Cohort Profile

Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA)

Marike Gabrielson,[1]* Mikael Eriksson,[1] Mattias Hammarström,[1] Signe Borgquist,[2] Karin Leifland,[3] Kamila Czene[1] and Per Hall[1]

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [2]Division of Oncology and Pathology, Department of Clinical Sciences, Lund University, Lund, Sweden and [3]Department of Medical Imaging, Stockholm South General Hospital, Stockholm, Sweden

*Corresponding author. Department of Medical Epidemiology and Biostatistics, Nobels väg 12A, Karolinska Institutet, SE-171 77, Stockholm, Sweden. E-mail: marike.gabrielson@ki.se
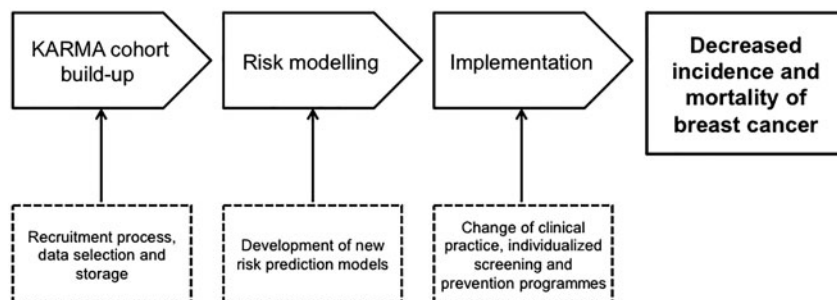
Why was the cohort set up?

The overarching goal of the KARMA (KARolinska MAmmography Project for Risk Prediction of Breast Cancer) study is to reduce the incidence and mortality of breast cancer by focusing on individualized prevention and screening (Figure 1). The primary objectives of the KARMA study are: (i) to determine the roles and interaction of mammographic density, genetic determinants and lifestyle in the risk of breast cancer through a large-scale population cohort study; (ii) to build a comprehensive risk prediction model including all known and established risk factors for breast cancer; (iii) to discover new biomarkers for early detection and prevention of breast cancer using genomic and proteomic approaches; and (iv) to establish a comprehensive database of information linked to biological specimens, which will become a resource for further scientific studies.

The number of women diagnosed with breast cancer is increasing globally. In Europe more than 360 000 women were diagnosed in 2012 and 90 000 died from the disease.[1] In Sweden, one woman in eight will be diagnosed with breast cancer during her lifetime, corresponding to nearly one woman being diagnosed with breast cancer every hour.[2]

To improve identification of individual breast cancer risk, a variety of breast cancer risk prediction models have been developed.[3,4] The majority of the established prediction models focuses on family history of breast cancer, but newer prediction models includes additional risk factors.[3,5–7] The Tyrer-Cuzick risk prediction model is based on classical phenotypic factors including age, family history of breast cancer, age at menopause and menarche, and parity.[8] Although this model is comprehensive, it currently does not include mammographic density or genetic determinants for risk prediction.

Following age, mammographic density is considered one of the strongest risk factors for breast cancer.[9–11] Mammographic density reflects the composition of fibro-glandular and fat tissue of the breast. Radiologically dense tissues, such as connective and epithelial tissue, appear light on the mammogram, whereas radiologically fat lucent tissue appears dark.[12] Women with high mammographic density have a 4–6-fold increased breast cancer risk as compared with women with low or no density.[10,13] A number of other breast cancer risk factors have been well described, notably factors related to hormonal exposure and reproductive patterns, lifestyle factors and genetic determinants.[12,14–18]

Figure 1. Overview of the KARMA (KARolinska MAmmography Project for Risk Prediction of Breast Cancer) study. The KARMA study aims to study and determine the roles and interactions of mammographic density, genetic determinants and lifestyle in the risk of breast cancer to build a comprehensive risk prediction model including all known and established risk factors for breast cancer. The overarching goal of the KARMA study is to reduce the incidence and mortality of breast cancer by focusing on individualized prevention and screening in the clinic as well as participation in community-based prevention programmes using the KARMA risk prediction model.

Who is in the cohort?

The KARMA project is a prospective cohort study comprising women attending mammography screening or clinical mammography at four hospitals in Sweden. Since 1994, all women in Sweden, aged 40–74 years, are invited every 18–24 months to the national mammography screening programme. About 80% of eligible women attend screening regularly, and attendance is among the highest recorded in the world.[19,20] All women invited for screening from January 2011 to March 2013, at the four hospitals, were invited to participate in the study. Additionally, women who had a clinical mammography at any of the participating mammography screenings centres during this time were also invited. During the recruitment period, a total of 210 233 women were invited to participate in the KARMA study and 70 877 women (34%) joined the study (Figure 2).

After performing their scheduled mammogram, participants entered the KARMA cohort through signing the informed consent at the KARMA Test Centre. The participants were recruited from the Stockholm South General Hospital (50%), Helsingborg Hospital (27%), Skåne University Hospital, Lund, (14%) and Landskrona Hospital (9%). The majority of the participants were recruited during year 2012 (54%). The mean ages of invited and recruited women were 53.7 [standard deviation (SD) 9.9] years and 54.6 (SD 10.0) years, respectively. The ages of the participants ranged from 21 to 95 years at study entry. Most participants were aged between 40 and 49 years when recruited (35%) (Table 1 and Supplementary Figure S1A, available as Supplementary data at *IJE* online), and approximately 53% of the recruited women were postmenopausal. The proportion of women agreeing to participate was highest in the age group 65–69 years (39% recruited) and lowest among women in the age group 40–44 years (30% recruited) (Supplementary Figure S1B, available as Supplementary data at *IJE* online).
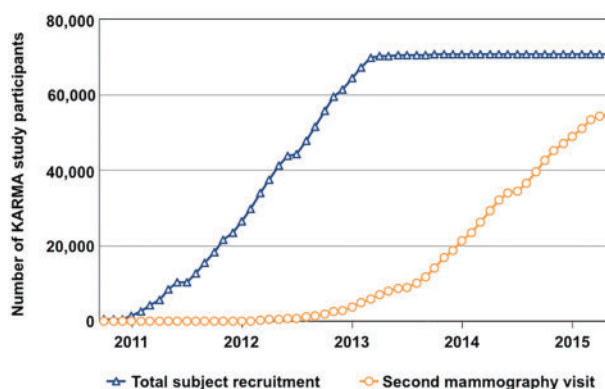
Figure 2. Subject recruitment and follow-up flow chart of the KARMA study. Since study start, 70 877 women have been recruited to KARMA (triangle). At recruitment, the participants answered a comprehensive questionnaire including information on known breast cancer risk factors, and biospecimens and mammograms were collected. Follow-up mammograms are continuously collected each time the participants attend their screening mammography every 18–24 months (circle). All participants are continuously matched to a number of Swedish health care registers including the National Swedish Cancer and Prescription Registers.

The informed consent is inclusive and allows research on both risk and prognosis of breast cancer. It includes collection and storage of questionnaire data, mammograms, matched health care register data, and biological samples. The consent also allows KARMA to share data with external collaborators and commercial companies. Information regarding the consent can be found at: [karmastudy.org]. The ethical board at the Karolinska Institutet approved the study.

How often have they been followed up?

Collection of mammograms and biospecimens

Mammograms are collected continuously each time a KARMA participant undergoes mammography, both screening and clinical mammography, at the participating hospitals. Collection of mammograms is continuing and

Table 1. Potential breast cancer risk factors within the KARMA study sample *(n = 70 773[a])*

Risk factor	Results
Age (years), mean (SD)[b]	54.6 (10.0)
Age groups, % (*n*)	
20 to 29	0 (21)
30 to 39	1.3 (942)
40 to 49	35.3 (24980)
50 to 59	28.6 (20257)
60 to 69	27.4 (19369)
70 to 79	7.2 (5112)
80 to 89	0.1 (91)
90 to 99	0 (1)
BMI (kg/m^2), mean (SD)[b]	25.3 (4.2)
Missing, % (*n*)	7.1 (5016)
Age at menarche (years), mean (SD)[b]	13.1 (1.5)
Missing, % (*n*)	9.6 (6798)
Parity, % (*n*)[b]	
0	11.7 (8299)
1	13.6 (9597)
2	44.0 (31116)
≥ 3	22.7 (16078)
Missing, % (*n*)	8.0 (5683)
Age at first childbirth (years), mean (SD)[b]	27.1 (5.3)
Missing, % (*n*)	0.0 (25)
Menopausal status, % (*n*)[b]	
Premenopausal	36.8 (26073)
Perimenopausal/unknown	3.3 (2324)
Postmenopausal	52.5 (37140)
Missing, % (*n*)	7.4 (5236)
Age at menopause (years), mean (SD)[b]	49.8 (5.3)
Missing, % (*n*)	49.7 (18454)
Ever oral contraceptive use, % (*n*)[b]	77.3 (54715)
Missing, % (*n*)	9.0 (6401)
HRT use, % (*n*) [b]	
Never	68.2 (48 271)
Previous user	14.1 (9 994)
Current user	3.4 (2 407)
Missing, % (*n*)	14.3 (10 101)
Benign breast disorder, % (*n*)[b]	21.1 (14 926)
Missing, % (*n*)	8.7 (6 166)
Family history of breast cancer, % (*n*)[b]	11.9 (8 456)
Missing, % (*n*)	10.8 (7 609)
Family history of ovarian cancer, % (*n*)[b]	3.3 (2 352)
Missing, % (*n*)	12.1 (8 555)
Smoking, pack/years	5.6 (8.8)
Missing, % (*n*)	8.3 (5 872)
Alcohol, g/week	50.2 (60.7)
Missing, % (*n*)	8.8 (6 205)

BMI, body mass index; HRT, hormone replacement therapy.

[a]Women who actively requested to be excluded from the KARMA study are not included in the table.

[b]Risk factor included in the Tyrer-Cuzick risk prediction model.

mammograms from approximately 200 examinations are stored daily. As of June 2016, the average number of examinations per woman was 2.5. A total of 6546 (9%) women had one examination, 25 637 (36%) two examinations, 32 650 (46%) had three examinations and 5952 (8%) women had four examinations.

Baseline blood samples were collected from the vast majority of the KARMA participants (69 440; 98%). During the initial recruitment period, a total of 17 599 (25%) participants had another round of screening and chose to answer a follow-up questionnaire and donate blood.

Questionnaire data and data from national registers

All study participants responded to the KARMA online questionnaire at baseline, and of these 68 426 (97%) participants completed the questionnaire. A second follow-up of the web-based questionnaire started in May 2016 and is currently ongoing. To further enrich the data of the cohort, we have used national registers available in Sweden.[21] Matching to the registers is done using Swedish Personal Identity Numbers, a 10-digit number that is unique for every individual and used in all health-based registers.[22] All KARMA participants are biannually matched to the registers. A summary of all registers included in the KARMA study is shown in Table 2.

All cancers and medical diagnoses are classified according to the International Classification of Disease system.[23] Information regarding cancer diagnoses, tumour characteristics and therapy is collected from the Register for Breast Cancer (Information Network for Cancer treatment, INCA), the Stockholm Regional Register for Breast Cancer and the Swedish National Cancer Register. The National Cancer Register includes all types of cancers, and an estimated 99% of all tumours in the National Cancer Register have also been morphologically verified.[24]

Individual data on dispensed drugs are obtained from the Swedish Prescribed Drug Register according to the Anatomical Therapeutic Chemical (ATC) classification system. The Swedish Prescribed Drug Register includes information on dates of all prescribed and dispensed drugs and is reported to be complete, with less than 0.3% of entries having missing patient identity data.[25]

The Swedish In- and Outpatient Registers provide information on main and secondary diagnoses at hospital admission, including diagnostic and surgical codes, dates of hospitalization and information on outpatient visits. The National Cause of Death Register covers date of death, histopathological verification and underlying and contributory causes of death.

Lost to follow-up and deceased

During follow-up, a smaller fraction of participants has actively requested to be excluded from the KARMA study

Table 2. Summary of the Swedish medical quality registers used for data collection for all KARMA study participants

| Register	Year of initiation	Extent	Example of information extracted from register	Reference
Stockholm Regional Register for Breast Cancer	1977	Regional	Merged with Swedish National Cancer Register and INCA in 2007	[33]
Information Network for Cancer treatment (INCA Register for Breast Cancer)	2007–	National	Pre-surgical diagnostics, tumour characteristics (grade, differentiation, receptor status, HER2 status), treatment, local and distant recurrences	[34]
Swedish National Cancer Register	1958–	National	Includes all types of cancer diagnoses, date of diagnosis, mode of diagnosis, location of tumour, clinical stage of tumour, and histological type	[35]
Swedish Prescribed Drug Register	2005–	National	Anatomic Therapeutic Chemical (ATC) codes, brand name, prescribed dose.	[36]
National Inpatient Register	1964–	National	Main and secondary diagnoses for hospitalization, dates of hospitalization, surgical codes for both inpatient and day surgery	[37]
National Outpatient Register	2001–	National	Main and secondary diagnoses for the visit, surgical codes for day surgery	[37]
National Cause-of-Death Register	1952–	National	Date of death, underlying and contributory causes of death, information from autopsies, histological verification of disease	[38]

[99 women (0.1%; mean age 58.4 years)]. Another 375 participants (0.5%; mean age 63.7 years) have died since entering the study.

What has been measured?

All KARMA cohort data and IT infrastructure are managed through the KARMA Research Platform: [karmastudy.org]. The platform currently holds variables and information on lifestyle factors through questionnaires, mammographic density from mammograms, genetic variants through genotyping, extensive information from health care registers and research results.

The KARMA Research Platform provides researchers with a versatile online inventory to browse the content of the cohort. The web interface provides visualization capability with distribution and scatter plots for any of the variables in the system. This gives immediate basic insights into the data of interest before detailed data analysis. An example of accessible data from the KARMA Research Platform web interface is visualized in Figure 3.

Upon request, researchers can get access to non-identifiable data. The KARMA Research Platform was launched in December 2011 and has, as of 1 November 1
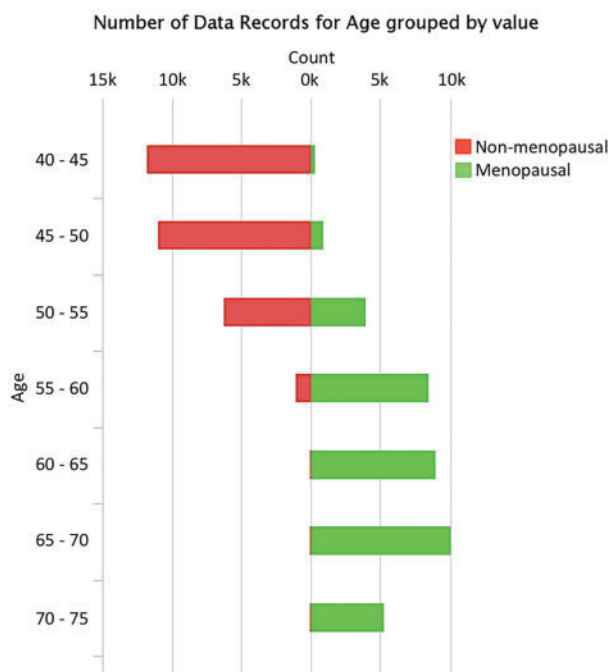
Figure 3. Self-reported menopausal status by age in 68 002 KARMA participants. This is an example of data visualizations using the KARMA Research Platform web tool. The web interface gives immediate basic insights into the data of interest before detailed data analysis.

2015, facilitated more than 1000 data requests of which 25% were from external research groups.

Mammograms and mammographic density assessment

The KARMA project continuously collects mammograms from cranial-caudal and mediolateral oblique views by full-field digital mammography using five different models from three manufactures (General Electrics Medical Systems, Philips Healthcare, Sectra Imtec AB and FUJI). For most participants, eight mammograms from each examination have been collected. The total number of stored raw and processed mammograms was $\sim$ 1.8 million as of June 2016.

Breast size and mammographic density can be measured two-dimensionally as an area and three-dimensionally as a volume. The automated Volpara system estimates the breast volume and the volume of dense tissue in the breast based on raw mammograms. Volpara does not measure the true volume, but takes the X-ray attenuation into account to compute the estimated volume of dense tissue in the breast at each image pixel, and creates a 3D density map.

The weakness of Volpara is that raw mammograms are needed. In the clinical setting, normally only processed mammograms are stored. We have therefore developed the STRATUS program that analyses both raw and processed mammograms and estimates the breast and density areas based on mammographic textures. Fifteen threshold techniques are used to identify the textures, and each pattern segment is analysed for several statistical features including pattern area, circumference, intensity, positioning, relation to other areas, and shape. This quantified texture structure of the breast is compared with a reference library of matching breast texture-density level pairs. The reference library was created using the penalized machine learning method. Figure 4 illustrates processed mammograms of breasts with low or high mammographic density measured by STRATUS. The estimated mammographic density of all KARMA participants measured by STRATUS, as absolute dense area (cm^2) and percent density, is shown in Supplementary Table S1. Mammographic volumetric density has been measured on all raw mammograms in the KARMA cohort using Volpara. Analysis by the STRATUS program is ongoing and all images will finally be measured.

Baseline questionnaire

All study participants responded to the KARMA online questionnaire upon study entry; of these, 68 426 (97%) participants completed the questionnaire. The comprehensive questionnaire included more than 250 questions covering breast cancer-related topics, i.e. reproductive health, alcohol and tobacco use, medical history, menopausal status, use of oral contraceptives, hormone replacement therapy and other medications, family history of cancer, quality of life, physical activity and diet. Potential breast cancer risk factors are summarized in Table 1. Body mass index was calculated based on self-reported height and weight. During the KARMA study recruitment period,
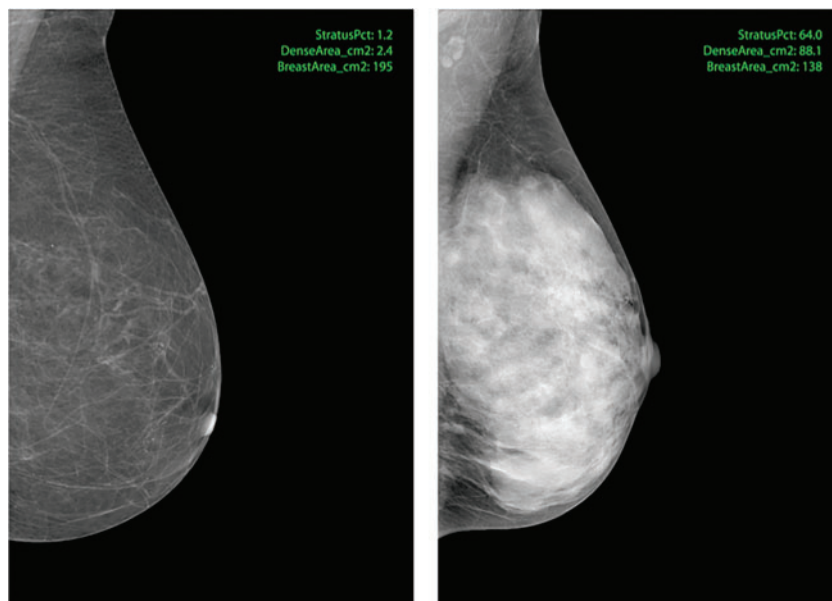
Figure 4. Examples of processed mammograms from the KARMA cohort. Percent density (proportion of the entire breast that is dense) was measured in the left (1.2%) and right (63.8%) image using STRATUS.

25% of the participants returned for a new screening mammogram. At that visit, participants answered a comprehensive follow-up questionnaire. The follow-up questionnaire includes questions on reproductive health, hormone replacement therapy, other medications, alcohol and tobacco use, diseases and treatments, heredity, quality of life, physical activity and diet.

Biospecimens and blood pressure

All KARMA participants donated 24 ml blood upon study entry, unless phlebotomy was not possible. In total, EDTA blood samples from 69 440 (98%) study participants were collected. All blood samples were handled in accordance with a strict 30-h cold-chain protocol. The blood samples were processed in the Karolinska Institutet high-throughput biobank and aliquoted into 16 plasma aliquots, one aliquot of extracted DNA and two aliquots of whole blood for back-up. Approximately 15 000 participants have been genotyped within the Collaborative Oncological Gene-environment Study [COGS] project: [http://www.nature.com/icogs], in collaboration with the Breast Cancer Association Consortium. During the KARMA study recruitment period, 25% of the participants donated a second blood sample when returning for a new screening mammogram. Upon study entry, baseline blood pressure was measured by KARMA personnel for 70 306 (99%) participants.

Breast cancer risk assessment

The individual breast cancer risk was calculated for each participant using the Tyrer-Cuzick risk prediction model. The Tyrer-Cuzick model is based on lifestyle factors and reproductive and family history,[8] and risk was calculated using information from the baseline questionnaire. The model generates the individual risk of being diagnosed with breast cancer within 5 and 10 years, in addition to lifetime risk. Breast cancer risk factors included in the Tyrer-Cuzick risk prediction model are presented in Table 1.

Stratification of KARMA participants using mammographic density and estimated Tyrer-Cuzick lifetime breast cancer risk enables identification of the individual risk of breast cancer and also gives the reason for the individual risk estimate. Our aim is to amalgamate lifestyle factors, mammographic density and genetic information into one comprehensive risk model (Eriksson *et al.*, submitted for publication).

What has the KARMA study found? Key findings and publications

A complete list of publications from the KARMA study is available at: [karmastudy.org]. Key findings across important topics are presented below.

Identification of novel breast cancer risk loci

KARMA is part of the Breast Cancer Association Consortium. In the European Commission-funded COGS project, > 29 000 single nucleotide polymorphisms (SNPs) were genotyped in more than 45 000 cases and 41 000 controls of European ancestry.[16] In all, 41 new breast cancer susceptibility loci were identified and it was suggested that more than 1000 additional loci are influencing the risk of breast cancer. Using the currently known loci identifies a genetic profile for which 5% of the female population has a risk that is $\sim$ 2.3-fold higher than the population average and for which 1% of the population has a risk that is $\sim$ 3-fold higher. This is the largest genetic association study in cancer so far, and the results more than double the number of known susceptibility loci for breast cancer.

Novel genetic markers of mammographic density

To explore the genetic variations of mammographic density, we performed a large-scale genetic association study including 8419 women of whom 4025 were KARMA participants.[26] Genome-wide significant associations with mammographic density were observed for two variants at 6q25.1, in the *TAB2* gene and in the *CCDC170/ESR1* region, both of which are known breast cancer susceptibility loci. Both regions have been implicated in estrogen receptor signalling, with TAB2 being a potential regulator of tamoxifen response. These findings underscore the importance of 6q25.1 as a susceptibility region and provide more insight into the mechanisms through which mammographic density influences breast cancer risk. We also validated the previously known associations between *ZNF365* and *AREG* with mammographic density, and *MLK1* and *NTN4* with non-dense volume.[26,27]

Heritability of mammographic density

Using the Personal Identity Number and by linking the KARMA cohort to the national Multi-Generation Register for retrieval of sister-relations, we have studied the heritability of volumetric mammographic density in 955 sibling-pairs in KARMA. The Multi-Generation Register enables linking the nuclear family of all Swedes born after 1931. Our results confirmed the high heritability of mammographic density.[27] These data support the notion that mammographic density is a risk factor under strong genetic influence that may partially explain the familial aggregation of breast cancer.

Lifestyle influences the mammographic density

We used self-reported alcohol consumption and volumetric mammographic density in 53 060 participants to show that

higher alcohol consumption was associated with increased mammographic density.[28] After stratifying women into background risk of breast cancer using the Tyrer-Cuzick score, we found that the background risk modified the effect of alcohol on mammographic density. Only those 15% of women at highest risk of breast cancer had an increased mammographic density that was associated with alcohol intake.

In contrast, we have demonstrated that physical activity may decrease breast cancer risk through reducing mammographic density, and that the physical activity needed to reduce mammographic density may depend on background risk of breast cancer.[29] KARMA has also been used to evaluate the effect of statin use on volumetric density in a study including 41 102 participants. Overall, we observed no effect of statin use on mammographic density in terms of absolute dense volume.[30]

What are the main strengths and weaknesses?

The KARMA study is one of the world's best-characterized breast cancer cohorts aiming to reduce the incidence and mortality of breast cancer through translational research focusing on breast cancer prevention (Figure 1). The KARMA study has several strengths worth pointing out. The project takes advantage of the tax-funded health care system in Sweden and the initial recruitment phase was therefore comparatively cheap; further, women were invited when conducting either screening or clinical mammography at any of four large mammography screening units in Sweden, and participants attending the mammography screening programme continue to visit the four units where mammograms are stored at each new visit. The design of the KARMA study thus allows easy and continuous collection of follow-up mammograms of all participants attending screening mammography.

The web-based questionnaire, completed by 97% of the participants, was developed over the past decade and is very comprehensive. The diet and physical activity parts of the questionnaire have previously been validated.[31,32] Since the screening interval in Sweden is 18–24 months and the KARMA recruitment period was 36 months, quite a number of KARMA participants were invited for a new round of screening. In all, 25% of the participants chose to answer the follow-up questionnaire and donate blood. During the spring of 2016, all participants were invited to again answer the follow-up questionnaire.

For 98% of the KARMA cohort, plasma and DNA have been extracted. Currently ∼ 15 000 women have had their DNA genotyped and ∼ 3 000 have had plasma analysed.

Both DNA and plasma are of high quality. The aim is to genotype the entire cohort. A logistical strength is the use of the Karolinska Institutet Biobank, where blood and plasma for all participants are stored and handled, including withdrawal of samples. The Biobank is completely robotized, thus ensuring swift and correct handling of samples.

A further strength of the KARMA study is the useage of the unique Personal Identification Number given to all Swedish citizens at birth, which enables easy and regular linkage to Swedish, government-administered health care registries, including the unique Multi-Generation Register.[22] Matching to registers generates information on a number of relevant factors such as incident cancers, date and cause of death, in- and outpatient visits, drug prescriptions, sister-relations, etc.

All KARMA cohort data and IT infrastructure are managed through the KARMA Research Platform: [karmastudy.org]. The platform collects and holds data on consents, mammograms, biosamples, genotypes, health care register data, questionnaire data and research results. When data are entered into the platform, they are automatically linked to all other data on study-person level using coded IDs. The platform is governed by automated cleaning and recoding rules and creates readily useable datasets with minimal extra coding needed by the scientist. The platform also has a web interface for scientists that provides visualization capabilities with distribution and scatter plots for any of the variables in the system. This gives immediate basic insights into the data of interest before detailed data analysis.

KARMA has some weaknesses worth noting. We have a two-stage selection bias. First, only women attending mammography screening or a clinical mammogram could be members of the cohort. Second, women visiting the mammography units actively chose to participate in the KARMA cohort. As a consequence of the two-stage selection process, KARMA participants are more highly educated and more likely to have a family history of breast cancer than the average Swedish population.

Another important concern is that we had difficulty in (Color online) obtaining tumour samples for incident breast cancers. This is a problem currently addressed and solved for parts of the cohort. Furthermore, the National Swedish Cancer Register is almost complete but there is a 2-year delay in releasing new versions of the register. In contrast, the INCA breast cancer register is continually updated with only approximately 3 months of delay.

Can I get hold of the data? Where can I find out more?

As stated, a substantial amount of time and resources has been spent in creating the KARMA Research Platform:

[karmastudy.org]. Any person interested in learning about the KARMA cohort can browse the website. When an application has been successfully filed, users are given a password and user name. This enables access to aggregated data and some interactive activities such as creating simple frequency tables. Browsing does not enable identification of individuals.

Researchers interested in getting access to KARMA data and in-depth analyses can apply for access to phenotype and/or biomaterial. This is done by submitting a KARMA Material and Data Transfer Agreement form to: [karmastudy@ki.se]. It should be underlined that the applicants may only use the material/data for the purposes listed in the description of the research project. Another prerequisite is that all results generated using KARMA data must, after an appropriate time for publication is allowed, be added to the database for other researchers to use. Requests for phenotype data will be decided directly by the Principal Investigator (PI) and the co-PIs. Requests for biomaterial will further involve the KARMA Data Access Committee.

Profile in a nutshell

- KARMA is a cohort created with the overarching goal to reduce the incidence and mortality of breast cancer by focusing on individualized prevention and screening.
- A total of 70 877 women (21–95 years of age) were recruited between 2011 and 2013 when attending their mammographic screening programme or clinical mammography at four hospitals in Sweden.
- The dataset comprises a wide range of questionnaire-based data, mammograms, Tyrer-Cuzick risk prediction, biological samples, genetic information and linkage to several nationwide health care registers.
- New mammograms are continuously collected each time a participant is examined. Register-based information is updated continuously every 6 months, and data collection by a follow-up questionnaire was initiated in the spring of 2016.
- More information about KARMA, including details of the questionnaires, is found browsing the KARMA Research Platform: [karmastudy.org]. Application to access and use the datasets can be completed at this platform.

Supplementary Data

Supplementary data are available at *IJE* online.

References

1. Ferlay J, Soerjomataram I, Dikshit R *et al*. (2014). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;**136**: E359–86.
2. The National Board of Health and Welfare, Sweden (SE) (Socialstyrelsen). Cancer database 2015 [Internet]; Mars 2015 [cited Mars 15 2015]. http://www.socialstyrelsen.se/statistik/statistikdatabas/cancer.
3. Antoniou AC, Easton DF. Risk prediction models for familial breast cancer. *Future Oncol* 2006;**2**:257–74.
4. Jacobi CE, de Bock GH, Siegerink B, van Asperen CJ. Differences and similarities in breast cancer risk assessment models in clinical practice: which model to choose?. *Breast Cancer Res Treat* 2009;**115**:381–90.
5. Engel C, Fischer C. Breast cancer risks and risk prediction models. *Breast Care (Basel)* 2015;**10**:7–12.
6. Fischer C, Kuchenbacker K, Engel C *et al*. Evaluating the performance of the breast cancer genetic risk models BOADICEA, IBIS, BRCAPRO and Claus for predicting BRCA1/2 mutation carrier probabilities: a study based on 7352 families from the German Hereditary Breast and Ovarian Cancer Consortium. *J Med Genet* 2013;**50**:360–67.
7. Gail MH, Brinton LA, Byar DP *et al*. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;**81**: 1879–86.
8. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004;**23**:1111–30.
9. Boyd NF, Byng JW, Jong RA *et al*. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J Natl Cancer Inst* 1995;**87**:670–75.
10. Boyd NF, Guo H, Martin LJ *et al*. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;**356**: 227–36.
11. Byrne C, Schairer C, Wolfe J *et al*. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *J Natl Cancer Inst* 1995;**87**:1622–29.
12. Boyd NF, Lockwood GA, Martin LJ *et al*. Mammographic densities and breast cancer risk. *Breast Dis* 1998;**10**:113–26.

13. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006;**15**:1159–69.

14. Easton DF, Pooley KA, Dunning AM *et al*. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;**447**:1087–93.

15. Key TJ, Appleby PN, Reeves GK *et al* Endogenous Hormones and Breast Cancer Collaborative Group. Sex hormones and risk of breast cancer in premenopausal women: a collaborative re-analysis of individual participant data from seven prospective studies. *Lancet Oncol* 2013;**14**:1009–19.

16. 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, 61e1–2.

17. Russo J, Moral R, Balogh GA, Mailo D, Russo IH. The protective role of pregnancy in breast cancer. *Breast Cancer Res* 2005;**7**:131–42.

18. Schonfeld SJ, Pfeiffer RM, Lacey JV Jr *et al*. Hormone-related risk factors and postmenopausal breast cancer among nulliparous versus parous women: An aggregated study. *Am J Epidemiol* 2011;**173**:509–17.

19. Olsson S, Andersson I, Karlberg I, Bjurstam N, Frodis E, Hakansson S. Implementation of service screening with mammography in Sweden: from pilot study to nationwide programme. *J Med Screeng* 2000;**7**:14–18.

20. Swedish Organized Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. *Cancer Epidemiol Biomarkers Prev* 2006;**15**:45–51.

21. Emilsson L, Lindahl B, Koster M, Lambe M, Ludvigsson JF. Review of 103 Swedish Healthcare Quality Registries. *J Intern Med* 2015;**277**:94–136.

22. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number:possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009;**24**:659–67.

23. International Statistical Classification of Diseases and Related Health Problems. 10th Revision. World Health Organization. 2015 Version [Internet]; Nov 2015 [cited November 9 2015]; http://apps.who.int/classifications/icd10/browse/2015/en.

24. Helgesson O, Bengtsson C, Lapidus L, Merck C, Sparen P. Malignant disease observed in a cohort of women. A validation of Swedish Cancer Registry data. *Scand J Soc Med* 1994;**22**:46–49.

25. Wettermark B, Hammar N, Fored CM *et al*. The new Swedish Prescribed Drug Register – opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007;**16**:726–35.

26. Brand JS, Li J, Humphreys K *et al*. Identification of two novel mammographic density loci at 6Q25.1. *Breast Cancer Res* 2015;**17**:75.

27. Brand JS, Humphreys K, Thompson DJ *et al*. Volumetric mammographic density: heritability and association with breast cancer susceptibility loci. *J Natl Cancer Inst* 2014;**106:**doi: 10.1093/jnci/dju334.

28. Trinh T, Christensen SE, Brand JS *et al*. Background risk of breast cancer influences the association between alcohol consumption and mammographic density. *Br J Cancer* 2015;**113**:159–65.

29. Trinh T, Eriksson M, Darabi H *et al*. Background risk of breast cancer and the association between physical activity and mammographic density. *Breast Cancer Res* 2015;**17**:50.

30. Skarping I, Brand JS, Hall P, Borgquist S. Effects of statin use on volumetric mammographic density: results from the KARMA study. *BMC Cancer* 2015;**15**:435.

31. Bonn SE, Bergman P, Trolle Lagerros Y, Sjolander A, Balter K. A validation study of the web-based physical activity questionnaire Active-Q against the GENEA Accelerometer. *JMIR Res Protoc* 2015;**4**:e86.

32. Christensen SE, Moller E, Bonn SE *et al*. Relative validity of micronutrient and fiber intake assessed with two new interactive meal- and Web-based food frequency questionnaires. *J Med Internet Res* 2014;**16**:e59.

33. Stockholm Regional Register for Breast Cancer (RCC STHLM), Sweden (SE) [Internet]. August 2015 [cited August 5 2015]; http://www.cancercentrum.se/ stockholmgotland/.

34. Information Network for Cancer Treatment (INCA), Sweden (SE) [Internet]. August 2015 [cited August 5 2015]; http://www.cancercentrum.se/inca/.

35. The National Board of Health and Welfare. Swedish National Cancer Register (NCR) (SE) [Internet]. August 2015 [cited August 5 2015]; http://www.socialstyrelsen.se/register/halsoda taregister/ cancerregistret/inenglish.

36. The National Board of Health and Welfare. Swedish Prescribed Drug Register (SPDR) (SE) [Internet]. August 2015 [cited August 5 2015]; http://www.socialstyrelsen.se/register/halsodataregis ter/lakemedelsregistret.

37. The National Board of Health and Welfare. National Patient Register (NPR), Sweden (SE) [Internet]. August 2015 [cited August 5 2015]; http://www.socialstyrelsen.se/register/halsoda taregister/patientregistret/inenglish.

38. The National Board of Health and Welfare. National Cause of Death Register, Sweden (SE) [Internet]. August 2015 [cited August 5 2015]; http://www.socialstyrelsen.se/register/dodsorsaksregistret.