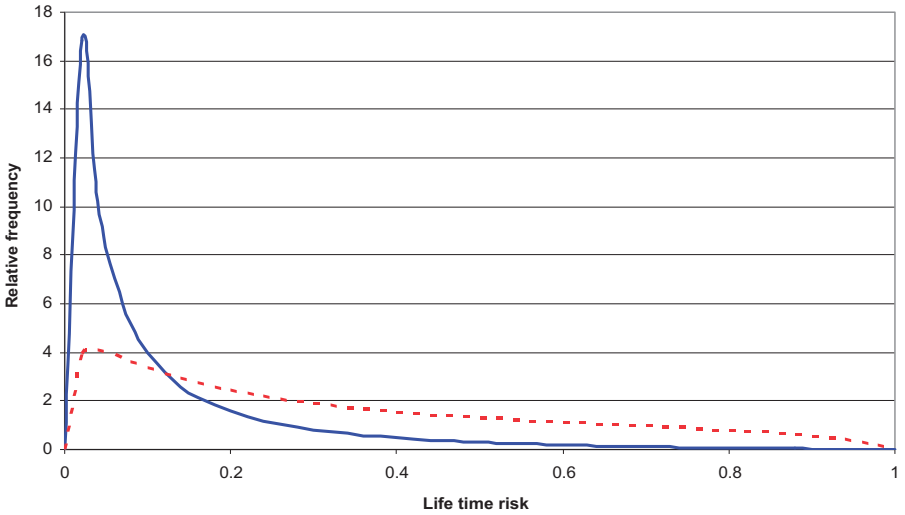


Figure 1. Comparison of the theoretical distribution of familial risk profile (FRP) for women from the general population (blue line) and for those affected with either early-onset breast cancer or unaffected but with a strong family history of breast cancer, equivalent to a 3-fold increased risk (dotted red line), under a multiplicative, multifactorial, polygenic model. For details see [18,47].

cancer, even those with a strong family history, do not have causal mutations in the known genes, and large genome-wide association studies (GWAS) and now next-generation sequencing efforts are identifying additional genetic risk factors.[4–10]

Epidemiological evidence suggests that some environmental factors modify breast cancer risk for women with a family history. Most epidemiology studies, however, record only first-degree family history as a binary factor (e.g.[1,11,12]), which does not capture the potential importance of disease in second-degree and more distant relatives,[13–15] and rarely take into account the importance of age at diagnosis of affected relatives. The few studies that do[16,17] suggest greater environmental and genetic heterogeneity in risk.

One approach to studying gene-environment interactions is to consider a woman's underlying familial risk profile (FRP), representing her inherent lifetime risk due to familial determinants. A woman's FRP can be predicted from her multi-generational family history including the number of affected relatives and her relationship with each affected relative, their age(s) at diagnosis, and if known, her genetic risk status (including causal variants and markers associated with risk) and the genetic risk status of her relatives.

It is not well recognised that there must be very large variation in FRP. Given the increased risks associated with having a family history, mathematical models predict that, as a group, women in the top 25% of FRP must be at least 20 times more likely to develop breast cancer than women in the bottom 25% of FRP.[18,19] Nevertheless, unlike matching on age to control for its strong effect on cancer risk, epidemiological studies rarely match well by design or analysis on FRP, even though cases and controls differ greatly by FRP, especially in the upper tail.

Environmental and genetic effects may have different effects in women with increased FRP. Studies of such 'gene-environment interactions' for which controls are better matched to cases for FRP, and even for mutations in specific genes—either by design or by analyses that use good predictors of FRP—might be more informative, especially if both cases and controls are over-sampled for increased familial risk. They also have greater validity if they are prospective.[18,19] Few prospective studies of families exist. They include the Minnesota Breast Cancer Family Study (544 families), established in the 1950s,[20–23] and the USA-based Sisters Study (50 844 unaffected sisters of affected women aged 35–74 years).[24] Family-based cohorts are also important for novel behavioural, psychosocial and health care utilization research, such as attitudes and practices regarding screening and risk reduction, and for the translation of primary, secondary and tertiary prevention findings into clinical practice.[25–28]

In the mid-1990s,when two major susceptibility genes, *BRCA1* and *BRCA2*, were discovered, the Breast Cancer Family Registry (BCFR),[29] and the Kathleen Cuningham Foundation Consortium for research into Familial Breast cancer (kConFab)[30] (in 2001 the kConFab FUP started[31]) were established. Importantly, both the BCFR and the kConFab were designed from the outset so that they could generate cohorts from which data could be pooled; they used the same baseline questionnaire and have conducted regular active follow-up of families. In mid-2014, a systematic follow-up of both the BCFR and kConFab FUP was completed as part of an NIH-funded grant to study the following aims using a prospective family study cohort

(ProF-SC) (CA159868): (i) estimate age-specific absolute risks of breast cancer; (ii) estimate relative risks associated with modifiable factors and test whether these associations vary by FRP; and (iii) develop and internally validate comprehensive clinical breast cancer risk assessment models for women across the spectrum of breast cancer risk.

Who is in the cohort?

The ProF-SC includes all female participants in the BCFR or kConFab FUP who were enrolled before 30 June 2011 and completed a baseline questionnaire. A total of 31 640 women from 11 171 families, 11.4% of whom are Ashkenazi Jewish, completed the same baseline questionnaire.

The BCFR recruited families from six sites; one in Australia, one in Canada and four in the USA. The Australian, Canadian and Northern Californian sites recruited population-based case families using cancer registries. kConFab and the Australian BCFR recruited cases unselected for family history, over-sampling for early age at diagnosis, whereas the other two population-based sites used a two-stage sampling scheme, over-sampling for early age at diagnosis and/or having a family history or other predictors of a genetic predisposition. The Northern California site over-sampled racial/ethnic minority families. The New York, Philadelphia, Utah, Canadian and Australian sites recruited multiple-case families through family cancer clinics and community outreach.

kConFab recruited multi-generational, multiple-case families through cancer family clinics in Australia and New Zealand.[30,31] The eligibility criteria for recruitment of families evolved over time and were intended to maximise the number of living potentially high-risk women, including known *BRCA1* and *BRCA2* mutation carriers, whether or not they had been diagnosed with breast cancer.

Although ascertainment issues sometimes require separate analyses of population-based and clinic-ascertained families for retrospective studies, for prospective studies all unaffected family members can be combined into a single cohort because all families are being followed using the same methods and their members have been studied using the same protocol at baseline. A family cohort is in contrast to conventional cohort studies in which the vast majority of incident cases do not have a family history, at least not for first- or second-degree relatives. A large proportion of the ProF-SC cohort was affected at baseline, which also facilitates studies on risk factors for subsequent primary breast cancer, including contralateral disease.

Table 1 describes the baseline characteristics of ProF-SC, illustrating wide variation in demographics, family history and other risk factors. Table 2 summarises the distribution of the incident breast cancer cases diagnosed since baseline. There is a wide distribution in age at diagnosis, with cases diagnosed before age 50 years accounting for 34% in the sub-cohort of women unaffected at baseline and 28% in the sub-cohort of women affected at baseline, respectively. For both these sub-cohorts, a high proportion of incident cases have been confirmed through pathology records (78% and 71%, respectively) and a high proportion have DNA available (88% and 95%, respectively).

How often have they been followed up?

The family-based design facilitates follow-up primarily through tracing and updates of vital status using multiple informants, thereby increasing the validity of the cohort's data on outcomes and family cancer history.[32] Since baseline, there has been regular contact with families through BCFR and kConFab newsletters and websites. Vital and cancer statuses have been updated through phone interviews, mailed questionnaires, clinic visits and linkages to cancer registries. In addition, there have been systematic updates of risk factor and clinical outcomes data (see below for details).

High participation at follow-up is a critical issue for the validity of cohort studies, and we have demonstrated that this can be achieved by using a family-based design with multiple contacts typically available for each cohort member. Of the 31 640 women in the cohort, after an average of 9 years of follow-up, 11% were no longer living, 5% no longer wished to participate in follow-up and 14% have been lost to follow-up. Table 1 shows that baseline characteristics are similar for those lost to follow-up or no longer participating in active follow-up and those who have remained active. For those lost to follow-up and/or who dropped out of active follow-up, we have information on vital status, including cancer history, for 63% from their participating relatives.

What has been measured?

For all ProF-SC members, the BCFR and kConFab have collected detailed family history, demographic and risk factor data and biospecimens, regardless of their breast cancer history. For all women with breast cancer, pathology records, archived tumour tissue and self-reported information on cancer treatment have been sought (Table 3).

Family history/pedigrees. Pedigree information includes age at diagnosis of all cancers (except non-melanoma skin cancer) and deaths for first- and second-degree relatives of all participants (not just probands). This provides the most

Table 1. Baseline characteristics of ProF-SC participants, by breast cancer status at baseline and by loss to follow-up

	Total cohort ($n = 31640$)	Affected at baseline ($n = 12787$)	Unaffected at baseline ($n = 18853$)	Lost to follow-up ($n = 5728$)[a]
	N or % or mean ± SD (min, max)	*N* or % or mean ± SD (min, max)	*N* or % or mean ± SD (min, max)	*N* or % or mean ± SD (min, max)
BRCA1 mutation only	1514	782	732	174
BRCA2 mutation only	1219	633	586	138
BRCA1&*BRCA2* mutation	8	8	0	1
Year at recruitment				
1992–94	1.7	1.1	2.0	0.6
1995–99	47.1	47.8	46.6	38.0
2000_04	30.2	29.8	30.5	35.0
2005–09	18.3	17.9	18.6	23.3
2010–30 June 2011	2.7	3.3	2.3	3.1
Age at baseline (years)	49.8 ± 14.8 (18, 101)	52.9 ± 12.1 (21, 98)	47.8 ± 16.1 (18, 101)	48.8 ± 15.5 (18, 97)
Number of first-degree relatives with breast cancer	0.9 ± 0.8 (0, 6)	0.6 ± 0.8 (0, 6)	1.1 ± 0.7 (0, 6)	0.9 ± 0.8 (0, 6)
Number of second-degree relatives with breast cancer	0.7 ± 0.9 (0, 11)	0.6 ± 0.9 (0, 8)	0.8 ± 1.0 (0, 11)	0.7 ± 0.9 (0, 7)
Body mass index (BMI) (kg/m^2)	25.9 ± 5.6	26.2 ± 5.6	25.8 ± 5.6	26.2 ± 5.8
Age at menarche (years)	12.8 ± 1.6	12.7 ± 1.6	12.9 ± 1.6	12.8 ± 1.6
Smoking status				
Never	60.1	59.4	60.5	61.1
Former	26.9	28.3	25.9	23.3
Current	13.1	12.3	13.7	15.6
Alcohol intake				
Never	52.4	54.4	51.1	58.7
Former	14.3	13.4	14.9	14.1
Current	33.3	32.3	34.0	27.2
Menopausal hormone use				
Never	75.1	72.8	76.7	79.6
Former	16.9	24.5	11.8	13.9
Current	7.9	2.7	11.5	6.5
Hormonal contraceptive use				
Never	27.5	29.1	26.4	33.5
Former	64.1	69.3	60.5	58.5
Current	8.4	1.6	13.0	8.0
Race/ethnicity				
Non-Hispanic White	75.9	68.8	80.8	62.7
Non-Hispanic Black	6.2	8.6	4.6	8.7
Hispanic	9.4	11.2	8.1	15.1
Asian	5.8	8.8	3.8	10.2
Other	2.7	2.7	2.7	3.3
Education				
≤High school graduation / GED	34.4	34.6	34.3	40.4
Vocational or technical school / some college or university	36.8	35.6	37.7	33.7
Bachelor's or graduate degree	28.7	29.8	28.0	26.0
Benign breast disease				
Yes	31.0	36.3	27.4	26.7
No	69.0	63.7	72.6	73.3

(Continued)

Table 1. Continued

	Total cohort ($n = 31640$)	Affected at baseline ($n = 12787$)	Unaffected at baseline ($n = 18853$)	Lost to follow-up ($n = 5728$)[a]
	N or % or mean ± SD (min, max)	*N* or % or mean ± SD (min, max)	*N* or % or mean ± SD (min, max)	*N* or % or mean ± SD (min, max)
Parity				
0	21.1	17.7	23.5	23.2
1	11.7	13.0	10.8	13.3
2	30.0	33.4	27.7	27.8
3	20.3	20.5	20.2	17.8
≥4	16.8	15.5	17.7	17.9
Menopausal status				
Premenopausal	45.1	30.3	55.3	49.9
Postmenopausal	54.9	69.7	44.7	50.1

[a]Includes refusals and not located.

Table 2. Prospectively ascertained breast cancer cases among ProF-SC participants

	Unaffected at baseline	Affected at baseline
Number of women with self-reported new breast cancers	1093	1252
New breast cancers confirmed by pathology, *n* (%)	848 (78%)	883 (71%)
New breast cancers with blood/buccalsample collected, *n* (%)	961 (88%)	1184 (95%)
Age at diagnosis of new breast cancer (years), *n* (%)		
<40	117 (11%)	103 (8%)
40–44	107 (10%)	120 (10%)
45–49	147 (13%)	131 (10%)
50–54	128 (12%)	205 (16%)
55–59	132 (12%)	186 (15%)
60–64	135 (12%)	181 (14%)
65–69	98 (9%)	129 (10%)
≥70	199 (18%)	185 (15%)
Unknown	30 (3%)	12 (1%)

comprehensive description of family history of any epidemiological breast cancer study.

Risk factor questionnaires. The BCFR and kConFab used the same baseline questionnaire to collect data on menstrual and reproductive history, medical history and behavioral factors. The BCFR conducted a systematic follow-up beginning in 2007, and collected updated information on personal and family history of cancer, breast and ovarian surgeries and breast cancer risk factors collected at baseline. New items of interest were added, including screening behaviours such as use of magnetic resonance imaging (MRI), use of non-steroidal anti-inflammatory drugs (NSAIDs) and knowledge and understanding of genetic test results. The most recent systematic follow-up of the BCFR, conducted in 2011–14, updated some risk factor data in addition to family history and pedigree information. kConFab FUP surveyed participants every 3 years, using questionnaires that cover the same content as the BCFR follow-up questionnaires, with the exception of diagnostic radiation.

Biospecimen collection. At baseline, depending on relationship to the proband, most women were asked by the BCFR and kConFab to provide either a blood or a buccal sample. As a result, for 83% of ProF-SC there are banked DNA and plasma samples, and for an additional 2% there is DNA from buccal samples. There are no major differences between women who gave blood and those who did not with regard to the characteristics in Table 1 (data not shown).

BRCA1 and BRCA2 genotyping. Screening for germline *BRCA1* and *BRCA2* mutations and other known or putative susceptibility variations in other genes has been undertaken by the BCFR and kConFab, as previously described.[29,33,34] The cohort includes 1508 (844 *BRCA1*, 658 *BRCA2*, 6 both *BRCA1* and *BRCA2*) female carriers from the BCFR, and 1233 (670 *BRCA1*, 561 *BRCA2*, 2 both *BRCA1* and *BRCA2*) female carriers from kConFab.

Outcome information. We have collected pathology reports for 74% of prospectively ascertained (incident) cases to date. The BCFR collected self-reported treatment data using a validated questionnaire addressing stage and the type of initial breast cancer treatments (surgery, radiation treatment, endocrine treatment and chemotherapy).[35,36] The Australian, Canadian and Utah sites of the BCFR regularly link to

population-based cancer registries to validate cancers reported during follow-up. Linkage to death registries in Australia and Canada has been used to update vital status and related information (date and cause of death) as well as the National Death Index for the USA-based sites.

What has it found?

To empirically evaluate the differences in breast cancer risk estimates from different constructs of family history, we compared standard ways of defining family history with estimates using full family history pedigrees (Table 4). For the more than 18 000 women unaffected at baseline, we compared family history as typically defined by cohort studies [any affected first-degree relative(s); yes/no] with that of the number of affected first-degree relatives, and with the more comprehensive family history measure of FRP based on the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA)[47] model (10-year predicted risk). We fitted

Table 3. Overview of measurements made for ProF-SC participants

Type of construct	When collected	Details
Family history/pedigree	Baseline Annual follow-up 10-year follow-up ProF-SC follow-up	Multi-generational pedigree completed at baseline, 10-year and ProF-SC interviews; additional updates collected when families contacted annually
Epidemiological questionnaires	Baseline 10-year follow-up ProF-SC follow-up Every 3 years[a]	Reproductive history; personal medical history; behavioural risk factors
Biospecimen collection	Baseline	Blood and/or buccal sample
BRCA1 and *BRCA2* genotyping	Baseline Follow-up as new family mutations are identified	Youngest affected family member tested, other family members tested if youngest affected had mutation
Outcome information	At diagnosis After identified through personal or family report Linkage with cancer (Australia, California, Canada) and national death registries	Pathology report; pathology material Treatment questionnaire Linkage with cancer registries and linkage with National Death Index

[a]kConFab FUP only.

Table 4. Associations of family history measures as predictors of age-adjusted breast cancer incidence for the sub-cohort of 17 403 women in ProF-SC who were unaffected at baseline

Family history measure	Number of events	Person-time (yrs)	Hazard ratio estimate[a]	95% confidence interval	χ_1^2
Breast cancer in 1st-degree relative(s) (yes/no, binary categorical)	1070	175186	1.45	(1.22, 1.73)	18
Number of 1st-degree relatives with breast cancer (ordered categorical)	1070	175186	1.39	(1.29, 1.49)	81
BOADICEA 10-year risk (per 1% change, continuous)	947	154885	1.12	(1.11, 1.14)	523
Breast cancer in 1st-degree relative(s) (yes/no, binary categorical)	947	154885	1.22	(1.01, 1.47)	4
BOADICEA 10-year risk (per 1% change, continuous)			1.12	(1.11, 1.13)	503
Number of 1st-degree relatives with breast cancer (ordered categorical)	947	154885	1.17	(1.08, 1.27)	14
BOADICEA 10-year risk (per 1% change, continuous)			1.12	(1.11, 1.13)	428

[a]Each row represents a separate age-adjusted model; rows 4 and 5 report models in which two constructs of family history are simultaneously fitted.

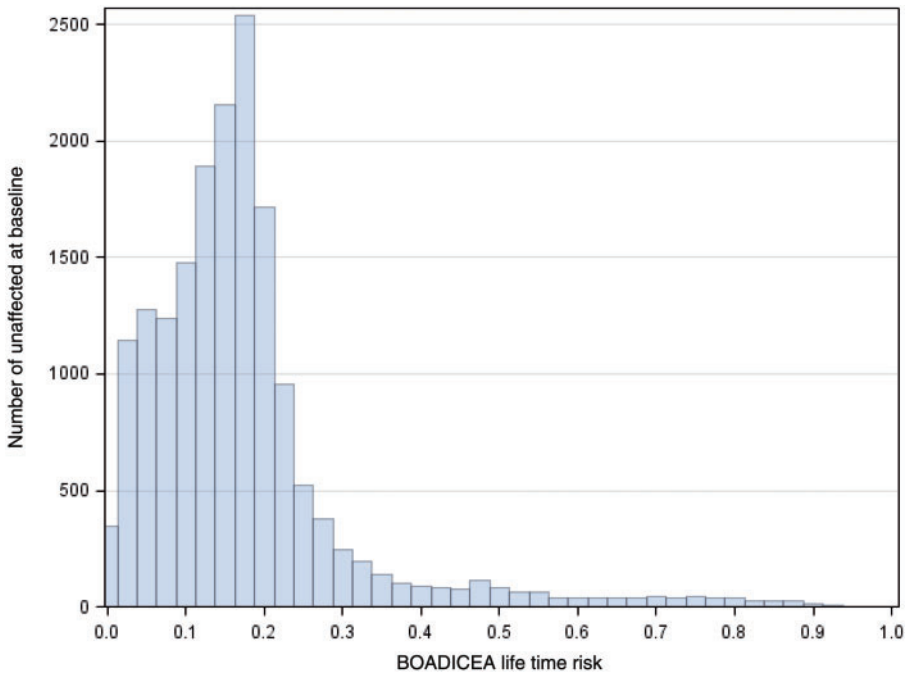


Figure 2. Remaining lifetime risks according to BOADICEA based on baseline characteristics, including family history, for the sub-cohort of 17 403 women in ProF-SC who were unaffected at baseline.

an age-adjusted proportional hazards model and found that all measures predicted risk. As we compare binary categorical, ordered categorical and continuous constructs of family history, the χ^2 statistics (based on the change in log likelihood which allows comparisons across constructs) showed that the BOADICEA score was clearly the best predictor of risk ($\chi^2 = 523$ for BOADICEA vs $\chi^2 = 18$ for ever/never family history) in that it captures more information on risk. After fitting the BOADICEA score, the strengths of the associations with the other two predictors were approximately halved, whereas the association with the BOADICEA score was virtually unchanged. Therefore, there is also scope for the BOADICEA model to be improved as a measure of FRP.

Figure 2 shows the predicted remaining lifetime risk based on BOADICEA for the sub-cohort of women unaffected at baseline. This illustrates the large range in risk and therefore why ProF-SC can be used to develop risk models which consider modification of risk by underlying FRP for women across the risk continuum.

These results suggest that the range of risk across ProF-SC participants is large, which will be essential for prospectively validating many of the retrospective findings from our families, including the importance of biomarkers that change over the life course (such as DNA repair phenotype, telomere length, oxidative stress and DNA methylation markers) in high-risk women.[37,38] We have also investigated environmental modifiers of risk for carriers of *BRCA1* and *BRCA2* mutations using retrospective data, and found a positive association with smoking[39] but no association with alcohol intake,[40] oral contraceptive use[41,42] or medical diagnostic radiation.[43] These studies suggest that it might be misleading to extrapolate findings about cancer risk factors from studying the general population to *BRCA1* and *BRCA2* mutation carriers or other sub-populations of women at increased familial/genetic risk. We are currently working to prospectively validate these finding using ProF-SC, with the aim of identifying modifiable risk factors for women across the full spectrum of risk.[44]

A key goal of ProF-SC is to validate and extend risk assessment models that predict breast cancer risk and are used in clinics and elsewhere. Most models, such as the BCRAT or Gail model,[45] have been developed for average risk populations and do not incorporate extensive data on family history of breast cancer or *BRCA1* and *BRCA2* mutation status. Exceptions are the International Breast Cancer Intervention Study (IBIS, or the Tyrer-Cuzick model)[46] and the BOADICEA.[47] Using data from the New York site of the BCFR, we have observed large discordances across models,[48,50] for example predictions from the IBIS model were generally closer to the observed number of events than predictions using BCRAT even for the women considered to be at average risk (e.g. those with no family history and no *BRCA1* or *BRCA2* mutation).[48] Using data from the Australian site of the BCFR, we have shown that BOADICEA is well calibrated and has good discrimination and accuracy at the individual level.[49]

What are the main strengths and weaknesses?

As shown above, a family-based cohort over-sampled for increased familial risk has many strengths including: (i) enrichment on outcome so that fewer individuals need be followed, and/or for a shorter time, in order to have the same statistical power as a cohort unselected for risk; (ii) ensuring the cohort covers a large range of risk; (iii) ability to identify environmental and genetic modifiers of risk for women at higher than average risk; and (iv) better retention through having multiple family contacts. From a practical perspective, however, family studies can be challenging because additional layers of protocol and privacy need to be considered. Care also has to be taken to ensure that information is not inadvertently passed to other family members. These issues, however, can be handled through study protocols and training, and we strongly believe that the benefits of a family cohort far outweigh its limitations and that this design should be considered when conducting aetiological research focused on environmental modifiers across the risk spectrum.

Can I use the data? Where can I find out more?

For information on how to collaborate with the ProF-SC cohort in making further use of the data and resources, and also with the BCFR, please see [http://www.bcfamilyregistry.org/]. For access to kConFab resources, see [www.kconfab.org].

Profile in a nutshell

- ProF-SC is a prospective cohort study of 31 640 women from 11 171 families ascertained through population-based and clinic-based sampling and who cover the full spectrum of familial risk. The study is designed to use pedigree and genetic data to predict the underlying familial risk profile (FRP) of each participant and determine if risk associations differ according to FRP, thereby informing targeted risk modification and prevention.
- Recruitment commenced in 1992 and families were from the USA, Canada and Australia. At baseline this cohort of women over the age of 18 years included 18 853 unaffected and 12 787 affected with breast cancer, of whom 2741 are *BRCA1* or *BRCA2* mutation carriers.
- The most recent follow-up was completed in 2014 and included updating family cancer histories from multiple sources; 5% have withdrawn and 14% were lost to follow-up. Over on average 9 years, we have identified 1093 and 1252 women with incident breast cancers among those at baseline who were unaffected and affected, respectively.
- Multi-generational cancer family histories were sought from all participants who were administered the same baseline epidemiology questionnaire. Blood samples have been collected from 83%, and another 2% have DNA available through buccal samples, of participants and used for extensive genetic characterisation.
- For collaboration with the ProF-SC cohort, see [http://www.bcfamilyregistry.org/] and [www.kconfab.org].

Funding

This work was supported by an award from NIH grants R01 CA159868 and UM1 CA164920 from the USA National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the BCFR, nor does mention of trade names, commercial products or organisations imply endorsement by the USA Government or the BCFR. K.A.P. is an Australian National Breast Cancer Foundation Fellow. kConFab is supported by grants from the National Breast Cancer Foundation, the National Health and Medical Research Council (NHMRC) and the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. J.L.H. is an NHMRC Senior Principal Research Fellow. I.L.A. holds the Anne and Max Tanenbaum Chair in Molecular Medicine at Mount Sinai Hospital and the University of Toronto.

Acknowledgements

We would like to gratefully acknowledge the team of the ProF-SC project coordinators, data managers and analysts and interviewers including: Yuyan Liao, Ann Johnston Cloud, Richard Buschbaum, Diane Levy, Ashley Thai, Melissa White, Gord Glendon, Nayana Weerasooriya, Tom Conner, Enid Satariano, Daisy Lubag, Connie Cady, Jocelyn Koo, Lisa Bealin, Nina Galpern, Prue Weideman, Charmaine Smith, Sandra Picken, Lucy Stanhope, Kelly Aujard, Gillian Dite and Amanda Torosidis. We would also like to acknowledge other co-investigators who have been instrumental in leading the laboratory analyses and/or setting up the original cohorts including: Regina Santella, Dee West, Ruby Senie, Norman Boyd, Melissa Southey and Graham Giles. We thank the kConFab research nurses and staff, the heads and staff of the Australian and New Zealand Family Cancer Clinics and the kConFab investigators responsible for establishing the resource and the follow-up study, including: Michael Friedlander, Sue-Anne MacLachlan, Roger Milne and Heather Thorne, and especially the many families enrolled in the BCFR and kConFab.

Conflict of interest: None declared.

References

1. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women

with breast cancer and 101 986 women without the disease. *Lancet* 2001;**358:**1389–99.
2. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet* 2008;**40:**17–22.
3. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. *Mol Oncol* 2010;**4:** 174–91.
4. Campa D, Barrdahl M, Tsilidis KK *et al*. A genome-wide "pleiotropy scan" does not identify new susceptibility loci for estrogen receptor negative breast cancer. *PloS One* 2014;**9:**e85955.
5. Garcia-Closas M, Couch FJ, Lindstrom S *et al*. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* 2013;**45:**39298.
6. Milne RL, Benitez J, Nevanlinna H *et al*. Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. *J Natl Cancer Inst* 2009;**101:** 1012–18.
7. Warren H, Dudbridge F, Fletcher O *et al*. 9q31.2-rs865686 as a susceptibility locus for estrogen receptor-positive breast cancer: evidence from the Breast Cancer Association Consortium. *Cancer Epidemiol Biomarkers Prev* 2012;**21):**1783–91.
8. Antoniou AC, Kuchenbaecker KB, Soucy P *et al*. Common variants at 12p11, 12q24, 9p21, 9q31.2 and in ZNF365 are associated with breast cancer risk for BRCA1 and/or BRCA2 mutation carriers. *Breast Cancer Res* 2012;**14:**R33.
9. Couch FJ, Gaudet MM, Antoniou AC *et al*. Common variants at the 19p13.1 and ZNF365 loci are associated with ER subtypes of breast cancer and ovarian cancer risk in BRCA1 and BRCA2 mutation carriers. *Cancer Epidemiol Biomarkers Prev* 2012;**21:** 645–57.
10. Michailidou K, Hall P, Gonzalez-Neira A *et al*. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;**45:**353–61.
11. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer. *Lancet* 1997;**350:**1047–59.
12. Hamajima N, Hirose K, Tajima K *et al*. Alcohol, tobacco and breast cancer – collaborative reanalysis of individual data from 53 epidemiological studies, including 58 515 women with breast cancer and 95 067 women without the disease. *Br J Cancer* 2002;**87:**1234–45.
13. Colditz GA, Willett WC, Hunter DJ *et al*. Family history, age, and risk of breast cancer. Prospective data from the Nurses' Health Study. *JAMA* 1993;**270:**338–43.
14. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. *Int J Cancer* 1997;**71:**800–09.
15. Houlston RS, McCarter E, Parbhoo S, Scurr JH, Slack J. Family history and risk of breast cancer. *J Med Genet* 1992;**29:**154–57.
16. Becher H, Schmidt S, Chang-Claude J. Reproductive factors and familial predisposition for breast cancer by age 50 years. A case-control-family study for assessing main effects and possible gene-environment interaction. *Int J Epidemiol* 2003;**32:**38–48.
17. Delgado-Cruzata L, Wu HC, Liao Y, Santella RM, Terry MB. Differences in DNA methylation by extent of breast cancer family history in unaffected women. *Epigenetics* 2014;**9:**243–48.
18. Hopper JL, Carlin JB. Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. *Am J Epidemiol* 1992;**136:** 1138–47.
19. Hopper JL. Disease-specific prospective family study cohorts enriched for familial risk. *Epidemiol Perspct Innov* 2011;**8:**2.
20. Sellers TA, Anderson VE, Potter JD *et al*. Epidemiologic and genetic follow-up study of 544 Minnesota breast cancer families: design and methods. *Genet Epidemiol* 1995;**12:**417–29.
21. Sellers TA, King RA, Cerhan JR *et al*. Fifty-year follow-up of cancer incidence in a historical cohort of Minnesota breast cancer families. *Cancer Epidemiol Biomarkers Prev* 1999;**8:**1051–57.
22. Cerhan JR, Grabrick DM, Vierkant RA *et al*. Interaction of adolescent anthropometric characteristics and family history on breast cancer risk in a Historical Cohort Study of 426 families (USA). *Cancer Causes Control* 2004;**15:**1–9.
23. Vachon CM, Cerhan JR, Vierkant RA, Sellers TA. Investigation of an interaction of alcohol intake and family history on breast cancer risk in the Minnesota Breast Cancer Family Study. *Cancer* 2001;**92:**240–48.
24. Xu Z, Bolick SC, DeRoo LA, Weinberg CR, Sandler DP, Taylor JA. Epigenome-wide association study of breast cancer using prospectively collected sister study samples. *J Natl Cancer Inst* 2013;**105:**694–700.
25. Price MA, Butow PN, Charles M *et al*. Predictors of breast cancer screening behavior in women with a strong family history of the disease. *BreastCancer Res Treat* 2010;**124:**509–19.
26. Antill YC, Reynolds J, Young MA *et al*. Screening behavior in women at increased familial risk for breast cancer. *Fam Cancer* 2006;**5:**359–68.
27. Collins IM, Milne RL, Weideman PC *et al*. Preventing breast and ovarian cancers in high-risk BRCA1 and BRCA2 mutation carriers. *Med J Aust* 2013;**199:**680–83.
28. Kiely BE, Jenkins MA, McKinley JM *et al*. Contralateral risk-reducing mastectomy in BRCA1 and BRCA2 mutation carriers and other high-risk women in the Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab). *Breast Cancer Res Treat* 2010;**120:**715–23.
29. John EM, Hopper JL, Beck JC *et al*. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res* 2004;**6:**R375–89.
30. Osborne RH, Hopper JL, Kirk JA, Chenevix-Trench G, Thorne HJ, Sambrook JF. kConFab: a research resource of Australasian breast cancer families. Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer. *Med J Aust* 2000;**172:**463–64.
31. Phillips KA, Butow PN, Stewart AE *et al*. Predictors of participation in clinical and psychosocial follow-up of the kConFab breast cancer family cohort. *Fam Cancer* 2005;**4:**105–13.
32. Tehranifar P, Wu HC, Shriver T, Cloud AJ, Terry MB. Validation of Family Cancer History Data in High-Risk Families: The Influence of Cancer Site, Ethnicity, Kinship Degree, and Multiple Family Reporters. *Am J Epidemiol* 2015; **181:**204–12
33. Andrulis IL, Anton-Culver H, Beck J *et al*. Comparison of DNA- and RNA-based methods for detection of truncating BRCA1 mutations. *Hum Mutat* 2002;**20:**65–73.

34. Neuhausen SL, Ozcelik H, Southey MC *et al*. BRCA1 and BRCA2 mutation carriers in the Breast Cancer Family Registry: an open resource for collaborative research. *Breast Cancer Res Treat* 2009;**116:**379–86.
35. Barisic AG, Weerasooriya N, Andrulis IL, Knight JA. Agreement between self-reported breast cancer information and medical records among women in the Ontario Familial Breast Cancer Registry. *J Cancer Epidemio* 2012;**3:**10804.
36. Phillips KA, Milne RL, Buys S *et al*. Agreement between self-reported breast cancer treatment and medical records in a population-based Breast Cancer Family Registry. *J Clin Oncol* 2005;**23:** 4679–86.
37. Delgado-Cruzata L, Wu HC, Perrin M *et al*. Global DNA methylation levels in white blood cell DNA from sisters discordant for breast cancer from the New York site of the Breast Cancer Family Registry. *Epigenetics* 2012;**7:**868–74.
38. Wu HC, Wang Q, Yang HI, Tsai WY, Chen CJ, Santella RM. Global DNA methylation levels in white blood cells as a biomarker for hepatocellular carcinoma risk: a nested case-control study. *Carcinogenesis* 2012;**33:**1340–45.
39. Smoking and risk of breast cancer in carriers of mutations in BRCA1 or BRCA2 aged less than 50 years. *Breast Cancer Res Treat* 2008;**109:**67–75.
40. McGuire V, John EM, Felberg A *et al*. No increased risk of breast cancer associated with alcohol consumption among carriers of BRCA1 and BRCA2 mutations ages <50 years. *Cancer Epodemiol Biomarkers Prev* 2006;**15:**1565–67.
41. Milne RL, Knight JA, John EM *et al*. Oral contraceptive use and risk of early-onset breast cancer in carriers and noncarriers of BRCA1 and BRCA2 mutations. *Cancer Epodemiol Biomarkers Prev* 2005;**14:**350–56.
42. Haile RW, Thomas DC, McGuire V *et al*. BRCA1 and BRCA2 mutation carriers, oral contraceptive use, and breast cancer before age 50. *Cancer Epodemiol Biomarkers Prev* 2006;**15:** 1863–70.
43. John EM, Phipps AI, Knight JA *et al*. Medical radiation exposure and breast cancer risk: findings from the Breast Cancer Family Registry. *Int J Cancer* 2007;**121:**386–94.
44. Cloud AJ, Thai A, Liao Y, Terry MB. The impact of cancer prevention guideline adherence on overall mortality in a high-risk cohort of women from the New York site of the Breast Cancer Family Registry. *Breast Cancer Res Treat* 2015;**149:**537–46.
45. Gail MH, Greene MH. Gail model and breast cancer. *Lancet* 2000;**355:**1017.
46. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004; **23:**1111–30.
47. Antoniou AC, Pharoah PP, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer* 2004;**91:**1580–90.
48. Quante AS, Whittemore AS, Shriver T, Strauch K, Terry MB. Breast cancer risk assessment across the risk continuum: genetic and nongenetic risk factors contributing to differential model performance. *Breast Cancer Res* 2012;**14:**R144.
49. MacInnis RJ, Bickerstaffe A, Apicella C *et al*. Prospective validation of the breast cancer risk prediction model BOADICEA and a batch-mode version BOADICEACentre. *Br J Cancer* 2013; **109:**1296–301.
50. Quante AS, Whittemore AS, Shriver T, Hopper JL, Strauch K, Terry MB. Practical problems with clinical guidelines for breast cancer prevention based on remaining lifetime risk. *J Natl Cancer Inst* 2015;doi: 10.1093/jnci/djv124.