



D4.1 - Description of database for environmental risk factors

Standardization and integration of multiple sources of data across contributing studies is essential for comprehensive data analyses and the development of risk and prognostication models of molecular subtypes of breast cancer.

Through previous work in BCAC under the EU project COGS; we have developed a data dictionary for environmental risk factors that accounted for the different instruments used to collect data in the studies (coordinated by DKFZ), which is the basis of the current risk factor data base (see D4.1_attachment1_BCAC_Extended_Data_Dictionary_v1.90_Oct2015_MASTER). Relevant risk factors include reproductive history, exogenous hormone use, body weight and height /body mass index (BMI), physical activity, alcohol use, tobacco smoking, occupational radiation exposure and family history.

Since the completion of COGS project, 29 new studies, which have been genotyped using the Infinium OncoArray 500K-Beadchip, have participated in BCAC. Additionally, new study participants from 46 existing studies were genotyped and also added to the BCAC risk factor database. At the last data freeze (end of 2015), the database contained a total of 362,002 subjects from **97** studies. This means an approximately additional 150,000 new subjects in the risk factor database. The studies included 16% prospective (or nested case-control) studies, 72% case-control studies (population-based, hospital-based, mixed) and 12% patient cohorts (cases only).

To use the risk factor database for a first assessment of established risk factors for invasive breast cancer, we included 294,654 subjects from 70 studies, excluding 18,540 subjects with non-invasive breast cancer or breast cancer with unknown invasiveness. Additionally, patient cohorts as well as case-control studies with missing age information (diagnosis or recruitment) or with less than 200 cases/controls having age information were excluded. Based on this data set, we found the expected associations between environmental risk factors and breast cancer risk. The number of subjects (in total and separately for cases and controls) for each of the risk variables assessed is shown in D4.1_attachment2_TableRiskFactors.

Collection of additional risk factor data

In order to plan the collection and harmonization of data on additional risk factors not currently included in the BCAC risk factor database, an online survey was conducted regarding availability of information on mammographic density, medical history of benign breast disease, mode of breast cancer detection, comorbidity (such as diabetes and cardiovascular disease, occupation (shift work), and medication use (see D4.2_attachment_WP4_survey_final). Based on the responses from 72 studies, a data dictionary for additional variables was developed (see D4.1_attachment3_BCAC_additional_Data_Request_Dictionary_18Aug2016), based on which standard procedure/algorithms will be developed for plausibility check, cross-variable check and quality control.

D4.1_attachment1_BCAC_Extended_Data_Dictionary_v1.90_Oct2015_MASTER

BCAC EXTENDED COVARIATE DATA DICTIONARY - COMPLETE RISK FACTOR VARIABLES					
Version 1.90 (09 October 2015)					
Category	Variable	Label	Coding	Variable type	Comment
ID	UniqueID	BCAC unique person identifier			
	PersonID	Person identified unique within original study	assigned by study	qualitative (nominal)	Concatenation of Study Acronym, "-", and PersonID
	Study	BCAC acronym for study	assigned by study	qualitative (nominal)	
	contrType	Type of control	1=population-based, 2=hospital-based, 3=family-based, 4=blood donor, 5=nested case-control, 6=BRCA1/2 carrier without bc, 777=NA=not applicable (for cases), 888=DK=don't know	qualitative	
	Status	Case-control status	0=control, 1=invasive case, 2=in-situ case, 3=case unknown invasiveness, 9=excluded sample	qualitative	The status of the tumor should be the status of the INDEX tumor (in line with variables DateDiagIndex and AgeDiagIndex)
	matchId	ID of pair or matched case-control set	777=NA, 888=DK	qualitative (nominal)	
	SubStudy	Identifier for within study strata (e.g. multi-ethnic cohort, different recruitment groups, or other strata)	777=NA, 888=DK	qualitative (nominal)	Only relevant for some studies; extracted from data supplied
	Studytype	Identifier for within study strata: especially for identification of cohorts with familial cases	0='sporadic' (population or hospital based), 1='familial' (clinical genetic centre based), 2=other, 777=control, 888=DK	qualitative	If you coded 'other', please provide additional information in your email
	StudytypeOt	details of study type if Studytype = 'other'		text	
	Exclusion	Reason for exclusion	0=include, 3=male individual, 4= gender discordance, 5=no phenotypic data, 6=other, 7=non-breast carcinoma (e.g. sarcoma), 8=duplicate sample, 888=DK	qualitative	For new studies this should be irrelevant as we do not want ineligible subject data sent. Most of these we have excluded after data submission having got details like tumour morphology updated, and identified duplicates etc.
Age	birthDate	Date of birth	dd/mm/yyyy, 08/08/8000=DK	date	If only year of birth is available, use the first of July as a surrogate for the true day and month; if only month and year of birth are available, use the 15 as a surrogate for the true day.
	birthDate_known	Marker for date of birth	DMY=day,month and year known; MY=only month and year known; Y=only year known; NA=all unknown	qualitative	This marker allows the identification of true and surrogate dates.
	bDay	Day of birth	day; DD; 888=DK	quantitative	
	bMonth	Month of birth	month; 888=DK	quantitative	If month is known and can be provided
	bYear	Year of birth	year; YYYY; 888=DK	quantitative	If only year is known or complete DOB cannot be given
	AgeInt	Age at interview/questionnaire for controls and cases	years, 888=DK	quantitative	
	intDate	Date at interview/questionnaire for cases and controls	dd/mm/yyyy, 08/08/8000=DK	date	If only year of interview is available, use the first of July as a surrogate for the true day and month; if only month and year of interview are available, use the 15 as a surrogate for the true day.
	refMonth	Month of diagnosis of breast cancer for cases and month of completing interview/questionnaire for controls	month; 888=DK	quantitative	For data protection reasons we are only asking for reference month (variable refMonth) and reference year (variable refYear). If reference month is still considered to be too much information then please simply give the reference year (see refYear)
	refYear	Year of diagnosis of breast cancer for cases and year of completing interview/questionnaire for controls	year, YYYY; 888=DK	quantitative	
	AgeDiagIndex	Age at diagnosis of index breast cancer for cases	years, 888=DK	quantitative	
Sex	Sex	M=male, F=female, U=unknown		qualitative (nominal)	
Genetics	brca12	BRCA12 status	0=non-carrier, 1=BRCA1, 2=BRCA2, 3=BRCA1 and BRCA2, 4=not tested, 888=DK	qualitative	
	testDate_BRCA1	Date of genetic testing of BRCA1	dd/mm/yyyy, 07/07/7000=NA, 08/08/8000=DK	date	If only year of genetic testing is available, use the first of July as a surrogate for the true day and month; if only month and year of genetic testing are available, use the 15 as a surrogate for the true day.
	testDate_BRCA1_known	Marker for date of genetic testing of BRCA1	DMY=day,month and year known; MY=only month and year known; Y=only year known; NA=all unknown	qualitative	This marker allows the identification of true and surrogate dates.
	testDate_BRCA2	Date of genetic testing of BRCA2	dd/mm/yyyy, 07/07/7000=NA, 08/08/8000=DK	date	If only year of genetic testing is available, use the first of July as a surrogate for the true day and month; if only month and year of genetic testing are available, use the 15 as a surrogate for the true day.
	testDate_BRCA2_known	Marker for date of genetic testing of BRCA2	DMY=day,month and year known; MY=only month and year known; Y=only year known; NA=all unknown	qualitative	This marker allows the identification of true and surrogate dates.
	dnaAge	Age at blood/buccal sample	years, 777=NA, 888=DK	quantitative	
		TestSensBRCA1	Testing sensitivity BRCA1	0, 1, 2, 3, 4, 5, 6, 777=NA, 888=DK	qualitative
	TestSensBRCA2	Testing sensitivity BRCA2	0, 1, 2, 3, 4, 5, 6, 777=NA, 888=DK	qualitative	See sheet 'TestSensBRCA codes' for description of the coding. If additional methods have been used in your study, please provide this information along with your codes (i.e. code, brief description, and detailed description -> cp. sheet 'testing sensitivity') used for these methods
Ethnicity	EthnicityClass	Ethnic origin	1=European (for EthnicitySubClass 1, 2, 3, 4, 5), 2=Hispanic American (for EthnicitySubClass 6), 3=African (for EthnicitySubClass 7, 8, 9), 4=Asian Subcontinent (for EthnicitySubClass 10, 11, 12), 5=South-East Asian (for EthnicitySubClass 13, 14, 15), 6=Other (including 'mixed race'), 888=DK	qualitative	

Category	Variable	Label	Coding	Variable type	Comment
Alcohol	AlcRegStartAge	Age at starting regular alcohol drinking before reference date	years, 777=NA (i.e. no regular alcohol drinking), 888=DK	quantitative	
	alcoholFreq	Frequency alcoholic drinks in last year before reference date	glasses / week, 888=DK	quantitative	Calculated as (AlcBeer + AlcWine + AlcFWine + AlcSpirits)
	alcoholCum	Cumulative lifetime gms/day alcohol	grams / day, 888=DK	quantitative	
	AlcType	Type of alcohol in last year before reference date	1=beer, 2=wine, 3=fortified wine, 4=spirits, 5=mixed, 777=NA, 888=DK	qualitative	
	AlcBeer	No of glasses of beer in last year before reference date	glasses / week, 888=DK	quantitative	
	AlcWine	No of glasses of wine in last year before reference date	glasses / week, 888=DK	quantitative	
	AlcFWine	No of glasses of fortified wine in last year before reference date	glasses / week, 888=DK	quantitative	
	AlcSpirits	No of glasses of spirits in last year before reference date	glasses / week, 888=DK	quantitative	
	AlcGram	Grams of alcohol per week in last year before reference date	grams / week, 888=DK	quantitative	
Smoking	smokingEver	Cigarette smoking	0=never, 1=past, 2=current, in last year before reference date (year before diagnosis for cases, year before questionnaire for controls), 888=DK	qualitative	
	smokingCurrent	Current cigarette smoking	0=no, 1=yes, current, 888=DK	qualitative	
	SmokeYrs	Total duration of smoking	years, 888=DK	quantitative	
	SmokeAvg	Frequency of smoking on average	cig/day, 888=DK	quantitative	
	SmoRegStartAge	Age at starting regular smoking	years, 777=NA, 888=DK	quantitative	
	SmoRegStopAge	Age at stopping regular smoking	years, 666=current smoker, 777=NA (never smoker), 888=DK	quantitative	
	packyrs	Smoking pack-years	pack-years, 888=DK	quantitative	A pack year is defined as twenty cigarettes smoked everyday for one year, usually calculated as: (average number of cigarettes per day)*(number of years smoked)/20
Physical activity	physActRec	Frequency of physical activity in the last year before reference date (date of diagnosis for cases, date of questionnaire for controls)	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	physActBefore30	Frequency of physical activity until age 30	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	physAct30_50	Frequency of physical activity at age 30-50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	physAct50plus	Frequency of physical activity at age >50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	modPhysActRec	Frequency of moderate physical activity in the last year before reference date (date of diagnosis for cases, date of questionnaire for controls)	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	modPhysActBefore30	Frequency of moderate physical activity until age 30	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	modPhysAct30_50	Frequency of moderate physical activity at age 30-50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	modPhysAct50plus	Frequency of moderate physical activity at age >50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	strPhysActRec	Frequency of strenuous physical activity in the last year before reference date (date of diagnosis for cases, date of questionnaire for controls)	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	strPhysActBefore30	Frequency of strenuous physical activity until age 30	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	strPhysAct30_50	Frequency of strenuous physical activity at age 30-50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	strPhysAct50plus	Frequency of strenuous physical activity at age >50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	walkingRec	Frequency of walking in the last year before reference date (date of diagnosis for cases, date of questionnaire for controls)	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	walkingBefore30	Frequency of walking until age 30	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	walking30_50	Frequency of walking at age 30-50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	walking50plus	Frequency of walking at age >50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	cyclingRec	Frequency of cycling in the last year before reference date (date of diagnosis for cases, date of questionnaire for controls)	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	cyclingBefore30	Frequency of cycling until age 30	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	cycling30_50	Frequency of cycling at age 30-50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	cycling50plus	Frequency of cycling at age >50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	sittingRec	Frequency of sitting in the last year before reference date (date of diagnosis for cases, date of questionnaire for controls)	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	sittingBefore30	Frequency of sitting until age 30	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	sitting30_50	Frequency of sitting at age 30-50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	sitting50plus	Frequency of sitting at age >50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	houseworkRec	Frequency of doing housework in the last year before reference date (date of diagnosis for cases, date of questionnaire for controls)	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	houseworkBefore30	Frequency of doing housework until age 30	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	housework30_50	Frequency of doing housework at age 30-50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	housework50plus	Frequency of doing housework at age >50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	gardeningRec	Frequency of gardening in the last year before reference date (date of diagnosis for cases, date of questionnaire for controls)	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	gardeningBefore30	Frequency of gardening until age 30	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
	gardening30_50	Frequency of gardening at age 30-50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding

Category	Variable	Label	Coding	Variable type	Comment
	gardening50plus	Frequency of gardening at age >50	hours / week, 777=NA, 888=DK	quantitative	If hours/week is available in categories only, please provide categories along with the respective coding
Medication	NSAIDs1Ever	Use of NSAIDs (group 1) regularly (≥2 times/week) before reference date	0=No, 1=Yes, 888=DK	qualitative	NSAIDs, group 1: ASS (ATC code: N02BA01, B01AC06, B01AC30, M03BA53, N02BA51, N02BA71)
	NSAIDs1Current	Current use of NSAIDs (group 1) regularly (≥2 times/week)	0=No, 1=Yes, 888=DK	qualitative	NSAIDs, group 1: ASS (ATC code: N02BA01, B01AC06, B01AC30, M03BA53, N02BA51, N02BA71)
	NSAIDs1Start	Age when use of NSAIDs (group 1) started	years, 777=NA, 888=DK	quantitative	NSAIDs, group 1: ASS (ATC code: N02BA01, B01AC06, B01AC30, M03BA53, N02BA51, N02BA71)
	NSAIDs1Yrs	Total duration of use of NSAIDs (group 1)	years, 888=DK	quantitative	NSAIDs, group 1: ASS (ATC code: N02BA01, B01AC06, B01AC30, M03BA53, N02BA51, N02BA71)
	NSAIDs1Dose	Weekly dose of NSAIDs (group 1)	times/week, 888=DK	quantitative	NSAIDs, group 1: ASS (ATC code: N02BA01, B01AC06, B01AC30, M03BA53, N02BA51, N02BA71)
	NSAIDs2Ever	Use of NSAIDs (group 2) regularly (≥2 times/week) before reference date	0=No, 1=Yes, 888=DK	qualitative	NSAIDs, group 2: Paracetamol (ATC code: N02BE01, A03DB04)
	NSAIDs2Current	Current use of NSAIDs (group 2) regularly (≥2 times/week)	0=No, 1=Yes, 888=DK	qualitative	NSAIDs, group 2: Paracetamol (ATC code: N02BE01, A03DB04)
	NSAIDs2Start	Age when use of NSAIDs (group 2) started	years, 777=NA, 888=DK	quantitative	NSAIDs, group 2: Paracetamol (ATC code: N02BE01, A03DB04)
	NSAIDs2Yrs	Total duration of use of NSAIDs (group 2)	years, 888=DK	quantitative	NSAIDs, group 2: Paracetamol (ATC code: N02BE01, A03DB04)
	NSAIDs2Dose	Weekly dose of NSAIDs (group 2)	times/week, 888=DK	quantitative	NSAIDs, group 2: Paracetamol (ATC code: N02BE01, A03DB04)
	NSAIDs3Ever	Use of NSAIDs (group 3) regularly (≥2 times/week) before reference date	0=No, 1=Yes, 888=DK	qualitative	NSAIDs, group 3: mixed group ASS/Paracetamol/other NSAIDs (ATC code: N02AA59, N02BE51)
	NSAIDs3Current	Current use of NSAIDs (group 3) regularly (≥2 times/week)	0=No, 1=Yes, 888=DK	qualitative	NSAIDs, group 3: mixed group ASS/Paracetamol/other NSAIDs (ATC code: N02AA59, N02BE51)
	NSAIDs3Start	Age when use of NSAIDs (group 3) started	years, 777=NA, 888=DK	quantitative	NSAIDs, group 3: mixed group ASS/Paracetamol/other NSAIDs (ATC code: N02AA59, N02BE51)
	NSAIDs3Yrs	Total duration of use of NSAIDs (group 3)	years, 888=DK	quantitative	NSAIDs, group 3: mixed group ASS/Paracetamol/other NSAIDs (ATC code: N02AA59, N02BE51)
	NSAIDs3Dose	Weekly dose of NSAIDs (group 3)	times/week, 888=DK	quantitative	NSAIDs, group 3: mixed group ASS/Paracetamol/other NSAIDs (ATC code: N02AA59, N02BE51)
	NSAIDs4Ever	Use of NSAIDs (group 4) regularly (≥2 times/week) before reference date	0=No, 1=Yes, 888=DK	qualitative	NSAIDs, group 4: other NSAIDs (ATC code: M01A*, N02BB*, A03DC, N02AG, A07EC*)
	NSAIDs4Current	Current use of NSAIDs (group 4) regularly (≥2 times/week)	0=No, 1=Yes, 888=DK	qualitative	NSAIDs, group 4: other NSAIDs (ATC code: M01A*, N02BB*, A03DC, N02AG, A07EC*)
	NSAIDs4Start	Age when use of NSAIDs (group 4) started	years, 777=NA, 888=DK	quantitative	NSAIDs, group 4: other NSAIDs (ATC code: M01A*, N02BB*, A03DC, N02AG, A07EC*)
	NSAIDs4Yrs	Total duration of use of NSAIDs (group 4)	years, 888=DK	quantitative	NSAIDs, group 4: other NSAIDs (ATC code: M01A*, N02BB*, A03DC, N02AG, A07EC*)
	NSAIDs4Dose	Weekly dose of NSAIDs (group 4)	times/week, 888=DK	quantitative	NSAIDs, group 4: other NSAIDs (ATC code: M01A*, N02BB*, A03DC, N02AG, A07EC*)
	StatinsEver	Use of Statins regularly (≥2 times/week) before reference date	0=No, 1=Yes, 888=DK	qualitative	ATC code: C10AA
	StatinsCurrent	Current use of Statins regularly (≥2 times/week)	0=No, 1=Yes, 888=DK	qualitative	ATC code: C10AA
	StatinsStart	Age when use of Statins started	years, 777=NA, 888=DK	quantitative	ATC code: C10AA
	StatinsYrs	Total duration of use of Statins	years, 777=NA, 888=DK	quantitative	ATC code: C10AA
	StatinsDose	Weekly dose of Statins	times/week, 777=NA, 888=DK	quantitative	ATC code: C10AA
Family history	sisters	Total number of sisters	888=DK	quantitative	Including half-sisters
	brCancerSis	Number of sisters with breast cancer	integer, 888=DK	quantitative	Including half-sisters; if sisters=0 then brCancerSis should be 0
	ovCancerSis	Number of sisters with ovarian cancer	integer, 888=DK	quantitative	Including half-sisters; if sisters=0 then ovCancerSis should be 0
	daughters	Total number of daughters	integer, 888=DK	quantitative	
	brCancerDau	Number of daughters with breast cancer	integer, 888=DK	quantitative	If daughters=0 then brCancerDau should be 0
	ovCancerDau	Number of daughters with ovarian cancer	integer, 888=DK	quantitative	If daughters=0 then ovCancerDau should be 0
	brCancerMom	Mother had breast cancer	0=No, 1=Yes, 888=DK	qualitative	
	ovCancerMom	Mother had ovarian cancer	0=No, 1=Yes, 888=DK	qualitative	
	brCancerDad	Father had breast cancer	0=No, 1=Yes, 888=DK	qualitative	
	FHisFstBC	First degree female family members with breast cancer	0=No, 1=Yes, 888=DK	qualitative	First degree: mother, sisters, daughters
	FHisFstBCNr	Number of first degree female family members with breast cancer	integer, 888=DK	quantitative	If FHisFstBC=0 then FHisFstBCNr should be 0
	FHisSecBC	Second degree female family members with breast cancer	0=No, 1=Yes, 888=DK	qualitative	Second degree: grandmothers, sisters of mother/father, daughters of sisters/brothers
	FHisSecBCNr	Number of second degree female family members with breast cancer	integer, 888=DK	quantitative	If FHisSecBC=0 then FHisSecBCNr should be 0
	FHisFstOC	First degree female family members with ovarian cancer	0=No, 1=Yes, 888=DK	qualitative	First degree: mother, sisters, daughters
	FHisFstOCNr	Number of first degree female family members with ovarian cancer	integer, 888=DK	quantitative	If FHisFstOC=0 then FHisFstOCNr should be 0
	FHisSecOC	Second degree female family members with ovarian cancer	0=No, 1=Yes, 888=DK	qualitative	Second degree: grandmothers, sisters of mother/father, daughters of sisters/brothers
	FHisSecOCNr	Number of second degree family members with ovarian cancer	integer, 888=DK	quantitative	If FHisSecOC=0 then FHisSecOCNr should be 0
fam1grBC50	Number of first degree female family members with breast ca before age 50	integer, 888=DK	quantitative	If FHisFstBC=0 then fam1grBC50 should be 0	
fam1grOC50	Number of first degree family members with ovarian ca before age 50	integer, 888=DK	quantitative	If FHisFstBC=0 then fam1grOC50 should be 0	
FamHist	Family history of breast cancer in a first degree relative (0=no, 1=yes)	0=No, 1=Yes, 888=DK	qualitative		
Fhnumber	Number of affected (breast cancer) first degree relatives	integer, 888=DK	quantitative		
Fhscore	Family history score	1 for each first degree affected relative, 0.5 for second degree, 0.25 for third degree (not ovarian for any relative)	integer, 888=DK	quantitative	Not ovarian for first, second or third degree relative

D4.1_attachment2_TableRiskFactors

Table 1: Description of selected lifestyle/environmental risk factors of breast cancer in the BCAC risk factor database: Numbers of invasive breast cancer cases and controls from 70 studies included

	Cases	Controls	Total
Total	146813	147623	294436
Age at menarche	97260	102627	199887
Ever Parous	101195	115559	216754
Number of full-term births	100319	114178	214467
Age at first full-term birth	74029	83771	157800
Ever breastfed	72080	73598	145678
Height	90586	91027	181613
Usual BMI	90586	91026	181612
Ever use of oral contraceptives	77562	85451	163013
Current use of menopausal hormone therapy (HRT)	73104	74356	147460
Current use of estrogen-progesterone HRT	62849	65350	128199
Current use of estrogen only HRT	62520	64981	127501
Cumulative lifetime grams/day alcohol	23703	30456	54159
Current smoking	74311	86961	161272
Pack-years	61159	73418	134577
Family history first degree of breast cancer	100735	85301	186036
Recent physical activity	26112	31256	57368

D4.1_attachment3_BCAC_additional_Data_Request_Dictionary_18Aug2016

	A	B	C	D	E	F	G
1	Additional Data Request Dictionary						
2							
3	Category	Variable	Label	Coding	Variable type	Comment	
4	ID	MDC_Id	Person identified within MD consortium	assigned by study	qualitative (nominal)		
5		PersonID	Person identified within original study	assigned by study	qualitative (nominal)		
6		UniqueID	BCAC unique person identifier	assigned by study	qualitative (nominal)		
7		Status	Case-control status	0=control, 1=invasive case, 2=in-situ case, 3=case unknown invasiveness, 9=excluded sample	qualitative		
8	Age	birthDate	Date of birth	dd/mm/yyyy, 08/08/8000=DK	date		
9	Diagnostic mammographic density	Mam_dt	Date of mammogram	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date	Refers to the BCAC index tumour	
10		MamSide_diag	Corresponding side of mammogram (left vs. right vs. average of both)	0=Left; 1=Right; 2=Both; 888=DK	qualitative		
11		MamView_diag	MLO vs. CC	0=MLO; 1=CC; 2=Other; 888=DK	qualitative		
12		MamType_diag	Digital vs Analogue	0=Digital; 1=Analogue; 888=DK	qualitative		
13		MamProcess_diag	Raw ("For Processing") vs Processed ("For Presentation")	0=Raw; 1=Processed; 2=Not Applicable (analogue); 888=Don't know	qualitative		
14		MamManufacturer_diag	Manufacturer of mammography machine used	0=GE; 1=Siemens; 2=Hologic; 3=Phillips; 4=Fuji; 5=Konica; 6=Sectra; 7=Agfa; 8=Other; 777=Not applicable; 888=Don't know	qualitative	State analogue images as 888 if do not know which manufacturer it was	
15		MamAnalogScanner_diag	Scanner used for analog images		text		
16		MamCanSide_diag	On which breast the cancer is located	0=Left; 1=Right; 888=DK	qualitative		
17							
18		PDArea_diag	Percent density from thresholding method on corresponding image	0-100, 777=NA, 888=DK	qualitative		
19		DenArea_diag	Absolute dense area from mammogram	cm ² , 777=NA, 888=DK	qualitative		
20		NonDArea_diag	Absolute nondense area from mammogram (Total area-dense area)	cm ² , 777=NA, 888=DK	qualitative		
21							
22		PDVol_diag	Percent density from volumetric method on corresponding image	0-100, 777=NA, 888=DK	qualitative		
23		DenVol_diag	Absolute dense volume from mammogram	cm ³ , 777=NA, 888=DK	qualitative		
24		NonDVol_diag	Absolute nondense volume from mammogram (Total vol-dense vol)	cm ³ , 777=NA, 888=DK	qualitative		
25							
26		PDArea_Stratus_diag	Percent density from Stratus method on corresponding image	0-100, 777=NA, 888=DK	qualitative		
27		DenArea_Stratus_diag	Absolute dense area from Stratus method on mammogram	cm ² , 777=NA, 888=DK	qualitative		
28		NonDArea_Stratus_diag	Absolute nondense area from Stratus method on mammogram (Total area-dense area)	cm ² , 777=NA, 888=DK	qualitative		
29							
30		BIRADS_Class_diag	BI-RADS density categories	1=1st category; 2=2nd category; 3=3rd category; 4=4th category; 5=5th category; 888=DK	qualitative	refers to breast composition (density) categories and not malignancy scale	
31		BIRADS_edition_diag	Edition of BIRADS used	1=1st edition, 2=2nd edition, 3=3rd edition, 4=4th edition, 5=5th edition	qualitative		
32	Non-diagnostic mammographic density	Mam_date_prediag	Date of mammogram	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
33		MamSide_prediag	Corresponding side of mammogram (left vs. right vs. average of both)	0=Left; 1=Right; 2=Both; 888=DK	qualitative		
34		MamView_prediag	MLO vs. CC	0=MLO; 1=CC; 2=Other; 888=DK	qualitative		
35		MamType_prediag	Digital vs Analogue	0=Digital; 1=Analogue; 888=DK	qualitative		
36		MamProcess_prediag	Raw ("For Processing") vs Processed ("For Presentation")	0=Raw; 1=Processed; 2=Not Applicable (analogue); 888=Don't know	qualitative		
37		MamManufacturer_prediag	Manufacturer of mammography machine used	0=GE; 1=Siemens; 2=Hologic; 3=Phillips; 4=Fuji; 5=Konica; 6=Sectra; 7=Agfa; 8=Other; 777=Not applicable; 888=Don't know	qualitative	State analogue images as 888 if do not know which manufacturer it was	
38		MamAnalogScanner_prediag	Scanner used for analog images		text		
39		MamCanSide_prediag	On which breast the cancer is located	0=Left; 1=Right; 777=NA; 888=DK	qualitative		
40							
41		PDArea_prediag	Percent density from thresholding method on corresponding image	0-100, 777=NA, 888=DK	qualitative		

	A	B	C	D	E	F	G
42		DenArea_prediag	Absolute dense area from mammogram	cm^2, 777=NA, 888=DK	qualitative		
43		NonDArea_prediag	Absolute nondense area from mammogram (Total area-dense area)	cm^2, 777=NA, 888=DK	qualitative		
44							
45		PDVol_prediag	Percent density from volumetric method on corresponding image	0-100, 777=NA, 888=DK	qualitative		
46		DenVol_prediag	Absolute dense volume from mammogram	cm^3, 777=NA, 888=DK	qualitative		
47		NonDVol_prediag	Absolute nondense volume from mammogram (Total vol-dense vol)	cm^3, 777=NA, 888=DK	qualitative		
48							
49		PDArea_Stratus_prediag	Percent density from Stratus method on corresponding image	0-100, 777=NA, 888=DK	qualitative		
50		DenArea_Stratus_prediag	Absolute dense area from Stratus method on mammogram	cm^2, 777=NA, 888=DK	qualitative		
51		NonDArea_Stratus_prediag	Absolute nondense area from Stratus method on mammogram (Total area-dense area)	cm^2, 777=NA, 888=DK	qualitative		
52							
53		BIRADS_Class_prediag	BI-RADS density categories	1=1st category; 2=2nd category; 3=3rd category; 4=4th category; 5=5th category; 888=DK	qualitative	refers to breast composition (density) categories and not malignancy scale	
54		BIRADS_edition_prediag	Edition of BIRADS used	1=1st edition, 2=2nd edition, 3=3rd edition, 4=4th edition, 5=5th edition	qualitative		
55	Biopsies	Biopsies_number	Number of breast biopsies, excluding biopsies leading to diagnosis of breast cancer in cases	0=0; 1=1; 2=2; 3=3; 4=4; 5=5; 6=6 or more; 777=Not applicable; 888=DK	qualitative		
56	Benign Breast Disease	BBC_data_collect	Date BBD data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
57		BBD_history	History of benign breast disease	0=No, 1=Yes, 888=DK	qualitative		
58		BBD_number	Number of benign breast disease diagnoses	0=0, 1=1, 2=2, 3=3, 4=4 or more; 777=NA, 888=DK	qualitative		
59		BBD_type1	Type of benign breast disease for first BBD diagnosis	0=non-proliferative disease; 1=proliferative disease without atypia; 2=atypical hyperplasia; 777=Not Applicable; 888=DK	qualitative		
60		BBD_date1	Date of diagnosis of first benign breast disease diagnosis	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	qualitative		
61		BBD_age1	Age at diagnosis of first benign breast disease diagnosis	years; 777=Not applicable; 888=Don't know	qualitative		
62		BBD_side1	Corresponding side of first benign breast disease diagnosis	0=Left; 1=Right; 888=DK	qualitative		
63		BBD_type2	Type of benign breast disease for second BBD diagnosis	0=non-proliferative disease; 1=proliferative disease without atypia; 2=atypical hyperplasia; 777=Not Applicable; 888=DK	qualitative		
64		BBD_date2	Date of diagnosis of second benign breast disease diagnosis	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	qualitative		
65		BBD_age2	Age at diagnosis of second benign breast disease diagnosis	years; 777=Not applicable; 888=Don't know	qualitative		
66		BBD_side2	Corresponding side of second benign breast disease diagnosis	0=Left; 1=Right; 888=DK	qualitative		
67		BBD_type3	Type of benign breast disease for third BBD diagnosis	0=non-proliferative disease; 1=proliferative disease without atypia; 2=atypical hyperplasia; 777=Not Applicable; 888=DK	qualitative		
68		BBD_date3	Date of diagnosis of third benign breast disease diagnosis	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	qualitative		
69		BBD_age3	Age at diagnosis of third benign breast disease diagnosis	years; 777=Not applicable; 888=Don't know	qualitative		
70		BBD_side3	Corresponding side of third benign breast disease diagnosis	0=Left; 1=Right; 888=DK	qualitative		
71		BBD_type4	Type of benign breast disease for fourth BBD diagnosis	0=non-proliferative disease; 1=proliferative disease without atypia; 2=atypical hyperplasia; 777=Not Applicable; 888=DK	qualitative		
72		BBD_date4	Date of diagnosis of fourth benign breast disease diagnosis	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	qualitative		
73		BBD_age4	Age at diagnosis of fourth benign breast disease diagnosis	years; 777=Not applicable; 888=Don't know	qualitative		
74		BBD_side4	Corresponding side of fourth benign breast disease diagnosis	0=Left; 1=Right; 888=DK	qualitative		
75	Detection	Detection_screen	Mode of detection of first breast cancer diagnosis	0=routine breast cancer screening; 1=other methods of detection; 888=DK	qualitative		
76		Detection_detailed	Method of detection of first breast cancer	0=mammography; 1=sonography; 2=self-detected by palpation; 3=medical exam, 888=DK	qualitative		

	A	B	C	D	E	F	G
77		Screen_ever	Ever attended breast cancer screening	0=no; 1=yes; 888=DK	qualitative		
78		Last_screen_year	Year of last breast cancer screening prior to the first diagnosis of breast cancer	years; 777=Not applicable; 888=DK	qualitative	First diagnosis of breast cancer refers to the index tumour reported in BCAC dataset	
79		Last_screen_month	Month of last breast cancer screening prior to the first diagnosis of breast cancer	month; 777=Not applicable; 888=DK	qualitative		
80	Medication use	MetforminEver	Use of metformin regularly (≥2 times/week) before reference date	0=No; 1=Yes; 888=DK	qualitative	Reference date is date of diagnosis of breast cancer for cases and month of completing interview/questionnaire for controls. Metformin, ATC code: A10BA02, A10BD03, A10BD05, A10BD07, A10BD08, A10BD10, A10BD11, A10BD13, A10BD14, A10BD15, A10BD16, A10BD17, A10BD18, A10BD20	
81		MetforminCurrent	Current use of metformin regularly (≥2 times/week)	0=No; 1=Yes; 888=DK	qualitative	Metformin, ATC code: A10BA02, A10BD03, A10BD05, A10BD07, A10BD08, A10BD10, A10BD11, A10BD13, A10BD14, A10BD15, A10BD16, A10BD17, A10BD18, A10BD20	
82		MetforminStart	Age when use of metformin started	years; 777=Not applicable; 888=DK	qualitative	Metformin, ATC code: A10BA02, A10BD03, A10BD05, A10BD07, A10BD08, A10BD10, A10BD11, A10BD13, A10BD14, A10BD15, A10BD16, A10BD17, A10BD18, A10BD20	
83		MetforminYrs	Total duration of use of metformin	years; 888=DK	qualitative	Metformin, ATC code: A10BA02, A10BD03, A10BD05, A10BD07, A10BD08, A10BD10, A10BD11, A10BD13, A10BD14, A10BD15, A10BD16, A10BD17, A10BD18, A10BD20	
84		MetforminDose	Weekly dose of metformin	times/week; 888=DK	qualitative	Metformin, ATC code: A10BA02, A10BD03, A10BD05, A10BD07, A10BD08, A10BD10, A10BD11, A10BD13, A10BD14, A10BD15, A10BD16, A10BD17, A10BD18, A10BD20	
85		InsulinEver	Use of insulin regularly (≥2 times/week) before reference date	0=No; 1=Yes; 888=Don't know	qualitative	Reference date is date of diagnosis of breast cancer for cases and month of completing interview/questionnaire for controls. Insulin, ATC code: A10AB*, A10AC*, A10AD*, A10AE*, A10AF*	
86		InsulinCurrent	Current use of insulin regularly (≥2 times/week)	0=No; 1=Yes; 888=Don't know	qualitative	Insulin, ATC code: A10AB*, A10AC*, A10AD*, A10AE*, A10AF*	
87		InsulinStart	Age when use of insulin started	years; 777=Not applicable; 888=Don't know	qualitative	Insulin, ATC code: A10AB*, A10AC*, A10AD*, A10AE*, A10AF*	
88		InsulinYrs	Total duration of use of insulin	years; 888=Don't know	qualitative	Insulin, ATC code: A10AB*, A10AC*, A10AD*, A10AE*, A10AF*	
89		InsulinDose	Weekly dose of insulin	times/week; 888=Don't know	qualitative	Insulin, ATC code: A10AB*, A10AC*, A10AD*, A10AE*, A10AF*	
90	Comorbidities	Hypertension	Hypertension	0=No, 1=Yes, 888=DK	qualitative		
91		Hypertension_date	Date hypertension data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
92		CirculatoryDisHeart	Circulatory disorders in the heart (angina pectoris)	0=No, 1=Yes, 888=DK	qualitative		
93		CirculatoryDisHeart_date	Date circulatory heart disorders data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
94		HeartAttack	Heart attack	0=No, 1=Yes, 888=DK	qualitative		
95		HeartAttack_date	Date heart attack data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
96		Stroke	Stroke	0=No, 1=Yes, 888=DK	qualitative		
97		Stroke_date	Date stroke data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
98		Chronic_pulmonary_disease	Chronic pulmonary disease (asthma, bronchitis)	0=No, 1=Yes, 888=DK	qualitative		
99		Chronic_pulmonary_disease_date	Date chronic pulmonary disease data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
100		Chronic_liver_disease	Chronic liver disease (chronic hepatitis, cirrhosis)	0=No, 1=Yes, 888=DK	qualitative		
101		Chronic_liver_disease_date	Date chronic liver disease data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
102		Chronic_bowel_disease	Chronic bowel disease	0=No, 1=Yes, 888=DK	qualitative		
103		Chronic_bowel_disease_date	Date chronic bowel disease data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
104		Chronic_bladder_kidney_disease	Chronic bladder and kidney diseases	0=No, 1=Yes, 888=DK	qualitative		

	A	B	C	D	E	F	G
105		Chronic_bladder_kidney_disease_date	Date chronic bladder and kidney diseases were collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
106		Chronic_bronchitis	Chronic bronchitis	0=No, 1=Yes, 888=DK	qualitative		
107		Chronic_bronchitis_date	Date chronic bronchitis was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
108		Inflammatory_diseases	Inflammatory joint diseases (arthritis, rheumatism)	0=No, 1=Yes, 888=DK	qualitative		
109		Inflammatory_diseases_date	Date inflammatory joint diseases were collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
110		Osteoporosis	Osteoporosis	0=No, 1=Yes, 888=DK	qualitative		
111		Osteoporosis_date	Date osteoporosis data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
112		Diabetes	Diabetes	0=No, 1=Yes, 888=DK	qualitative		
113		Diabetes_date	Date diabetes data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		
114		Other_disease	Other disease		text		
115		Other_disease_date	Date other disease data was collected	dd/mm/yyyy; 07/07/7000=Not applicable; 08/08/8000=DK	date		