



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Diplomarbeit

Studiendepartment Life Sciences
Studiengang Gesundheit

**Untersuchung des Selektionsbias einer Fall-Kontroll-
Studie zu Brustkrebsrisiko und Hormontherapie
anhand einer Kurzbefragung unter Nicht-Teilnehmern**

Anja Köchel

19. Februar 2007

Hochschule für Angewandte Wissenschaften (HAW), Hamburg

Fakultät *Life Sciences*

Studiendepartment Gesundheitswissenschaften

Studiengang Gesundheit

Lohbrügger Kirchstraße 65

21033 Hamburg

Verfasser: Anja Köchel
Matrikelnummer: 1547295
Abgabedatum: 19. Februar 2007

1. Prüfer: Prof. Dr. Ralf Reintjes
Department Gesundheitswissenschaften
Fakultät *Life Sciences*
Hochschule für angewandte Wissenschaften (HAW), Hamburg

2. Prüfer: Priv.-Doz. Dr. Dieter Flesch-Janys
MARIE-Studie/Arbeitsgruppe Epidemiologie
Institut für medizinische Biometrie und Epidemiologie (IMBE)
Universitätskliniken Eppendorf der Universität (UKE), Hamburg

Ich danke allen, die mich zu und während meiner Arbeit ermutigten.

*Insbesondere danke ich Herrn Priv.-Doz. Dr. Flesch-Janys
für die freundliche und konstruktive Zusammenarbeit seit dem Jahr 2001
und für die Überlassung des interessanten Themas.*

*Herrn Professor Dr. Reintjes danke ich
für die freundliche und geduldige Betreuung meiner Arbeit.*

Inhalt

1	Einleitung	1
1.1	Beschreibung des Forschungsvorhabens.....	1
1.2	Vollständigkeit und Repräsentativität.....	3
1.3	Selektionsbias.....	4
1.4	Ziel der Untersuchung	5
1.5	Hypothesenstellung	6
2	Methodik	7
2.1	Daten	7
2.1.1	Datenquellen.....	7
2.1.2	Datenschutz.....	8
2.1.3	Datenerhebung	9
2.1.4	Erhebungsinstrumente.....	11
2.1.5	Einschlusskriterien für Teilnehmerinnen	13
2.1.6	Datacleaning.....	14
2.1.7	Auswahl der Datensätze für die Untersuchung des Selektionsbias	17
2.2	Deskription der Arbeitsdatei.....	19
2.2.1	Organisatorische Variablen	19
2.2.2	Die abhängige Variable	20
2.2.3	Unabhängige Variablen	20
2.3	Bivariate Analyse	38
2.3.1	Test der stetigen Variablen (T-Test, Levene-Test)	38
2.3.2	Test der kategorialen Variablen (Chi-Quadrat-Test).....	40
2.3.3	Auswahl der erklärenden Variablen.....	42
2.4	Multivariate Analyse.....	44
2.4.1	Binäre logistische Regression.....	44
2.4.2	Auswahl des Modells	48

3	Ergebnisse	49
3.1	Modell 1: ohne Berücksichtigung von Interaktionseffekten	49
3.1.1	Modell 1a: Die Hormonvariable ist „Hormoneinnahme ever, never“	49
3.1.2	Modell 1b: Hormonvariable ist „Hormoneinnahme current, past, never“ ...	52
3.1.3	Modell 1c: Hormonvariable ist „Dauer der Hormoneinnahme in Klassen“	54
3.2	Modell 2: Modell mit Betrachtung von Interaktionseffekten bezüglich der Studienregion.....	56
4	Diskussion	58
4.1	Methode	58
4.1.1	Stichproben	58
4.1.2	Erhebungsinstrumente.....	58
4.2	Ergebnisse.....	60
4.2.1	Deskription: Univariate Analyse	60
4.2.2	Bivariate Analyse	61
4.2.3	Multivariate Analyse.....	62
4.2.4	Forschungsfragen und Hypothesen.....	63
5	Schlussfolgerung	65
	Literaturverzeichnis	66
	Anhänge	69
	Anhang 1: Tabellen der logistischen Regression	69
	Anhang 2: Kurzfragebogen zur telefonischen Befragung	77
	Anhang 3: Kurzfragebogen zum Selbstauffüllen	78
	Anhang 4: Versionsbeschreibung des Kurzfragebogens	79
	Anhang 5: Hauptfragebogen der MARIE (reduziert auf vergleichbare Items)	81

Abbildungsverzeichnis

Abb. 1:	Box-and-Whiskers-Plot der Geburtsjahrgänge nach Region und Fragebogenteilnahme	22
Abb. 2:	Box-and-Whiskers-Plot des Referenzalters nach Region und Fragebogenteilnahme	24
Abb. 3:	Histogramm – Referenzalter nach Teilnahme NR/MARIE	25
Abb. 4:	Balkendiagramm – Höchster Schulabschluss nach Studienregion	26
Abb. 5:	Box-and-Whiskers-Plot der Hormoneinnahme (kumuliert) nach Region und Fragebogenteilnahme	32
Abb. 6:	Histogramm – Alter bei Beginn der Hormoneinnahme nach Teilnahme NR/MARIE	34
Abb. 7:	Histogramm – Alter bei Ende der Hormoneinnahme nach Teilnahme NR/MARIE	35

Tabellenverzeichnis

Tabelle 1:	Beispiel für eine Veränderung des Odds Ratios durch einen Selektionsbias	4
Tabelle 2:	Erhebungsphase der MARIE (Teilnehmer des Hauptfragebogens)	10
Tabelle 3:	Erhebungsphase des Kurzfragebogens (Nicht-Teilnehmer der MARIE)	11
Tabelle 4:	Gewichtsklassifikation bei Erwachsenen anhand des BMI (nach WHO, 2000 EK IV)	16
Tabelle 5:	Verteilung der Teilnehmerinnen beider Fragebögen auf die Studienregionen	19
Tabelle 6:	BMI in Klassen (aktuelles Gewicht) nach Teilnahme NR/MARIE	21
Tabelle 7:	Schulabschluss nach Teilnahme NR/MARIE	26
Tabelle 8:	Erwerbstätig current nach Teilnahme NR/MARIE	27
Tabelle 9:	Anzahl der Kinder nach Studienregion	28
Tabelle 10:	Kontrazeptiva nach Teilnahme NR/MARIE und Studienregion	29
Tabelle 11:	Hormoneinnahme (ever/never) nach Teilnahme NR/MARIE und Studienregion	30
Tabelle 12:	Hormoneinnahme (current/past/never) nach Teilnahme NR/MARIE	31
Tabelle 13:	Hormoneinnahmedauer nach Teilnahme NR/MARIE	33
Tabelle 14:	Einnahme pflanzlicher und homöopathischer Präparate (ever/never) nach Teilnahme NR/MARIE	36
Tabelle 15:	Levene-Test auf Varianzhomogenität	38
Tabelle 16:	T-Test für die Mittelwertgleichheit bei unabhängigen Stichproben	39
Tabelle 17:	Chi-Quadrat-Unabhängigkeitstest	40
Tabelle 18:	Erklärende Variablen	43

Tabellen im Anhang:

Tabelle 19:	Basismodell der Rückwärts-Elimination 1a	69
Tabelle 20:	Endmodell der Rückwärts-Elimination 1a	70
Tabelle 21:	Basismodell der Rückwärts-Elimination 1b	71
Tabelle 22:	Endmodell der Rückwärts-Elimination 1b	72
Tabelle 23:	Basismodell der Rückwärts-Elimination 1c	73
Tabelle 24:	Endmodell der Rückwärts-Elimination 1c	74
Tabelle 25:	Basismodell der Rückwärts-Elimination 2	75
Tabelle 26:	Endmodell der Rückwärts-Elimination 2	76
Tabelle 27:	Kurzfragebogen Version 1	79
Tabelle 28:	Kurzfragebogen Version 2	79
Tabelle 29:	Kurzfragebogen Version 3	80
Tabelle 30:	Kurzfragebogen Version 4	80
Tabelle 31:	Kurzfragebogen Version 5	80

Zusammenfassung

Keywords: selection bias, nonresponse, case-control-study, hormone replacement therapy, short questionnaire, epidemiologic methods, data collection, breast cancer

Einleitung

In einer Fall-Kontroll-Studie (MARIE) zu Brustkrebsrisiko und Hormontherapie in Deutschland (Studienregionen Hamburg und Rhein-Neckar-Karlsruhe) bei 50- bis 75-jährigen Frauen wurde unter den Nicht-Teilnehmerinnen eine Kurzbefragung (*short questionnaire*) durchgeführt. Untersucht wurden teilnehmende Kontrollen mit einem vergleichbaren Erhebungszeitraum von 2003 bis 2006 (*participants short questionnaire* n=3021, *participants main questionnaire* n=6508). Ziel war die Ermittlung und Untersuchung eines Selektionsbias der Fall-Kontroll-Studie.

Methodik

Merkmale beider Stichproben wurden auf signifikante Unterschiede untersucht. Einflüsse der verschiedenen Prädiktoren (Geburtsjahrgang, *Body Mass Index*, Schulbildung, Erwerbstätigkeit, Anzahl der Kinder, Ernährungsverhalten, Rauchverhalten, Einnahme von Kontrazeptiva und Einnahme pflanzlicher/homöopathischer bzw. hormoneller Präparate gegen Wechseljahresbeschwerden) auf die Teilnahme bzw. Nicht-Teilnahme an der Fall-Kontroll-Studie wurden im Hinblick auf Stärke und Richtung bi- und multivariat (*stepwise backward elimination*) untersucht.

Ergebnisse

Die Prädiktoren Hormoneinnahme, Einnahme von Kontrazeptiva und pflanzlichen Präparaten sowie Schulbildung (Trend in Richtung höherer Bildung) haben einen signifikanten schwach-negativen Einfluss auf eine Nicht-Teilnahme. Die Studienregion zeigte keine Interaktionseffekte.

Tabelle: Effektkoeffizienten und Signifikanz (CI=95%) der Prädiktoren

Prädiktor	Exp(B)	p
Hormone	0,43	0,000
Pflanzliche/homöopathische Mittel	0,34	0,000
Kontrazeptiva	0,65	0,000
Schulbildung		
mittlerer Abschluss	0,61	0,000
höherer Abschluss	0,49	0,000

Diskussion

Obwohl es Mängel in der Methodik gab, konnte ein Selektionsbias festgestellt werden. Da einer der Einflussfaktoren auf eine Teilnahme an der Studie die Hormoneinnahme ist, die in der Fall-Kontroll-Gruppe als Exposition untersucht wird, ist bezüglich der Ergebnisse der MARIE zu den Einflüssen einer Hormontherapie auf Brustkrebs durch Selektionsbias einer Unterschätzung des *Odds Ratios* zu vermuten.

Schlussfolgerung

Die Methode der Kurzbefragung eignet sich hervorragend zur Erhebung von wichtigen Charakteristika und Einflussfaktoren unter Nicht-Teilnehmern. Die hohe Zahl von Teilnehmern aus frühen Geburtsjahrgängen oder mit gesundheitlichen Einschränkungen bieten die Möglichkeit, mit einem solchen Instrument besonders derartige Teilnehmergruppen anzusprechen. Das Ergebnis einer Kurzbefragung kann zur Diskussion (und eventuell zur Korrektur) von Studienergebnissen einer Fall-Kontroll-Studie beitragen und erhöht die Repräsentativität von Fall-Kontroll-Studien..

1 Einleitung

1.1 Beschreibung des Forschungsvorhabens

Der MARIE (Mammakarzinom-Risikofaktoren-Erhebung) fühle ich mich sehr verbunden, da ich schon seit dem Jahr 2000 an deren Planungsphase teilnehmen durfte und in vielen Bereichen der Studie (Planung und Entwicklung des Kurzfragebogens, Erstellung der Erhebungsinstrumente, Organisation, Datenerhebung, Durchführung der Umfragen mittels Kurz- und Hauptfragebogen, Qualitätssicherung und Auswertung) während des gesamten Erhebungszeitraumes beschäftigt war.

Bei der MARIE handelt es sich um eine Fall-Kontroll-Studie, die ihren Fokus auf die Untersuchung des Brustkrebsrisikos bei Frauen mit einer Hormontherapie während der Postmenopause (HRT) setzt. Aber auch andere Risiko- und Einflussfaktoren wurden erhoben, denn die Entstehung von Brustkrebs ist ein multifaktorielles Geschehen.

Brustkrebs ist die häufigste maligne Tumorerkrankung unter der weiblichen Bevölkerung Deutschlands: Die Inzidenz liegt laut Robert-Koch-Institut jährlich bei etwa 47500 Frauen [RKI 2005, pp 7-8]. Um primärpräventiv eingreifen zu können, ist es ein gesundheitspolitisches Anliegen, Risikofaktoren und andere Einflüsse auf das Krankheitsgeschehen zu ermitteln.

Bekannte Risikofaktoren sind das Alter (mit einem Erkrankungsgipfel zwischen dem 60. und 65. Lebensjahr), ethnische und genetische Faktoren (Mutationen der Tumorsuppressorgene BRCA-1 und -2 u.a. oder eine familiäre Häufung von Brustkrebs bei weiblichen Verwandten ersten Grades) und – da Brustkrebs ein hormonsensitiver Tumor ist – gynäkologische Faktoren (Kinderlosigkeit oder eine sehr späte Erstgeburt in einem Alter über 30 Jahren, ein großer Fertilitätszeitraum oder Vorerkrankungen der Brust) [RKI 2005, pp 14-15].

Weitere Einflussfaktoren, wie ein gesundheitsbewusstes Verhalten (z. B. Körperpflege, körperliche Aktivität, Vorsorgeverhalten) und Lebensstilfaktoren (Ernährungsverhalten, Adipositas, Rauchen und Alkoholkonsum), scheinen indirekt durch soziodemographische Faktoren wie Bildung, Erwerbstätigkeit, Einkommensverhältnisse beeinflusst zu werden, die aus diesem Grund auch durch die MARIE erhoben wurden [Lancaster et al. 1995].

Soziodemographische Merkmale charakterisieren nicht nur verschiedene soziale Gruppierungen, sondern wirken sich auch auf die Akzeptanz der Akteure des Gesundheitswesens aus, wie sich in der Häufigkeit von Arztbesuchen, im Vorsorgeverhalten und in der Compliance zeigt. Auch die Bereitschaft, an einer wissenschaftlichen Studie teilzunehmen, und nicht zuletzt die Erreichbarkeit der Teilnehmer an einer solchen Umfrage (Berufstätigkeit, Telefonanschluss) wird durch Merkmale der verschiedenen Bevölkerungsgruppen beeinflusst.

Um wichtige Informationen auch über die Nichtbeteiligten an einer Erhebung auswerten zu können, wurde den Nicht-Teilnehmerinnen an der MARIE die Gelegenheit geboten, einen Kurzfragebogen zu beantworten. In diesen wurden einige der oben genannten Risiko- und Einflussfaktoren gezielt übernommen, um auch von diesen Probandinnen Aussagen über die Exposition gegenüber solchen Faktoren zu erhalten und so genauere Risikoschätzer zu erhalten [Voigt et al 2003]. Gleichzeitig bietet diese zusätzliche Befragung die Möglichkeit, Informationen des Kurzfragebogens mit denen des MARIE-Hauptfragebogens in Bezug auf soziodemographische Merkmale und auf die Exposition gegenüber Risiko- und Einflussfaktoren zu vergleichen und beide Stichproben (Teilnehmer der MARIE und mit dem Kurzfragebogen befragte Nicht-Teilnehmer) auf Merkmale von Bevölkerungsgruppen zu untersuchen, die einen Einfluss auf die Teilnahme bzw. Nicht-Teilnahme an einer wissenschaftlichen Studie haben. Durch den direkten Vergleich kann auf diese Weise ein eventuell bestehender Auswahlfehler (Selektionsbias) zwischen Teilnehmern und Nicht-Teilnehmern ermittelt und beschrieben werden.

Mit dieser Arbeit soll untersucht werden, ob ein derartiger Selektionsbias besteht und welche Charakteristika der teilnehmenden Bevölkerungsgruppen auf ihn Einfluss nehmen. Die Public-Health-Relevanz bei der Beschreibung des Selektionsbias und der einflussnehmenden Faktoren besteht darin, derartige Faktoren bei der Durchführung künftiger Studien zu berücksichtigen, mögliche Selektionsfehler zu ermitteln und gegebenenfalls zur Korrektur der Ergebnisse zu nutzen.

1.2 Vollständigkeit und Repräsentativität

Eine Studie ist repräsentativ, wenn die Erfassungswahrscheinlichkeiten in allen Bevölkerungsgruppen der Grundgesamtheit, aus der Fälle und Kontrollen gezogen wurden, gleich groß ist. Die Repräsentativität hängt entscheidend von der Vollständigkeit einer Teilnahme aller gewählten Fälle und Kontrollen ab.

Eine hohe Vollständigkeit unter den Fällen wird durch eine Kombination sensitiver und spezifischer Erfassungskriterien erzielt. Im optimalen Fall soll eine Fall-Erhebung sowohl sensitiv (d. h. es werden alle Fälle erfasst) als auch spezifisch (d. h. es werden nur gesicherte Fälle erfasst) sein. Im Design der MARIE wurde versucht, die Vollständigkeit der Meldungen durch einen Abgleich der Meldungen von teilnehmenden Ärzten, Praxen und Kliniken mit den Meldungen bei den zuständigen Krebsregistern zu optimieren. Die Diagnose wurde durch einen histologischen Befundbogen gesichert. Da es sich bei Brustkrebs um eine populationsbezogene Erkrankung aus allen sozialen Schichten und Bevölkerungsgruppen handelt, ist unter den Fällen eine Repräsentativität für die gesamte Bevölkerung der Studienregionen zu unterstellen. Die Vollständigkeit der Teilnehmer unter den Fällen ist groß, denn die Erkrankung motiviert zur Teilnahme.

Um ein näherungsweise genaues Modell der Grundgesamtheit mit all ihren Charakteristiken als Referenzpopulation zu erhalten, wird die Gruppe der nicht erkrankten Vergleichspersonen (Kontrollen) als Zufallsstichprobe – populationsbezogen und randomisiert über die Einwohnermeldeämter – aus der Grundgesamtheit gezogen. Die Kontrollen wurden durch mehrfache, ausführlich erklärende Anschreiben und ein dem Anschreiben beigelegtes Incentive zur Teilnahme motiviert [Hoffmann et al 2004].

1.3 Selektionsbias

Der Selektionsbias ist der Auswahlfehler einer Studie. Wenn die Studienstichprobe nicht repräsentativ für eine Bevölkerung ist, so ist dies auch das Ergebnis der Studie nicht. Leider korrelieren häufig einflussnehmende Faktoren für die Teilnahme mit der Exposition von Risikofaktoren. Werden z. B. alle Fälle erhoben, aber beispielsweise durch einen Selektionsbias eine höhere Anzahl exponierter als nicht exponierter Kontrollen, so errechnet sich – vorausgesetzt dieser Selektionsbias besteht nur bei den Kontrollen und nicht auch bei den Fällen – ein halb so großes *Odds Ratio* für die Erkrankung (Tabelle 1).

Tabelle 1: Beispiel für eine Veränderung des Odds Ratios durch einen Selektionsbias

Beispiel				repräsentative Erhebung		biased	
Krankheit				Krankheit		Krankheit	
				+	-	+	-
Exposition	+	A	B	9	6	9	12
	-	C	D	3	9	3	9
Odds ratio		=(A*D)/(C*B)		=4,5		=2,25	

Der häufigste Grund für einen Selektionsbias ist die Ablehnung der Teilnahme an einer Studie. Erkrankte sind oft leichter zu einer Teilnahme zu motivieren als zufällig aus der Bevölkerung gezogene Kontrollen. Ferner kann die Motivation zur Teilnahme an einer Studie mit dem Bildungsstatus (Interesse an wissenschaftlichen Untersuchungen), der Erwerbssituation und anderen soziodemographischen Faktoren zusammenhängen. Auch das gesundheitsbewusste Verhalten, die Akzeptanz der Akteure, die Compliance gegenüber den Interviewern oder die Einstellung zu den untersuchenden Institutionen spielen eine Rolle. Und nicht zuletzt müssen Sprachbarrieren und möglicher gesundheitlicher Zustand (vor allem bei einer älteren Stichprobe) beachtet werden. Diese Merkmale können zu einer Unterrepräsentanz verschiedener Bevölkerungsgruppen führen und so zu einem Verlust der Repräsentativität eines Forschungsvorhabens.

Um zu beurteilen, ob die Nicht-Teilnehmer einer Studie zu einer systematischen Verzerrung geführt haben, müssen über eine zusätzliche Kurzbefragung studienrelevante Merkmale erhoben werden. [Hoffmann et al 2004]

1.4 Ziel der Untersuchung

Ziel dieser Untersuchung ist es, beide Stichproben auf vergleichbare Erfassungswahrscheinlichkeiten in allen Bevölkerungsgruppen zu prüfen und auf diese Weise zu beurteilen, ob es eine systematische Verzerrung in der Stichprobe der MARIE-Teilnehmer gab oder ob sie repräsentativ ist. Hierfür wurden anhand einer Kurzbefragung unter den Nicht-Teilnehmern der MARIE beide Stichproben verglichen, die Charakteristika der Bevölkerungsgruppen (soweit erfragt: biometrische und soziodemographische Merkmale, Hormoneinnahme in den Wechseljahren und gesundheitsbewusstes Verhalten) auf Unterschiede untersucht und bestehende Unterschiede dargestellt. Unterschiede der Charakteristika beider Stichproben würden auf eine Verzerrung durch einen Auswahlfehler infolge unterschiedlicher Teilnahme aus verschiedenen Bevölkerungsgruppen, also einen Selektionsbias, deuten.

Die Forschungsfragen lauten daher:

1. Gibt es zwischen den Teilnehmerinnen des Kurzfragebogens (absagende Probandinnen) und den Teilnehmerinnen der MARIE Unterschiede bezüglich der erhobenen Merkmale, die auf eine unterschiedliche Erfassung von Bevölkerungsgruppen schließen lassen, d. h. gibt es eine Verzerrung im Sinne eines Selektionsbias?
2. Welche Merkmale der Teilnehmerpopulationen haben den größten Einfluss auf den Selektionsbias?

1.5 Hypothesenstellung

Eine Hypothese lässt sich nur für die erste Forschungsfrage formulieren.

Die Nullhypothese H_0 bezüglich der ersten Forschungsfrage lautet daher:
Zwischen den Stichproben bestehen keine Unterschiede der erhobenen Merkmale, somit besteht kein Selektionsbias.

Die Alternativhypothese H_1 lautet:
Zwischen den Stichproben bestehen signifikante Unterschiede der erhobenen Merkmale, somit besteht ein Selektionsbias.

Die zweite Forschungsfrage soll explorativ beantwortet werden.

2 Methodik

2.1 Daten

2.1.1 Datenquellen

Die MARIE (Mammakarzinom-Risikofaktoren-Erhebung) ist eine Fall-Kontroll-Studie zur Hormonsubstitution, zu anderen Brustkrebs-assoziierten Risikofaktoren und zur Brustkrebserkrankung selbst. Die MARIE wurde bizen­trisch in der Region Nord (Erhebungsgebiet Hansestadt Hamburg, Studienzentrum Hamburg) durch das Universitätskrankenhaus Eppendorf (UKE), Hamburg, und in der Region Süd (Erhebungsgebiet Rhein-Neckar-Karlsruhe, Studienzentrum Heidelberg) durch das Deutsche Krebsforschungszentrum (DKFZ) Heidelberg erhoben. Auftraggeber der Studie ist die Deutsche Krebs­hilfe e.V. in Bonn.

Geplant war, 3500 Fälle – in der Region Nord sowohl retrospektiv (erhoben vom Hamburger Krebsregister mit einer Erst-Diagnose an Brustkrebs ab 2001) als auch prospektiv (erhoben von Kliniken und Praxen), in der Region Süd ausschließlich prospektiv – und 7000 nach Geburtsjahrgang und Häufigkeit im Verhältnis 1:2 gematchte Kontrollen zu interviewen.

Nach Erhebungsende lagen 11154 gültige Interviews von 3813 Fällen (2074 aus der Region Nord, 1739 aus der Region Süd) und von 7341 Kontrollen (4008 aus der Region Nord, 3333 aus der Region Süd) vor.

Zusätzlich zu den Fragebogendaten wurden für die Fälle die histologischen Pathologiedaten und seit dem 1. Juli 2003 auch Blutwerte zur genetischen Analyse erhoben.

Die Daten des Kurzfragebogens wurden von Dipl.-Gesundheitswirt Daniel Engel und Dipl.-Dokumentarin Irene Walter im Rahmen ihrer Diplomarbeit [Walther 2006] mit einer Eingabemaske digitalisiert.

2.1.2 Datenschutz

Die MARIE erhebt Individualdaten, die wegen eines geplanten Follow-up's individuell kenntlich gemacht wurden und die auch eine gewisse Zeit gespeichert werden müssen.

Die Speicherung persönlicher Daten führt in der Bevölkerung häufig zu Verunsicherungen, kann zu einer Abnahme der Akzeptanz und einer geringeren Response Rate der Studie führen und sich so auf die Datenqualität auswirken. Außerdem unterliegt die Speicherung personenbezogener Daten verschiedenen Gesetzen: Am 24. Oktober 1995 wurde die „Richtlinie zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten“ vom Europäischen Parlament [Richtlinie 95/46/EG] verabschiedet und in verschiedene Bundes- und Landesgesetze übernommen. So stellt die Erhebung, Verarbeitung und Speicherung personenbezogener Daten nach Bundesdatenschutzgesetz §§4,5 [Bundesministerium der Justiz: BDSG], Hamburgischem Datenschutzgesetz §§5,7 [HmbDSG], Hamburgischem Gesundheitsdienstgesetz §§24-28 [HmbGDG], Landesdatenschutzgesetz Baden-Württemberg §§ 4,6,13-15 [LDSG] und dem Sozialgesetzbuch §35 (Sozialgeheimnis) [SGB] einen Eingriff in die Persönlichkeitsrechte dar und unterliegt einem strengen Reglement.

In Absprache mit den Datenschutzbeauftragten und der Ethikkommission wurde festgelegt, dass die persönlichen Daten nach erfolgter Eingabe und Verschlüsselung bei einem Notar hinterlegt und spätestens 2010 gelöscht werden. Andernfalls müssen die Probanden erneut um eine schriftliche Einwilligung gebeten werden. Nach Eingabe, Qualitätskontrolle und Plausibilitätsprüfungen soll nur noch mit verschlüsselten Daten gearbeitet werden.

Deshalb wurden die Probandinnen (in Absprache mit dem Ethikrat und den Datenschutzbeauftragten) vor weiterer Kontaktaufnahme umfassend über den Inhalt und den Zweck der Befragung informiert und dann telefonisch oder per Anschreiben um eine Teilnahme gebeten.

Die Einwilligung zur Verwendung persönlicher Daten erfolgte schriftlich und war auf das dargestellte Forschungsvorhaben beschränkt. Die Einwilligung zur Blutentnahme wurde gesondert erfragt und konnte separat erteilt werden. Die Probandinnen wurden auf ihr Recht des Widerrufs hingewiesen.

2.1.3 Datenerhebung

2.1.3.1 Datenerhebung des MARIE-Hauptfragebogens

Die MARIE erhob Individualdaten. Fälle und Kontrollen wurden auf unterschiedliche Weise kontaktiert:

Die Fälle wurden der MARIE von unabhängigen Meldern (Ärzte, Mitarbeiter von Studien) in einem Passiv-System gemeldet. Alle Melder wurden mit einem unkomplizierten, standardisierten Erfassungsinstrument (Meldebogen) ausgestattet und bezüglich der Faldefinition und der Eingabe in das Erfassungsinstrument geschult. Um die Melder zu motivieren, wurde ständig mit ihnen Kontakt gehalten und erste Ergebnisse der MARIE im Sinne einer Rückkoppelung in Interviewer- und Melder-Treffen kommuniziert. Gab eine Probandin das Einverständnis zur Teilnahme an der Hauptbefragung, wurden die Daten in einem face-to-face-Interview erhoben. Im Falle einer Absage wurde der Kurzfragebogen von den Meldern nicht erhoben.

Waren die Fälle für die Melder nicht mehr erreichbar, wurden sie von den jeweiligen Studienzentren angeschrieben und um eine Teilnahme gebeten. Dem dritten Anschreiben wurde der Kurzfragebogen für Nicht-Teilnehmer beigelegt, da laut Brögger et al,xx

Fälle wurden also nicht standardisiert mit dem Kurzfragebogen befragt.

Die Kontrollen wurden randomisiert und populationsbezogen aus den Einwohnermeldeämtern der jeweiligen Region gezogen, vorab angeschrieben, über die Studie informiert und um ein Einverständnis zur Teilnahme gebeten. Waren die Kontrollen in einem öffentlichen Verzeichnis für Telefonanschlüsse verzeichnet, wurden sie telefonisch kontaktiert.

Wenn dies nicht gelang oder die Kontrollen nicht mit ihrer Telefonnummer verzeichnet waren, wurden sie noch zweimal angeschrieben und um eine Teilnahme an der MARIE gebeten. Auch hier wurde dem dritten Anschreiben der Kurzfragebogen für Nicht-Teilnehmer beigelegt.

Gab eine Probandin das Einverständnis zur Teilnahme an der Hauptbefragung, wurden die Daten in einem face-to-face-Interview erhoben.

Die Erhebungsphase der MARIE dauerte von 2002 bis 2006 (Tabelle 2).

Tabelle 2: Erhebungsphase der MARIE (Teilnehmer des Hauptfragebogens)

Erhebung		Region Nord		Region Süd	
		Fälle	Kontrollen	Fälle	Kontrollen
Beginn	retrospektiv	02.01.2001			
	prospektiv	01.09.2002	04.12.2002	01.08.2002	16.12.2002
Ende		30.09.2005	15.02.2006	31.07.2005	02.03.2006

2.1.3.2 Datenerhebung des MARIE-Kurzfragebogens

Der Kurzfragebogen wurde am 30.4.2003 in der Teamsitzung der Region Nord beschlossen. Die Vorgabe war die Entwicklung eines kurzen und leicht verständlichen Fragebogens, mit dem bei Nicht-Teilnehmerinnen der MARIE einige Items erfragt werden sollten, die Kernfragen der fünf Themengebiete des Hauptfragebogens enthielten.

Als Pretest habe ich am 5. Mai 2003 zehn Probandinnen mit einem offenen Fragebogen nach Hormoneinnahmen, Kindern, Schulbildung und Rauchverhalten befragt. Da hierbei alle zehn Nicht-Teilnehmerinnen der MARIE bereitwillig Auskunft gaben, habe ich in Absprache mit der Studienleitung Hamburg einen standardisierten Kurzfragebogen entwickelt, der seit dem 7. Mai 2003 eingesetzt wurde (Tabelle 3).

Tabelle 3: Erhebungsphase des Kurzfragebogens (Nicht-Teilnehmer der MARIE)

Erhebung		Region Nord		Region Süd	
		Fälle	Kontrollen	Fälle	Kontrollen
Beginn	retrospektiv	22.01.2001			
	prospektiv	06.11.2003	07.05.2003		28.08.2003
Ende		26.08.2005	08.01.2006		13.03.2006

Die Daten wurden telefonisch durch geschulte Interviewer(innen) erhoben. Diese waren spezifisch in der Durchführung telefonischer Kontaktaufnahme, im Vertrauensaufbau und in der Interviewführung geschult und nachgeschult worden.

2.1.4 Erhebungsinstrumente

2.1.4.1 Der MARIE-Fragebogen

Der standardisierte MARIE-Fragebogen wurde schon seit dem Planungsbeginn der Studie im Jahr 1999 von Dipl.-Gesundheitswirtin Zita Schillmöller zur Erhebung von Primärdaten entwickelt, mit einem Pretest im Jahr 2002 getestet, im Jahr 2003 validiert und im Jahr 2004 reliabilisiert. Im Laufe der Erhebungsphase wurde die Formulierung einiger Fragen geringfügig variiert.

Der MARIE-Fragebogen umfasst in 98 Items neben organisatorischen Angaben fünf Themengebiete:

- Biometrische Maße
- Soziodemographische Faktoren
- Gynäkologische Anamnese
- Familienanamnese
- Weitere Einflussfaktoren (Rauchen, Alkoholkonsum, körperliche Aktivität, Arbeitsplatz- und Umweltbelastungen)

Die mit dem Kurzfragebogen vergleichbaren Items können im Anhang (Anhang 4) eingesehen werden.

Der MARIE-Fragebogen wurde in einem face-to-face-Interview erhoben. Die Interviewer(innen) waren vor Studienbeginn geschult, während der Studie einem Monitoring unterzogen, nachgeschult und einer Qualitätskontrolle (Aufnahme und Prüfung des Interviewstils und der standardisierten Durchführung) unterzogen worden.

In einem separaten 32-seitigen standardisierten Fragebogen konnten von den Teilnehmerinnen Angaben zu einem weiteren Einflussfaktor, dem Ernährungsverhalten (in Bezug auf Art und Zubereitung der Speisen, Portionsgröße und Häufigkeit des Verzehrs) gemacht werden.

2.1.4.2 Kurzfragebogen für Nicht-Teilnehmer

Der Kurzfragebogen beinhaltet einige Fragen zu relevanten Brustkrebs-assoziierten Risikofaktoren in Anlehnung an die in der MARIE verwendeten Items. In vier Themengebieten wurden folgende HRT-assoziierte Fragestellungen [Duetz et al. 2000] erfragt:

- Biometrische Maße: Größe, Gewicht
- Soziodemographische Faktoren: Geburtsdatum, Schulbildung und Erwerbstätigkeit
- Gynäkologische Anamnese:
 - Anzahl der Kinder
 - Kontrazeptiva
 - Hormoneinnahme
- Weitere Einflussfaktoren:
 - Rauchen
 - Verzehr von Salat und Gemüse

Die Familienanamnese wurde im Kurzfragebogen nicht erhoben.

Es gibt fünf Fragebogen-Versionen, in denen im Laufe der Erhebungszeit die Art und Anzahl der Items veränderte: Der erste Kurzfragebogen beinhaltete nur fünf Items, die aber schon nach einem Monat auf die endgültigen acht Items ausgeweitet wurden. Andere Items wurden teilweise in der Formulierung der Fragen erheblich verändert (Tabellen 27 bis 31, Anhang).

Der Kurzfragebogen wurde mit einer telefonisch zu erhebenden Version (Anhang 2) erhoben. Wurde eine Probandin aber bis zum dritten Anschreiben nicht erreicht, wurde eine Version zum selbst ausfüllen (Anhang 3) mit Rückumschlag beigelegt.

2.1.5 Einschlusskriterien für Teilnehmerinnen

Für eine Teilnahme an der MARIE wurden folgende standardisierte Kriterien festgelegt, um eine einheitliche Datenqualität zu gewährleisten:

- Weibliches Geschlecht
- Wohnort zum Diagnosezeitpunkt in einer der Studienregionen
- Fähigkeit, dem Interview in deutscher Sprache zu folgen
- Ausreichender Gesundheitszustand, am Interview teilzunehmen

Weitere Kriterien für Fälle:

- Diagnose Brustkrebs: Die Fälle der Studie wurden durch das Krebsregister bzw. die meldenden Ärzte und Kliniken nach ICD erfasst und die Diagnose durch histologische Befunde gesichert.
- Alter zum Diagnosezeitpunkt zwischen 50 und weniger als 75 Jahren.

Weitere Kriterien für Kontrollen:

- Die Kontrollen wurden nach Geburtsjahrgang und Studienregion mit einer Häufigkeit von 1:2 gematcht und per Zufallsauswahl aus dem Einwohner-Meldereister der jeweiligen Region gezogen.
- Alter zum Befragungszeitpunkt (durch Meldeverzögerungen bei den Fällen) zwischen 50 und weniger als 80 Jahren.
- Zum Zeitpunkt des Interviews ist keine Brustkrebserkrankung bekannt.

2.1.6 Datacleaning

2.1.6.1 MARIE-Arbeitsdatei

Da die Datensätze der MARIE-Teilnehmerinnen zu diesem Zeitpunkt noch bearbeitet werden, verwende ich hier eine Datei mit dem Stand vom 11. Dezember 2006. Sie enthält 11154 Datensätze:

3813 von Fällen (2074 aus der Region Nord, 1739 aus der Region Süd), und 7341 von Kontrollen (4008 aus der Region Nord und 3333 aus der Region Süd). Weiter wurden für die Variablenbildung die Dateien „*referenzdatum.sav*“ (Stand 28.11.2006), „*ht_use.sav*“ (Stand 23.10.2006) und „*SalatGemuese.sas7bdat*“ (Stand 07.12.2006) verwendet.

2.1.6.2 Arbeitsdatei Kurzfragebogen

Vom Studienzentrum Süd wurde am 14. August 2006 eine endgültige Access-Datei „*2006-08-14_NR_HD.mdb*“ mit 1397 Datensätzen geschickt.

Aus der Studienregion Nord liegt seit dem 23.11.2006 die Access-Datei „*2006_11_23_NR_HH.mdb*“ mit den beiden Tabellen „*Tab_Haupt*“ (3115 Datensätze) und „*Tab_Zusagen_Ausschlüsse*“ (165 Datensätze) vor.

Diese drei Dateien wurden in Variablennamen, -werten und -labeln einander angepasst und zu einer gemeinsamen Datei zusammengeführt.

Nach Ausschluss einiger Fragebögen, die

- nicht ausgefüllt waren,
- die gleiche Studiennummer aufwiesen, oder die von Teilnehmerinnen stammten, die
- sich zusätzlich für eine Teilnahme an der MARIE entschieden hatten,
- zu jung waren,
- im Kommentarfeld Hinweis auf ein Ausschlusskriterium gegeben hatten (aus medizinischen Gründen nicht befragbar, prämenopausal, schon verstorben und Ersatzinterview mit einem Verwandten),

resultierte eine Arbeitsdatei („*nr_arbeitsdatei*“, 3168 Datensätze, 157 Variablen) mit Teilnehmerinnen, die den Kurzfragebogen mit mindestens einer Frage beantwortet hatten.

Variablenamen, -werte und -label, sowie fehlende Werte wurden an die Datei der MARIE angepasst. Verschiedene Ausprägungen der fehlenden Werte („nicht gefragt“, „fehlt“, „keine Angabe/weiß nicht“) wurden vereinheitlicht und auf den Wert „unbekannt“ gesetzt. Insbesondere wurden die Angaben aus den Freitextfeldern berücksichtigt, Medikamente nach vorliegender Code-Liste codiert und die Hormonangaben bereinigt.

2.1.6.3 Vergleichbare Items beider Dateien

Da der Kurzfragebogen sich zwar am Hauptfragebogen orientiert, aber nicht die gleichen Fragen stellt, mussten vergleichbare Items geschaffen werden.

Das Geburtsdatum der Teilnehmerinnen der MARIE liegt im Datumsformat vor. Bei den Teilnehmerinnen des Kurzfragebogens wurde das Datum teilweise imputiert und auf den 1. Juli des Geburtsjahres gesetzt. Zwei Teilnehmerinnen aus der Studienregion Süd, für die sich dadurch ein Alter von 49 Jahren errechnete, wurden auf ein Alter von 50 Jahren gesetzt.

In der MARIE wird nach der Anzahl der Schwangerschaften und ihrem Ergebnis gefragt. Aus diesen Angaben wurde die Anzahl der lebendgeborenen Kinder errechnet, um sie mit der Anzahl der Kinder aus dem Kurzfragebogen vergleichen zu können.

Im Hauptfragebogen wird die Angabe der Kontrazeptiva differenziert in Art und Grund der Einnahme. Aus diesen Angaben wurde die Einnahme oraler Kontrazeptiva, die jemals (egal aus welchem Grund) eingenommen wurden, herausgefiltert, um sie mit der Angabe des Kurzfragebogens vergleichen zu können.

Die Einnahme von postmenopausalen Präparaten wurde im Hauptfragebogen nur allgemein erfragt, die Differenzierung nach hormonellen und pflanzlichen Präparaten erfolgte erst nach Auswertung der verschiedenen namentlichen Angaben in verschiedenen Einnahmezeiträumen. Im Kurzfragebogen wurde gezielt nach Einnahme hormoneller und/oder pflanzlicher Mittel gefragt.

Im Hauptfragebogen konnte die Dauer der Hormoneinnahme aus den einzelnen Einnahmezeiträumen errechnet werden, wogegen im Kurzfragebogen nur pauschal nach Beginn und Ende der Hormoneinnahme gefragt wurde. Eine abweichende Dauer (durch Einnahmepausen oder Einnahme pflanzlicher Präparate verursacht) wurde im Freitext vermerkt.

Die Angabe der aktuellen Hormoneinnahme wurde korrigiert, da im MARIE-Hauptfragebogen innerhalb der letzten sechs Monate eine beendete Hormoneinnahme als aktuelle Hormoneinnahme gilt. Im Kurzfragebogen wurde deshalb ein Ende der Hormoneinnahme innerhalb des Befragungsjahres als aktuelle Hormoneinnahme interpretiert.

Da im Kurzfragebogen die Angabe eines am längsten gehaltenen Gewichtes fehlt, wird der Body Mass Index (BMI), der Quotient aus Körpergewicht (in kg) und quadrierter Körperhöhe (in m²), in beiden Dateien aus dem aktuellen Gewicht errechnet. Seine Einteilung erfolgt sowohl nach WHO in sechs Klassen [WHO 1997] (Tabelle 4) als auch vergleichbar mit einer Variablen der MARIE-Auswertung dichotom (+/-25).

Tabelle 4: Gewichtsklassifikation bei Erwachsenen anhand des BMI

Kategorie	BMI	Risiko für Begleiterkrankungen
Untergewicht	< 18.5	niedrig
Normalgewicht	18.5 – < 25	durchschnittlich
Übergewicht	>= 25	
Präadipositas	25 – < 30	gering erhöht
Adipositas Grad I	30 – < 35	erhöht
Adipositas Grad II	35 – < 40	hoch
Adipositas Grad III	>= 40	sehr hoch

Die Frage nach der Erwerbstätigkeit wurde in beiden Fragebögen gleichlautend gestellt, benötigte also keine Korrekturen.

Das Rauchverhalten konnte durch Veränderung der Fragestellung in den verschiedenen Versionen des Kurzfragebogens nur für regelmäßiges Rauchen angegeben werden. Die entsprechenden Angaben des Hauptfragebogens wurden gebildet.

Bei den Teilnehmern der MARIE wurde das Ernährungsverhalten in einem 32-seitigen, selbst auszufüllenden Fragebogen angegeben. Aus diesen sehr differenzierten Angaben einzelner Nahrungsmittel, ihrer Portionsgröße und der Häufigkeit ihres Verzehrs wurde versucht, ein Äquivalent für die Angaben des Kurzfragebogens zu bilden. Dieser erhob pauschal lediglich die durchschnittliche Verzehrhäufigkeit von Salat und Gemüse pro Woche in fünf Kategorien.

Da die Angaben aus beiden Erhebungen auffällig stark voneinander abwichen, musste angenommen werden, dass hier Fehler aufgrund der unterschiedlichen Fragestellung vorlagen. Die Angaben zum Verzehr von Salat und Gemüse wurden deshalb nicht für die Untersuchung des Selektionsbias hinzugezogen.

Nach Zusammenfügen von Haupt- und Kurzfragebögen ergaben sich 14322 Datensätze in 52 Variablen.

2.1.7 Auswahl der Datensätze für die Untersuchung des Selektionsbias

Bei einer ersten Berechnung schien die Tatsache erkrankt zu sein (ein Fall zu sein) oder auch der Studienregion Nord anzugehören, die Entscheidung einer Probandin, am Haupt- oder Kurzfragebogen teilzunehmen, zu beeinflussen: Da die Fälle (an Brustkrebs erkrankte Teilnehmerinnen) nicht standardisiert mit dem Kurzfragebogen kontaktiert worden waren, war die Anzahl der an Brustkrebs erkrankten Frauen, die am Kurzfragebogen teilnahmen, gering. Dadurch entstand der irreführende Eindruck, dass an Brustkrebs erkrankte Frauen eher am Hauptfragebogen teilgenommen hätten.

Da der Kurzfragebogen in der Region Nord entwickelt wurde, begann seine Erhebungsphase hier sechs Monate früher als in der Region Süd. Deshalb entstand der fälschliche Eindruck, dass im Verhältnis zur MARIE mehr Teilnehmerinnen der Region Nord als aus der Region Süd am Kurzfragebogen teilgenommen hätten.

Um diese Einflüsse zu eliminieren, wurde eine Anpassung der Daten vorgenommen: An Brustkrebs erkrankte Teilnehmerinnen wurden wegen der oben erwähnten Verzerrung und der kleinen Fallzahl unter den Teilnehmerinnen des Kurzfragebogens von der weiteren Untersuchung ausgeschlossen.

Ebenso wurden alle Kontrollen ausgeschlossen, die vor dem Erhebungszeitraum des Kurzfragebogens in der jeweiligen Studienregion befragt wurden.

Auf diese Weise entfielen:

- 147 Datensätze der Arbeitsdatei Kurzfragebogen (Fälle),
- 3813 Datensätze der MARIE-Arbeitsdatei (Fälle) und
- 833 Datensätze der MARIE-Arbeitsdatei (Kontrollen, die außerhalb des Erhebungszeitraumes des Kurzfragebogens befragt wurden).

Es verblieb eine gemeinsame Arbeitsdatei mit 9529 Datensätzen (Deskription unter 2.2).

2.2 Deskription der Arbeitsdatei

Die endgültige Datei „*ARBEITSDATEI_MARIE_NR_FINAL_co_dipl.sav*“ beinhaltet 9529 Datensätze in 22 Variablen, ausschließlich von nicht erkrankten Teilnehmerinnen (Kontrollen), die im gleichen Erhebungszeitraum befragt wurden. Die Deskription erfolgt ausführlich, da die oben genannte Arbeitsdatei meiner Diplomarbeit vor Publikation der Ergebnisse der MARIE nicht beigelegt werden darf.

Zur Deskription der Daten wurden jeweils die Median-Werte für die stetigen Variablen, die absoluten und prozentualen Häufigkeiten für die kategorialen Variablen angegeben. Dabei wurden teilweise Abweichungen nach Studienregion berücksichtigt.

2.2.1 Organisatorische Variablen

Die Variable „*StuDiennummer*“ (ID) ist eine in der Studienregion Nord siebenstellige, in der Studienregion Süd eine bis zu fünfstelligen fallbezogene, eindeutige Identifikationsnummer ohne fehlende Werte.

Die Variable „*StuDienzentrum*“ (zentrum) ist eine dichotome Variable für die Studienregion mit den Ausprägungen: „Region Nord (Hamburg, Studienzentrum Hamburg)“ n=5411 (56,78%) und „Region Süd (Rhein-Neckar-Karlsruhe, Studienzentrum Heidelberg)“ n=4118 (43,22%).

Tabelle 5: Verteilung der Teilnehmerinnen beider Fragebögen auf die Studienregionen

Teilnahme NR/MARIE			
	NR	MARIE	Gesamt
Region Nord	1701 (17,95%)	3710 (38,93%)	5411 (56,78%)
Region Süd	1320 (13,85%)	2798 (29,36%)	4118 (43,22%)
Gesamt	3021 (31,70%)	6508 (68,30%)	9529 (100%)

2.2.2 Die abhängige Variable

Die abhängige Variable in einem statistischen Modell ist diejenige Zielvariable, deren Ausprägungen durch eine oder mehrere andere Variablen erklärt oder vorhergesagt werden soll.

Für das zu berechnende Modell der logistische Regression (Logit-Modell) ist die Variable „*Teilnahme NR/MARIE*“ (NR_Marie) die abhängige Variable. Sie ist eine dichotome (binäre) Variable mit den Ausprägungen: „NR“ (Non-Responder) für eine Teilnehmerin am Kurzfragebogen, n=3021 (37,7%) bzw. „MARIE“ für eine Teilnehmerin am MARIE-Hauptfragebogen, n=6508 (68,3%). In der Arbeitsdatei bezeichnet der Wert „1“ die Teilnahme am Kurzfragebogen, der Wert „2“ die MARIE-Teilnahme.

Im Logit-Modell wird die Teilnahme an der MARIE als Basiswert definiert (auf 0 gesetzt). Zielvariable ist die Teilnahme am Kurzfragebogen, d. h. alle Merkmale werden daraufhin untersucht, ob und mit welcher Stärke sie mit einer Absage am Hauptfragebogen zusammenhängen. Die errechneten *Odds Ratio*'s setzen abhängige und unabhängige Variablen ins Verhältnis. Im vorliegenden Modell gibt das errechnete *Odds Ratio* für eine Merkmalsträgerin – im Vergleich zu einer Trägerin des Referenzmerkmals – an, nicht an der MARIE teilzunehmen (z. B. wie hoch die Chance einer untergewichtigen im Verhältnis zu einer normalgewichtigen Frau ist, nicht an der MARIE teilzunehmen).

2.2.3 Unabhängige Variablen

In einem statistischen Modell ist eine unabhängige Variable diejenige Variable, aus deren Werten die Ausprägungen einer abhängigen Variablen erklärt oder vorhergesagt werden soll. Um einen ersten Eindruck von den erklärenden Variablen selbst und ihrem Zusammenhang zur abhängigen Variablen (den beiden Teilnehmergruppen) zu gewinnen, werden sie kurz beschrieben. Es werden erste bivariate Betrachtungen durchgeführt und Kontingenztafeln für die abhängige Variable „*Teilnahme NR/MARIE*“ (NR_Marie) und die in Frage kommenden erklärenden Variablen gebildet. Die weitere bivariate Analyse wird unter 2.3. weiter ausgeführt.

2.2.3.1 Biometrische Merkmale

Die Variable „BMI in Klassen (aktuelles Gewicht)“ (BMI_Klassen) ist eine kategoriale Variable mit sechs Ausprägungen nach der Einteilung der WHO [WHO 1997] in Untergewicht ($<18,5 \text{ kg/m}^2$), Normalgewicht ($18,5-<25 \text{ kg/m}^2$), Präadipositas ($25-<30 \text{ kg/m}^2$), Adipositas Grad1 ($30-<35 \text{ kg/m}^2$), Grad2 ($35-<40 \text{ kg/m}^2$) und Grad3 ($\geq 40 \text{ kg/m}^2$). Der Body Mass Index wurde auf der Grundlage des aktuellen Gewichtes zum Befragungsdatum errechnet, die Variable enthält 9362 gültige und 167 fehlende Werte. 139 (1,5%) Probandinnen sind untergewichtig, 4129 (44,1%) sind normalgewichtig und 5094 (54,4%) Probandinnen sind übergewichtig (hier wurden alle Unterkategorien von Übergewicht zusammengefasst). Bei der Differenzierung nach Region haben Teilnehmerinnen aus der Region Süd einen etwas höheren Body Mass Index: In der Region Nord sind 48,0% der Teilnehmerinnen normalgewichtig und 49,8% übergewichtig, in der Region Süd sind 39,0% der Teilnehmerinnen normalgewichtig und 59,9% übergewichtig.

Bezüglich der abhängigen Variablen scheint die Variable „BMI in Klassen (aktuelles Gewicht)“ keine starke Abhängigkeit vom Teilnahmestatus zu haben.

Tabelle 6: BMI in Klassen (aktuelles Gewicht) nach Studienregion

BMI in Klassen [kg/m^2] (aktuelles Gewicht)		Studienzentrum		Gesamt
		Region Nord	Region Süd	
<18,5: Untergewicht	Anzahl	98	41	139
	% von Studienzentrum	1,9%	1,0%	1,5%
18,5-<25: normalgewichtig	Anzahl	2542	1587	4129
	% von Studienzentrum	48,0%	39,0%	44,1%
25-<30: Präadipositas	Anzahl	1798	1523	3321
	% von Studienzentrum	34,0%	37,4%	35,5%
30-<35: Adipositas Grad 1	Anzahl	626	676	1302
	% von Studienzentrum	11,8%	16,6%	13,9%
35-<40: Adipositas Grad 2	Anzahl	176	186	362
	% von Studienzentrum	3,3%	4,6%	3,9%
≥ 40 : Adipositas Grad 3	Anzahl	55	54	109
	% von Studienzentrum	1,0%	1,3%	1,2%
Gesamt	Anzahl	5295	4067	9362
	% von Studienzentrum	100,0%	100,0%	100,0%

2.2.3.2 Soziodemographische Merkmale

Die Variable „Geburtsjahr“ (gebjahr) ist eine stetige Variable, die das Geburtsjahr der Teilnehmerinnen angibt, es gibt keine fehlenden Werte. Der Median liegt im Geburtsjahr 1940 (Spannweite 1926 bis 1955).

Untersucht man nach der abhängigen Variablen, so weichen beide Teilnehmergruppen im Median voneinander ab: der Median der MARIE-Teilnehmerinnen liegt im Geburtsjahrgang 1941, der der Kurzfragebogen-Teilnehmerinnen im Jahrgang 1938. Der folgende Boxplot (Abb. 1) zeigt, dass die Teilnehmerinnen des MARIE-Hauptfragebogens unabhängig von der Studienregion in etwas späteren Geburtsjahrgängen geboren sind.

Wird nach Region differenziert, sind Teilnehmerinnen der Region Nord in früheren Geburtsjahrgängen geboren.

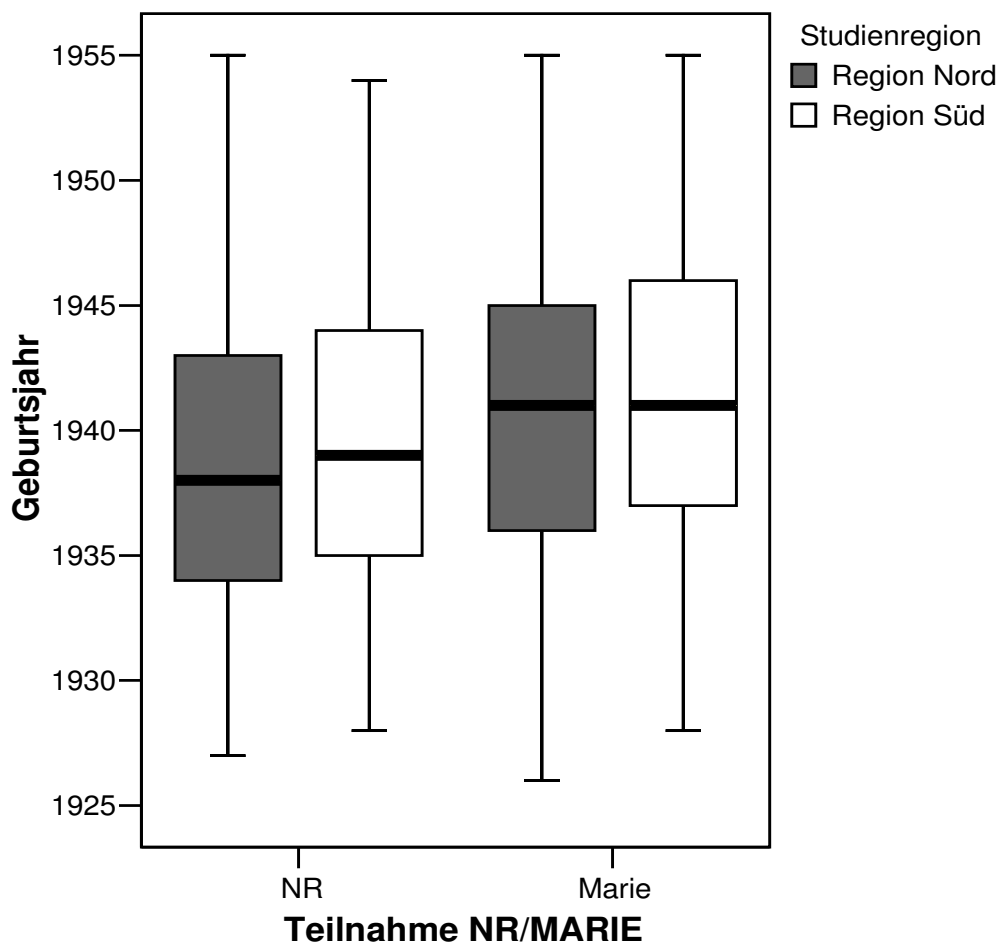


Abb. 1: Box-and-Whiskers-Plot der Geburtsjahrgänge nach Region und Fragebogenteilnahme

Die Variable „*Geburtsjahr in Klassen*“ (gebjahr_class) ist eine kategoriale Variable, die das Geburtsjahr (in Klassen von jeweils fünf Jahren gegliedert) angibt, es gibt keine fehlenden Werte. 1857 Probandinnen (19,5%) sind in den Geburtsjahren bis 1934 geboren, 2706 (28,4%) in den Jahren 1935 bis 1939, 2366 (24,8%) in den Jahren 1940 bis 1944, 1968 (17,6%) in den Jahren 1945 bis 1949 und 922 (9,7%) ab dem Geburtsjahr 1950.

Bei der Differenzierung nach Teilnahmepopulation fallen Unterschiede auf: Bei den MARIE-Teilnehmerinnen sind die Klassen „1935-1939“ (27,3%) und „1940-1944“ (27,0%) am stärksten besetzt, bei den Teilnehmerinnen des Kurzfragebogens sind es die Klassen „bis 1934“ (27,2%) und „1935-1939“ (30,9%).

Die Variable „*Referenzalter*“ (refalter) ist eine stetige Variable, die das auf ganze Zahlen trunkierte Alter der Teilnehmerinnen zum Referenzdatum angibt. Es gibt keine fehlenden Werte. Der Median liegt bei einem Alter von 64 Jahren (Spannweite 50 bis 79 Jahre).

Im Hinblick auf die abhängige Variable sind die Teilnehmerinnen der MARIE mit einem Median bei 63 Jahren etwas jünger als die des Kurzfragebogens mit einem Median bei 65 Jahren.

Diese Altersdifferenz beider Teilnehmergruppen zeigt sich auch in der regionalen Verteilung: Teilnehmerinnen beider Fragebögen sind in der Region Nord älter (siehe Abb. 2).

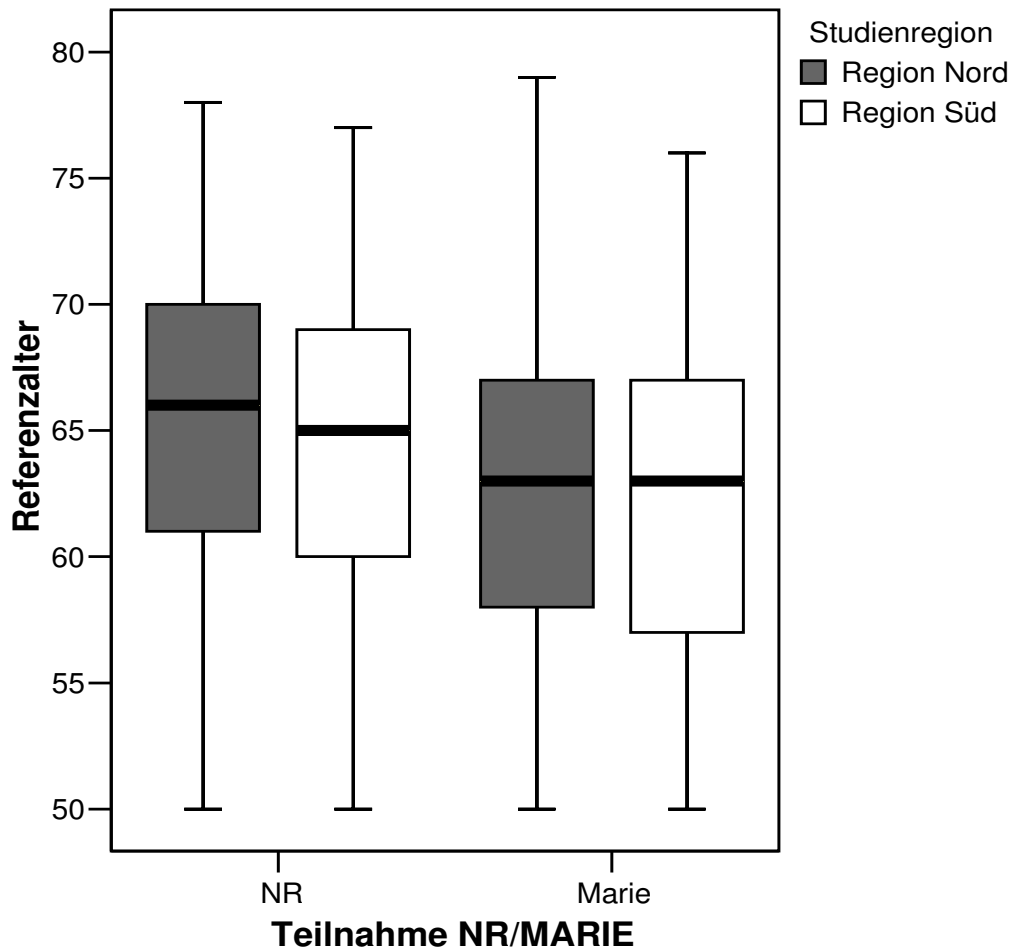


Abb. 2: Box-and-Whiskers-Plot des Referenzalters nach Region und Fragebogenteilnahme

Die Variable „Altersklassen Referenzalter“ (refalter_klass) ist eine kategoriale Variable und gibt das Referenzalter (in Klassen von jeweils fünf Jahren gegliedert) an. 1021 (10,7%) Probandinnen sind unter 55 Jahre alt, 1703 (17,9%) sind 55 bis 59 Jahre alt, 2434 (25,5%) sind 60 bis 64 Jahre alt, 2654 (27,9%) sind 65 bis 69 Jahre alt, 1467 (15,4%) sind 70 bis 74 Jahre alt und 250 (2,6%) Probandinnen sind älter als 74 Jahre.

Wie schon in der stetigen Variablen fallen auch hier Unterschiede im Hinblick auf die abhängige Variable auf: Bei den MARIE-Teilnehmerinnen sind die Klassen „60-64 Jahre“ (27,4%) und „65-69 Jahre“ (26,9%) am stärksten besetzt, bei den Teilnehmerinnen des Kurzfragebogens ist es die Klasse „65-69 Jahre“ (29,9%). Klasseneinteilung und Unterschiede nach Teilnehmergruppen ersieht man aus dem Histogramm (Abb. 3).

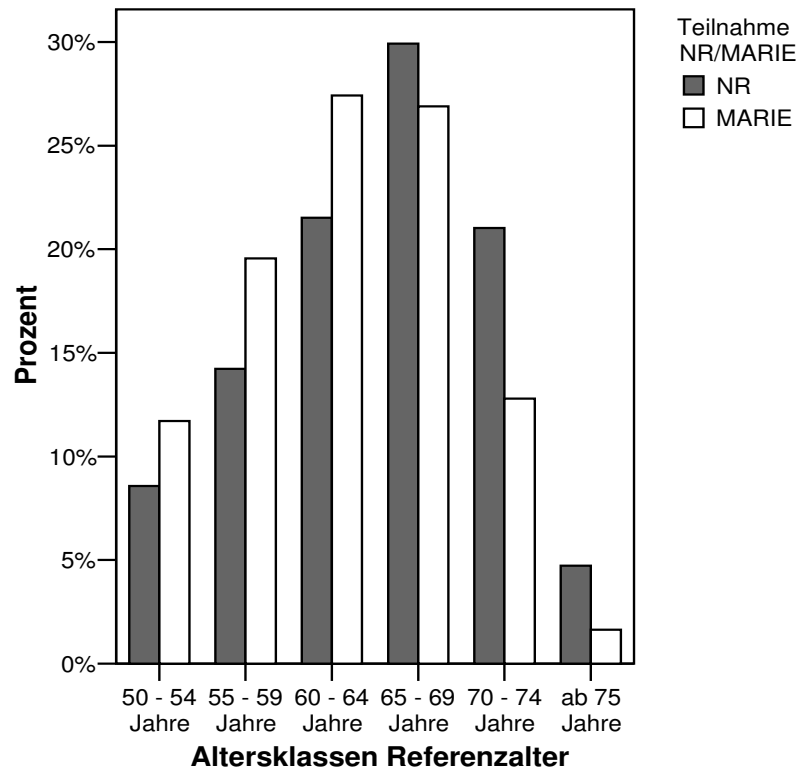


Abb. 3: Histogramm – Referenzalter nach Teilnahme NR/MARIE

Die Variable „*höchster Schulabschluss, 3 Kategorien*“ (schule) ist eine kategoriale Variable mit drei Ausprägungen: „Kein oder niedriger Schulabschluss“ entspricht keinem oder einem Haupt- oder Volksschulabschluss, „mittlerer Schulabschluss“ bezeichnet einen Realschul- oder einen polytechnischen Oberschulabschluss, „hoher Schulabschluss“ ist einer Hochschul- oder Fachhochschulreife zugeordnet. Schwer zu definierende Schulabschlüsse, die jedoch genauer beschrieben wurden (n=56), konnten bei dieser groben Unterteilung eingeordnet werden.

Die Variable enthält 9469 gültige und 60 fehlende Werte. 5733 (60,5%) Probandinnen verfügen über keinen oder nur einen niedrigen, 2552 (27,0%) über einen mittleren und 1184 (12,5%) über einen hohen Schulabschluss.

In Bezug auf die abhängige Variable ist bei beiden Teilnehmergruppen die Klasse „niedriger bzw. kein Schulabschluss“ am stärksten besetzt, allerdings fallen 71,1% der Teilnehmerinnen des Kurzfragebogens und nur 55,7% der MARIE-Teilnehmerinnen in diese Kategorie.

Tabelle 7: Schulabschluss nach Teilnahme NR/MARIE

höchster Schulabschluss, 3 Kategorien		Teilnahme NR/MARIE		Gesamt
		NR	MARIE	
niedriger/kein Schulabschluss	Anzahl	2108	3625	5733
	% Teilnahme NR/Marie	71,1%	55,7%	60,5%
mittlerer Schulabschluss	Anzahl	617	1935	2552
	% Teilnahme NR/Marie	20,8%	29,7%	27,0%
höherer Schulabschluss	Anzahl	238	946	1184
	% Teilnahme NR/Marie	8,0%	14,5%	12,5%
Gesamt	Anzahl	2963	6506	9469
	% Teilnahme NR/Marie	100,0%	100,0%	100,0%

Nach Studienregion differenziert erkennt man, dass in der Region Süd deutlich seltener ein mittlerer oder höherer Bildungsabschluss erreicht wird: In die Klasse „niedriger oder kein Bildungsabschluss“ fallen in der Region Süd 69,5%, in der Region Nord 53,7%. Einen mittleren Schulabschluss erreichen in der Region Süd 20,5%, in der Region Nord 31,9% und einen höheren Schulabschluss erreichen in der Region Süd 10%, in der Region Nord 14,4% der Teilnehmerinnen, wie auch das folgende Balkendiagramm (Abb. 4) zeigt:

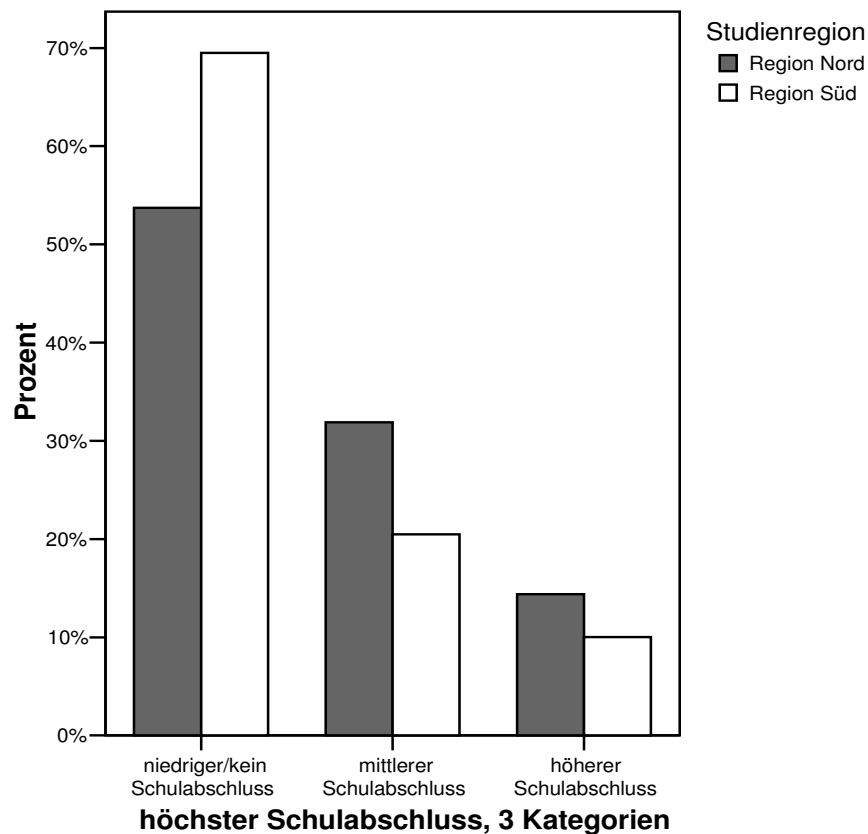


Abb. 4: Balkendiagramm – Höchster Schulabschluss nach Studienregion

Die Variable „*Erwerbstätig current*“ (erwerb) ist eine dichotome Variable für „ja“ und „nein“. Die Variable enthält 9490 gültige und 39 fehlende Werte. Insgesamt sind 6821 (71,9%) Probandinnen aktuell nicht (mehr) erwerbstätig, 2669 (28,1%) gehen noch einer Erwerbstätigkeit nach. Dies hängt sicher mit dem Altersspektrum der gewählten Teilnehmer an der Erhebung zusammen, das den Zeitpunkt einer „normalen“ Beendigung der Erwerbstätigkeit, das „Rentenalter“, beinhaltet.

Im Vergleich der Teilnehmergruppen sind MARIE-Teilnehmerinnen häufiger aktuell erwerbstätig (31,1%) als Teilnehmerinnen des Kurzfragebogens (21,7%). Hier könnte die Altersverteilung beider Stichproben eine Rolle spielen: Ab einem Alter von 65 Jahren überwiegen in den Kategorien prozentual die Teilnehmer des Kurzfragebogens gegenüber den Teilnehmern der MARIE (vgl. Abb. 3).

Nach Studienregion differenziert fallen bezüglich der aktuellen Erwerbstätigkeit keine wesentlichen Unterschiede der Teilnehmergruppen auf.

Tabelle 8: Erwerbstätig current nach Teilnahme NR/MARIE

Erwerbstätig current		Teilnahme NR/MARIE		Gesamt
		NR	MARIE	
nein	Anzahl	2337	4484	6821
	% Teilnahme NR/Marie	78,3%	68,9%	71,9%
ja	Anzahl	646	2023	2669
	% Teilnahme NR/Marie	21,7%	31,1%	28,1%
Gesamt	Anzahl	2983	6507	9490
	% Teilnahme NR/Marie	100,0%	100,0%	100,0%

2.2.3.3 Gynäkologische Merkmale

Die Variable „Kinder 3+“ (kinder_bis3) ist eine kategoriale Variable. Sie enthält 9524 gültige und 5 fehlende Werte. Bis zu zwei Kinder werden genau gezählt, drei oder mehr Kinder werden in einer Kategorie „3+“ zusammengefasst.

1501 (15,8%) der Probandinnen sind kinderlos, 2474 (26,0%) haben ein Kind, 3600 (37,8%) zwei und 1949 (20,5%) drei oder mehr Kinder. Differenziert man nach der Studienregion, finden sich in der Region Süd weniger kinderlose Probandinnen (11,3%) und mehr Probandinnen mit mehr als zwei Kindern (26,0%) als in der Region Nord. Dort sind es mehr kinderlose Probandinnen (19,2%) und weniger Probandinnen mit mehr als zwei Kindern (16,2%), siehe Tabelle 9.

In Bezug auf die abhängige Variable ist die Kategorie „zwei Kinder“ bei beiden Teilnehmerpopulationen am stärksten besetzt (MARIE 35,5%, Kurzfragebogen 38,9%), Teilnehmerinnen des Kurzfragebogens scheinen allerdings häufiger mehr als zwei Kinder zu haben (19,4% MARIE, 22,7% Kurzfragebogen).

Tabelle 9: Anzahl der Kinder nach Studienzentrum

Kinder 3+		Studienzentrum		Gesamt
		Region Nord	Region Süd	
keine Kinder	Anzahl	1038	463	1501
	% Studienzentrum	19,2%	11,3%	15,8%
1 Kind	Anzahl	1462	1012	2474
	% Studienzentrum	27,0%	24,6%	26,0%
zwei Kinder	Anzahl	2031	1569	3600
	% Studienzentrum	37,5%	38,1%	37,8%
>=3	Anzahl	879	1070	1949
	% Studienzentrum	16,2%	26,0%	20,5%
Gesamt	Anzahl	5410	4114	9524
	% Studienzentrum	100,0%	100,0%	100,0%

Die Variable „*orale Kontrazeptiva ever/never*“ (Kontrazeptiva) ist eine dichotome Variable für „niemals“ und „jemals“. Die Variable enthält 9437 gültige und 92 fehlende Werte. 3501 (37,1%) Probandinnen haben nie und 5936 (62,9%) haben jemals orale Kontrazeptiva eingenommen. Auch in dieser Variablen finden sich regionale Unterschiede: In der Region Süd werden weniger (58,6%) orale Kontrazeptiva eingenommen als in der Region Nord (66,2%).

Bei Differenzierung nach der abhängigen Variablen erkennt man, dass MARIE-Teilnehmerinnen häufiger (68,6%) orale Kontrazeptiva eingenommen haben als Teilnehmerinnen des Kurzfragebogens (50,3%). Zur Veranschaulichung hier eine nach Region und Teilnahmestatus differenzierte Tabelle:

Tabelle 10: Kontrazeptiva nach Teilnahme NR/MARIE und Studienregion

Region	orale Kontrazeptiva ever/never		Teilnahme NR/MARIE		Gesamt
			NR	MARIE	
Region Nord	orale Kontrazeption	Anzahl	733	1067	1800
	nie	% Teilnahme NR/Marie	45,2%	28,8%	33,8%
	orale Kontrazeption	Anzahl	887	2642	3529
	ever	% Teilnahme NR/Marie	54,8%	71,2%	66,2%
	Gesamt	Anzahl	1620	3709	5329
		% Teilnahme NR/Marie	100,0%	100,0%	100,0%
Region Süd	orale Kontrazeption	Anzahl	723	978	1701
	nie	% Teilnahme NR/Marie	55,1%	35,0%	41,4%
	orale Kontrazeption	Anzahl	589	1818	2407
	ever	% Teilnahme NR/Marie	44,9%	65,0%	58,6%
	Gesamt	Anzahl	1312	2796	4108
		% Teilnahme NR/Marie	100,0%	100,0%	100,0%

Die Variable „*Hormoneinnahme ever/never (>3Mon., ohne Estriol-Präparate)*“ (ht_ever_gt3mon) ist eine dichotome Variable mit den Ausprägungen „niemals“ und „jemals“.

In der MARIE-Hauptdatei wurden Präparateinnahmen mit einer Dauer von weniger als drei Monaten nicht berücksichtigt. Im Kurzfragebogen wurde die Dauer der Hormoneinnahme in Jahren erfragt. Aufgrund zusätzlicher Aussagen im Freitext, die auf eine Einnahmedauer unter drei Monaten schließen ließen, konnten derartige Hormonangaben ebenfalls gelöscht werden. Gleichfalls unberücksichtigt blieben in beiden Dateien niedrig dosierte Estriole, die nur lokal appliziert wurden. Die Variable enthält 9452 gültige Werte und 77 fehlende Werte. 4582 (48,5%) Probandinnen haben nie und 4870 (51,5%) haben Hormone (HRT) länger als drei Monate eingenommen.

Differenziert man nach Region, haben auch hier Teilnehmerinnen aus der Studienregion Nord häufiger (54,7%) Hormone eingenommen als Teilnehmerinnen aus der Region Süd (47,4%).

Wie Variable „*Hormoneinnahme ever, never (>3Mon., ohne Estriol-Präparate)*“ (ht_ever_gt3mon) unter Berücksichtigung der abhängigen Variablen zeigt, haben MARIE-Teilnehmerinnen häufiger (58,5%) Hormone eingenommen als Teilnehmerinnen des Kurzfragebogens (36,3%).

Tabelle 11: Hormoneinnahme (ever/never) nach Teilnahme NR/MARIE und Studienregion

Studienzentrum	Hormoneinnahme (ever/never) (>3Mon., ohne Estriol-Präparate)		Teilnahme		Gesamt
			NR	MARIE	
Region Nord	never	Anzahl	1028	1405	2433
		% Teilnahme NR/Marie	61,4%	38,0%	45,3%
	ever >3 Mon.	Anzahl	646	2291	2937
		% Teilnahme NR/Marie	38,6%	62,0%	54,7%
	Gesamt	Anzahl	1674	3696	5370
	% Teilnahme NR/Marie	100,0%	100,0%	100,0%	
Region Süd	never	Anzahl	869	1280	2149
		% Teilnahme NR/Marie	66,7%	46,0%	52,6%
	ever >3 Mon.	Anzahl	433	1500	1933
		% Teilnahme NR/Marie	33,3%	54,0%	47,4%
	Gesamt	Anzahl	1302	2780	4082
	% Teilnahme NR/Marie	100,0%	100,0%	100,0%	

Die Variable „*Hormoneinnahme current, past, never*“ (ht_use) ist eine kategoriale Variable mit drei Ausprägungen: „aktuelle Hormoneinnahme“, „frühere Hormoneinnahme“ und „niemals Hormone eingenommen“. Sie enthält 9371 gültige und 158 fehlende Werte. Hier wurde die „Hormoneinnahme ever“ aus der Variablen „Hormoneinnahme ever, never“ in aktuelle Hormoneinnahme und Hormoneinnahme in der Vergangenheit differenziert: 4582 (48,9%) Probandinnen gaben an, niemals Hormone (HRT) eingenommen zu haben. Dagegen nahmen 2678 (28,6%) früher und 2111 (22,5%) aktuell Hormone (HRT) ein.

In Bezug auf die abhängige Variable ist der Unterschied zwischen den Teilnehmerpopulationen noch stärker. In beiden Studienregionen nehmen im Vergleich zu den Teilnehmern des Kurzfragebogens (13,7%) prozentual doppelt so viele MARIE-Teilnehmerinnen (26,5%) aktuell Hormone ein. Betrachtet man die Hormoneinnahme in der Vergangenheit, so sind auch hier starke Unterschiede zwischen den Teilnehmerpopulationen zu beobachten: In der Region Süd nahmen 21,1% der Teilnehmerinnen des Kurzfragebogens und 31,9% der MARIE früher Hormone ein (siehe Tabelle 12).

Tabelle 12: Hormoneinnahme (current/past/never) nach Teilnahme NR/MARIE

Hormoneinnahme (current/past/never)		Teilnahme NR/MARIE		Gesamt
		NR	MARIE	
never user <= 3 Mon.	Anzahl	1897	2685	4582
	% Teilnahme NR/Marie	65,2%	41,6%	48,9%
past user >6 Mon.	Anzahl	615	2063	2678
	% Teilnahme NR/Marie	21,1%	31,9%	28,6%
current use <=6 Mon.	Anzahl	398	1713	2111
	% Teilnahme NR/Marie	13,7%	26,5%	22,5%
Gesamt	Anzahl	2910	6461	9371
	% Teilnahme NR/Marie	100,0%	100,0%	100,0%

Die Variable „Kumulierte Dauer der Hormoneinnahme“ (tab_jahre_cum) ist eine stetige Variable. Wegen der unterschiedlichen Fragestellung liegen von MARIE-Teilnehmerinnen Werte mit zwei Dezimalstellen vor, von Teilnehmerinnen des Kurzfragebogens nur auf ganze Jahre gerundete Werte (sofern im Freitext nicht anders angegeben). Diese Variable wird nur zur Übersicht der Verteilung der Werte beibehalten. Die Variable enthält nur 4684 gültige und 4845 fehlende Werte, weil never-user nicht aufgeführt werden. Der Median liegt bei einer Einnahmedauer von 9 Jahren (Spannweite 0,25 bis 47 Jahre), siehe Abb. 5.

Teilnehmerinnen der Region Süd nehmen Hormone im Mittel nicht so lange ein wie Teilnehmerinnen der Region Nord, die größten Unterschiede scheint es unter den Teilnehmern des Kurzfragebogens zu geben. Der folgende Box-and-Whiskers-Plot (Abb. 5) gibt auch Ausreißer mit auffällig hoher Einnahmedauer wieder.

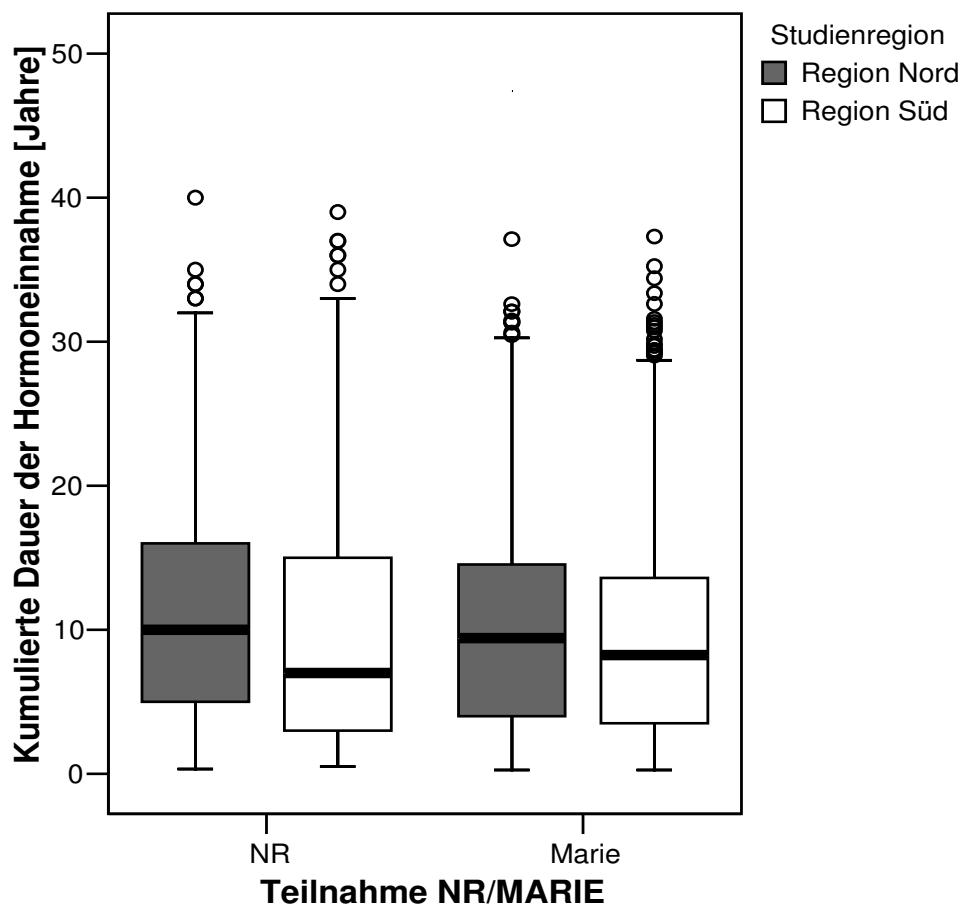


Abb. 5: Box-and-Whiskers-Plot der Hormoneinnahme (kumuliert) nach Region und Fragebogenart

Die Variable „Einnahmedauer in Klassen“ (Hormone_Dauer_Klassen) ist eine kategoriale Variable für die Einnahmedauer (in Klassen von jeweils fünf Jahren gegliedert). Sie enthält auch never-user und deshalb 9266 gültige und nur 263 fehlende Werte. 4582 (49,4%) Probandinnen geben an, niemals Hormone (HRT) eingenommen zu haben. 1456 (15,7%) Probandinnen nahmen Hormone kürzer als fünf Jahre, 1096 (11,8%) fünf bis neun Jahre und 1052 (11,4%) zehn bis vierzehn Jahre lang ein. 1080 (11,7%) der Frauen berichten über eine Hormoneinnahmedauer von fünfzehn und mehr Jahren.

Nach Teilnehmerpopulation differenziert gibt die Variable „Einnahmedauer in Klassen“ an, dass MARIE-Teilnehmerinnen häufiger und länger Hormone einnahmen als Teilnehmerinnen des Kurzfragebogens (Tabelle 13).

Unter den Teilnehmern der verschiedenen Studienregionen fällt bezüglich der Einnahmedauer von Hormonen kein wesentlicher Unterschied auf.

Tabelle 13: Hormoneinnahmedauer nach Teilnahme NR/MARIE

Einnahmedauer klassiert		Teilnahme NR/MARIE		Gesamt
		NR	MARIE	
keine Einnahme	Anzahl	1897	2685	4582
	% Teilnahme NR/Marie	68,0%	41,5%	49,4%
bis <5 Jahre	Anzahl	244	1212	1456
	% Teilnahme NR/Marie	8,7%	18,7%	15,7%
5 - <10 Jahre	Anzahl	207	889	1096
	% Teilnahme NR/Marie	7,4%	13,7%	11,8%
10 -<15 Jahre	Anzahl	189	863	1052
	% Teilnahme NR/Marie	6,8%	13,3%	11,4%
≥15 Jahre	Anzahl	253	827	1080
	% Teilnahme NR/Marie	9,1%	12,8%	11,7%
Gesamt	Anzahl	2790	6476	9266
	% Teilnahme NR/Marie	100,0%	100,0%	100,0%

Die Variable „Alterklassen bei Beginn der Hormoneinnahme“ ist eine kategoriale Variable mit sechs Ausprägungen. Sie gibt das Alter bei erster Hormoneinnahme an (in Klassen von jeweils fünf Jahren gegliedert). Sie enthält 9190 gültige und 339 fehlende Werte. 4582 (49,9%) Probandinnen gaben an, niemals Hormone eingenommen zu haben. Die Klassen „45 -49 Jahre“ (16,3%) und „50-54 Jahre“ (17,1%) waren am stärksten besetzt.

In dieser Variablen sind starke Unterschiede beider Teilnehmergruppen zu beobachten: Bei MARIE-Teilnehmerinnen ist die Klasse „45-49 Jahre“ (20,2%) und „50-54 Jahre“ (19,6%), bei Teilnehmerinnen des Kurzfragebogens die Klasse „50-54 Jahre“ (11,3%) am stärksten besetzt. Letztere begannen später mit der Hormoneinnahme, wie das folgende Histogramm (Abb. 6) zeigt:

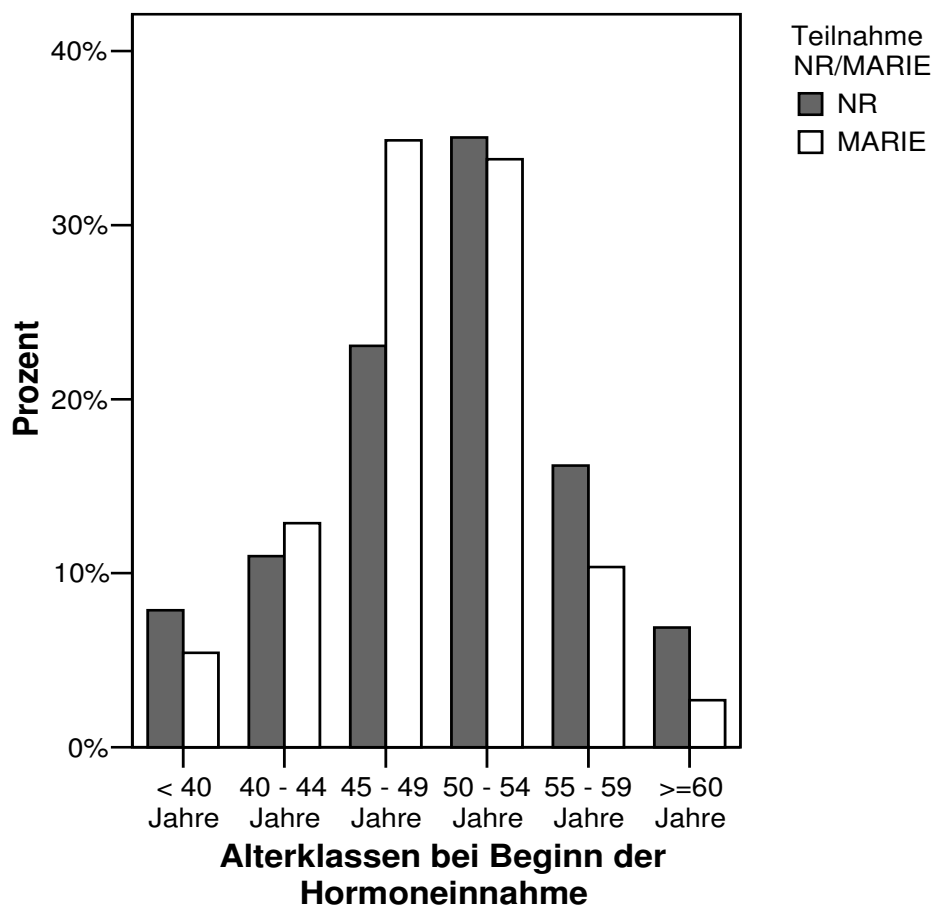


Abb. 6: Histogramm – Alter bei Beginn der Hormoneinnahme (ohne Darstellung der never-user)

Die Variable „*Altersklassen bei Absetzen der Hormoneinnahme*“ ist ebenfalls eine kategoriale Variable mit sechs Ausprägungen. Sie gibt das Alter bei Ende der letzten Hormoneinnahme an (in Klassen von jeweils fünf Jahren gegliedert). Sie enthält 8856 gültige und 673 fehlende Werte. Die Klasse „größer gleich 60 Jahre“ ist am stärksten besetzt, fast die Hälfte der Frauen, die Hormone einnehmen, setzen die Therapie erst ab dem 60. Lebensjahr wieder ab (siehe Histogramm, Abb. 7).

Im Hinblick auf die abhängige Variable erkennt man, dass im Alter unter 55 Jahren prozentual mehr Teilnehmerinnen des Kurzfragebogens als des Hauptfragebogens die Hormontherapie beenden. Ab einem Alter von 55 Jahren beenden prozentual mehr Teilnehmerinnen der MARIE als des Kurzfragebogens.

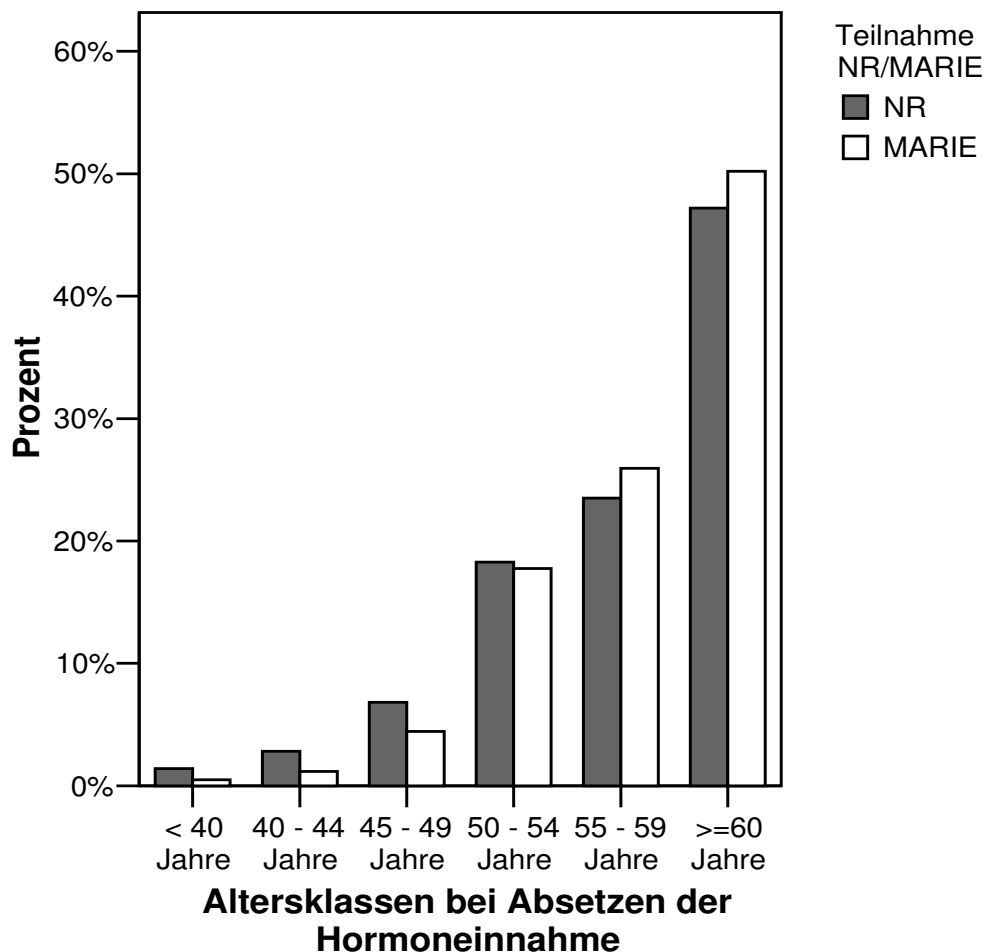


Abb. 7: Histogramm – Alter bei Ende der Hormoneinnahme (ohne Darstellung der never-user)

Die Variable „Einnahme pflanzl./homöop. Präp. ever/never“ (pfl_evernever) ist eine dichotome Variable für „jemals“ und „niemals“. Die Variable enthält 9478 gültige und 51 fehlende Werte. 8401 (88,6%) Probandinnen gaben an, niemals pflanzliche oder homöopathische Präparate bei klimakterische Beschwerden eingenommen zu haben, während nur 1077 (11,4%) eine Einnahme bejahten.

Nach Teilnahmestatus differenziert nehmen MARIE-Teilnehmerinnen häufiger (14,5%) als Teilnehmerinnen des Kurzfragebogens (4,6%) pflanzliche oder homöopathische Präparate ein (Tabelle 14).

Tabelle 14: Einnahme pflanzlicher/homöopathischer Präparate (ever/never) nach Teilnahme NR/MARIE

Einnahme pflanzl./homöop. Präp. (ever/never)		Teilnahme NR/MARIE		Gesamt
		NR	MARIE	
never	Anzahl	2840	5561	8401
	% Teilnahme NR/Marie	95,4%	85,5%	88,6%
ever	Anzahl	136	941	1077
	% Teilnahme NR/Marie	4,6%	14,5%	11,4%
Gesamt	Anzahl	2976	6502	9478
	% Teilnahme NR/Marie	100,0%	100,0%	100,0%

2.2.3.4 Merkmale für gesundheitsbewusstes Verhalten

Die Variable „Regelmäßiges Rauchen ever“ (rauchen) ist eine dichotome Variable für „jemals“ und „niemals“. Sie enthält 9507 gültige und 22 fehlende Werte. 5537 (58,2%) Teilnehmerinnen rauchten niemals regelmäßig, 3970 (41,8%) gaben an, schon regelmäßig geraucht zu haben. Differenziert man nach Studienregion, rauchen Teilnehmerinnen der Region Nord häufiger (46,8%) als Teilnehmerinnen der Region Süd (35,1%).

Aus der Variablen „Regelmäßiges Rauchen ever“ (rauchen) kann man in Bezug auf die abhängige Variable ablesen, dass Teilnehmerinnen des MARIE-Hauptfragebogens geringfügig häufiger (43,0%) rauchen als Teilnehmerinnen des Kurzfragebogens (39,1%).

Die Variable „*Salat und Gemüse*“ (Salat_Gemuese) ist eine kategoriale Variable mit sieben Ausprägungen für den Verzehr von Salat oder Gemüse mit einer Häufigkeit von „niemals“ bis „mehrmals täglich“. Die beobachteten starken Unterschiede zwischen beiden Stichproben sind durch unterschiedliche Fragestellungen in den Fragebögen vermutlich erhebungsbedingt.

Diese Variable enthält 8177 gültige und 1352 fehlende Werte. 132 (1,6%) Teilnehmerinnen gaben an, Salat und Gemüse nur einmal pro Woche oder seltener zu verzehren. 813 (9,9%) gaben einen Verzehr von zwei- bis dreimal pro Woche, 1854 (22,7%) einen Verzehr von vier- bis sechsmal pro Woche, 1945 (23,8%) einen einmaligen und 3433 (42%) einen mehrmaligen Verzehr pro Tag an.

Bezüglich der abhängigen Variablen scheinen MARIE-Teilnehmerinnen viel häufiger Salat und Gemüse zu verzehren als Teilnehmerinnen des Kurzfragebogens: Bei MARIE-Teilnehmerinnen ist die Klasse „mehrmals täglich“ (63,4%) und bei Teilnehmerinnen des Kurzfragebogens die Klasse „täglich“ (46,7%) die am stärksten besetzte Kategorie.

2.3 Bivariate Analyse

Um eine Vorauswahl derjenigen Faktoren zu treffen, welche einen signifikanten Einfluss auf die abhängige Variable (Teilnahme am Haupt- bzw. Kurzfragebogen) haben, wurde im ersten Schritt eine statistische bivariate Zusammenhangsanalyse aller erklärenden Variablen durchgeführt:

Alle stetigen Variablen wurden mit dem T-Test auf Mittelwert-Gleichheit, alle kategorialen Variablen wurden mit dem Chi-Quadrat-Unabhängigkeitstest auf einen Zusammenhang mit der abhängigen Variablen getestet. Hierfür wurde das Programmpaket SPSS für Windows (Version 12) verwendet.

2.3.1 Test der stetigen Variablen (T-Test, Levene-Test)

Für stetige Variablen wurde der Levene-Test auf Varianzhomogenität und der T-Test auf Mittelwertgleichheit gewählt und mit dem Programm SPSS 12 gerechnet.

Beim Levene-Test auf Varianzhomogenität besagt die Nullhypothese die Gleichheit der Varianzen beider Stichproben. Wie Tabelle 15 zeigt, kann die Nullhypothese für alle drei Variablen verworfen werden, die Varianzen der untersuchten Merkmale Geburtsjahr, Referenzalter und kumulierte Einnahmedauer von Hormonen sind in Bezug auf die Teilnahme am Kurz- oder am Hauptfragebogen signifikant verschieden.

Tabelle 15: Levene-Test auf Varianzhomogenität

		F	Signifikanz
	Die Varianzen sind		
Geburtsjahr	gleich	6,004	,014
	nicht gleich		
Referenzalter	gleich	4,221	,040
	nicht gleich		
Kumulierte Dauer der Hormoneinnahme	gleich	33,285	,000
	nicht gleich		

Darauf wurde der T-Test auf Mittelwertgleichheit durchgeführt. Anhand des T-Tests kann die Nullhypothese, wonach es keine signifikanten Unterschiede der Mittelwerte beider Stichproben gibt, geprüft werden. Tabelle 16 zeigt, dass die Nullhypothese für alle drei Variablen verworfen werden kann:

Zwischen Teilnehmerinnen des MARIE-Hauptfragebogens und des Kurzfragebogens unterscheiden sich das mittlere Alter, das mittlere Geburtsjahr und die mittlere kumulierte Einnahmedauer für Hormone signifikant.

Tabelle 16: T-Test für die Mittelwertgleichheit bei unabhängigen Stichproben

		T	df	Sig. (2- seitig)	Mittlere Differenz	Stan- dard- fehler	95% Konfidenzintervall	
Varianzen sind							Untere	Obere
Geburts- jahr	gleich	-14,537	9527	,000	-1,995	,137	-2,264	-1,726
	nicht gleich	-14,265	5621,454	,000	-1,995	,140	-2,269	-1,721
Referenz- alter	gleich	14,865	9527	,000	2,049	,138	1,779	2,320
	nicht gleich	14,630	5660,515	,000	2,049	,140	1,775	2,324
Kumulierte	gleich	2,937	4682	,003	,78310	,267	,26032	1,30589
Dauer der Hormon- einnahme	nicht gleich	2,649	1209,653	,008	,78310	,296	,20303	1,36318

Nach Durchführung der Tests in Bezug auf die abhängige Variable sind alle drei stetigen Merkmale Geburtsjahr, Referenzalter und kumulierte Einnahmedauer von Hormonen auf einem Signifikanzniveau von 0,05 signifikant unterschiedlich in Mittelwert und Varianz.

2.3.2 Test der kategorialen Variablen (Chi-Quadrat-Test)

Der Chi-Quadrat-Unabhängigkeitstest wird benutzt, um kategoriale Variablen bezüglich einer abhängigen Variablen auf überzufällige Zusammenhänge zu prüfen. Dabei wird auf einem 0,05 Signifikanzniveau die Nullhypothese, ob zwei Variablen unabhängig sind, geprüft [Backhaus et al. 2005].

Da der Stichprobenumfang ausreichend groß war, wurden keine der Approximationsbedingungen dieses Test verletzt (Zum einen muss die erwartete Häufigkeit jeder Zelle der Kontingenztabelle größer als Null sein, zum anderen dürfen nicht mehr als 20% der Zellen eine erwartete Häufigkeit von kleiner 5 aufweisen): In den Zellen der gebildeten Kreuztabellen war keine der erwarteten Häufigkeiten gleich Null, sie waren immer weit größer als 5.

Die Ergebnisse der Chi-Quadrat-Tests sind in Tabelle 17 aufgeführt:

Tabelle 17: Chi-Quadrat-Unabhängigkeitstest

Variable	Chi	Signifikanz (p-Wert)
Body Mass Index	4,662	0,458
Geburtsjahr in Klassen	232,562	0,000
Referenzalter in Klassen	252,541	0,000
Schulabschluss	209,057	0,000
Kinder	19,592	0,000
Orale Kontrazeptiva	287,562	0,000
Hormoneinnahme ever, never	405,329	0,000
Hormoneinnahme past, current, never	457,733	0,000
Einnahmedauer in Klassen	563,190	0,000
Alterklassen bei Beginn der Hormoneinnahme	589,734	0,000
Alterklassen bei Absetzen der Hormoneinnahme	1003,078	0,000
Einnahme pflanzlicher und homöopathischer Mittel	198,768	0,000
Regelmäßiges Rauchen	12,422	0,000
Verzehr von Salat und Gemüse	3274,823	0,000

Für alle erklärenden Variablen – bis auf den Body Mass Index – wurde die Nullhypothese bezüglich der Unabhängigkeit der Variablen auf dem 0,05-Signifikanzniveau verworfen, d. h. ab einem p-Wert kleiner als 0,05 kann davon ausgegangen werden, dass es einen signifikanten Zusammenhang zwischen der Variablen „Teilnahme NR/MARIE“ (NR_Marie) und den einzelnen erklärenden Variablen gibt.

Da einige Merkmale auch einen starken Effekt der Studienregion aufzeigen, wurde der Chi-Quadrat-Test auch nach Studienregion getrennt durchgeführt.

Bezüglich des Body Mass Index, der bei beiden Teilnehmergruppen gemeinsam einen nur schwachen Zusammenhang mit der abhängigen Variable zeigt, findet sich bei Differenzierung nach Studienregion in der Region Süd (Chi-Quadrat-Wert 10,313, $p=0,067$) ein wesentlich signifikanterer Zusammenhang als in der Region Nord (Chi-Quadrat-Wert 4,770, $p=0,445$).

Bezüglich der Variablen „Kinder 3+“ findet sich in der Region Nord (Chi-Quadrat-Wert 7,263, $p=0,064$) ein schwächerer Zusammenhang als in der Region Süd (Chi-Quadrat-Wert 13,284, $p=0,004$).

Bezüglich der Variablen zum Rauchverhalten „rauchen“ gibt es minimale Unterschiede zwischen den Regionen (Region Nord mit einem Chi-Quadrat-Wert 5,922 und einer Signifikanz $p=0,015$ gegenüber Region Süd mit einem Chi-Quadrat-Wert 6,212 und einer Signifikanz von $p=0,013$).

Alle erklärenden Variablen zeigen in der bivariaten Analyse einen signifikanten Zusammenhang ($p<0,05$) mit der abhängigen Variablen und wurden in die engere Auswahl für die multivariate Analyse genommen. Für die hier gefundenen Zusammenhänge sind Stärke und Richtung des Zusammenhangs nicht bekannt und wurden im logistischen Regressionsmodell untersucht.

2.3.3 Auswahl der erklärenden Variablen

Die biometrische Variable „BMI“ wird trotz der mangelnden Signifikanz im Chi-Quadrat-Test als einzige Biometrische Variable in das Modell genommen, das Normalgewicht (2: „18,5- <25 kg/m²“) wird als Referenzkategorie definiert.

Unter den demographischen Variablen wird zum einen das Geburtsjahr in das Modell aufgenommen und eine mittlere Klasse (3: „1940-1944“) als Referenzkategorie definiert. Das Referenzalter als ähnliche Variable wurde abgelehnt, da es aus organisatorischen Gründen Verzerrungen aufweisen könnte: Teilnehmer des Kurzfragebogens wurden entweder sehr früh (bei Erstkontakt) oder sehr spät (nach Erhalt des dritten Anschreibens) befragt. Das Geburtsjahr erschien hier die geeignete Variable zu sein. Zum anderen wurden die Variablen „Schulbildung“ und „Erwerbstätigkeit“ ausgewählt. Die Referenzkategorien bilden bei beiden Variablen die Nullwerte „kein oder niedriger Schulabschluss“ bzw. „keine aktuelle Erwerbstätigkeit“.

Unter den gynäkologischen Variablen wurden die Variablen „Kinder“, „Kontrazeptiva“, „Hormoneinnahme ever, never“ und die „Einnahme pflanzlicher bzw. homöopathischer Präparate“ ausgewählt. Da mehrere Variablen zur Hormoneinnahme wegen der Interaktionen nicht im gleichen Modell berücksichtigt werden können, wurde im ersten Modell exemplarisch die Variable „Hormoneinnahme ever, never“ ausgewählt. Bezüglich dieser Variable wird ein Haupteffekt erwartet. Das Logit-Modell wurde aber beispielhaft auch mit anderen Variablen zur Hormoneinnahme („Hormoneinnahme current, past, never“ und „Einnahmedauer in Klassen“) berechnet.

Die Studienregion wird für ein zweites Modell zusätzlich ausgewählt und auf Interaktionseffekte untersucht.

Tabelle 18 fasst die erklärenden Variablen für das Ausgangsmodell der logistischen Regression zusammen. Die Referenzkategorien sind hervorgehoben.

Tabelle 18: Erklärende Variablen

Variablenname	Kategorien	Modell
Die folgenden Variablen sind die Einschlussvariablen und dürfen nicht eliminiert werden		
Geburtsjahr in Klassen	bis 1934 1935-1939 1940-1944 1945-1949 ab 1950	1a, 1b, 1c, 2
Studienregion	Region Nord (Hamburg) Region Süd (Rhein-Neckar-Karlsruhe)	1a, 1b, 1c, 2
Die folgenden Variablen werden für das <i>stepwise backward Modell</i> genutzt		
BMI in Klassen (aktuelles Gewicht)	<18,5: Untergewicht 18,5-<25: Normalgewicht 25-<30: Präadipositas 30-<35: Adipositas Grad 1 35-<40: Adipositas Grad 2 >=40: Adipositas Grad 3	1a, 1b, 1c, 2
Kinder in 3 Kategorien	keine Kinder ein Kind zwei Kinder >=3 Kinder	1a, 1b, 1c, 2
Orale Kontrazeptiva	orale Kontrazeption nie orale Kontrazeption ever	1a, 1b, 1c, 2
Hormoneinnahme ever, never	never (<= 3 Mon.) ever (>3 Mon.)	1a, 2
Hormoneinnahme current, past, never	never user (<= 3 Mon.) past user (>6 Mon.) current use (<=6 Mon.)	1b
Einnahmedauer von Hormonen	keine Einnahme bis < 5 Jahre 5 -<10 Jahre 10 -<15 Jahre >=15 Jahre	1c
Einnahme pflanzlicher oder homöopathischer Mittel	never ever	1a, 1b, 1c, 2
Schulbildung	niedriger/kein Schulabschluss mittlerer Schulabschluss höherer Schulabschluss	1a, 1b, 1c, 2
Aktuelle Erwerbssituation	nein ja	1a, 1b, 1c, 2
Rauchen	never ever	1a, 1b, 1c, 2
Studienregion als Interaktionsterm	mit allen anderen Variablen	2

2.4 Multivariate Analyse

2.4.1 Binäre logistische Regression

Die binäre logistische Regression ist ein nützliches Verfahren zur Modellierung der Beziehung zwischen einer binären abhängigen Variablen und mindestens einer erklärenden Variablen [Hosmer et al. 1989, Menard et al. 1995]. Die binäre abhängige Variable hat zwei Ausprägungen, z. B. „ja“ ($Y=1$) oder „nein“ ($Y=0$). Die erklärenden X-Variablen können ein beliebiges Skalenniveau aufweisen.

Eine logistische Regression untersucht die Frage, wie die Eintrittswahrscheinlichkeit der abhängigen Variablen von den Ausprägungen der unabhängigen erklärenden Variablen beeinflusst wird. In dieser Arbeit wird die logistische Regression dazu dienen, herauszufinden, wie die Wahrscheinlichkeit, nicht an der MARIE teilzunehmen, durch eine Reihe Merkmale der Teilnehmerinnen erklärt werden kann.

Unter der Annahme, dass das Modell p erklärende Variablen enthält, die durch den Vektor $x' = (x_1, x_2, \dots, x_p)$ angegeben werden, liefert eine einfache lineare Regression die Beziehung zwischen der abhängigen Variablen Y und den unabhängigen X-Variablen durch die Gleichung:

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Ein Problem bei der Anwendung solcher Modelle für eine dichotom abhängige Variable ist jedoch, dass die Variable Y nicht nur Werte zwischen 0 und 1, sondern zwischen $-\infty$ und $+\infty$ annehmen kann. Für dieses Problem bietet die logistische Regression eine Lösung: Die bedingte Wahrscheinlichkeit, dass ein interessierendes Ereignis (hier: nicht an der MARIE teilzunehmen) unter einem bestimmten Kovariatenmuster eintritt, kann mit $\pi(x) = P(Y = 1|x)$ ausgedrückt werden. Diese Wahrscheinlichkeit kann Werte zwischen 0 und 1 annehmen. Die Beziehung zwischen der Eintrittswahrscheinlichkeit und den X-Variablen kann dann mit Hilfe einer logistischen Verteilung, die ebenfalls nur Werte zwischen 0 und 1 annehmen kann, in folgender Gleichung angegeben werden:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}$$

Zunächst wird das Verhältnis zwischen π_i und $1 - \pi_i$ betrachtet. Dieses Verhältnis wird als „Odds“ gekennzeichnet. Diese Odds können Werte zwischen 0 und $+\infty$ annehmen.

$$\text{Odds}(Y = 1) = \frac{\pi_i}{1 - \pi_i}$$

Werden die Odds logarithmiert, erhält man den sogenannten „Logit“.

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Mit dieser Transformation wird erreicht, dass der Beziehungs-Parameter, der zwischen den X-Variablen und der Eintrittswahrscheinlichkeit des Ereignisses besteht, linear wird. Dieser Ausdruck kann Werte zwischen $-\infty$ und $+\infty$ annehmen. Der Logit vermittelt eine Verbindung zwischen der dichotomen abhängigen Variablen und dem linearen Regressionsausdruck auf der rechten Seite der Gleichung.

Wie im Fall einer linearen Regression werden die Parameter β geschätzt. Dies wird normalerweise durch ein iteratives Maximum-Likelihood Verfahren erreicht. Die geschätzten Parameter geben an, wie sich die Logits verändern, wenn sich die unabhängigen Variablen um eine Einheit erhöhen. Einfacher zu interpretieren ist jedoch der Effektkoeffizient e^{β_j} , auch *Odds Ratio* (OR) genannt. In den Modelltabellen des Programms SPSS wird der Effektkoeffizient mit dem Kürzel Exp(B) bezeichnet. Dieser gibt an, um welchen Faktor sich die Odds (die Chancen, dass die Zielvariable Y den Wert 1 annimmt) vervielfachen, wenn sich die unabhängige Variable um eine Einheit erhöht. Oder anders ausgedrückt: Eine Erhöhung der unabhängigen Variablen um eine Einheit bewirkt eine Veränderung des Odds um den Faktor e^{β_j} . Für die Interpretation des Effektkoeffizienten (*Odds Ratio* oder Exp(B)) siehe Kapitel 2.4.1.3.

Mit dem Programm SPSS werden umfangreiche Modelltabellen erstellt, deren wichtigste Kenngrößen für die inhaltliche Interpretation der erklärenden Variablen (Regressionskoeffizienten B, Wald-Werte, Signifikanz des Wald-Wertes und Exp(B)-Werte) kurz erläutert werden:

2.4.1.1 Wald-Wert

Die Wald-Werte sind in der Beurteilung der relativen Wichtigkeit einer Einflussvariablen in Bezug auf das Gesamtmodell wichtig. Etwas vereinfacht gesagt, kann bei Anwendung der automatischen Auswahlprozedur „*stepwise model selection by backward elimination*“ davon ausgegangen werden, dass diejenige Einflussvariable mit dem größten Wald-Wert den relativ größten Einfluss auf das Modell (nicht auf die Zielvariable) hat. Die Vereinfachung ergibt sich dadurch, dass bei kategorialen Variablen neben dem Wald-Wert auch die Anzahl an Freiheitsgraden bei der Beurteilung des relativen Einflusses der Variablen auf das Modell berücksichtigt werden muss.

2.4.1.2 Signifikanz p

Die Signifikanzangaben in den SPSS-Tabellen mit erklärenden Variablen beziehen sich auf die Wald-Werte. Es wird die Signifikanz angegeben, mit der sich der angegebene Wald-Wert von Null unterscheidet. Bei der „*stepwise model selection by backward elimination*“ wird jeweils die unabhängige Variable mit der geringsten Signifikanz vom Modell ausgeschlossen.

2.4.1.3 Regressionskoeffizient B

Die Regressionskoeffizienten sind in zweierlei Hinsicht wichtig. Erstens ist dem Vorzeichen des jeweiligen Regressionskoeffizienten zu entnehmen, wie die Wirkrichtung der jeweiligen Einflussvariablen aussieht. Dabei ist die Codierung der Zielvariablen zu beachten. Im berechneten Modell ist die abhängige Variable im Modell mit „0“ und „1“ codiert.

- Ein positiver Regressionskoeffizient erhöht die Wahrscheinlichkeit der mit „1“ codierten Merkmalsausprägung der „Nicht-Teilnahme an der MARIE“.
- Ein negativer Regressionskoeffizient erhöht die Wahrscheinlichkeit der mit „0“ codierten Merkmalsausprägung „Teilnahme an der MARIE“.

2.4.1.4 Effektkoeffizient Exp(B)

Exp(B)-Werte geben Auskunft über die Stärke des Wirkeinflusses der jeweiligen erklärenden Variablen. Exp(B)-Werte lassen sich als *Odds Ratio*'s (OR) in Bezug auf die jeweilige Zielvariable interpretieren. Ist der Wert Exp(B) kleiner als 1, verringern sich die Chancen gegenüber der Referenzkategorie, dass die abhängige Variable den Wert 1 annimmt. Wenn dieser Wert größer als 1 ist, erhöhen sich diese Chancen. Vereinfacht gilt folgende Einteilung:

- Ein *Odds Ratio* zwischen 0 und 0,25 bedeutet einen stark-negativen Einfluss der erklärenden Variablen.
- Ein *Odds Ratio* von 0,25 bis 0,5 bedeutet einen mäßig-negativen Einfluss des erklärenden Merkmals.
- Ein *Odds Ratio* zwischen 0,5 und 1 zeigt einen schwach-negativen Einfluss der erklärenden Variablen an.
- Ein *Odds Ratio* von 1 bedeutet, es keinen Unterschied der *Odds* der Referenzkategorie und der betrachteten Kategorie in Bezug auf die abhängige Variable, bzw. es gibt keinen Einfluss der im Modell aufgeführten Kategorie der erklärenden Variablen.
- Ein *Odds Ratio* zwischen 1 und 2 bedeutet einen gering-positiver Einfluss der erklärenden Variablen.
- Ein *Odds Ratio* von 2 bis 4 bedeutet einen mäßig-positiver Einfluss der erklärenden Variablen.
- Ein *Odds Ratio* größer als 4 bedeutet einen stark-positiver Einfluss der erklärenden Variablen.

Auf die Interpretation von Wald-Wert, Regressionskoeffizient und Exp(B)-Wert werden sich die Ausführungen in Kapitel 3 immer wieder beziehen. Sie sind in den Tabellen 19 bis 26 (Anhang) für die jeweiligen Ausgangs- und Endmodelle angegeben.

2.4.2 Auswahl des Modells

In der logistischen Regression werden auch Wechselwirkungen zwischen einzelnen erklärenden Merkmalen berücksichtigt. Um nach den stärksten Zusammenhängen zu suchen, wurde als Vorgehensweise die schrittweise Modell-Selektion durch Rückwärts-Elimination (*stepwise model selection by backward elimination*, kurz: *backward elimination*) gewählt. Sie wurde ebenfalls mit dem Programmpaket SPSS für Windows (Version 12) berechnet. Dieses Modell ist ein automatisches Auswahlverfahren zur Wahl einflussreicher Variablen für das Modell.

Zunächst wurde bei der *stepwise model selection by backward elimination* ein logistisches Ausgangsmodell mit allen Variablen, die in die Analyse aufgenommen werden sollen, berechnet. Die Variable, die die geringste Signifikanz des Wald-Wertes aufweist, wird aus diesem Modell herausgenommen und in einem weiteren Schritt ein zweites Logit-Modell berechnet. Anschließend wird mittels eines Likelihood-Ratio-Tests geprüft, ob sich das zweite Modell im ersten wiederfindet.

Wenn keine signifikante Veränderung auftritt, ist es nicht weiter notwendig, das Modell mit allen Variablen zu rechnen, denn die Variable mit dem geringsten Einfluss kann aus dem Modell entfernt werden. Diese Prozedur wird in mehreren Schritten so oft wiederholt, bis das Endmodell nicht mehr im Modell des vorherigen Schrittes wiederzufinden ist.

Die Modelle 1a bis 1c wurde mit verschiedenen Variablen zur Hormoneinnahme berechnet. Im Modell 2 wurde zusätzlich ein interagierender Einflussfaktor betrachtet: Bei der oben durchgeführten Deskription der einzelnen Variablen wurde häufig ein Interaktionseffekt mit der Variablen „Studienregion“ beobachtet. Für das Modell 2 wurden jeweils Interaktionsterme zwischen der Studienregion und jeder erklärenden Variable gebildet und separat auf ihren Einfluss untersucht.

3 Ergebnisse

Vorab sei erwähnt, dass alle Effektkoeffizienten der erklärenden Variablen in den jeweiligen Konfidenzintervallen liegen, die aus diesem Grund nicht in den Tabellen 19 bis 26 (Anhang) angegeben wurden. Die vier Modelle (1a bis 2) sind in allen Schritten, insbesondere für das jeweilige Ausgangs- und Endmodell signifikant (die Chi-Quadrat-Werte der Modelle sind ausnahmslos $p=0,000$), folglich können die gefundenen Parameter interpretiert werden.

3.1 Modell 1: ohne Berücksichtigung von Interaktionseffekten

3.1.1 Modell 1a: Die Hormonvariable ist „Hormoneinnahme ever, never“

Wie erwähnt wurde im Modell 1a die Variable „Hormoneinnahme ever, never“ als einzige der Variablen zur Hormoneinnahme ausgewählt. Für das Modell wurden 9189 (96,4%) Fälle in die Analyse einbezogen, es gab 340 (3,6%) fehlende Fälle. Die Klassifizierungstabelle gibt den richtigen Vorhersagewert für die Teilnahme an der Marie im Nullmodell mit 70,1% an.

3.1.1.1 Modell „Einschluss“

Im Modell Einschluss sieht man, dass durch Ergänzung der Variablen „Geburtsjahrgang“ und „Studienregion“ eine signifikante Verbesserung des Nullmodells entsteht (Chi-Quadrat-Wert 217,041, $p=0,000$).

Bezüglich der einzelnen Geburtsjahrgänge weisen die Jahrgangsklassen unterhalb der Referenzkategorie einen hochsignifikanten Trend auf: Im Vergleich zur Referenzkategorie „1940-1944“ zeigen die Geburtsjahrgängen „1935-1939“ ($\text{Exp}(B)=1,485$, $p=0,000$) eine 1,5-fach höhere, in der Kategorie „bis 1934“ ($\text{Exp}(B)=2,291$, $p=0,000$) sogar eine 2,3-fach höhere Chance auf, nicht an der MARIE teilzunehmen. Die Vorhersage lautet:

Je niedriger die Geburtsjahrgangsklasse, desto mehr Teilnehmerinnen werden nicht teilnehmen. Die beiden anderen Geburtsjahrgangsklassen sind nicht signifikant.

Die Variable „Studienregion“ ist nicht signifikant und würde aussagen, dass die Tatsache, eine Teilnehmerin aus der Region Nord zu sein, einen schwach-negativen Einfluss auf eine Nicht-Teilnahme hätte (sie also eher teilnehmen würde).

3.1.1.2 Modell „Rückwärts Schrittweise (Likelihood-Quotient)“

Nach Ergänzung der Regressionsgleichung um die unabhängigen Variablen ergibt sich in der Klassifizierungstabelle ein Prozentsatz der richtig vorhergesagten Fälle von 71,4%, also eine geringe Verbesserung gegenüber dem Nullmodell. Im Basismodell des Modells „Rückwärts Schrittweise“ sieht man, dass durch Ergänzung aller erklärender Variablen eine signifikante Verbesserung des Nullmodells entsteht (Chi-Quadrat-Wert 932,638, $p=0,000$). In den folgenden drei Schritten ändert sich der Chi-Quadrat-Wert kaum und ist jeweils hochsignifikant.

In das Basismodell (Tabelle 19, Anhang) wurden die in Tabelle 18 genannten Variablen aufgenommen. Durch schrittweise Elimination der zwei Variablen mit den am wenigsten signifikanten Wald-Werten, der Variablen „Anzahl der Kinder“ und „Rauchen“, erhält man das Endmodell der Schätzergebnisse (Tabelle 20, Anhang).

Die Auswertung des Endmodells ergibt (entsprechend den Wald-Werten) eine Rangfolge der Wichtigkeit der Variablen für das Modell. In absteigender Reihenfolge fallen (als stärkste Einflussfaktoren auf das Modell) die „Hormoneinnahme ever, never“, „Einnahme pflanzlicher Präparate“, „Schulbildung“ und „Kontrazeptiva“ auf:

Die „Hormoneinnahme ever, never“ ($\text{Exp}(B)=0,431$, $p=0,000$) geht mit negativem Vorzeichen in die Regressionsgleichung ein. Folglich besteht ein mäßig-negativer Einfluss auf eine Nicht-Teilnahme an der MARIE-Studie, d. h. eine Teilnehmerin, die Hormone einnahm, ist mit einer 0,4-fachen Chance eine Nicht-Teilnehmerin der MARIE (oder im Umkehrschluss mit einer 2,3-fachen Chance eine Teilnehmerin).

Es folgt die Variable „Einnahme pflanzlicher Präparate“ ($\text{Exp}(B)=0,344$, $p=0,000$), die ebenfalls einen hochsignifikanten, mäßig-negativen Einfluss auf die Nicht-Teilnahme aufweist: Eine Frau, die pflanzliche Präparate einnahm, nimmt mit einer 2,9-fachen Chance eher an der MARIE teil als eine Frau, die keine pflanzlichen Präparate einnahm.

Bei der Variablen „Schulbildung“ („mittlerer Schulabschluss“ $\text{Exp}(B)=0,609$, $p=0,000$ und „höherer Schulabschluss“ $\text{Exp}(B)=0,489$, $p=0,000$) zeigt sich ebenfalls ein hochsignifikanter negativer Einfluss (ein mittlerer Schulabschluss hat einen schwach-negativen Effekt, ein hoher Schulabschluss einen mäßig-negativen Effekt) auf die Nicht-Teilnahme an der MARIE. Bezüglich der Bildung lässt sich also ein gerichteter Effekt (Trend) ablesen: Mit zunehmender Bildung steigt die Chance, an der MARIE teilzunehmen.

Die Einnahme von Kontrazeptiva ($\text{Exp}(B)=0,648$, $p=0,000$) hat nur einen hochsignifikanten schwach-negativen Einfluss auf die Nicht-Teilnahme, d. h. eine Frau, die Kontrazeptiva einnahm, hat eine 1,5-fache Chance, an der MARIE teilzunehmen.

Die Geburtsjahrgangsklassen sind in den höheren Klassen nicht signifikant. Bezüglich der unteren Geburtsjahrgangsklassen kann die Aussage getroffen werden: Je höher die Klasse des Geburtsjahrgangs, desto eher eine Teilnahme an der MARIE. Im Verhältnis zur mittleren Kategorie „1940-1944“ haben Geburtsjahrgänge „bis 1934“ ($\text{Exp}(B)=1,408$) einen schwach-positiven, Geburtsjahrgänge von „1935 bis 1939“ ($\text{Exp}(B)=1,180$) einen sehr schwach-positiven Einfluss auf eine Nicht-Teilnahme.

Aus den Effektkoeffizienten des Body Mass Index lässt sich ein Trend ablesen, wobei nur die Klassen der „Adipositas Grad 1“ und „Adipositas Grad 2“ signifikant sind: „Adipositas Grad 1“ zeigt einen schwach-negativen ($\text{Exp}(B)=0,802$, $p=0,003$) und „Adipositas Grad 2“ ($\text{Exp}(B)=0,690$, $p=0,004$) einen etwas stärkeren, aber immer noch schwach-negativen Effekt auf die Nicht-Teilnahme. Unter Berücksichtigung der nicht signifikanten $\text{Exp}(B)$ -Werte für die anderen BMI-Klassen könnte man folgende Tendaussage treffen: Je größer der BMI, desto eher eine Teilnahme an der MARIE, je kleiner, desto eher eine Teilnahme am Kurzfragebogen. Unter Berücksichtigung der fehlenden Signifikanz in der bivariaten Analyse scheint eine solche Aussage eher fraglich.

Bei den Studienregionen zeigt sich ein äußerst schwach-positiver Effekt der Region Nord auf die Nicht-Teilnahme ($\text{Exp}(B)=1,107$, $p=0,040$).

Den geringsten Einfluss auf das Modell hat nach der Rangfolge nach Wald die Variable „Erwerbstätigkeit“. Sie ist nicht signifikant (man könnte ansonsten interpretieren, dass berufstätige Frauen eher an der MARIE teilnehmen. Diese Annahme könnte mit dem Alter und dem Kinderreichtum korrelieren, was hier nicht weiter untersucht wird).

Um zu ermitteln, ob weitere Hormonvariablen andere Einflüsse ergeben, wurde das Modell versuchsweise geringfügig verändert und die Hormonvariable ausgetauscht, siehe Kapitel 3.1.2 und 3.1.3).

3.1.2 Modell 1b: Hormonvariable ist „Hormoneinnahme current, past, never“

Im Modell 1b wurde statt der Variablen zu Hormoneinnahme „Hormoneinnahme ever, never“ die Variable „Hormoneinnahme current, past, never“ in das Modell aufgenommen. Für das Modell wurden 9141 (95,9%) Fälle in die Analyse einbezogen, es gab 388 (4,1%) fehlende Fälle. Die Klassifizierungstabelle gibt den richtigen Vorhersagewert für die Teilnahme an der MARIE im Nullmodell mit 70,3% an.

3.1.2.1 Modell „Einschluss“

Im Modell Einschluss sieht man, dass durch Ergänzung der Variablen „Geburtsjahrgang“ und „Studienregion“ eine signifikante Verbesserung des Nullmodells entsteht (Chi-Quadrat-Wert 222,207, $p=0,000$).

Das gesamte Modell Einschluss bleibt nahezu unverändert, die Aussagen des Modells 1a (Kapitel 3.1.1.1) behalten ihre Gültigkeit.

3.1.2.2 Modell „Rückwärts Schrittweise (Likelihood-Quotient)“

Nach Ergänzung der Regressionsgleichung um die unabhängigen Variablen ergibt sich in der Klassifizierungstabelle ein Prozentsatz der richtig vorhergesagten Fälle von 71,6%, also eine geringe Verbesserung gegenüber dem Nullmodell. Im Basismodell des Modells Rückwärts Schrittweise sieht man, dass durch Ergänzung aller erklärender Variablen eine signifikante Verbesserung des Nullmodells entsteht (Chi-Quadrat-Wert 958,426, $p=0,000$). In den folgenden drei Schritten ändert sich der Chi-Quadrat-Wert kaum und ist jeweils hochsignifikant. Man sieht schon hier, dass das Modell gegenüber Modell 1a kaum Veränderungen aufweist.

In das Basismodell (Tabelle 21, Anhang) wurden die in Tabelle 18 genannten Variablen aufgenommen. Auch in diesem Modell wurden die schon bekannten Prädiktoren mit den am wenigsten signifikanten Wald-Werten, die Variable „Anzahl der Kinder“ und „Rauchen“ ausgeschlossen. Auf diese Weise erhält man das Endmodell der Schätzergebnisse (Tabelle 22, Anhang).

Die Auswertung des Endmodells ergibt (entsprechend den Wald-Werten) die schon aus Modell 1a bekannte Rangfolge des Modelleinflusses der Prädiktoren „Hormoneinnahme ever, never“, „Einnahme pflanzlicher Präparate“, „Schulbildung“ und „Kontrazeptiva“. Das gesamte Modell bleibt nahezu unverändert. Lediglich der Modelleinfluss der Studienregion wird deutlich geringer (Wald-Wert) und ist nicht mehr signifikant.

Die Variable „Hormontherapie current, past, never“ ist die wichtigste Variable für dieses Modell und zeigt einen gerichteten, mäßig-negativen Effekt: „past user“ ($\text{Exp}(B)=0,465$, $p=0,000$) zeigen einen mäßig-negativen, „current user“ ($\text{Exp}(B)=0,366$, $p=0,000$) einen etwas stärkeren mäßig-negativen Einfluss auf die Zielvariable „Nicht-Teilnahme“ an der MARIE. Dies bedeutet, dass (im Gegensatz zu Frauen, die nie Hormone eingenommen haben) Frauen, die eine Hormoneinnahme beendet haben, eine 2,2-fache Chance und Frauen, die aktuell noch Hormone einnehmen, sogar eine 2,7-fache Chance haben, an der MARIE teilzunehmen.

Die Variable „Hormoneinnahme current, past, never“ ist eine Differenzierung der Variablen „Hormoneinnahme ever, never“: Frauen, die Hormone eingenommen hatten, wurden nach vergangener und aktueller Einnahme aufgeteilt. Deshalb zeigt sich im Vergleich zu Modell 1b, dass das Mittel beider Effektkoeffizienten – unter Berücksichtigung der etwas geringeren Zahl an „current usern“ gegenüber den „past usern“ – annähernd den Effektkoeffizienten der Variablen „Hormoneinnahme ever, never“ ergibt.

3.1.3 Modell 1c: Hormonvariable ist „Dauer der Hormoneinnahme in Klassen“

Im Modell 1c wurde als Variable zu Hormoneinnahme die Variable „Dauer der Hormoneinnahme“ in das Modell einbezogen. Für das Modell wurden 9050 (95,0%) Fälle in die Analyse einbezogen, es gab 479 (5,0%) fehlende Fälle. Die Klassifizierungstabelle gibt den richtigen Vorhersagewert für die Teilnahme an der Marie im Nullmodell mit 71,2% an.

3.1.3.1 Modell „Einschluss“

Im Modell Einschluss sieht man, dass durch Ergänzung der Variablen „Geburtsjahrgang“ und „Studienregion“ eine signifikante Verbesserung des Nullmodells entsteht (Chi-Quadrat-Wert 214,075, $p=0,000$).

Das gesamte Modell Einschluss bleibt nahezu unverändert, die Aussagen des Modells 1a (Kapitel 3.1.1.1) behalten ihre Gültigkeit.

3.1.3.2 Modell „Rückwärts Schrittweise (Likelihood-Quotient)“

Nach Ergänzung der Regressionsgleichung um die unabhängigen Variablen ergibt sich in der Klassifizierungstabelle ein Prozentsatz der richtig vorhergesagten Fälle von 72,5%, also eine geringe Verbesserung gegenüber dem Nullmodell. Im Basismodell des Modells Rückwärts Schrittweise sieht man, dass durch Ergänzung aller erklärender Variablen eine signifikante Verbesserung des Nullmodells entsteht (Chi-Quadrat-Wert 1029,577, $p=0,000$). In den folgenden drei Schritten ändert sich der Chi-Quadrat-Wert kaum und ist jeweils hochsignifikant. Man sieht auch in diesem veränderten Modell, dass es gegenüber Modell 1a kaum Veränderungen aufweist.

In das Basismodell (Tabelle 23, Anhang) wurden die in Tabelle 17 genannten Variablen aufgenommen. Auch bei diesem Modell wurden die schon bekannten Prädiktoren mit den am wenigsten signifikanten Wald-Werten, die Variable „Anzahl der Kinder“ und „Rauchen“ ausgeschlossen. Auf diese Weise erhält man das Endmodell der Schätzergebnisse (Tabelle 24, Anhang).

Es finden sich an vorderer Rangposition für den Modelleinfluss die bereits aus Modell 1a und 1b bekannten vier Prädiktoren „Hormoneinnahme ever“, „Einnahme pflanzlicher Präparate“, „Schulbildung“ und „Kontrazeptiva“. Abgesehen vom Signifikanzverlust einer Klasse der Geburtsjahrgänge und der Schwächung des Modelleinflusses der Geburtsjahrgänge insgesamt (Wald-Wert) sind in Trend, Stärke und Richtung der Effektkoeffizienten keine wesentlichen Abweichungen zum Modell 1a festzustellen. Die Veränderungen bezüglich der Variablen Geburtsjahrgang sind durch einen Zusammenhang zwischen Geburtsjahrgang der Teilnehmerinnen und der Einnahmedauer von Hormonen erklärbar.

Die Hormonvariable „Einnahmedauer von Hormonen“ zeigt – obwohl hochsignifikant – keinen Trend zwischen Einnahmedauer und einer Teilnahme oder Nicht-Teilnahme an der MARIE. Im Hinblick auf eine Nicht-Teilnahme zeigen alle Klassen der Hormoneinnahme einen mäßig-negativen Effekt auf die Nicht-Teilnahme: „bis 5 Jahre“ ($\text{Exp}(B)= 0,338$, $p=0,000$), „5-<10 Jahre“ ($\text{Exp}(B)= 0,394$, $p=0,000$), „10-<15 Jahre“ ($\text{Exp}(B)= 0,335$, $p=0,000$) und „ ≥ 15 Jahre“ ($\text{Exp}(B)= 0,416$, $p=0,000$).

Das Fazit des Austauschs der Hormonvariablen ist, dass nicht die Dauer der Einnahme, sondern die Tatsache, ob eine Frau überhaupt Hormone eingenommen hat und vor allem, ob sie aktuell noch Hormone einnimmt, entscheidend für einen positiven Effekt der Teilnahme an der Mammakarzinom-Risikofaktoren-Erhebung zu sein scheint.

3.2 Modell 2: Modell mit Betrachtung von Interaktionseffekten bezüglich der Studienregion

Im Modell 2 wurde basierend auf dem Modell 1a (mit der Variablen „Hormoneinnahme ever, never“) der Interaktionseffekt der Studienregion untersucht: Die Ausgangsbedingungen entsprechen dem Modell 1a, die Verbesserung des Einschlussmodells gegenüber dem Nullmodell (durch den Einschluss der Variablen „Geburtsjahrgang“ und „Studienregion“) unterscheidet sich kaum vom Modell 1a (Chi-Quadrat=217,041, $p=0,000$). Das Modell Einschluss wird deshalb hier nicht weiter beschrieben.

3.2.1.1 Modell „Rückwärts Schrittweise (Likelihood-Quotient)“

Man sieht im Modell 2 (Tabelle 25 zeigt das Basismodell, Tabelle 26 das Endmodell der Schätzergebnisse), dass aus mit der Variablen Studienregion und den anderen erklärenden Variablen sogenannte Interaktionsterme gebildet wurden.

Durch Hinzufügen der unabhängigen Variablen ergibt sich in der Klassifizierungstabelle ein Prozentsatz der richtig vorhergesagten Fälle von 71,4%, der zwar eine geringe Verbesserung gegenüber dem Nullmodell darstellt, aber keine Verbesserung gegenüber Modell 1a.

Die signifikante Verbesserung des Chi-Quadrat-Wertes zum Nullmodell ist etwas größer als im Modell 1a (hier: Chi-Quadrat-Wert 954,986, $p=0,000$). In den folgenden 11 Schritten ändert sich der Chi-Quadrat-Wert kaum und ist jeweils hochsignifikant.

Das Basismodell (Tabelle 25, Anhang) zeigt, dass keiner der Interaktionsterme auf dem 0,05-Niveau signifikant sind. Ferner wurden zu den schon bekannten eliminierten Prädiktoren „Kinder“ und „Rauchen“ alle Interaktionsterme eliminiert: Die Interaktionsterme der einzelnen erklärenden Variablen mit der Studienregion haben keinen Einfluss auf das Modell.

Die Auswertung des Endmodells (Tabelle 26, Anhang) ergibt nach Elimination der Interaktionsterme die bereits bekannte Rangfolge (entsprechend den Wald-Werten) der Wichtigkeit der Variablen für das Modell. Es finden sich in den vordersten Rangpositionen die bereits bekannten vier erklärenden Variablen „Hormoneinnahme ever, never“, „Einnahme pflanzlicher Präparate“, „Schulbildung“ und „Einnahme von Kontrazeptiva“.

Bei Trend, Stärke und Richtung aller Effektkoeffizienten ist nach Elimination der Interaktionsterme mit der Variablen „Studienregion“ keine wesentliche Abweichung zu Modell 1a festzustellen. Die Studienregion hat somit keinen interagierenden Effekt auf die unabhängigen Variablen.

4 Diskussion

4.1 Methode

4.1.1 Stichproben

Die teilnehmenden Fälle des Kurzfragebogens waren nicht standardisiert mit dem Kurzfragebogen befragt worden, d. h. viele erkrankte Frauen konnten absagen, ohne um die Teilnahme am Kurzfragebogen gebeten worden zu sein (vgl. Kapitel 2.1.7). Deshalb nahm nur eine geringe Anzahl der Fälle an der Kurzbefragung teil. Als erklärendes Merkmal war deshalb die Erkrankung an Brustkrebs nicht auswertbar und hätte zu einer Verzerrung der Studienergebnisse geführt. Zweitens war es auch aufgrund der geringen Anzahl der Fälle, die am Kurzfragebogen teilgenommen haben, nicht möglich, eine Schätzung des Brustkrebsrisikos unter den nicht teilnehmenden Fällen anzustellen. Ein drittes Problem für die Befragung der Fälle mit dem Kurzfragebogen waren datenschutzrechtlichen Bedenken.

Aus diesen Gründen wurden alle teilnehmenden Fälle von der Untersuchung des Selektionsbias ausgeschlossen.

Auch die Kontrollen führten zu einer anfänglichen Verzerrung des Modells, weil es unterschiedliche Erhebungszeiträume des Kurzfragebogens in den Studienregionen gab (vgl. 2.1.7). Im logistischen Modell ist noch ein schwacher Effekt der Region Nord auf die Nicht-Teilnahme zu erkennen, der durch die längere Erhebungsdauer in dieser Region zu erklären ist. Doch konnte durch die Einschränkung der Kontrollen mit gleichem Erhebungszeitraum größere Effekte der Studienregion auf das Modell gering gehalten werden.

4.1.2 Erhebungsinstrumente

Bei Entwicklung des Kurzfragebogens stand vor allem die einfache Struktur der aus dem MARIE-Fragebogen extrahierten Fragen und eine knappe Formulierung im Vordergrund, um die Akzeptanz der Befragung zu erhöhen. Die dabei gemachten Umformulierungen haben zu einer Abweichung der Items geführt.

Diese hätte im Vorfeld stärker beachtet werden müssen, hat sie doch dazu geführt, dass ein wichtiger Indikator für gesundheitsbewusstes Verhalten (der Verzehr von Salat und Gemüse) bei der Auswertung nicht berücksichtigt werden konnte.

Da sich nach einmonatiger Erhebungszeit herausstellte, dass das telefonische Interview selten abgerochen wurde, die Akzeptanz des Erhebungsinstrumentes also hoch war, wurden nach und nach weitere Items in den Kurzfragebogen aufgenommen. Obwohl studentisch verständlich, führten diese verschiedenen Versionen dazu, dass Variablen nicht (z. B. aktuelles Rauchen) oder nur eingeschränkt ausgewertet werden konnten (z. B. konnten zum Rauchverhalten nach Häufigkeit die Kategorien „nur einmal im Leben“ und „Gelegenheitsraucherin“ nicht berücksichtigt werden, lediglich „regelmäßiges Rauchen ever“ wurde beibehalten).

Folglich ist eine der Stärken dieser Kurzbefragung, dass das teilnehmende Klientel sehr robust auf die Anzahl der gestellten Fragen reagierte: Es gab kaum Probandinnen, die den Fragebogen abgebrochen haben. Künftige Erhebungen könnten deshalb von Anfang an umfangreicher ausfallen.

Eine weitere Stärke der Untersuchung war die Feststellung von Ausschlusskriterien unter Nicht-Teilnehmern, wie z. B. die Brustkrebserkrankung in der Anamnese einer Probandin. Dadurch war es möglich, auch Nicht-Teilnehmer – sofern sie am Kurzfragebogen teilgenommen hatten – nach den Ein- und Ausschlusskriterien der Studie zu beurteilen. Vorteil einer solchen Kurzbefragung ist es also, auch unter absagenden Probandinnen Ausschlüsse identifizieren zu können.

Allerdings wurde bewusst auf die Erhebung eines besonderen Einschlusskriteriums der MARIE verzichtet: Die MARIE untersucht das Risiko unter postmenopausalen Frauen. Da die Definition des Postmenopausenstatus nicht einfach zu treffen ist und im Hauptfragebogen eine ganze Reihe von Items für die Erfassung vorgesehen war, wurde für den Kurzfragebogen bewusst darauf verzichtet. Es ist also anzunehmen, dass sich unter den Teilnehmerinnen des Kurzfragebogens noch prä- und perimenopausale Frauen befinden.

4.2 Ergebnisse

4.2.1 Deskription: Univariate Analyse

In der Deskription zeigen sich bei fast allen erklärenden Variablen deutliche Unterschiede bezüglich der abhängigen Variablen. Die größten Unterschiede finden sich bezüglich der Einnahme von Präparaten:

Sehr große Unterschiede findet man bei der Einnahme von Kontrazeptiva: 68% der MARIE-Teilnehmerinnen und nur 50,3% der Nicht-Teilnehmerinnen haben mit der „Pille“ verhütet. Ebenfalls sehr große Unterschiede findet man bei der Hormoneinnahme: Im Gegensatz zu nur 36,3% der Nicht-Teilnehmerinnen haben 58,8% der MARIE-Teilnehmerinnen Hormone eingenommen, unter ihnen ist der Anteil der „current user“ deutlich höher. Sie haben auch früher mit der Einnahme begonnen und später mit der Einnahme aufgehört als die Nicht-Teilnehmerinnen. Auch pflanzliche Präparate wurden von Teilnehmerinnen (14,5%) häufiger eingenommen als von Nicht-Teilnehmerinnen (4,6%).

Ebenfalls auffällige Unterschiede zeigt die Schulbildung, denn Nicht-Teilnehmerinnen haben zu 71,1% einen niedrigen Schulabschluss, MARIE-Teilnehmerinnen dagegen nur zu 55,7%. Auch bei der Erwerbstätigkeit zeigen sich starke Unterschiede, denn nur 21,7% der Nicht-Teilnehmer sind im Verhältnis zu 31,1% der MARIE-Teilnehmerinnen berufstätig. Ferner scheinen Teilnehmer des Kurzfragebogens aus niedrigeren Geburtsjahrgängen zu stammen bzw. älter zu sein. Und schließlich hatten bezüglich der Kinderzahl hatten Nicht-Teilnehmerinnen mehr Kinder als Teilnehmerinnen.

Geringe Unterschiede sieht man bezüglich der Variablen für gesundheitsbewusstes Verhalten (MARIE-Teilnehmerinnen rauchen nur wenig häufiger als Nicht-Teilnehmerinnen) und beim Body Mass Index, der hinsichtlich der abhängigen Variablen keine Auffälligkeiten zeigt.

Die beobachteten großen Unterschiede zugunsten der MARIE-Teilnehmer beim Verzehr von Salat und Gemüse sind eher artifiziell.

Allerdings fiel bei der Deskription des Datensatzes sehr häufig ein Einfluss der Studienregion auf: Besonders in Bezug auf die Geburtsjahrgänge schienen in der Region Nord niedrigere Geburtsjahrgangsklassen aufzutreten. Dies kann durch das Matching der Kontrollen mit den retrospektiven Fällen der Region Nord, die deshalb niedrigeren Geburtsjahrgängen als in der Region Süd angehören, verursacht sein. Der Body Mass Index scheint unter Frauen der Region Süd höher zu sein. Insgesamt sind haben mehr Teilnehmer eine geringe Schulbildung, dies scheint in der Region Süd stärker ausgeprägt zu sein. Kinderlosigkeit tritt in der Region Nord häufiger auf, mehr als zwei Kinder gibt es eher in der Region Süd. Hier wurden auch weniger Kontrazeptiva (58,6%) eingenommen als in der Region Nord (66,2%). Dies zeigt sich auch bei anderen Präparateinnahmen: Es haben deutlich weniger Teilnehmerinnen aus der Region Süd (47,4%) Hormone eingenommen als in der Region Nord (54,7%), und die Einnahmedauer war kürzer. In der Region Süd gibt es deutlich weniger Raucher (35,1%) gegenüber 46,8% im Norden.

Es werden also nach Deskription Interaktionseffekte der Studienregion erwartet.

4.2.2 Bivariate Analyse

In der Bivariaten Analyse bestätigen sich die vermuteten Zusammenhänge im Hinblick auf die abhängige Variable: Alle erklärenden Variablen – außer dem BMI – zeigen einen hochsignifikanten Unterschied bezüglich der abhängigen Variablen auf. Die getrennte Auswertung nach Studienregion zeigt – wie nach der univariaten Analyse vermutet – in der bivariaten Analyse einen Interaktionseffekt bezüglich der Variablen „Body Mass Index“, „Kinder“ und „Rauchverhalten“.

4.2.3 Multivariate Analyse

Die Modelle bestätigten den ersten Eindruck aus der uni- und bivariaten Analyse: Hochsignifikanten Einfluss haben „Hormoneinnahme ever, never“ ($\text{Exp}(B)=0,431$, $p=0,000$), „Einnahme pflanzlicher Präparate“ ($\text{Exp}(B)=0,344$, $p=0,000$), „Einnahme von Kontrazeptiva“ ($\text{Exp}(B)=0,648$, $p=0,000$) und „Schulbildung“ („mittlerer Schulabschluss“ $\text{Exp}(B)=0,609$, $p=0,000$ und „höherer Schulabschluss“ $\text{Exp}(B)=0,489$, $p=0,000$). In den Geburtsjahrgängen zeigen nur die niedrigen Geburtsjahrgänge einen schwach-positiven Effekt auf die Nichtteilnahme. Nach Studienregionen zeigt sich ein äußerst schwach-positiver Effekt der Region Nord auf die Nichtteilnahme ($\text{Exp}(B)=1,107$, $p=0,040$).

Der Body Mass Index ist im Hinblick auf seine Einflüsse schwer zu beurteilen: Bei der bivariaten Analyse zeigte er keinen signifikanten Unterschied in beiden Teilnehmergruppen, so auch in zwei Kategorien bei der logistischen Regression. Die Aussagen der signifikanten Kategorien (je größer der Body Mass Index, desto eher eine Teilnahme an der MARIE) ist deshalb nicht aussagekräftig.

Die Variablen „Rauchen“ und „Kinder“ wurden aus dem Modell eliminiert, und die „Erwerbstätigkeit“ zeigte kein signifikantes Ergebnis.

In den folgenden Modellvarianten (Modell1b und Modell 1c) wurden die Hormonvariablen ausgetauscht, das Modell bewies sich aber als sehr stabil und veränderte sich kaum: Meiner Einschätzung nach stellt Modell 1a (Tabelle 20) den aussagekräftigsten Eindruck der Ergebnisse dar.

Ein möglicher Interaktionseffekt für die Studienregion, der nach der Deskription sicher schien, ließ sich nicht bestätigen, da im logistischen Modell 2 alle Interaktionsterme mit der Studienregion eliminiert wurden.

Dieses Ergebnis erstaunt nach den starken Eindrücken der Deskription.

4.2.4 Forschungsfragen und Hypothesen

Die Ergebnisse der bivariaten und multivariaten Analyse haben gezeigt, dass die Nullhypothese H_0 bezüglich der ersten Forschungsfrage dieser Untersuchung: „Zwischen den Stichproben bestehen keine Unterschiede der erhobenen Merkmale, somit besteht kein Selektionsbias“ verworfen werden kann.

Allerdings muss die Alternativhypothese H_1 wegen der während der Untersuchung erfolgten Beschränkung auf die Kontrollen noch angepasst werden und lautet nun:

„Es bestehen bezüglich der nicht an Brustkrebs erkrankten Referenzpopulation zwischen den Stichproben signifikante Unterschiede der erhobenen Merkmale, somit besteht ein Selektionsbias unter den Kontrollen.“

Die zweite Forschungsfrage lautete: „Welche Merkmale der Teilnehmerpopulationen haben den größten Einfluss auf den Selektionsbias?“ und sollte explorativ beantwortet werden:

Hier zeigt sich, dass die stärksten und hochsignifikanten Einflussfaktoren auf eine Nichtteilnahme an der MARIE folgende Merkmale der Teilnehmerpopulationen sind (die Odds Ratios sind dem meiner Meinung nach aussagekräftigsten Modell entnommen, dem Endmodell 1a, Tabelle 20):

„Hormoneinnahme“ (OR=0,431 (CI95 0,391-0,475)), „Einnahme pflanzlicher bzw. homöopathischer Präparate“ (OR 0,344 (CI95 0,283-0,419)), „Einnahme von Kontrazeptiva“ (OR 0,648 (CI95 0,584-0,719) und „Schulbildung“: „mittlerer Schulabschluss“ OR 0,609 (CI95 0,543-0,684), „höherer Schulabschluss“ (OR=0,489 (CI95 0,414-0,578)). Letztere weist einen Trend auf: Je höher der Bildungsstatus, desto wahrscheinlicher eine Teilnahme an der MARIE.

Nach obigen Ergebnissen hat die Präparateinnahme einen positiven Einfluss auf die Teilnahme an der MARIE. Die gefundenen Zusammenhänge sind in allen untersuchten Ausprägungen, unabhängig von Art (Kontrazeptiva, vergangener oder aktueller Hormoneinnahme oder Einnahme pflanzlicher Mittel) hochsignifikant.

Unter den Teilnehmerinnen der MARIE wird also die Einnahme hormoneller Präparate überschätzt, weil ein hochsignifikanter Zusammenhang zwischen Präparateinnahme und Teilnahmebereitschaft besteht. Dieses Ergebnis deckt sich mit den Ergebnissen von Moorman et al., die in ihrer Studie bei der Analyse von Unterschieden der Teilnehmer eines face-to-face-Interviews und eines telefonischen Kurzfragebogens einen ähnlichen Zusammenhang gefunden haben [Moorman et al. 2000], während Strom et al. in ihrer Studie zu dem Ergebnis kommen, dass die Hormoneinnahmen zwischen Nonrespondern und Respondern annähernd gleich sind [Strom et al. 2004].

In Bezug auf die Ergebnisse der MARIE führt der gefundene Selektionbias zu einer möglichen Verzerrung des zu errechnenden *Odds Ratio* zwischen Hormoneinnahme und Brustkrebs (siehe Kapitel 1.3).

5 Schlussfolgerung

Die Befragung mittels des Kurzfragebogens war eine im Verhältnis zur Hauptbefragung mit geringem Aufwand durchzuführende Erhebung. Die hier gefundenen Ergebnisse zeigen, dass sich dieser Aufwand gelohnt hat, da der gefundene Selektionsbias (mit der Folge eines einer Unterschätzung des *Odds Ratios* für die Exposition Hormoneinnahme in der Fall-Kontroll-Studie) zu einer Diskussion und eventuellen Korrektur der Studienergebnisse der Fall-Kontroll-Studie genutzt werden kann [Lyles et al 2002].

Weitere Vorteile einer Kurzbefragung unter Nichtteilnehmern sehe ich vor allem für Forschungsvorhaben mit einer erwarteten höheren Non-Response, für die eine solche Befragung ein wertvolles Mittel zur Erhebung wichtiger Stichprobencharakteristika bieten könnte.

Für geplante Erhebungen innerhalb bestimmter Bevölkerungsgruppen (Probanden früherer Geburtsjahrgänge, mit hohem Alter, mit gesundheitlichen Beeinträchtigungen, niedrigem Sozialstatus) könnte ein Kurzfragebogen eine zusätzliche Ergänzung sein:

Die Ergebnisse der Untersuchung haben gezeigt, dass die Akzeptanz des Kurzfragebogens bei älteren Teilnehmern und solchen mit geringerem Bildungsniveau höher war.

Die hier durchgeführte Untersuchung bietet außerdem interessante Ansatzpunkte für weitere Forschungsvorhaben:

- Möglicherweise besteht eine Korrelation zwischen der Einnahme von Medikamenten, der Einstellung gegenüber Ärzten und der Akzeptanz einer wissenschaftlichen gesundheitsbezogenen Erhebung wie der MARIE. Dies konnte mit dem vorliegenden Material nicht untersucht werden.
- Unbeantwortet bleibt auch die Frage, ob es möglicherweise Interaktionen zwischen den verschiedenen Präparateinnahmen (Kontrazeptiva, Hormone ever (current, past) und pflanzlichen Präparaten gibt.

Literaturverzeichnis

Arbeitsgruppe Epidemiologische Methoden der Deutschen Arbeitsgemeinschaft Epidemiologie (DAE), Hrsg.: Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP); 1999

Backhaus K, Erichson B, Plinke W und Weiber R: Multivariate Analysemethoden; Springer 2005, Berlin

BDSG :Bundesministerium der Justiz: Bundesdatenschutzgesetz, §§4,5,
http://www.gesetze-im-internet.de/bdsg_1990/index.html

Bortz J, Döring N: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, 4. Auflage; Springer 2006, Berlin

Brennecke R, Greiser E, Paul HA., Schach E: Datenquellen für Sozialmedizin und EpidemiologieTitel; Springer 1981, Berlin

Csismadi I, Friedenreich CM, Bryant HE et al.: An Analysis of the effect of selection bias on the association of hormone replacement therapy and breast cancer; Chronic diseases in Canada 2005, Vol 26(2/3)

de Gruyter (Hg.): Pschyrembel Klinisches Wörterbuch, 257. Auflage 1994, Berlin

Duetz M, Abel T, Meier C, Niemann S: Self-rated health, life satisfaction and personal characteristics of post-menopausal women under estrogen replacement therapy; Maturitas 2000, Vol 35(1): 71-79

Fahrmeir L, Tutz G: Statistik. der Weg zur Datenanalyse; Springer 1999, Berlin

Frey JH, Kunz G, Lüschen G: Telefonumfragen in der Sozialforschung. Methoden, Techniken, Befragungspraxis; Westdeutscher Verlag 1990, Opladen, Wiesbaden

Gabler S, Häder S, Hoffmeyer-Zlotnik (Hrsg.): Telefonstichproben in Deutschland; Westdeutscher Verlag 1998, Opladen, Wiesbaden

Heinemann L, et al.: Epidemiologische Arbeitsmethoden; Gustav-Fischer-Verlag 1994, Stuttgart

HmbDSG :Senat Hamburg: Hamburgisches Datenschutzgesetz (1990), §§5,7,
http://hh.juris.de/hh/gesamt/DSG_HA_1990.htm#DSG_HA_1990_G5

HmbGDG: Senat Hamburg: Hamburgisches Gesundheitsdienstgesetz (2001), §§24-28 ,
http://hh.juris.de/hh/GesDG_HA_rahmen.htm

Hoffmann W, Terschüren C, Holle R et al: Zum Problem der Response in epidemiologischen Studien in Deutschland (Teil II); Gesundheitswesen 2004, Vol 66: 482-491

Hosmer D, Lemeshow S: Applied Logistic Regression; John Wiley & Sons 1989, New York

Kreienbrock L, et al.: Epidemiologische Methoden; Gustav Fischer Verlag 1997, Stuttgart

Lancaster T, Surman G, Lawrence M et al.: Hormone replacement therapy: characteristics of users and non-users in a British general practice cohort identified through computerised prescribing records; Journal of Epidemiologic Community Health 1995, Vol49: 398-394

LDStG: Gesetz zum Schutz personenbezogener Daten (Landesdatenschutzgesetz) Baden-Württemberg (2000, zuletzt geändert 2004), §§ 4,6,13-15,
<http://www.baden-wuerttemberg.datenschutz.de/recht/ldsg/default.htm>

Lorenz RJ: Grundbegriffe der Biometrie; Gustav Fischer Verlag 1996, Stuttgart

Lyles RH, Allen AS: Estimating Crude or Common Odds Ratios in Case-Control Studies with informatively Missing Exposure Data; American Journal of Epidemiology 2002, Vol155(1): 274-281

Madigan MP, Troisi R, Potischman N et al.: Case-Control Studies in Clinical Research: Mechanism and Prevention of Selection Bias; International Journal of Epidemiologie 2000, Vol 29: 793-798

Mayer HO: Interview und schriftliche Befragung. Entwicklung, Durchführung und Auswertung; Oldenbourg Wissenschaftsverlag 2002, München

Menard S: Applied Logistic Regression; Sage University Papers Series on Quantitative Applications in the Social Sciences no.106 1995, Thousand Oaks, CA: Sage.

Moorman PG, Kuwabara H, Millikan RC et al: Menopausal hormones and breast cancer in a biracial population; American Journal of Public Health 2000 Vol90: 966-971

Morabia A: Case-Control Studies in Clinical Research: Mechanism and Prevention of Selection Bias; Preventive Medicine 1997, Vol26: 674-677)

Richtlinie 95/46/EG: Europäisches Parlament und Rat der EU: Richtlinie 95/46/EG vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr.

Robert-Koch-Institut, Gesundheitsberichterstattung des Bundes, Heft 25; Berlin Mai 2005

Rogers A, Murtaugh MA, Edwards S et al.: Contacting Controls: Are We Working Harder for Similar Response Rates, and Does It Make a Difference?; American Journal of Epidemiology 2004, Vol160: 85-90

Sachs L: Angewandte Statistik: Anwendung statistischer Methoden; Springer 1997, Berlin

SGB: Sozialgesetzbuch Abschnitt 1 (1975, zuletzt geändert 2006) §35: Sozialgeheimnis,
<http://db03.bmgs.de/Gesetze/gesetze.htm>

Strom BL, Schinnar R: Participants and refusers in a telephone interview about hormone replacement therapy were equally likely to be taking it; Journal of Clinical Epidemiology 2004, Vol57: 624-626

Voigt L, Koepsell TD, Daling JR: Characteristics of Telephone Survey Respondents According to Willingness to Participate; American Journal of Epidemiology 2003, Vol157: 66-73

Walther I.: Nonresponder-Analyse der Mammakarzinom-Risikofaktoren-Erhebung (MARIE), Diplomarbeit im Studiengang Biowissenschaftliche Dokumentation an der Fachhochschule Hannover, 25.5.2006

World-Health-Organisation (Hsg.): Obesity - Preventing and managing the global epidemic; WHO Consultation on Obesity 1997: 1-276

Eidesstattliche Erklärung

Ich versichere an Eides statt, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel verwendet habe. Wörtlich oder dem Sinn nach entnommene Stellen sind unter Angabe der Quelle kenntlich gemacht.

Anja Köchel

Anhang 1: Tabellen der logistischen Regression

Tabelle 19: Basismodell der Rückwärts-Elimination 1a

Hormoneinnahme ever					
	Regressions- koeffizient B	Standard- fehler	Wald	Sig.	Exp(B)
gebjahr_class			20,999	0,000	
gebjahr_class(1)	0,341	0,076	20,183	0,000	1,406
gebjahr_class(2)	0,167	0,069	5,829	0,016	1,181
gebjahr_class(3)	0,062	0,084	0,535	0,465	1,063
gebjahr_class(4)	0,134	0,103	1,670	0,196	1,143
zentrum(1)	0,104	0,050	4,323	0,038	1,110
BMI_Klassen			18,360	0,003	
BMI_Klassen(1)	0,148	0,197	0,564	0,453	1,160
BMI_Klassen(2)	-0,105	0,055	3,596	0,058	0,901
BMI_Klassen(3)	-0,225	0,075	9,003	0,003	0,798
BMI_Klassen(4)	-0,378	0,130	8,506	0,004	0,685
BMI_Klassen(5)	-0,404	0,227	3,164	0,075	0,667
kinder_bis3			4,285	0,232	
kinder_bis3(1)	0,013	0,078	0,027	0,870	1,013
kinder_bis3(2)	-0,059	0,074	0,620	0,431	0,943
kinder_bis3(3)	0,075	0,083	0,816	0,366	1,078
kontrazeptiva(1)	-0,437	0,054	66,409	0,000	0,646
ht_ever_gt3mon(1)	-0,838	0,050	282,803	0,000	0,432
pfl_evernever(1)	-1,061	0,100	112,894	0,000	0,346
schule			112,516	0,000	
schule(1)	-0,491	0,059	68,382	0,000	0,612
schule(2)	-0,719	0,086	70,058	0,000	0,487
erwerb(1)	-0,120	0,070	2,938	0,086	0,887
rauchen	0,032	0,051	0,402	0,526	1,033
Konstante	0,006	0,102	0,003	0,957	1,006

* grau: die Merkmale, die bei der backward selection entfernt wurden

* fett: die Hormonvariablen, die in den verschiedenen Modellen separat betrachtet wurden

Tabelle 20: Endmodell der Rückwärts-Elimination 1a

Hormoneinnahme ever					
	Regressions- koeffizient B	Standard- fehler	Wald	Sig.	Exp(B)
gebjahr_class			21,425	0,000	
gebjahr_class(1)	0,342	0,075	20,582	0,000	1,408
gebjahr_class(2)	0,166	0,069	5,845	0,016	1,180
gebjahr_class(3)	0,061	0,084	0,536	0,464	1,063
gebjahr_class(4)	0,132	0,103	1,635	0,201	1,141
zentrum(1)	0,102	0,050	4,237	0,040	1,107
BMI_Klassen			17,927	0,003	
BMI_Klassen(1)	0,159	0,197	0,648	0,421	1,172
BMI_Klassen(2)	-0,103	0,055	3,477	0,062	0,902
BMI_Klassen(3)	-0,221	0,075	8,709	0,003	0,802
BMI_Klassen(4)	-0,371	0,129	8,266	0,004	0,690
BMI_Klassen(5)	-0,384	0,227	2,868	0,090	0,681
kontrazeptiva(1)	-0,434	0,053	66,869	0,000	0,648
ht_ever_gt3mon(1)	-0,842	0,050	285,655	0,000	0,431
pfl_evernever(1)	-1,066	0,100	114,084	0,000	0,344
schule			114,767	0,000	
schule(1)	-0,496	0,059	70,324	0,000	0,609
schule(2)	-0,715	0,085	70,201	0,000	0,489
erwerb(1)	-0,120	0,070	2,977	0,084	0,887
Konstante	0,017	0,080	0,046	0,830	1,017

* fett: die Hormonvariablen, die in den verschiedenen Modellen separat betrachtet wurden

Tabelle 21: Basismodell der Rückwärts-Elimination 1b

Hormoneinnahme current, past, never statt Hormoneinnahme ever					
	Regressions- koeffizient B	Standard- fehler	Wald	Sig.	Exp(B)
gebjahr_class			19,728	0,001	
gebjahr_class(1)	0,335	0,076	19,318	0,000	1,398
gebjahr_class(2)	0,163	0,069	5,502	0,019	1,177
gebjahr_class(3)	0,084	0,085	0,983	0,321	1,088
gebjahr_class(4)	0,138	0,105	1,734	0,188	1,148
zentrum(1)	0,080	0,051	2,538	0,111	1,084
BMI_Klassen			18,764	0,002	
BMI_Klassen(1)	0,128	0,199	0,411	0,522	1,136
BMI_Klassen(2)	-0,103	0,056	3,409	0,065	0,902
BMI_Klassen(3)	-0,224	0,075	8,820	0,003	0,799
BMI_Klassen(4)	-0,407	0,131	9,662	0,002	0,666
BMI_Klassen(5)	-0,397	0,228	3,035	0,081	0,672
kinder_bis3			4,152	0,245	
kinder_bis3(1)	0,015	0,079	0,034	0,854	1,015
kinder_bis3(2)	-0,067	0,075	0,794	0,373	0,935
kinder_bis3(3)	0,061	0,083	0,543	0,461	1,063
kontrazeptiva(1)	-0,441	0,054	66,753	0,000	0,644
ht_use			304,174	0,000	
ht_use(1)	-0,766	0,059	168,854	0,000	0,465
ht_use(2)	-1,004	0,068	219,678	0,000	0,366
schule			107,733	0,000	
schule(1)	-0,481	0,060	64,764	0,000	0,618
schule(2)	-0,713	0,087	67,758	0,000	0,490
pfl_evernever(1)	-1,067	0,100	113,503	0,000	0,344
erwerb(1)	-0,130	0,070	3,404	0,065	0,878
rauchen	0,030	0,052	0,329	0,566	1,030
Konstante	0,026	0,103	0,065	0,799	1,027

* grau: die Merkmale, die bei der backward selection entfernt wurden

* fett: die Hormonvariablen, die in den verschiedenen Modellen separat betrachtet wurden

Tabelle 22: Endmodell der Rückwärts-Elimination 1b

Hormoneinnahme current, past, never statt Hormoneinnahme ever					
	Regressions- koeffizient B	Standard- fehler	Wald	Sig.	Exp(B)
gebjahr_class			20,079	0,000	
gebjahr_class(1)	0,337	0,076	19,672	0,000	1,400
gebjahr_class(2)	0,162	0,069	5,487	0,019	1,176
gebjahr_class(3)	0,085	0,085	1,002	0,317	1,088
gebjahr_class(4)	0,137	0,104	1,720	0,190	1,147
zentrum(1)	0,079	0,050	2,523	0,112	1,082
BMI_Klassen			18,494	0,002	
BMI_Klassen(1)	0,139	0,199	0,488	0,485	1,149
BMI_Klassen(2)	-0,101	0,055	3,321	0,068	0,904
BMI_Klassen(3)	-0,220	0,075	8,594	0,003	0,802
BMI_Klassen(4)	-0,402	0,130	9,510	0,002	0,669
BMI_Klassen(5)	-0,380	0,228	2,782	0,095	0,684
kontrazeptiva(1)	-0,439	0,053	67,606	0,000	0,645
ht_use			306,962	0,000	
ht_use(1)	-0,769	0,059	170,516	0,000	0,463
ht_use(2)	-1,007	0,068	221,435	0,000	0,365
schule			109,794	0,000	
schule(1)	-0,485	0,059	66,518	0,000	0,616
schule(2)	-0,709	0,086	67,892	0,000	0,492
pfl_evernever(1)	-1,071	0,100	114,604	0,000	0,342
erwerb(1)	-0,131	0,070	3,456	0,063	0,878
Konstante	0,031	0,081	0,152	0,696	1,032

* fett: die Hormonvariablen, die in den verschiedenen Modellen separat betrachtet wurden

Tabelle 23: Basismodell der Rückwärts-Elimination 1c

Einnahmedauer von Hormonen statt Hormoneinnahme ever					
	Regressions- koeffizient B	Standard- fehler	Wald	Sig.	Exp(B)
gebjahr_class			15,984	0,003	
gebjahr_class(1)	0,309	0,078	15,689	0,000	1,362
gebjahr_class(2)	0,138	0,071	3,752	0,053	1,148
gebjahr_class(3)	0,083	0,086	0,914	0,339	1,086
gebjahr_class(4)	0,113	0,107	1,114	0,291	1,119
zentrum(1)	0,110	0,052	4,574	0,032	1,116
BMI_Klassen			17,058	0,004	
BMI_Klassen(1)	0,093	0,205	0,204	0,651	1,097
BMI_Klassen(2)	-0,090	0,057	2,505	0,114	0,914
BMI_Klassen(3)	-0,201	0,077	6,896	0,009	0,818
BMI_Klassen(4)	-0,414	0,133	9,621	0,002	0,661
BMI_Klassen(5)	-0,436	0,234	3,479	0,062	0,646
kinder_bis3			5,576	0,134	
kinder_bis3(1)	0,064	0,080	0,634	0,426	1,066
kinder_bis3(2)	-0,042	0,077	0,299	0,585	0,959
kinder_bis3(3)	0,102	0,085	1,452	0,228	1,108
kontrazeptiva(1)	-0,445	0,055	65,782	0,000	0,641
Hormone_Dauer_Klassen			374,725	0,000	
Hormone_Dauer_Klassen(1)	-1,084	0,081	180,366	0,000	0,338
Hormone_Dauer_Klassen(2)	-0,926	0,088	110,370	0,000	0,396
Hormone_Dauer_Klassen(3)	-1,088	0,091	142,183	0,000	0,337
Hormone_Dauer_Klassen(4)	-0,874	0,083	112,018	0,000	0,417
schule			102,606	0,000	
schule(1)	-0,478	0,061	61,737	0,000	0,620
schule(2)	-0,708	0,088	64,373	0,000	0,493
pfl_evernever(1)	-1,072	0,103	108,101	0,000	0,342
erwerb(1)	-0,125	0,072	3,039	0,081	0,883
rauchen	0,030	0,053	0,322	0,571	1,030
Konstante	-0,010	0,105	0,010	0,922	0,990

* grau: die Merkmale, die bei der backward selection entfernt wurden

* fett: die Hormonvariablen, die in den verschiedenen Modellen separat betrachtet wurden

Tabelle 24: Endmodell der Rückwärts-Elimination 1c

Einnahmedauer von Hormonen statt Hormoneinnahme ever					
	Regressions- koeffizient B	Standard- fehler	Wald	Sig.	Exp(B)
gebjahr_class			16,224	0,003	
gebjahr_class(1)	0,310	0,078	15,928	0,000	1,364
gebjahr_class(2)	0,137	0,071	3,722	0,054	1,147
gebjahr_class(3)	0,083	0,086	0,917	0,338	1,086
gebjahr_class(4)	0,109	0,107	1,056	0,304	1,116
zentrum(1)	0,106	0,051	4,399	0,036	1,112
BMI_Klassen			16,662	0,005	
BMI_Klassen(1)	0,102	0,205	0,247	0,619	1,107
BMI_Klassen(2)	-0,087	0,056	2,401	0,121	0,916
BMI_Klassen(3)	-0,197	0,076	6,644	0,010	0,821
BMI_Klassen(4)	-0,409	0,133	9,463	0,002	0,664
BMI_Klassen(5)	-0,416	0,233	3,180	0,075	0,660
kontrazeptiva(1)	-0,443	0,054	66,176	0,000	0,642
Hormone_Dauer_Klassen			377,880	0,000	
Hormone_Dauer_Klassen(1)	-1,084	0,081	180,355	0,000	0,338
Hormone_Dauer_Klassen(2)	-0,932	0,088	112,002	0,000	0,394
Hormone_Dauer_Klassen(3)	-1,094	0,091	143,946	0,000	0,335
Hormone_Dauer_Klassen(4)	-0,878	0,083	113,214	0,000	0,416
schule			105,932	0,000	
schule(1)	-0,486	0,061	64,173	0,000	0,615
schule(2)	-0,709	0,088	65,433	0,000	0,492
pfl_evernever(1)	-1,077	0,103	109,423	0,000	0,341
erwerb(1)	-0,126	0,071	3,114	0,078	0,882
Konstante	0,029	0,082	0,122	0,727	1,029

* fett: die Hormonvariablen, die in den verschiedenen Modellen separat betrachtet wurden

Tabelle 25: Basismodell der Rückwärts-Elimination 2

Hormoneinnahme ever, Studienregion als Interaktionsterm					
	Regressions- koeffizient B	Standard- fehler	Wald	Sig.	Exp(B)
gebjahr_class			21,664	0,000	
gebjahr_class(1)	0,347	0,076	20,768	0,000	1,414
gebjahr_class(2)	0,165	0,069	5,670	0,017	1,179
gebjahr_class(3)	0,065	0,084	0,587	0,444	1,067
gebjahr_class(4)	0,144	0,104	1,921	0,166	1,155
zentrum(1)	0,350	0,172	4,136	0,042	1,419
BMI_Klassen			12,313	0,031	
BMI_Klassen(1)	0,108	0,354	0,093	0,761	1,114
BMI_Klassen(2)	-0,054	0,084	0,411	0,521	0,947
BMI_Klassen(3)	-0,219	0,108	4,097	0,043	0,803
BMI_Klassen(4)	-0,579	0,189	9,345	0,002	0,561
BMI_Klassen(5)	-0,015	0,306	0,003	0,960	0,985
kinder_bis3			4,401	0,221	
kinder_bis3(1)	0,095	0,133	0,511	0,475	1,100
kinder_bis3(2)	0,077	0,127	0,372	0,542	1,080
kinder_bis3(3)	0,234	0,132	3,127	0,077	1,264
kontrazeptiva(1)	-0,546	0,078	48,874	0,000	0,579
ht_ever_gt3mon(1)	-0,727	0,076	92,200	0,000	0,483
pfl_evernever(1)	-1,292	0,166	60,175	0,000	0,275
schule			27,985	0,000	
schule(1)	-0,400	0,098	16,712	0,000	0,670
schule(2)	-0,563	0,140	16,203	0,000	0,569
erwerb(1)	-0,115	0,097	1,407	0,235	0,892
rauchen	0,098	0,079	1,529	0,216	1,103
zentrum * BMI_Klassen			7,313	0,198	
zentrum(1) by BMI_Klassen(1)	0,049	0,427	0,013	0,908	1,050
zentrum(1) by BMI_Klassen(2)	-0,092	0,112	0,679	0,410	0,912
zentrum(1) by BMI_Klassen(3)	-0,003	0,151	0,000	0,985	0,997
zentrum(1) by BMI_Klassen(4)	0,411	0,261	2,488	0,115	1,508
zentrum(1) by BMI_Klassen(5)	-0,878	0,477	3,387	0,066	0,416
zentrum * kinder_bis3			2,634	0,452	
zentrum(1) by kinder_bis3(1)	-0,116	0,165	0,498	0,480	0,890
zentrum(1) by kinder_bis3(2)	-0,203	0,157	1,664	0,197	0,817
zentrum(1) by kinder_bis3(3)	-0,252	0,172	2,161	0,142	0,777
zentrum(1) by kontrazeptiva(1)	0,195	0,104	3,554	0,059	1,216
zentrum(1) by ht_ever_gt3mon(1)	-0,191	0,100	3,662	0,056	0,826
zentrum(1) by pfl_evernever(1)	0,365	0,208	3,072	0,080	1,440
zentrum * schule			3,005	0,223	
zentrum(1) by schule(1)	-0,150	0,123	1,482	0,224	0,861
zentrum(1) by schule(2)	-0,258	0,177	2,130	0,144	0,772
zentrum(1) by erwerb(1)	-0,015	0,118	0,017	0,897	0,985
zentrum(1) by rauchen	-0,113	0,103	1,193	0,275	0,893
Konstante	-0,153	0,145	1,114	0,291	0,858

* grau: die Merkmale, die bei der backward selection entfernt wurden

* fett: die Hormonvariablen, die in den verschiedenen Modellen separat betrachtet wurden

Tabelle 26: Endmodell der Rückwärts-Elimination 2

Hormoneinnahme ever, Studienregion als Interaktionsterm					
	Regressions- koeffizient B	Standard- fehler	Wald	Sig.	Exp(B)
gebjahr_class			29,313	0,000	
gebjahr_class(1)	0,367	0,074	24,442	0,000	1,443
gebjahr_class(2)	0,186	0,068	7,578	0,006	1,205
gebjahr_class(3)	0,016	0,080	0,039	0,844	1,016
gebjahr_class(4)	0,073	0,097	0,557	0,455	1,075
zentrum(1)	0,101	0,050	4,160	0,041	1,106
BMI_Klassen			17,547	0,004	
BMI_Klassen(1)	0,162	0,197	0,676	0,411	1,176
BMI_Klassen(2)	-0,102	0,055	3,433	0,064	0,903
BMI_Klassen(3)	-0,218	0,075	8,464	0,004	0,804
BMI_Klassen(4)	-0,366	0,129	8,075	0,004	0,693
BMI_Klassen(5)	-0,378	0,227	2,784	0,095	0,685
kontrazeptiva(1)	-0,436	0,053	67,569	0,000	0,646
ht_ever_gt3mon(1)	-0,841	0,050	285,191	0,000	0,431
pfl_evernever(1)	-1,068	0,100	114,553	0,000	0,344
schule			118,748	0,000	
schule(1)	-0,500	0,059	71,897	0,000	0,606
schule(2)	-0,727	0,085	73,207	0,000	0,483
Konstante	-0,009	0,079	0,014	0,907	0,991

* fett: die Hormonvariablen, die in den verschiedenen Modellen separat betrachtet wurden

Anhang 2: Kurzfragebogen für Nicht-Teilnehmer 77

(telefonische Version)

Studiennr |_____| |_____| |_____| |_____| |_____| |_____| |_____| |_____| Geburtsdatum |_____| |_____| |_____| |_____| |_____| |_____| |_____|
nur in HH komplett erfragt

Interviewer-Kürzel |_____| |_____| Datum der Befragung: |_____| |_____| |_____| |_____| |_____| |_____|

Mamma CA ? nein keine Angabe/weiß nicht ja |_____| |_____| |_____| |_____| Jahr
nur in HH erfragt

Teilnahme: ja nein Grund für die Nichtteilnahme am Kurzfragebogen:

Einleitungstext (immer vorlesen):

„Schade, dass sie nicht teilnehmen möchten. Ich habe noch eine Bitte: Ich würde Ihnen gerne am Telefon einige kurze Fragen stellen. Die Beantwortung dauert 1-2 Minuten.“

1. Haben Sie Kinder?

nein keine Angabe / weiß nicht ja → wie viele Kinder haben Sie? |_____| |_____|

2. Haben Sie jemals orale Verhütungsmittel („die Pille“) eingenommen?

nein keine Angabe / weiß nicht ja

3. Haben Sie jemals zur Behandlung von Wechseljahresbeschwerden Hormone (Östrogene, Gestagene...) oder pflanzliche / homöopathische Mittel eingenommen?

(Bitte denken Sie daran, dass diese Präparate auch nach Entfernung beider Eierstöcke oder zur Vorbeugung bestimmter Erkrankungen (z.B. Osteoporose) verordnet werden und das es Hormone nicht nur in Tablettenform sondern auch als Pflaster, Scheidencreme u.a. gibt.)

nein keine Angabe / weiß nicht ja → Hormonpräparat
→ pflanzliches/homöop. Präparat

Name/n: * ← → keine Angabe / weiß nicht
*Nutzen Sie bitte ggf. das Kommentarfeld unten (Mehrfachangaben möglich)

Nur wenn es ein Hormonpräparat war:

→ wann haben Sie mit der Einnahme begonnen ? |_____| |_____| |_____| |_____| Jahr

→ nehmen Sie das Präparat zur Zeit noch ein? ja nein → bis wann? |_____| |_____| |_____| |_____|

4. Wie groß sind Sie und wie viel wiegen Sie ? (bitte aufrunden)

Jahr

Körpergröße |_____| |_____| |_____| cm Gewicht |_____| |_____| |_____| Kg keine Angabe / weiß nicht

5. Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?

- keinen keine Angabe / weiß nicht
 Hauptschulabschluss / Volksschulabschluss Realschulabschluss (Mittl. Reife)
 Polytechnische Oberschule Fachhochschulreife
10. Klasse (vor 1965 8. Klasse)
 Abitur, allg. Hochschulreife, erweiterte Oberschule (EOS) mit Facharbeiterabschluss
 anderer Abschluss:.....

6. Sind Sie zur Zeit erwerbstätig?

nein keine Angabe / weiß nicht ja
↳ falls nein: ich war nie erwerbstätig oder ich war erwerbstätig bis |_____| |_____| |_____| |_____|

7. Wie oft essen Sie durchschnittlich Salat oder Gemüse pro Woche?

Jahr

weniger als 1 mal 1 mal 2-3 mal 4-6 mal täglich mehrmals täglich
 esse ich gar nicht keine Angabe / weiß nicht

8. Haben Sie jemals regelmäßig geraucht?

nein keine Angabe / weiß nicht ja → ich rauche gegenwärtig
→ ich habe früher geraucht

„Herzlichen Dank für Ihre Bereitschaft, unsere Fragen zu beantworten!“

Kommentar:
.....

Bitte füllen Sie den folgenden Fragebogen aus, indem Sie die vorgesehenen Kästchen ankreuzen bzw. Angaben in den dafür vorgesehenen Feldern machen.

Geburtsdatum |__|_|.|__|_|.|__|_|_|_| Datum: |__|_| . |__|_| . |__|_|_|_|

1. Hatten Sie jemals Brustkrebs?

nein keine Angabe / weiß nicht ja → wann? |__|_|_|_| Jahr

2. Haben Sie Kinder?

nein keine Angabe / weiß nicht ja → wie viele Kinder haben Sie? |__|_|

3. Haben Sie jemals orale Verhütungsmittel („die Pille“) eingenommen?

nein keine Angabe / weiß nicht ja

4. Haben Sie jemals zur Behandlung von Wechseljahresbeschwerden, nach Entfernung beider Eierstöcke oder zur Vorbeugung bestimmter Erkrankungen Hormone (Östrogene, Gestagene...) oder pflanzliche / homöopathische Mittel eingenommen?

(Bitte denken Sie daran, dass es Hormone nicht nur in Tablettenform sondern auch als Pflaster, Scheidencreme u.a. gibt.)

nein keine Angabe / weiß nicht ja → Hormonpräparat
→ pflanzliches/homöop. Präparat

Name/n: * ← → keine Angabe / weiß nicht
Nutzen Sie bitte ggf. das Kommentarfeld unten (Mehrfachangaben möglich)

Wenn es ein Hormonpräparat war:

→ wann haben Sie mit der Einnahme begonnen? |__|_|_|_| Jahr

→ nehmen Sie das Präparat zur Zeit noch ein? ja nein → bis wann? |__|_|_|_|

5. Wie groß sind Sie und wie viel wiegen Sie? (bitte aufrunden)

Jahr

Körpergröße |__|_|_| cm Gewicht |__|_|_| Kg keine Angabe / weiß nicht

6. Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?

- | | |
|--|--|
| <input type="checkbox"/> keine | <input type="checkbox"/> keine Angabe / weiß nicht |
| <input type="checkbox"/> Hauptschulabschluss / Volksschulabschluss | <input type="checkbox"/> Realschulabschluss (Mittl. Reife) |
| <input type="checkbox"/> Polytechnische Oberschule
10. Klasse (vor 1965 8. Klasse) | <input type="checkbox"/> Fachhochschulreife |
| <input type="checkbox"/> Abitur, allg. Hochschulreife, erweiterte Oberschule (EOS) mit Facharbeiterabschluss | |
| <input type="checkbox"/> anderer Abschluss:..... | |

7. Sind Sie zur Zeit erwerbstätig?

nein keine Angabe / weiß nicht ja

↳ falls nein: ich war nie erwerbstätig oder ich war erwerbstätig bis
|__|_|_|_| Jahr

8. Wie oft essen Sie durchschnittlich Salat oder Gemüse pro Woche?

weniger als 1 mal 1 mal 2-3 mal 4-6 mal täglich mehrmals
täglich

esse ich gar nicht keine Angabe / weiß nicht

9. Haben Sie jemals regelmäßig geraucht?

nein keine Angabe / weiß nicht ja → ich rauche gegenwärtig
→ ich habe früher geraucht

Herzlichen Dank für Ihre Bereitschaft, unsere Fragen zu beantworten!

Kommentar:

.....

Die erhobenen Daten werden anonymisiert, d.h. ohne Bezug zu Ihrem Namen und Ihrer Anschrift, in Datenverarbeitungsanlagen der am Forschungsprojekt beteiligten Institutionen gespeichert werden. Die Auswertung der Daten geschieht ausschließlich anonymisiert.

In den folgenden Tabellen werden die Erhebungszeiträume, Besonderheiten und die Testversionen bis zur endgültigen Version wiedergegeben, sowie die Anzahl der mit dieser Version befragten Probandinnen. Hervorgehoben sind die in dieser Version erfragten Items:

Version1: 5 Items

- **Kinder**
- Kontrazeptiva fehlt
- **HRT**
 - Current/Dauer fehlt
- Größe/Gewicht fehlt
- **Schulabschluss,**
- **Erwerbstätigkeit current**
 - Bis wann/nie fehlt
- Salat/Gemüse fehlt
 - Häufigkeit fehlt
- **Rauchen jemals**
 - **Regelmäßigkeit nur als Unterpunkt erfragt**

Tabelle 27: Kurzfragebogen Version 1

Erhebungszeitraum gesamt	07.05.2003-06.06.2003	Version	Testversionen	N
HH tel. Interview	07.05.2003-06.06.2003	1c	1a, 1b	132
HH Selbstausfüller	--	--		
HD tel. Interview	--	--		
HD Selbstausfüller	--	--		
				132

Version2: 8 Items, alle Hauptfragekomplexe vorhanden
(es fehlen MaCa, Gebdat, HRT current/Dauer)

- **Kinder**
- **Kontrazeptiva**
- **HRT**
 - Current/Dauer fehlt
- **Größe/Gewicht**
- **Schulabschluss**
- **Erwerbstätigkeit current**
 - **nie/bis wann**
- **Salat/Gemüse**
 - **Häufigkeit**
- **Rauchen regelmäßig**
 - **current/past**

Tabelle 28: Kurzfragebogen Version 2

Erhebungszeitraum gesamt	10.06.03 - 26.06.03	Version	Testversionen	N
HH tel. Interview	10.06.03 - 26.06.03	2g	2a, 2b, 2c, 2d, 2e	103
HH Selbstausfüller	--		2f	
HD tel. Interview	--			
HD Selbstausfüller	--			
				103

Version3: 8 Items, alle Hauptfragekomplexe vorhanden
(es fehlen MaCa, Gebdat)

- **Kinder**
- **Kontrazeptiva**
- **HRT**
 - **Current/Dauer**
- **Größe/Gewicht**
- **Schulabschluss**
- **Erwerbstätigkeit current**
 - **nie/bis wann**
- **Salat/Gemüse**
 - **Häufigkeit**
- **Rauchen regelmäßig**
 - **current/past**

Tabelle 29: Kurzfragebogen Version 3

Erhebungszeitraum gesamt	03.07.2003-13.03.2006	Version	Besonderheiten	N
HH tel. Interview	03.07.2003-07.07.2003	3a	Langer HRT-Text	13
HH tel. Interview	07.07.2003-18.07.2003	3c	Kurzer HRT-Text	49
Ausreißer	22.10.2003	3c		1
HD tel. Interview	28.08.2003-28.02.2006	3d	Kurzer HRT-Text	1258
HD Selbstausfüller	10.03.2005-13.3.2006	3b	Langer HRT-Text	138
				1459

Version4: 8 Items, alle Hauptfragekomplexe vorhanden,
wie Version 3, zusätzlich Geburtsdatum erfragt (nur HH)

Tabelle 30: Kurzfragebogen Version 4

Erhebungszeitraum gesamt	09.07.2003-07.03.2005	Version	Besonderheiten	N
HH tel. Interview	18.07.2003-02.04.2004	4a	kurzer HRT-Text	1127
Ausreißer	05.04.2004-08.07.2004			65
Ausreißer	07.02.2005-21.07.2005			6
HH Selbstausfüller	09.07.2003-07.03.2005	4b	langer HRT-Text	109
Ausreißer	09.06.2005-02.12.2005			17
HD tel. Interview	--			
HD Selbstausfüller	--			
				1324

Version 5: 8 Items, alle Hauptfragekomplexe vorhanden,
wie Version 3; zusätzlich Geburtsdatum und MammaCa erfragt (nur HH)
in der schriftlichen Version wurde eine Brustkrebserkrankung als 9. Item erfragt

Tabelle 31: Kurzfragebogen Version 5

Erhebungszeitraum gesamt	02.04.2004-08.01.2006	Version	Testversionen	N
HH tel. Interview	02.04.2004-11.01.2006	5a		1555
HH Selbstausfüller	02.06.2005-08.01.2006	5b		147
HD tel. Interview	--			
HD Selbstausfüller	--			
				1702

Anhang 5: Hauptfragebogen MARIE
(nur mit dem Kurzfragebogen vergleichbare Items)

Studiennummer: |_|_|_|_|_|_|_|_|

Interviewer-Nummer: |_|_|_|_|_|_|_|_|



Mammakarzinom-Risikofaktoren-Erhebung

Datum des Interviews: |_|_| . |_|_| . |_|_|_|_|_|

Uhrzeit Beginn: |_|_|:|_|_|

Uhrzeit Ende: |_|_|:|_|_|

Ort des Interviews:

- Klinik
- zu Hause bei Probandin
- anderer Ort:

falls Interview nicht zu Ende geführt werden konnte, Grund:

Vielen Dank für Ihre Bereitschaft, an unserer Studie teilzunehmen. Ich möchte Ihnen Fragen stellen zu Ihrer medizinischen Vorgeschichte und einigen Lebensstilfaktoren.

Wir versichern Ihnen, dass alle Daten bei uns vertraulich behandelt werden und nur getrennt von Ihrem Namen im Computer gespeichert werden.

Ihre Teilnahme ist freiwillig. Obwohl es sehr wichtig ist, dass wir von jeder Teilnehmerin zuverlässige Antworten erhalten, müssen Sie Fragen nicht beantworten, die Sie nicht beantworten möchten.

Sie haben die Studieninformation für Teilnehmerinnen erhalten. Haben Sie hierzu noch Fragen?

Ich beantworte diese gerne jetzt vor Beginn des Interviews.

Allgemeine Fragen

Ich möchte das Interview beginnen mit allgemeinen Fragen zu Ihrer Person.

4. Wann wurden Sie geboren? |_|_| . |_|_| . |_|_|_|_|
7. Wie groß sind Sie? |_|_|_| cm
8. Wie viel wiegen Sie zur Zeit? |_|_|_| kg

Ausbildung und Beruf

13. Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?

- keinen
- Hauptschulabschluss / Volksschulabschluss
- Realschulabschluss (Mittlere Reife)
- Polytechnische Oberschule, 10. Klasse (vor 1965 8. Klasse)
- Fachhochschulreife
- Abitur, allg. Hochschulreife, erweiterte Oberschule (EOS) mit Facharbeiterabschluss
- anderer Abschluss:
- keine Angabe / weiß nicht

15. Sind Sie zur Zeit (bei Patientinnen vor der Erkrankung) erwerbstätig?

Unter Erwerbstätigkeit wird jede bezahlte bzw. mit einem Einkommen verbundene Tätigkeit verstanden, egal welchen zeitlichen Umfang sie hat.

- nein → bis wann waren Sie erwerbstätig?: |_|_|_|_| Jahr
- ich war nie erwerbstätig
- ja, und zwar (Art der Erwerbstätigkeit entfällt)

Schwangerschaft

19. Waren Sie jemals schwanger?

(einschließlich Schwangerschaftsabbrüche, Fehlgeburten und Totgeburten)

nein ja, wie viele Schwangerschaften hatten Sie? |__|__|

keine Angabe / weiß nicht

22. Ich möchte Sie nun zu jeder Ihrer Schwangerschaften im einzelnen befragen

	1. Schwangerschaft	2. Schwangerschaft
A. Wie endete Ihre 1. (2., 3.) Schwangerschaft? Lebendgeburt Totgeburt Fehlgeburt Abbruch (Abtreibung) keine Angabe/ weiß nicht	<input type="checkbox"/> <input type="checkbox"/> → c. <input type="checkbox"/> → c. <input type="checkbox"/> → c. <input type="checkbox"/> → c.	<input type="checkbox"/> <input type="checkbox"/> → c. <input type="checkbox"/> → c. <input type="checkbox"/> → c. <input type="checkbox"/> → c.
B. Welches Geschlecht hatte das Kind/ die Kinder? männlich weiblich 1 männlich / 1 weiblich 2 männlich 2 weiblich 1 männlich / 2 weiblich 2 männlich / 1 weiblich 3 männlich 3 weiblich	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
C. Wie viele Wochen / Monate dauerte die Schwangerschaft?	__ __ Wochen oder __ __ Monate	__ __ Wochen oder __ __ Monate
D. Wann wurde das Kind geboren? / Wann endete die Schwangerschaft?	__ __ __ __ __ __ Monat Jahr	__ __ __ __ __ __ Monat Jahr
E. Haben Sie das Kind / die Kinder gestillt? (nur bei Lebendgeburten) ja nein keine Angabe / weiß nicht	<input type="checkbox"/> <input type="checkbox"/> → nächste Schw. <input type="checkbox"/> → nächste Schw.	<input type="checkbox"/> <input type="checkbox"/> → nächste Schw. <input type="checkbox"/> → nächste Schw.
F. Wie lange haben Sie das Kind/ die Kinder gestillt?	__ __ Wochen __ __ Monate	__ __ Wochen __ __ Monate

usw. bis 9. Schwangerschaft

Empfängnisverhütung

23. Haben Sie jemals eine der folgenden Empfängnisverhütungsmethoden angewandt?

(Mehrfachnennungen möglich)

- Antibaby-Pille
- Implantat (Depotgelbkörperhormon im Unterhautfettgewebe des Oberarms)
- Spirale
- Spirale mit Gelbkörperhormon
- Dreimonatsspritze

- Sterilisation wann war das genau? |__|__| Monat |__|__|__|__| Jahr
- andere (z.B. Kondome, Diaphragma/Pessar, Temperaturkalender, Creme, Schaum)

- keines davon (nein)
- keine Angabe / weiß nicht

24. Haben Sie jemals aus anderen Gründen als zur Empfängnisverhütung eines der folgenden Mittel / Präparate angewendet?

Probing: z.B. wegen unregelmäßiger Periode, Unterleibsbeschwerden (stärkere Beschwerden von/bei/nach der Regelblutung oder zum Zeitpunkt des Eisprungs) oder Akne

(Mehrfachnennungen möglich)

- Antibaby-Pille
- Implantat (Depotgelbkörperhormon im Unterhautfettgewebe des Oberarms)
- Dreimonatsspritze
- keines davon (nein)
- keine Angabe / weiß nicht

Intervieweranweisung: Wenn bei Frage 23 und bei Frage 24 nur „keines davon“, „keine Angabe / weiß nicht“, „Sterilisation“ oder „andere“ angegeben ist, weiter mit Frage 26

Hormontherapie in den Wechseljahren

Einleitungstext (immer vorlesen):

Zur Linderung von Wechseljahresbeschwerden oder zur Vorbeugung bestimmter Erkrankungen (z.B. Osteoporose) können verschiedene Mittel angewendet werden. Manche Frauen bekommen diese Mittel auch nach der Entfernung beider Eierstöcke.

Hierzu zählen z. B. Östrogene, die es in Form von Tabletten, Pflaster, Gel, Scheidencreme oder -zäpfchen, Spritzen, Tropfen, Scheidenring oder Nasenspray gibt.

Außerdem gibt es auch Gelbkörperhormone, sogenannte Gestagene, die manchmal zusätzlich zu den Östrogenen verordnet werden oder auch alleine angewendet werden.

Auch pflanzliche Präparate, Milzextrakte oder homöopathische Mittel können in den Wechseljahren angewendet werden.

38. Haben Sie jemals Hormone (Östrogene, Gestagene...), oder pflanzliche / homöopathische Mittel angewendet?

Nicht gemeint sind Hormone bei Schilddrüsenerkrankungen.

- ja
 nein → Frage 43
 keine Angabe / weiß nicht → Frage 43

40. Ich möchte Sie nun als nächstes zu jedem einzelnen Zeitraum Ihrer Hormoneinnahme bzw. Ihrer Einnahme von pflanzlichen / homöopathischen Mitteln befragen.

	1. Zeitraum	2. Zeitraum																																												
A. Wann haben Sie zum 1. (2.,3.) Mal eines der Mittel / das 2. Mittel angewandt? In welchem Monat und Jahr?	<table style="width: 100%; border: none;"> <tr> <td style="border: none;"> _ _ _ _ _ _ _ </td> <td style="border: none;"> _ _ _ _ _ _ _ </td> </tr> <tr> <td style="border: none;">Monat Jahr</td> <td style="border: none;">Monat Jahr</td> </tr> </table>	_ _ _ _ _ _ _	_ _ _ _ _ _ _	Monat Jahr	Monat Jahr	<table style="width: 100%; border: none;"> <tr> <td style="border: none;"> _ _ _ _ _ _ _ </td> <td style="border: none;"> _ _ _ _ _ _ _ </td> </tr> <tr> <td style="border: none;">Monat Jahr</td> <td style="border: none;">Monat Jahr</td> </tr> </table>	_ _ _ _ _ _ _	_ _ _ _ _ _ _	Monat Jahr	Monat Jahr																																				
_ _ _ _ _ _ _	_ _ _ _ _ _ _																																													
Monat Jahr	Monat Jahr																																													
_ _ _ _ _ _ _	_ _ _ _ _ _ _																																													
Monat Jahr	Monat Jahr																																													
B. In welcher Form wandten Sie es an? Tabletten, Dragees, Kapseln Pflaster Gel Scheidenzäpfchen / Creme Spritzen Tropfen Beschichteter Scheidenring Nasenspray Anderes Welches? Keine Angabe / weiß nicht	<table style="width: 100%; border: none;"> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> </table>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<table style="width: 100%; border: none;"> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> </table>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
C. Sagen Sie mir bitte den genauen Namen des Mittels / Präparats? <i>(bitte Liste und Abbildungen der Präparate vorlegen zur Gedächtnisstütze, falls der Name nicht erinnerlich ist, bitte unbekannt eintragen).</i>	<table style="width: 100%; border: none;"> <tr><td style="border: none;">Name des Mittels</td><td style="border: none;">Name des Mittels</td></tr> <tr><td style="border: none;"> _ _ _ _ </td><td style="border: none;"> _ _ _ _ </td></tr> <tr><td style="border: none;">Code</td><td style="border: none;">Code</td></tr> </table>	Name des Mittels	Name des Mittels	_ _ _ _	_ _ _ _	Code	Code	<table style="width: 100%; border: none;"> <tr><td style="border: none;">Name des Mittels</td><td style="border: none;">Name des Mittels</td></tr> <tr><td style="border: none;"> _ _ _ _ </td><td style="border: none;"> _ _ _ _ </td></tr> <tr><td style="border: none;">Code</td><td style="border: none;">Code</td></tr> </table>	Name des Mittels	Name des Mittels	_ _ _ _	_ _ _ _	Code	Code																																
Name des Mittels	Name des Mittels																																													
_ _ _ _	_ _ _ _																																													
Code	Code																																													
Name des Mittels	Name des Mittels																																													
_ _ _ _	_ _ _ _																																													
Code	Code																																													
D. Falls der Name unbekannt ist, war es ein Hormonpräparat Pflanzliches / homöopathisches Präp. Keine Angabe / weiß nicht	<table style="width: 100%; border: none;"> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> </table>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<table style="width: 100%; border: none;"> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> <tr><td style="border: none;"><input type="checkbox"/></td><td style="border: none;"><input type="checkbox"/></td></tr> </table>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													
<input type="checkbox"/>	<input type="checkbox"/>																																													

Brusterkrankungen und Tumore

56. Wurde bei Ihnen jemals ein bösartiger Tumor (oder ein in-situ-Karzinom) der Brust diagnostiziert? (Bei Patientinnen ‚ja‘ ankreuzen, und die Frage überspringen)

- nein → Frage 60
 ja
 keine Angabe / weiß nicht → Frage 60

58. Bitte geben Sie an, wann die Operation des Brustkrebses (oder des in-situ-Karzinoms) war.

Datum der Operation / Biopsie: |_|_| . |_|_| . |_|_|_|_|
Tag Monat Jahr

Familienvorgeschichte

73. Wie viele leibliche Kinder haben Sie? (einschließlich mittlerweile verstorbene)
Intervieweranweisung: nicht noch mal fragen, sondern auf bereits genannte Kinder (Schwangerschaften) eingehen.

|_|_| Töchter |_|_| Söhne

Falls keine Töchter weiter mit Frage 76

Rauchen

86. Haben Sie jemals mindestens 1 Zigarette geraucht?

- nein → Frage 90
 ja
 keine Angabe / weiß nicht → Frage 90

87. Haben Sie entweder

- jemals regelmäßig, d.h. mindestens 1 Zigarette pro Tag mindestens 3 Monate lang geraucht?
 nur zu bestimmten Gelegenheiten oder nur ganz kurze Zeit geraucht? → Frage 89
 nur eine Zigarette im Leben probiert und nie wieder geraucht? → Frage 90

Der MARIE-Hauptfragebogen erhob insgesamt 102 Items auf 47 Seiten.

Der zusätzlich erhobene MARIE-Ernährungsfragebogen erhob weitere Items auf 32 Seiten.

Es wurde differenziert nach einzelnen Lebensmitteln (z. B. bestimmte Gemüsesorten in roher oder

gekochter Form, ihre Portionsgröße und die Verzehrhäufigkeit in Sommer- und Wintermonaten erfragt.

Für den Vergleich mit dem Kurzfragebogen wurde die Portionsgröße vernachlässigt, der Verzehr von

Sommer und Winter gemittelt.

Da aber verschiedene Salat- und Gemüsesorten erfragt und die Häufigkeiten addiert wurden, führte dies

im Vergleich zum Kurzfragebogen vermutlich zu überhöhten Werten.