Breast Cancer and Nutrition in the Kilimanjaro Region of Northern Tanzania: a case-control study

Dissertation
submitted to the
Faculty of Agricultural, Nutritional Sciences and Environmental Management,
Justus-Liebig-University Giessen, Germany
for the degree of Dr. oec. troph.

by

Dipl. oec. troph. Irmgard Jordan
Born in Hannover, Germany

Gießen, December 2011

Table of Contents

List of Tables 3

List of Figures 5

List of Abbreviations 6

1 Introduction 7

1.1 Rationale 8
1.2 Research partners and study site 9
1.3 Ethical clearance and outline 11

2 Validation of the food frequency questionnaire 11

2.1 Introduction 11
2.2 Methods 11
2.3 Results 14
2.4 Discussion 18

3 Breast cancer and nutrition 24

3.1 Introduction 24
3.2 Methods 24
3.3 Results 30
3.4 Discussion 58

4 Summary 69

5 Zusammenfassung 73

6 Acknowledgements 77

7 Annex 91

Erklärung 141

List of Tables

Table 1: Energy intake data (kcal) from 24-recall and food frequency questionnaire (t1; t2) stratified by interviewer 15

Table 2: Seasonal variability of food frequency questionnaire and 24h-recall stratified by interviewer based on results of Wilcoxon signed rank test* 16

Table 3: Seasonal variability of food frequency questionnaire stratified by interviewer based on results of a Wilcoxon signed rank test** 16

Table 4: Differences between food frequency questionnaire and 24-recall stratified by interviewer based on results of a Wilcoxon signed rank test*** 17

Table 5: Spearman correlation (r_s) and its level of significance (*P*) of the means of two food frequency questionnaires and two 24h-recalls 18

Table 6: Ethnic distribution in cases and controls..................30

Table 7: Socioeconomic indicators..................31

Table 8: Reproductive parameters and Body Mass Index (BMI) 33

Table 9: Energy and nutrient intake per day based on a food frequency questionnaire 34

Table 10: Summary of Principal Component Analysis with Varimax rotation retaining four dietary patterns based on 36 food groups (PCA 1)..................37

Table 11: Summary of Principal Component Analysis with Varimax rotation retaining six dietary patterns based on 34 food groups (PCA2) 39

Table 12: Spearman correlation coefficients of dietary patterns, socioeconomic parameters and BMI 41

Table 13: Logistic regression of basic breast cancer risk model 42

Table 14: Logistic regression for five groups with increasing lifelong lactation..................43

Table 15: Logistic regression of basic breast cancer risk model stratified for menopausal status..................44

Table 16: Logistic regression: Dietary patterns and breast cancer..................45

Table 17: Logistic regression: Dietary patterns (PCA1) and basic breast cancer risk model 46

Table 18: Logistic regression: Dietary patterns (PCA2), alcoholic beverages and basic breast cancer risk model..........48

Table 19: Logistic regression: Basic risk model, dietary patterns (PCA2) and alcoholic drinks in *premenopausal* women 50

Table 20: Logistic regression: Basic risk model, dietary patterns (PCA2) and alcoholic drinks in *postmenopausal* women 51

Table 21: Logistic regression: Basic risk model and dietary patterns (PCA2) in women with a body mass index below 24kg/m² 54

Table 22: Logistic regression: Basic risk model and dietary patterns (PCA2) in women with a body mass index of 24-26kg/m²..........55

Table 23: Logistic regression: Basic risk model and dietary patterns (PCA2) in women with a body mass index above 26kg/m² 56

List of Figures

Figure 1: Map of Tanzania 9

Figure 2: Maps of the districts of the Kilimanjaro Region, Tanzania 10

Figure 3: Pictogram to estimate the Body Mass Index (BMI) (modified after Stunkard *et al.* 1983) 27

Figure 4: Scree plot of principal component analysis 35

Figure 5: Scree plot of PCA and randomly generated data set................................38

Figure 6: Intake of fat, polyunsaturated and saturated fatty acids per day and the related odds and P/S ratios in quartiles of the Fatty Diet (PCA1) 47

Figure 7: Intake of fat, polyunsaturated and saturated fatty acids per day and the related odds and P/S ratios in quartiles of the Fruity Diet (PCA2) 49

Figure 8: BMI change over time: BMI at interview minus BMI age 20 years............52

Figure 9: Prevalence of hunger before the menarche and effect on body size........57

Figure 10: Prevalence of hunger after menarche and effect on body size.................57

List of Abbreviations

BRCA1	BReast CAncer gene type 1
BRCA2	BReast CAncer gene type 2
FFQ	Food Frequency Questionnaire
FNAC	Fine Needle Aspiration Cytology
IARC	International Agency for Research on Cancer
KCMC	Kilimanjaro Christian Medical Centre
KMO	Kaiser-Maier Olkin Measure
no alc	no alcoholic drinks
PAL	Physical Activity Level
PCA	Principal Component Analysis
t1	time 1 = first data assessment
t2	time 2 = second data assessment
TPSF	Tanzania Private Sector Foundation
WCRF	World Cancer Research Fund
WHO	World Health Organisation

1 Introduction

Breast cancer is the leading cause of death among women worldwide (WHO 2009b; WHO 2009a). Less than 10 % of women's breast cancer cases are caused by inherited mutations of either BReast CAncer 1 (BRCA 1) or BRCA 2 genes (Anderson *et al.* 1993; Blackwood *et al.* 1998). Although distinct mutations in these genes have been identified in black people in the US (Olopade *et al.* 2003), little is known of the prevalence of these mutations as well as of lifestyle factors associated with breast cancer in African populations (Parkin *et al.* 2008).

In Tanzania, breast cancer is currently regarded as the second most common cancer in women after tumours of the cervix uteri. In 2008 the national age-standardised breast cancer rate was estimated to be 10.1 per 100,000 (IARC 2008). Between 1998 and 2006 the World Health Organization (WHO) supported cancer registry based at the Kilimanjaro Christian Medical Centre (KCMC) counted annually an average of 36 new breast cancer cases in the region which results in an estimated crude breast cancer rate of 6.5 per100,000 in 2006. The real number was probably higher, because exclusive consultation with traditional healers and poor access to adequate health care are common. Therefore, a greater number of cancer patients may go undiagnosed. Whether the comparatively low incidence of breast cancer in Africa is related to a low prevalence of risk genes, a preventive lifestyle of rural African women, or a generally lower life expectancy has not been studied.

Established factors for breast cancer risk are age at menarche, age at menopause, age at first full-term pregnancy, breastfeeding, and alcohol consumption at all ages (Fioretti *et al*. 1999; Beral *et al*. 2002; Evans & Howell 2007; WCRF 2007). A high percentage of total body fat, and tall height at adulthood in postmenopausal women are considered to increase the risk (van den Brandt 2000; Key *et al.* 2002; Key *et al.* 2003; Ahlgren *et al.* 2004). The existing evidence mostly originates from studies in high-income countries, often showing that breast cancer rates increase with industrialisation and urbanisation (WCRF 2007; Sitas *et al.* 2008; M. Okobia *et al.* 2006).

Several studies have looked at possible linkages between single nutrient intake as well as foods or dietary patterns and breast cancer (Lee *et al.* 1991; Hunter *et al.* 1996; Wakai *et al.* 2000; Mattisson *et al.* 2007; Cho *et al.* 2003; Prentice *et al.* 2007; Engeset *et al.* 2009). However, the limited evidence suggests only that consumption of total dietary fat and special dietary patterns influence breast cancer risk and no internationally accepted conclusion has been reached to date except for alcoholic drinks (WCRF 2007; Bosetti *et al.* 2009; Brennan *et al.* 2010).

1.1 Rationale

In 2004, a pilot case control study was carried out in the Kilimanjaro Region of Northern Tanzania to identify if findings from high-income countries are relevant to low-income countries and whether indigenous factors affect the aetiology of breast cancer. The risk estimations showed an increased association between alcohol consumption and breast cancer and indicated that plant derived oils seem to decrease the risk (Hebestreit 2004). However, the data set of the pilot study was considered too small and did not have enough statistical power to detect moderate and small risks.

A larger case-control study was designed based on the results and the experience of the pilot-study. The case-control study design was chosen again because of a lack of demographic data and infrastructural deficits which did not allow a prospective study approach. The study aimed to provide additional data and to confirm existing information on the link between dietary intake and the occurrence of breast cancer. The study was eventually performed in a low income country like Tanzania because such communities face an increasing incidence of malignant diseases, including breast cancer. This poses a serious additional burden on the population and the health care system. Another reason to intensify research on cancer prevention in these areas is the fact that most patients affected in developing countries cannot afford any adequate treatment which is available in industrialised countries today.

The objective of this study was finally twofold:

1. The food frequency questionnaire (FFQ) which had already been tested in the pilot study had to be validated. FFQs are the primary method for measuring dietary intake in epidemiological studies. They are very practical in assessing average intake over a long-term period (Willett 1998). Based on a food list the study population is asked to remember over a certain period of time how often they eat each listed food item. However, climate, culture, food availability, education, socioeconomic status, and tradition influence individual food selection. Therefore it was necessary to adapt the food list to local consumer habits. To

ensure the food list covered the particular features of the local diet a validation study was conducted between July 2005 and February 2006.

2. Secondly, a case-control study was conducted with a focus on dietary patterns and breast cancer. This was done based on the hypothesis that nutrients are ingested within diets rather than as single nutrients. Based on results of the above mentioned pilot study it was assumed that the community in Tanzania shows a low risk lifestyle and reproductive behaviour. Thus, it is more likely to detect linkages between dietary intake and breast cancer in this community than in those of high income countries.

1.2 Research partners and study site

The study was carried out in collaboration with the KCMC in Moshi, Tanzania, and the Institute of Nutritional Sciences of the University of Giessen, Germany. The KCMC was founded in 1971 and serves as a referral hospital for the population in Northern Tanzania (Figure 1). The hospital is located in the foothills of Mount Kilimanjaro and has a capacity of over 450 beds in addition to treating hundreds of outpatients. Its department of histopathology was responsible for confirming the diagnosis of the breast cancer patients as well as the implementation of the data collection. The author of this thesis was the principal investigator and responsible for the study design, data management and analysis.

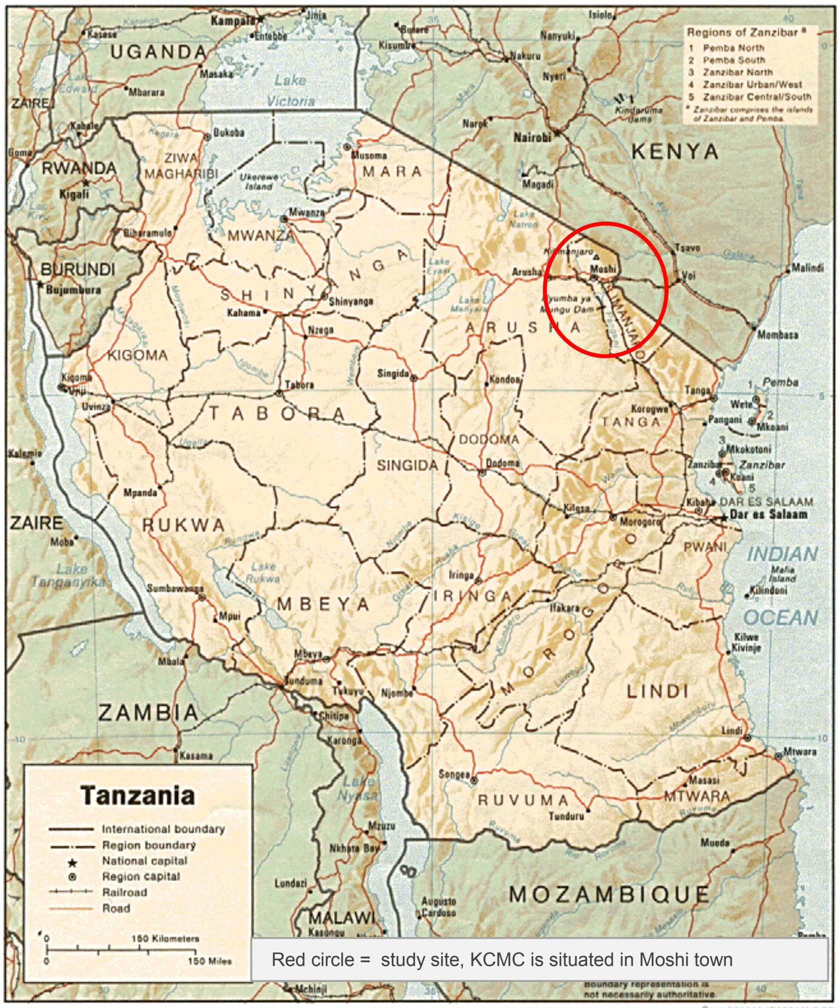


Figure 1: Map of Tanzania

The collaboration with the KCMC induced the selection of the Kilimanjaro Region as study site. The Region is characterized by the famous Mount Kilimanjaro and is divided into six districts covering an area of 13,209 km² (Figure 2). The census in 2002 estimated a population of 1,376,702 for the region with an average annual growth rate of 1.6 % which is lower than the national rate with 2.8 % (United Nations 2011; National Bureau of Statistics 2006; TPSF 2011).

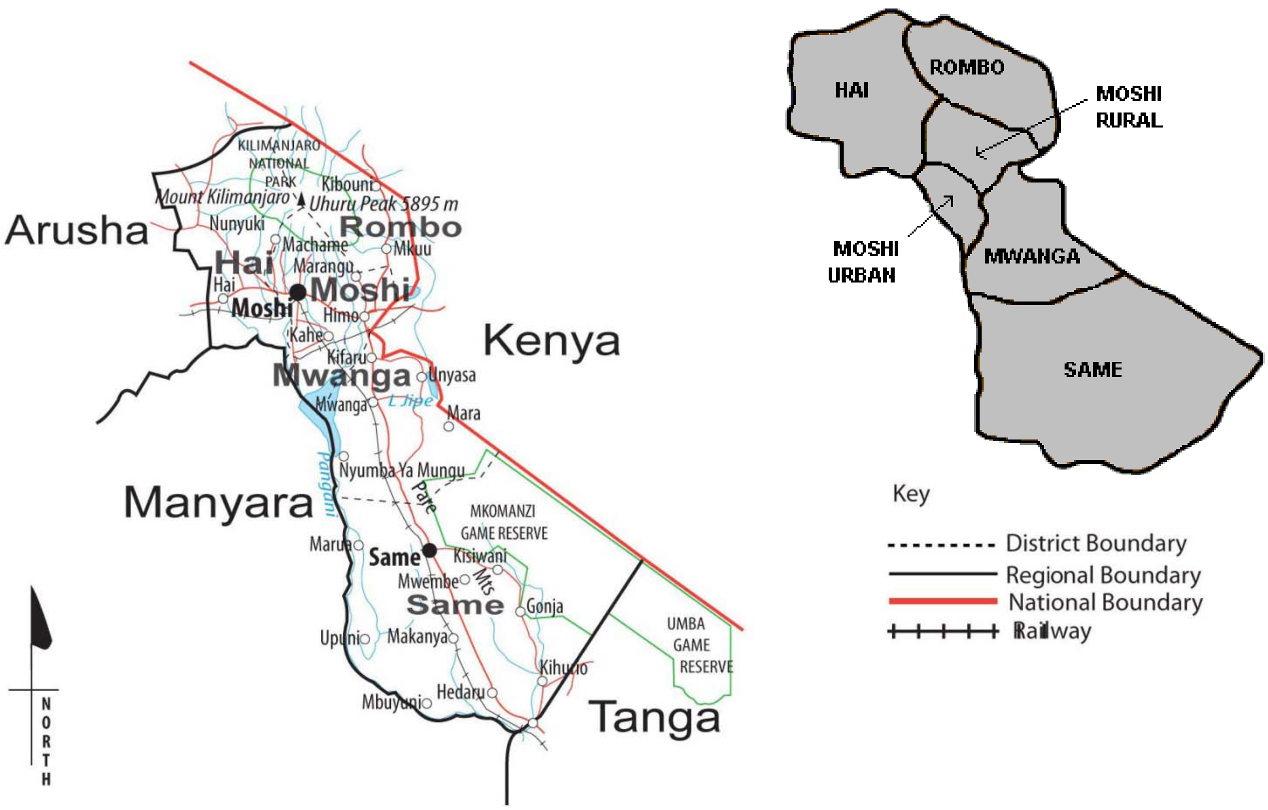


Figure 2: Maps of the districts of the Kilimanjaro Region, Tanzania

The area is inhabited by a number of different ethnic groups, such as the Chagga, the Maasai, and the Pare. Because of the long distances and the low level infrastructure, visits were conducted by the research team to other health centres in the Kilimanjaro and Northern Region in order to include as many patients as possible in the study. These outreach activities included the hospitals of Machame, Kibosho, Huruma, and Arusha. The latter belong to the neighbouring Arusha region but is also serving parts of the Kilimanjaro Region.

1.3 Ethical clearance and outline

Ethical clearance was obtained from the Research and Ethics Committee of the KCMC, Moshi, Tanzania, and the Institutional Review Board of the University of Giessen, Germany.

Details of the validation study design, its results and a discussion of the results are presented in chapter two. The case-control study is described in detail in chapter three. The chapter begins by bringing into focus already acknowledged breast cancer risk factors like lactation and reproductive parameters followed by a discussion of the results of the dietary behaviour assessment in Tanzania. The summary is presented in English in chapter 4 and in German in chapter 5.

2 Validation of the food frequency questionnaire

2.1 Introduction

There is no gold standard to validate a food frequency questionnaire (FFQ) (Margetts *et al.* 2006; Ahrens 2007). In the present study, the validation of the food frequency questionnaire (FFQ) was carried out using 24h-recalls. This method is used to collect dietary data of one day only. The advantage of the 24h-recall is the low recall bias as study participants have to recollect information from the previous day only. However, the one to one day variability of the 24h-recall is known to be high and it is therefore not used as an instrument to measure long-term dietary intake. The variability of a 24h-recall can only be avoided by repetition (Margetts *et al.* 2006). The 24h-recall was therefore conducted twice, covering two different seasons.

2.2 Methods

2.2.1 Study population

Only women living in the Kilimanjaro region were selected for the validation study in order to compare the results with the women participating in the case control study described in chapter 3. Four trained interviewers contacted women visitors of the KCMC in Moshi town and in the two districts of rural and urban Moshi. Selection criteria were adjusted to the characteristics of the case-control study. Thus, only women over 20 years of age were selected, they had to have lived in the region for more than five years of the past 10 years and have enough means to pay the fee for a consultation in a hospital like the KCMC.

2.2.2 Data collection

After informed consent about the study's objective, the women volunteered to sit twice for an interview of about 45 minutes in Swahili or English. The interviews were carried out in two different periods, July 2005 and February 2006. In each interview the enumerator completed a 24h-recall and a food frequency questionnaire (FFQ) with the respondent. During the 24h-recall the women were guided by the enumerator to recall their food intake of the previous day. The data was recorded according to meal sequence.

The food list of the semi quantitative FFQ consisted of 65 food items covering the food groups: beverages, meat, cereals and its products, vegetables, pulses, fruits, fats, oils and others. The interviewer read each item from the list aloud and the women were asked whether they usually consume the food *x*-times per day, week or month if at all. Especially for the assessment of the fruit intake the participants were asked additionally whether they eat the fruit throughout the year or only during the respective harvest season.

For both assessments, 24h-recall and FFQ, quantities were estimated using common household measurements, e.g. cups, spoons, customary packing size, and solid foods in pieces or slices. The cups had a volume capacity of 200 ml, 400 ml and 900 ml. Foods were prepared according to local standard recipes and weighed using household kitchen scales by the research staff. Countable foods such as onions, eggs or bananas were classified according to their size into small, medium and large. Samples of foods were obtained from the local market and mean weights were taken of each size. The matter of size was intensively discussed in interviewer training to assure a common standard. Samples of the different pulses were presented during the interviews since the names differ locally. A raw/ cooked coefficient was applied when large deviations between cooked and raw foods were expected after preparation, e.g. for dried cereals (pasta, rice) and dried legumes. The coefficients were calculated by cooking experiments done by the nutritionist but without calculating any loss of vitamins and minerals. Ten weighed records were used to follow up on recipes and quantities recorded in the FFQ and 24-recalls.

At the second data collection in February 2006, the women were asked in addition who is responsible for food supply and preparation in their household, where the food usually comes from, and whether they had experienced food shortage in the past year.

2.2.3 Statistical analysis

Data from the FFQ and 24h-recall responses were entered in NutriSurvey® 2004 for Windows and analysed with statistical package SAS® 9.1. and SPSS 12.0 for Windows. In the case the women consumed a fruit seasonally, the recorded quantity was multiplied by a mean seasonal factor to calculate the average intake per day independent of the season. The data from both FFQs were converted to gram intake per day for each food item. Finally the data sets from both, FFQs and 24h-recalls, were merged into six food groups to describe individual food intake:

1. cereals: bread, rolls, cereal products, grains, egg-free pasta
2. vegetables: vegetables, pulses, potatoes, mushrooms
3. animal products: eggs, dairy and cheese, meat, fish, poultry, sausages and other meat products
4. beverages: non-alcoholic beverages, coffee, tea, water, alcoholic beverages
5. fruits
6. fats: oil, fats, butter

Since the values of most variables were not normally distributed non parametric tests were carried out in the subsequent analysis. The studied population had a low educational level and considering the relative high number of interviewers in relation to the study population the data were tested for interviewer effects before any statistical analysis was performed.

Interviewer effects were measured using Kruskal-Wallis Test and Median-One-Way Test. The Wilcoxon signed rank test was used to test the 24h-recall and the FFQ for seasonal variability. It is a non parametric test equivalent to the paired t-test. In addition, the Wilcoxon signed rank test was used to test for differences in the results of the 24h-recall and FFQ. In addition spearman correlation coefficients for the means of both FFQs and both 24h-recalls were calculated to examine the relationship between the results of the two tools.

2.3 Results

2.3.1 Characteristics of the study population

Out of the 78 women who were interviewed in the first data assessment (rainy/ harvest season = t1), 25 women could not be contacted again for a second interview (dry/ preharvest season = t2). In addition 3 women did not want to repeat the interview. Most of the women who dropped out lived in remote areas. Finally 50 data sets were completed and made available for statistical analysis. This sample size reflects the statistically required 10-15 % of the expected sample size of the study population in the case-control study.

The mean age of the women who participated in both interview rounds was 40 years, with a range from 23 years to 70 years. Only 2 % of the interviewed women could not write, most women had finished primary school level (96 %); 28 % had also finished secondary school and 2 % had a university degree. More than half of the women (58 %) worked as civil servant and 30 % carried out a small scale business in the agricultural sector whilst 12 % were mainly housewives. Nearly all women (98 %) reported being responsible for food supply and food preparation in their household.

2.3.2 Interviewer effects

Following Good Epidemiological Practice (GEP) (DGEpi 2008; Ahrens 2007, pp 465) the data was tested for homogeneity between the interviewers using the Kruskal Wallis Test. The test showed interviewer effects in 100 % of the food groups confirmed by the Median One Way Test at a level of 83 %. Therefore, further analysis was carried out stratified by interviewer. Since the data set of interviewer four was too small for single analysis, the analysis continued with a total of 46 women; still enough considering the statistically required sample size (see above chapter 2.3.1).

2.3.3 Energy intake

Within the 24h-recall and FFQ there was no response for daily calorific intake below 600 kcal or above 4000 kcal, thus no outlier needed to be rejected. All questionnaires finally used in the analysis have been judged as valid.

The median energy intake measured with the 24h-recall at the first data assessment (t1) varied between the interviewers and was 1474 kcal, 1664 kcal and 2598 kcal respectively. At the second assessment (t2) the median energy intake measured was slightly higher for interviewer 1 and 2 and lower for interviewer 3 compared to t1. The interquartilsdistance (IQ) of the energy intake of the 24h-recalls of t1 was 249 kcal, 771 kcal or 839 kcal respectively. These results differed to t2 where the IQ was 531 kcal, 735 kcal and 516 kcal (.

Table 1). The median energy intake measured with the FFQ t1 was 1721 kcal, 1647 kcal to 1635 kcal and at t2 1790 kcal, 1738 kcal to 1928 kcal respectively. The median values at t2 were slightly lower than t1 (Table 1).

Table 1: Energy intake data (kcal) from 24-recall and food frequency questionnaire (t1; t2) stratified by interviewer

		Energy intake (kcal)			
		24h-recall		FFQ	
		t1	t2	t1	t2
	median	1474	1991	1721	1790
Interviewer 1	**min**	1064	968	980	1214
(n=13)	**max**	2377	2631	2233	3060
	IQ	249	531	468	381
	median	1664	2115	1647	1738
Interviewer 2	**min**	742	1014	1054	1229
(n=23)	**max**	2672	2830	2349	2291
	IQ	771	735	452	518
	median	2598	2480	1635	1928
Interviewer 3	**min**	992	1467	1077	1301
(n=10)	**max**	3969	3752	2089	3010
	IQ	953	516	756	556

2.3.4 Seasonal variability

Tanzania is affected by high seasonal food variability resulting in lower consumption when prices are high and food shortages. Food shortages in the year before the second data collection (February 2006) were experienced by 92 % of the respondents of interviewer 1, by 48 % of the respondents of interviewer 2, and 20 % of the respondents of interviewer 3. March to June were the months with the highest incidence of food shortage. An association between food shortage and food supply from own production was only found in the group of interviewer 1 (r_s = 0.7, P = 0.01), most women obtaining food from local markets (interviewer 1 and 2 = 100 %, interviewer 3 = 90 %) and smaller proportion buying food in supermarkets (interviewer 1 = 23 %, interviewer 2 = 9 %, interviewer 3 = 50 %). Nevertheless, seasonal variability, i.e. differences in the results of the two FFQ or two 24h-recalls due to seasonal food availability, was tested using the Wilcoxon signed rank test. The test indicates seasonal effects if the *P*-value is below 5 %. The test showed only isolated effects in the food groups “vegetables”, “fruits” and “oils and fats” but the *P*-values of the food group “beverages” were below the benchmark of 5 % in all interviewers indicating a seasonal variation (Table 2).

Table 2: Seasonal variability of food frequency questionnaire and 24h-recall stratified by interviewer based on results of Wilcoxon signed rank test*

	Interviewer 1 (n=13)		**Interviewer 2 (n=23)**		**Interviewer 3 (n=10)**	
	FFQ	**24h-recall**	**FFQ**	**24h-recall**	**FFQ**	**24h-recall**
Cereals	0.74	0.24	0.35	0.76	0.08	0.16
Vegetables	0.24	0.35	0.67	0.18	**0.002**	0.63
Animal products	0.29	0.85	0.08	0.08	0.56	0.66
Beverages	**0.001**	**0.01**	**<0.001**	**<0.001**	**0.002**	**0.03**
Fruits	0.20	0.86	**<0.001**	**0.003**	0.57	0.57
Oils and fats	0.14	0.59	0.43	0.08	**0.04**	0.29

* *P* > 0.05 = no seasonal effect

Further analysis within “beverages” revealed a seasonal effect only for water intake. Seasonal effects for beverages other than water and alcoholic drinks were only shown for one interviewer. Alcoholic drinks were not affected by seasonal variability (

Table 3).

Table 3: Seasonal variability of food frequency questionnaire stratified by interviewer based on results of a Wilcoxon signed rank test**

	Interviewer 1 (n=13)	**Interviewer 2 (n=23)**	**Interviewer 3 (n=10)**
Alcoholic drinks	0.63	0.09	0.82
Water	**0.0002**	**<0.0001**	0.05
All other beverages	0.79	**0.001**	0.03

** *P* >0.017 = no seasonal effects (after Bonferroni adjustment)

2.3.5 Differences between food frequency questionnaire and 24h-recall

The Wilcoxon signed rank test was also used to test whether the results of the FFQs differed from the results of the 24h-recalls. The global probability of error (*P*-value) was reduced from 5 % to 0.3 % to address the problem of multiple testing according to Bonferroni (Anon n.d.). The results are presented in table 4. Only one parameter out of the 18 tested, the food group "oils and fats" of interviewer 1, was below the benchmark of 0.3 % signalling differences between the results of the FFQ and the 24h-recall. All other *P*-values were above the benchmark.

Table 4: Differences between food frequency questionnaire and 24-recall stratified by interviewer based on results of a Wilcoxon signed rank test***

	Interviewer 1 (n=13)	Interviewer 2 (n=23)	Interviewer 3 (n=10)
Cereals	0.34	0.83	0.01
Vegetables	0.07	0.20	0.03
Animal products	0.70	0.28	0.57
Beverages	0.64	0.11	0.85
Fruits	0.89	0.02	0.77
Oils and fats	**0.0007**	0.22	0.79

*** *P* >0.003 = no statistical difference of the results of FFQ and 24h-recall

2.3.6 Correlation of food frequency questionnaire and 24h-recall

Table 5 shows the correlation coefficient (r_s) for the means of both FFQs and 24h-recalls. The analysis stratified by interviewer showed a variation within the food groups as well as between the interviewers. Correlation of interviewer 1 (n = 13) varied from -0.34 in the food group "vegetables" to 0.51 in the food group "beverages". The smallest correlation coefficient of interviewer 2 was 0.21 (fruits), the highest 0.53 (beverages). The results of interviewer 3 were lowest in the food group "oils and fats" (0.03) and highest in food group "fruits" (0.71).

In addition the Spearman correlation was calculated with all interviewers grouped together for comparison with other studies which did not report whether they checked for interviewer bias. The correlation coefficient (r_s) was highest in the food group “fruits” (r_s = 0.39, *p* = 0.01) followed by “cereals” (r_s = 0.38, *p* = 0.01), “beverages” (r_s = 0.33, *p* = 0.01) and the food groups “animal products” and “vegetables” (r_s =0.27 and r_s = 0.14 respectively, both values not significant). A negative correlation was found in the food group “oils and fats” (r_s = -0.22, p = 0.13). This was, however, not statistically significant.

Table 5: Spearman correlation (r_s) and its level of significance (*P*) of the means of two food frequency questionnaires and two 24h-recalls

Food groups		Interviewer 1 n=13	Interviewer 2 n=23	Interviewer 3 n=10	*All n=50*
Cereals	r_s	0.42	0.35	0.21	*0.38*
	P	0.16	0.10	0.56	***0.01***
Vegetables	r_s	-0.34	0.23	0.27	*0.14*
	P	0.26	0.30	0.45	*0.33*
Animal products	r_s	0.17	0.34	0.16	*0.27*
	P	0.59	0.11	0.65	*0.06*
Beverages	r_s	0.51	0.53	0.16	*0.33*
	P	0.07	**0.01**	0.65	***0.01***
Fruits	r_s	0.47	0.21	0.71	*0.39*
	P	0.11	0.34	**0.02**	***0.01***
Oils and fats	r_s	0.21	0.41	0.03	*-0.22*
	P	0.50	**0.05**	0.94	*0.13*

2.4 Discussion

Responses to questions regarding recall of subjective or personal information or those which require further probing like a 24h-recall might differ significantly by interviewer (Johannes *et al.* 1997; Brustad *et al.* 2003). This may cause differences in the results between the interviewers. Johannes *et al*. (1997) suggested therefore an exploration of possible interviewer effects should be incorporated into the initial phase of the data analysis. In this study the analysis showed that statistically significant interviewer bias could not be avoided. This might be due to the relatively low educational level of the study population, only 28 % had a secondary education or more. Little is known about the influence of educational level on the estimation of food intake assessed by a 24h-recall or a FFQ. Crispim *et al.* (2006) concluded in their study that groups with an average of 3.6 years of schooling (low educational level) in comparison to groups with an average schooling of 14.9 years (high educational level) have a tendency for poorer quantification of their dietary intake (Crispim *et al.* 2006). Posluna *et al.* (2009) argued in their review that although studies have reported low education as predictor for underreporting also higher education and increased awareness of “healthy food” might prompt the same response; often over reporting is less emphasized. However, the causal pathway for the interviewer bias could not be identified. Therefore, it was recommended to minimize the number of interviewers in the case-control study to reduce the risk of interviewer effects.

2.4.1 Seasonal effects

The inter-quartile range (IQ) of energy intake measured with the 24h-recalls tended to be higher than the IQ of the FFQ confirming the sensitivity of the 24h-recall to day to day variation. Whereas, the food frequency questionnaire (FFQ) showed consistent seasonal vulnerability in the food group “beverages”. This food group consisted of three subgroups: alcoholic drinks, drinking water, and all other beverages, e.g. tea, coffee, soda. After further testing only the water intake showed seasonal variability. The daily average temperature in July is around 25°C whereas it exceeds 30°C from December to March and drops again when the rains start. Water intake very much correlates positively with the outside temperature which may be the reason for the seasonal variability seen in this study (Sawka *et al.* 2005), thus, women tend to drink more water in times of high temperature.

The seasonal effects for “fruits” observed by interviewer 2 only, might be due to an interviewer bias or an estimation bias of the interviewed women themselves. Latter might be caused by memory lapses and/or problems in quantifying the portion size accurately (Gibson 2005, p. 111). Especially if foods are eaten rarely and/or in low quantities estimation of portion size is difficult as low intake levels seems to increase the risk of overestimation (Faggiano et al. 1992; MacIntyre *et al.* 2001). Fruits are often not available for purchase or are very expensive if off-season, thus not affordable for many people. Also fruit consumption is not very popular and fruit intake has reported being generally low in low income countries (Hall *et al.* 2009). However limited numbers of varieties may make recall and recording easier (MacIntyre *et al.* 2001). Seasonal fluctuations in fruit availability were only taken into account by asking the respondents about their access to seasonal fruits and if applicable including a seasonal factor for analysis. Problems in estimating portion sizes might be reduced by using two or three dimensional fruit models or applying an interactive recall method which special emphasis on estimation of portion sizes (Gibson & Ferguson 2008; Gibson 2005, p. 113).

Although there was no statistical difference between FFQ and 24h-recall for vegetable intake the low coefficients of the Wilcoxon test and the spearman correlation showed that the intake of vegetables was difficult to assess. In the Kilimanjaro Region vegetables are mostly cooked, mixed or stewed in gravy so it is not easy to quantify the consumed pieces or amounts of vegetables once the food is on the plate. In addition in poor households the consumption of collected, wild and/or traditional vegetables is relatively high (Weinberger *et al.* 2006). One could assume that the food list of the FFQ did not take into consideration enough of these traditional vegetables. Studies of questionnaires have frequently found that use of close categories biases responses as there is a heavy inclination to only use categories supplied (Singleton 2010). It is normally therefore suggested that categories are established through the use of open questions in a pilot study prior to the main study. This was taken into account by using the 24h-recall method as reference method which did not reveal vegetables other than those already on the food list of the FFQ.

Since water intake is not considered as a relevant parameter for breast cancer risk the FFQ was considered a reliable instrument to assess dietary intake at any time regardless of the season. However, a seasonal factor should be applied not only for fruit but also for vegetable intake assessments.

2.4.2 Differences between food frequency questionnaire and 24h-recall

There were no statistical differences in the results of the FFQ and the 24h-recalls apart from in oil and fat intake for one interviewer. The assessment of fat and oil intake with the 24h-recall was challenging. Often the women did not recall the amount of oil used for cooking or had great difficulties in estimating the amount they had used. The estimation of oil and fat intake within the FFQ was done differently by asking the women to recall the amount of oil they had purchased and how long it had lasted. This amount was then divided by the number of people usually taking part in meals. This method of estimation was cross checked and proved valid by individually weighed records. As an extra precaution however the interviewers in the case-control study underwent extra training to pay special attention while interviewing the study population regarding oil and fat intake.

2.4.3 Correlation of food frequency questionnaire and 24h-recall

Usually the validity of a FFQ is measured by calculating the correlation to its reference either using food groups or the calculated nutrients (Cade *et al.* 2002; Margetts *et al.* 2006). Hence, Spearman correlation coefficients were calculated first stratified for interviewers and then the complete set of the results to allow comparison with other studies where it is unknown whether they were tested for interviewer effects. The coefficients were low to modest. The results stratified by interviewer showed large variations within the food groups as well as between the interviewers. In the summarized analysis three of the six food groups showed a significant correlation from 0.33 to 0.39 (cereals, beverages and fruits). The other three food groups: vegetables, animal products, and oils and fats had a correlation of 0.14, 0.27 and -0.22 respectively, but not significant. The negative correlation for oils and fats might be explained by the difficulties in assessing the oil and fat consumption using the 24h-recall. Several studies carried out in rural and/or remote settings, reported comparable spearman correlation coefficients for food groups or nutrients from FFQs when the result for oils and fats are excluded. Their results ranged from 0.14-0.56 in a study in South Africa (MacIntyre *et al.* 2001) and 0.09-0.58 in a study in Mali (Torheim *et al.* 2001). Pearson correlation coefficients between 0.19-0.78 were reported in a study from Bangladesh (Yu Chen *et al.* 2004). The reference methods ranged from 7-day-weighed record and 2-day-weighed record to two 7-day food dairies. The high variation in correlation coefficients for the different food groups might be caused by under- or overestimation due to either high fluctuations in food availability or difficulties on the part of the respondents in estimating the quantities of the foods consumed. However, Parr *et al.* (2002) pointed out that these factors should not be directly linked to the questionnaire design, indicating the validity of the food list.

2.4.4 Limitations

Ambrosini *et al.* pointed out that correlation coefficients describe the relation between two results only but not an agreement (Ambrosini et al. 2001; 2003). They showed that correlation coefficients alone do not reflect the reliability and validity of a FFQ. Many studies in westernized countries used biomarkers in addition to 24h-recalls or weighed records to validate their FFQ (McNaughton *et al.* 2007; Mina *et al.* 2007; Pellegrini *et al.* 2007; Mahabir *et al.* 2006; Kabagambe *et al.* 2001; Kroke *et al.* 1999; Bohlscheid-Thomas 1999). Due to ethical, infrastructural and financial constraints biomarkers could not be used in this study.

Many FFQs have already been validated, most of them in westernized countries. The present validation study was carried out in a region where the use of the internet or self administered interviews as well as of biomarkers was not feasible. Gibney *et al*. (2004) recommended 24h-recalls for regions with populations for whom recording food intake is not practical as an appropriate method. In the present study 24h-recalls were administered by several interviewers resulting in an interviewer bias. However it is not known whether this was due to the interviewers themselves or due to the low educational level affecting the quantification capacities of the study participants. Another reason might be an unequal geographical distribution of the participants due to the selection process – interviewer selected the study participants on an individual level. This might also result in differences between the interviewers in regard to food intake of the participants.

The high drop out in the second season (about 36%) may have been due to poor travel facilities, low financial means of the respondents and a poor communication infrastructure in addition to a low level of willingness of the selected women to participate in a study with no direct incentive. This and the individual selection of the participants by the interviewer might have lead to a selection bias. However, individual selection by interviewer was done to avoid confounding errors due to different livelihood systems between the sample population of the validation study and the breast cancer study. In the absence of health insurance breast cancer patients in Tanzania need some financial means to get access to health facilities for diagnosis but only properly diagnosed women could participate in the breast cancer study.

2.4.5 Conclusion

There is no evidence for a seasonal effect in breast cancer relevant food groups if the FFQ is used. Differences in the intake of oils and fats assessed by the validated FFQ and its reference, the 24h recall, could only be shown by one interviewer. This might be due to low quantification capacities of the studied population especially in this respective food group and especially during the 24h-recall. However, special attention should be paid to the training of interviewers and especially to the assessment of the oil and fat intake. In addition, a calculation for seasonal variability in fruit and vegetable intake should be used where applicable.

There had been no consistent statistical differences between the FFQs and the 24h-recalls, thus the tested FFQ was considered a reliable instrument to assess dietary intake in the Kilimanjaro Region.

3 Breast cancer and nutrition

3.1 Introduction

Breast cancer and its progression are caused by many parameters (WCRF 2007). However, there is little consensus about the linkages of nutrition to breast cancer. Some research has shown that the diet during pregnancy and the levels of maternal pregnancy hormones is already effecting the risk of later breast cancer in the offspring (Lagiou *et al.* 2006; Ekbom *et al.* 1992; Ahlgren *et al.* 2003). However, breast cancer related to birth weight or its consequences might be primarily diagnosed in premenopausal women whereas adult weight gain for example is probably a cause at post menopause (WCRF 2007). Thus, if we look at possible causes for breast cancer in pre- and postmenopausal women we have to look at the lifelong lifestyle and nutrition behaviour of the patients.

3.2 Methods

3.2.1 Study population

Breast cancer patients and controls were recruited between year 2004 and 2007. The cases were identified using admission records in the surgery department and the list of biopsies of the department of pathology. Diagnosis of primary breast cancer was confirmed through fine needle aspiration cytology (FNAC) carried out by a medical doctor holding a degree in histopathology. Fine needle aspiration cytology is an acceptable and reliable procedure for the preoperative diagnosis of breast lesions, particularly in developing countries, and when used as part of the “triple test” (combined cytological, clinical and radiologic findings) (Chaiwun *et al.* 2007).

Based on the cases’ age and place of residence, controls were recruited in the ophthalmic clinic, the orthopaedic unit and among the visitors of the KCMC. The hospital- and visitor-based controls were age-matched (± 1.5 years) and had lived in the same district for at least five of the past ten years. The controls underwent a history and physical examination to exclude breast cancer. All eligible women were keen to be enrolled after being informed of the study’s objectives.

3.2.2 Data collection

Interviews

After obtaining their informed consent, cases and controls were interviewed by a trained nurse or a doctor in Swahili using a standardized questionnaire in English. Both interviewers were trained on how to conduct the interviews and conducted a pre-test of the questionnaire prior to the final data assessment and supervised by the principal investigator. After final revision the questionnaire included questions about the women's socioeconomic situation, reproductive and breastfeeding history, physical activity, hunger periods and dietary behaviour. Each interview took on average 45 minutes and was amended by anthropometric measurements and body mass index history assessment.

Socioeconomic situation

The socioeconomic situation was described by formal education level, main professional occupation and property level. The latter was estimated based on five objects which were given an abstract value in relation to its economic value: radio= 1, bicycle = 3, television = 10, motorbike = 13, and car = 25. The commonly used housing parameters were left out because of different perceptions towards housing by the different ethnic groups in the research area. The values of the objects available in the households were totalled to a "property index" which was grouped into three levels: low (0-3), medium (4-14), and high (>14). Consequently, women at low property level had access to a radio or a bicycle only, whereas women at medium property level might possess a motorbike and a radio or a television, radio and bicycle. The combinations motorbike and a bicycle, a television and/ or a car were classified as high property level.

Reproductive and breastfeeding history

The reproductive stage of the women was defined as premenopausal with regular or irregular menses and as postmenopausal with no menstrual period for the past year (Nelson 2008). Lactating women without a menstrual period were classified as premenopausal; women after ovarial surgery and hysterectomy were classified according to age.

The age at menarche was recorded as recalled by the women themselves. If they had children the age at birth of each child was recorded as well as the respective breastfeeding periods. Lifelong lactation was defined as the sum of all breastfeeding periods of the individual woman. No specific information was collected about exclusive breastfeeding periods.

Physical activity

The physical activity was assessed using a 24 hour history of a common working day. The activities were recorded in steps of 15 minutes and grouped into light, medium, and heavy work. So, for instance activities like sitting, resting or relaxing were considered as light work, housework, laundry, cooking or walking as medium work, and weeding, working in the field, collecting firewood or fodder, fetching water, walking long distances with load or milking a cow as heavy work. The average physical activity ratio of each group (light = 1.5, medium = 1.7, heavy = 2.0) was time weighted to obtain the physical activity level (PAL) (DGE 2002, p. 23).

Hunger periods

Hunger was assessed as an experienced threat affecting the body size. In case hunger was experienced the respective year and duration was recorded and cross-checked with secondary data on food availability and/or political instability to distinguish between a period of individual or overall experienced food insecurity.

Food frequency

The FFQ food list was prepared as described: chapter 2, page 11. The FFQ data were entered into the computer using NutriSurvey®, a nutrition assessment software package, which generated tables of the individual food and nutrient intake per day, latter based on food composition tables from Tanzania, Kenya, Senegal, Mali and Germany (Landig *et al.* 1998; Erhardt n.d.). The FFQ of the case-control study contained in total 65 food items. Quantities were assessed using common households measurements as described in chapter 2, page 12. Following the results of the validation study (chapter 2) seasonal food availability on individual level was assessed within the interview for fruits; a seasonal factor was applied accordingly.

Anthropometry

The women’s body weight was measured in minimum clothing with a calibrated digital scale (seca 862, Seca Ltd., Germany) and height was determined standing in an upright position without shoes using the scientific meter “person check” (Kirchner

& Wilhelm Ltd., Germany). Measurements followed established standard protocols (Marfell-Jones *et al.* 2006; WHO 1995).

In order to obtain information about BMI history, the women were shown a pictogram which had been modified to African settings (Figure 3) (Stunkard *et al.* 1983; Jordan *et al.* 2010). The drawings show women of different size. The participants were asked to indicate the drawing which they thought resembled them at the time of the interview, followed by the question as to how they might have looked at the age of 20 years or at marriage, and finally how their perceived size was before their menarche or while they were of primary school age. This assessment allows plain BMI-numbers without decimals only.

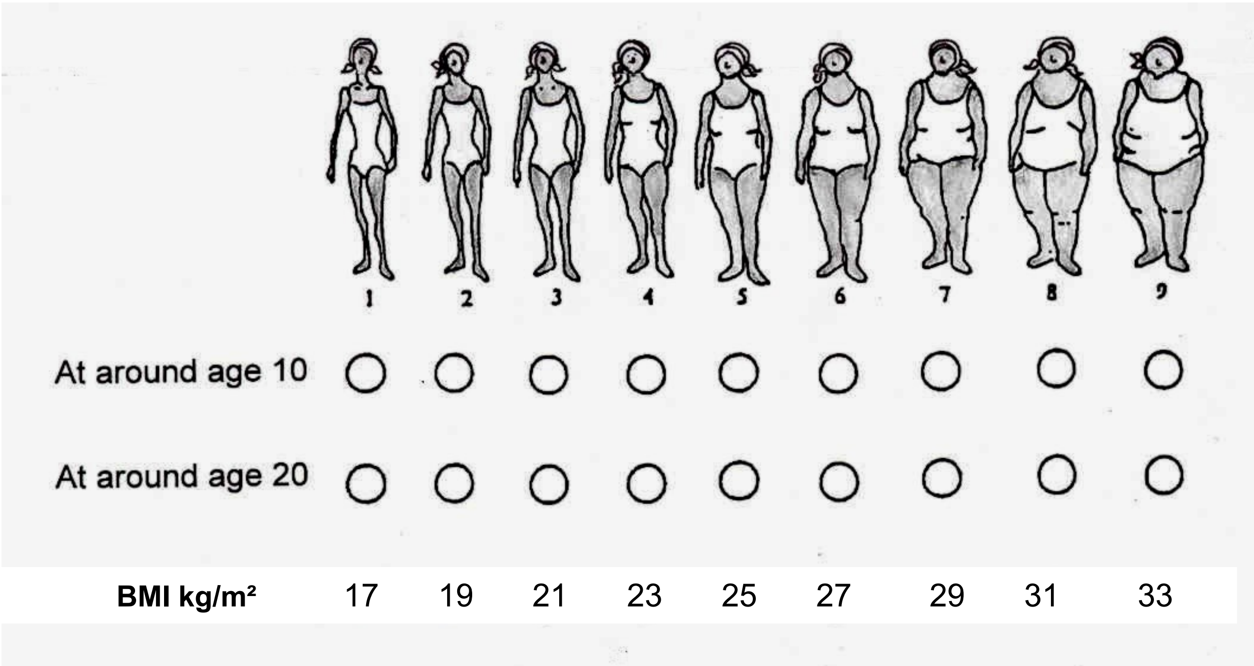


Figure 3: Pictogram to estimate the Body Mass Index (BMI) (modified after Stunkard *et al.* 1983)

3.2.3 Statistical Analysis

General questionnaire and anthropometry

The data on socioeconomic status, lifestyle and anthropometry was entered into an excel spreadsheet by two independent individuals and compared for inconsistencies after being imported as text files into SAS and SPSS editors. In order to detect missing values or errors ranges were set for all variables used in the logistic regression analysis described below. In case of inconsistency, missing values or if values laid outside the set ranges, the original questionnaire was looked up for verification and where necessary the data was corrected.

Descriptive statistics and logistic regression were performed using the statistical packages of SAS 9.1 and SPSS versions 17 to 19 (SPSS Inc.). At first the variables were tested for normal distribution, followed by their respective tests for statistically significant differences between cases and controls. The control group was analysed for differences between hospital- and visitor-based controls in their socioeconomic status using the Mann-Whitney *U* Test to check for selection bias.

The logistic regression model included the matching variables age, and place of living, and the acknowledged predictors in the aetiology of breast cancer from high-income countries i.e. socioeconomic status, age at menarche, age at first full-term pregnancy, BMI at the age of 20, current BMI, and lifelong lactation. If possible variables were entered as continuous variables. The variable "age at first full-term pregnancy" was categorised into three groups: first full-term pregnancy ≤ 20 years, > 20 years, and no pregnancy.

Cox & Snell's and Nagelkerke's R^2 were used for goodness of fit estimations of the logistic regression model. Both "pseudo"-R^2 statistics are based on the log-likelihood of the baseline and the new model showing how much the model improves after inclusion of the predictors variables. It can vary between 0, indicating that the predictors are useless at predicting the outcome variable and 1 indicating that the model predicts the outcome variable perfectly. However, Cox & Snell's R^2 never reaches its theoretical maximum of 1 which is different for Nagelkerke's R^2. Values above 0.2 are considered as acceptable, values above 0.4 or 0.5 as good or very good (Backhaus 2008, p.270; Field 2009, p.269).

Dietary intake

At first, the data on individual food intake of the FFQ used in the case-control study was calculated as food intake per day and controlled for outliers based on energy consumptions: < 600 kcal/ day or > 4000 kcal/ day. Mann-Whitney *U* Test was applied to test for group differences between cases and controls in regard of energy consumption and nutrient intake as well as alcohol intake.

Secondly, a principal component analysis (PCA) was performed with the purpose to identify groups or clusters, also called components, of food items that are collinear, thus, having a strong linear relationship. The components were defined as dietary patterns summarizing the available information on food item intake level. In preparation for the PCA the 65 food items and beverages listed in the FFQ were merged into 36 food groups at first based on their similarities, e.g. food items which are subsidising each other in the meals: e.g. gram intake per day of cassava and taro are summed up in one food group. Secondly, alcoholic beverages were excluded from the food group list and another PCA was performed using 34 good groups. The sampling adequacy of the food group variables for PCA was proofed using the Kaiser-Meyer-Olkin measure (KMO). The KMO statistics varies between 1 and 0. A value of 0 indicates diffusion in the pattern of correlations, hence, PCA is inappropriate. A KMO <0.5 is considered as unacceptable, values between 0.5 and 0.7 are mediocre, values between 0.7 and 0.8 are good, and values >0.8 are meritorious to marvellous (Cureton 1983; Hu *et al.* 1999; Backhaus 2008, p336; Field 2009, p647). Not all components are statistically important. Next to Kaiser's criterion, retaining all components with an eigenvalue greater than 1, scree plots and parallel analysis were used to quantify the number of components wanted (Backhaus 2008, p355; Costello & Osborne 2005; Horn 1965). The interpretations of the components were based on the factor loadings of the food groups after applying Varimax rotation which attempts to maximize the dispersion of loadings within the components. Positive loadings show that the respective food group is positively associated with the component and vice versa. Based on the factor loadings the components were interpreted as dietary patterns. Food groups with factor loadings between -0.4 and 0.4 were considered too low and though neglected in the interpretation (Field 2009, p644). Finally, factor scores were estimated based on a multiple regression. These factor scores describe the bond of each woman to each component. They are in the mean 0 with a standard deviation of 1. Is the factor score negative then is the individuum less affiliated with this component. A positive factor score means a strong relationship between the component and the individuum (Backhaus 2008, p374; Field 2009, p634).

However, a PCA is linked with a loss of information since the groups or cluster which are obtained usually do not explain the full variability of the original information (Backhaus 2008, p327). In addition there is a certain disagreement among statistical theorists about it (McCann *et al.* 2007; Costello & Osborne 2005; Hoffmann *et al.* 2004). Nevertheless, PCA was chosen for keeping the results comparable to other studies looking at dietary patterns and disease (Brennan *et al.* 2010; Handa & Kreiger 2007; Kim *et al.* 2005; Kim *et al.* 2004; Fung *et al.* 2001; Hu *et al.* 2000).

3.3 Results

3.3.1 Sample size and ethnic distribution of study population

The number of interviews per year increased with each year. Finally, interviews had been taken from 115 breast cancer patients and 230 controls within a period of 2.8 years.

Not all data sets were complete and thus, could not be included in the logistic regression models. For instance BMI was not assessed in case of pregnancy and therefore the respective women were left out in the risk estimation since BMI was considered as important predictor for breast cancer. Finally 333 women were included in the basic model.

About one third (33 %) of the women lived in the rural and 13 % in the urban area of Moshi. In Rombo lived 10 % and in Hai lived 14 % of the women. Both districts are in the North of Moshi. Nobody came from Mwanga and Same, both districts are in the South of Moshi whereas 17 % of the women came from rural and 11 % from the urban areas of Arusha. Only 2 % of the women came from outside the original study area. Pearson Chi Square Test showed that there was no difference in the distribution of place of living between cases and controls confirming that the matching criteria were met (*P* = 1).

About two thirds of the study population belonged to the Chagga (64.1 %), followed by Pare (11.6 %), Arusha (4.9 %), Massai (4.3 %), Meru (3.5 %), and others tribes (11.6 %). The distribution of the tribes was significantly different between cases and controls (*P* = 0.015) (Table 6).

Table 6: Ethnic distribution in cases and controls

Ethnic group	Cases %	Controls %	*P* value**
Chagga	73.9	59.1	0.015
Pare	5.2	14.8	
Maasai	2.6	5.2	
Meru	1.7	4.3	
Arusha	2.6	6.1	
Other	13.9	10.4	

* Pearson Chi-Square: differences between cases and controls

3.3.2 Socioeconomic characteristics

Socioeconomic characteristics of the study population are shown in Table 7. Overall, the educational level of the cases was lower than that of the controls: fewer controls (18 % versus 27 %) were illiterate or attended school for less than three years.

Table 7: Socioeconomic indicators

Variable	Cases n=115	Controls n=230	*P* value
Schooling (%)			0.119*
Less than 3 years	27	18	
Finished primary school	54	59	
Finished secondary school	17	23	
University degree	2	0	
Occupation (%)			0.206**
Farmer	47	40	
Dealer, trader, or salesperson	19	28	
Civil servant	22	23	
Other	12	9	
Property level (%)			<0.000*
Low	47	18	
Medium	45	65	
High	8	17	
Women with children (%)	92	94	0.515*

* Mann-Whitney *U*-Test: differences between cases and controls,
** Pearson Chi-Square: differences between cases and controls

More controls than cases (23 % versus 17 %) had secondary school education, but this was not statistically significant (P = 0.119). The main occupation of cases and controls was farming, followed by trading and working as civil servant.

The property level of the cases was significantly lower than that of the controls (P <0.001): almost 50 % of the cases and 18 % of the controls were found to have a low property level. No significant differences in property levels existed between hospital- and visitor-based controls (P = 0.54, Mann-Whitney *U* Test).

The level of schooling was significantly correlated with occupation, r_s = 0.59, and property level, r_s = 0.44 (all *P*s <0.001). The number of children per women corresponded negatively with level of schooling or occupation (r_s = -0.35 and r_s = -0.30, all *P*s <0.001). Mothers with fewer than three years of schooling had at median six, with primary and/ or secondary education four, and with a university degree only one to two children.

3.3.3 BMI and reproductive parameters

The results of the anthropometric assessment and the reproductive parameters are shown in Table 8. Although median BMI at the age of 20 years was 21 kg/m^2 in cases and controls, the distribution was different between the two groups (*P* = 0.003). At time of the interview the cases had a BMI ranging from 15.2 to 39.5 kg/m^2, while the controls had a BMI of 19.1 to 33.3 kg/m^2 (*P* = 0.008). The measured BMI at interview was significantly correlated with the estimated BMI at interview, r_s = 0.75 (*P* <0.001). Spearman correlation showed a significant association between BMI at interview and estimated BMI at 20 years, r_s = 0.40 (P <0.001). The association between estimated BMI at interview and at 10 years was also significant but low, r_s = 0.17 (*P* <0.01).

The mean and median age at menarche was 16 years in both cases and controls. More cases were nulliparous than controls (8 % vs. 6 %). Women with children in both groups had on average five children. The mother's median age at first pregnancy was equal to the median age at first full-term pregnancy. The youngest age at first pregnancy was 14 years in the cases and 13 years in the controls. The oldest age at first pregnancy was 35 years in the cases and 41 years in the controls.

Almost all mothers breastfed their children (cases 99.0 %, controls 99.5 %). The median total breastfeeding period per child was 23 months in both groups (min-max: cases 0.02 - 39 months; controls 12 - 34 months). Median lifelong lactation amongst the mothers was 96 months (cases) and 108 months (controls) constituting a significant difference (*P* <0.05). Lifelong lactation decreased with an increasing level of education due to the lower number of children (r_s = -0.3, *P* <0.001).

Table 8: Reproductive parameters and Body Mass Index (BMI)

	Cases			Controls			*P*
	Median	min/max	n	Median	min/max	n	value*
Age (years)	50	28/85	115	50	26/83	230	0.620
Age at menarche (years)	16	11/20	111	16	13/20	230	0.267
Number of children ⁂	5	1/10	106	5	1/9	217	0.219
Age at first full term pregnancy (years)	20	14/35	106	20	13/41	217	0.571
Breast feeding per child (months)	22	0/39	115	23	0/34	230	0.384
Lifelong lactation (months)	90	0/240	114	108	0/240	230	0.045
Body Mass Index (kg/m²)							
At 20 years	21	17/31	113	21	19/25	229	0.003
At interview	24.7	15.2/39.5	114	26.0	19.1/33.3	227	0.008

⁂ only parous women, * Mann-Whitney *U*-Test: differences between cases and controls

3.3.4 Nutrient intake and dietary patterns

Nutrient intake

Median energy consumption in all women was 1,714 kcal per day, the minimum 786 kcal and the maximum 3,928 kcal. No respondent needed to be rejected from further analysis since no FFQ was below 600 kcal energy and none above 4000 kcal intake per day.

Median intake of protein was 47 g/d (min 17 g/d; max 183 g/d), of fat 72 g/d (min 30 g/d; max 166 g/d) and of carbohydrates 188 g/d (min 85 g/d – max 537 g/d). Median percentage of food energy from protein was 12 %, from fat 39 % and from carbohydrates 46 %. Median alcohol intake from alcoholic drinks was 8.2 g/d (min 0 g/d; max 100 g/d). Main alcoholic drinks were *Mbege:* an often homemade, locally brewed beer (Hebestreit 2004), bottled beer and wine (median intake 57 g/d, min 0 g/d; max 298 g/d and 0 g/d, min 0 g/d; max 77 g/d respectively).

The energy and nutrient intake per day of cases and controls are presented in Table 9. Cases reported a significant higher energy consumption than controls (median 1914 kcal/d vs. 1687 kcal/d, *P* = 0.017). Protein and carbohydrate intake was also higher among cases than controls (50g/d vs 46g/d, *P* = 0.014 and 194g/d vs. 185g/d, *P* = 0.038 respectively). There was no significant difference between cases and control in total fat intake and saturated fatty acids but in polyunsaturated fatty acids (PUFA) (*P* = 0.003). Median and minimum values of percentage energy intake from protein were the same in cases and controls but the maximum values varied (*P* = 0.043).

Table 9: Energy and nutrient intake per day based on a food frequency questionnaire

	Cases n=115		Controls n=230		***P* value***
	Median	**min/max**	**Median**	**min/max**	
Energy (kcal/d)	1914	786/3928	1687	888/3514	**0.017**
Protein (g/d)	50	17/183	46	19/105	**0.014**
Fat (g/d)	72	34/166	72	30/152	0.472
Carbohydrates (g/d)	194	134/506	185	85/537	**0.038**
% Energy from protein (%)	12	6/32	12	6/19	**0.043**
% Energy from fat (%)	37	14/59	39	23/67	0.053
% Energy from carbohydrates (%)	46	26/67	46	27/65	0.646
Polyunsaturated fatty acids (g/d)	19	5/56	25	4/78	**0.003**
Saturated fatty acids (g/d)	18	7/53	19	7/41	0.986
Alcohol intake (g/d)	8	0/100	7	0/79	0.586

* Mann-Whitney *U* Test: differences between cases and controls

Dietary patterns

Primarily a principal component analysis (PCA) was conducted on 36 food groups with Varimax rotation. The Kaiser-Meyer-Olkin Measure verified the sampling adequacy for the PCA, KMO = 0.621, which is considered as mediocre (Backhaus 2008, p. 336). Following Kaiser's criterion retaining all components with eigenvalues greater than one, 14 components would have been useful for further analysis. However, the number of food groups with factor loadings < -0.4 or >0.4 varied between 0 to 11, thus, the results were not interpretable. Consequently, it was decided that the number of components should be retained based on a scree plot. For the scree plot the components were plotted in a graph against its eigenvalues in descending order. The cut-off point for selecting the components was the point of inflexion of the curve which is where the slope of the line changes dramatically (Figure 4-) (Backhaus 2008, p. 353; Field 2009, p. 639) which was in this analysis at component four.

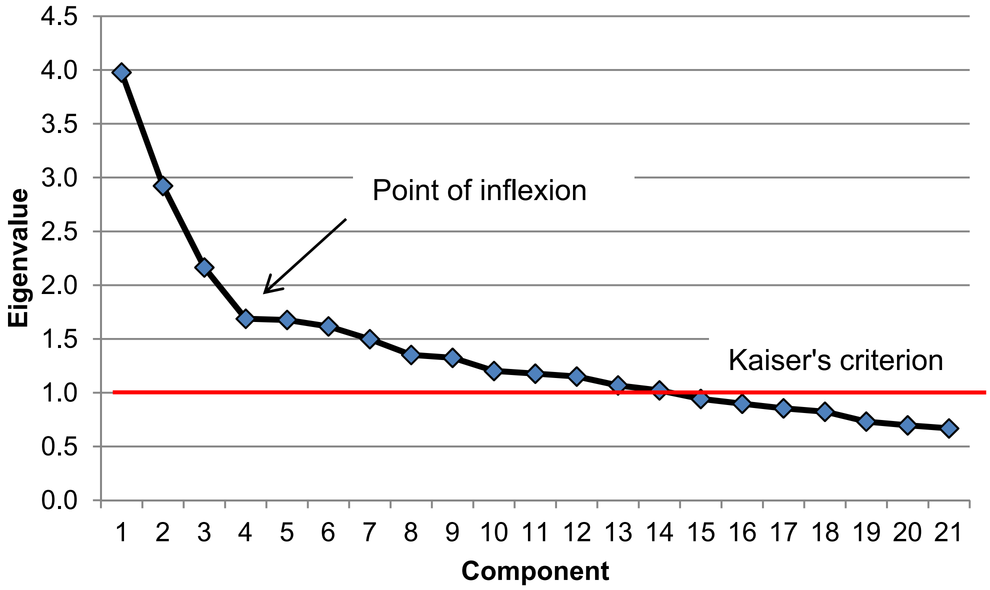


Figure 4: Scree plot of principal component analysis

The results of the PCA are presented in Table 10 showing the factor loadings for each food group (values within the range -0.4 and 0.4 were shaded since they were considered as too low). The four components described 29.9 % of the variance. The first component is characterized by rice, nuts, eggs, chapati (unleavened East African flat wheat bread), leguminous vegetables, bread, soda and red meat. This diet correlated significantly with property ($r_s = 0.37$, $P < 0.001$). Since most of these food items are usually purchased it was called the “Diet of the Rich”. Component two is characterized by *Mchicha*, cucumber, okra, onions, carrots, tomatoes, maize, fish and avocado. *Mchicha* is the Swahili name for amaranth leaf, a traditional food in Tanzania and often also used to name a dish consisting of amaranth leaves and e.g. onions, tomatoes and/or carrots in various amounts. This pattern was therefore named “Mchicha Diet”. Spearman correlation did not show a significant association to property. The third component is characterized by ripe and green banana, sugar, different fruits, tubers, pulses and mbege. The mountainous area of the Kilimanjaro Region is known for its various banana plants, thus, pattern three was called

“Banana Diet”. Property was negatively associated with this pattern at a low level, r_s = -0.16 (*P* = 0.004). Component four is characterized by a high consumption of milk, butter, lard, vegetable oils and fats and a low consumption of sunflower oil and tea. All of the positively loading food items relate to fat, thus we called this pattern “Fatty Diet”. This pattern was also negatively associated with property at a low level (r_s = -0.18, *P* = 0.001). Assuming that the women devoting this dietary pattern do not only eat fat it was analysed which other foods might characterize this pattern at an underlying level. Thus, the women were split into four groups of equal size according to their affiliation to the Fatty Diet. The explorative analysis showed that with increased affiliation to this pattern bread consumption decreased (1st Quartile median= 17 g bread/d, 4th Quartile median = 9 g bread/d; *P* for trend <0.001) and red meat consumption increased (1st Quartile: median 44 g/d, 4th Quartile: median = 52 g/d; *P* for trend = 0.09).

Table 10: Summary of Principal Component Analysis with Varimax rotation retaining four dietary patterns based on 36 food groups (PCA 1)

Food item	Rotated Factor Loadings			
	Diet of the Rich	*Mchicha* Diet	Banana Diet	Fatty Diet
Rice	**0.618**	0.205	-0.143	-0.170
Nuts	**0.587**	-0.006	0.124	-0.089
Egg	**0.557**	-0.039	0.162	0.043
*Chapati**	**0.556**	0.062	0.055	0.009
Leguminous vegetables	**0.537**	-0.093	0.006	-0.026
Bread	**0.503**	0.362	-0.220	-0.190
Soda drinks	**0.471**	0.108	-0.028	-0.155
Red meat	**0.453**	0.103	-0.037	0.367
Mchicha†	-0.017	**0.645**	0.029	0.110
Cucumber & okra	0.209	**0.581**	0.032	0.038
Onion	0.089	**0.579**	-0.042	0.138
Carrots & tomatoes	0.145	**0.516**	-0.096	-0.007
Maize	-0.180	**0.461**	0.135	-0.085
Fish	-0.018	**0.434**	0.337	-0.085
Avocado	-0.016	**0.413**	0.347	0.067
Banana	0.145	0.030	**0.667**	0.073
Green (cooking) banana	0.086	0.008	**0.616**	-0.176
Sugar	0.153	-0.103	**0.491**	-0.166
Watery fruits♯	0.085	0.189	**0.478**	-0.218
Starchy tubers	-0.275	-0.063	**0.461**	0.136
Mbege‡	-0.295	0.050	**0.442**	0.246
Pulses	-0.070	0.281	**0.415**	0.134
Sunflower oil	0.203	-0.207	-0.071	**-0.623**
Milk	0.264	-0.079	-0.042	**0.521**
Butter and lard	-0.213	-0.254	0.055	**0.457**
Mixed vegetable fats & oil	0.263	0.191	-0.115	**0.454**
Tea	0.055	0.013	0.366	**-0.410**
Variance explained (%)	9.1	7.9	7.6	5.3

Food groups with factor loadings < 0.4and >-0.4, thus, not presented: potatoes, juice, chicken meat, mango & papaya, cabbage (white), *mandazi* §, *uji* ∫, coffee, bottled beer and wine. Rotation method Varimax with Kaiser Normalization. Rotation converged in 7 iterations.
* unleavened East African flat wheat bread; † traditional Tanzanian food, synonymously used for a dish of amaranth leaves and e.g. onions, tomatoes or/ and carrots in various amounts; ‡ often homemade opaque beer from bananas and millet; § East African donuts; ∫ thin millet or maize based porridge; ♯ oranges, watermelon and pineapple

The Banana Diet includes *Mbege* - a local, often homemade opaque beer from bananas and millet (Hebestreit 2004) (see annex 8.1). Acknowledging that alcohol is an accepted risk factor for breast cancer and focusing the analysis on non-alcohol dietary elements, the PCA was repeated excluding the alcoholic beverages from the food group list. The KMO was 0.619, confirming the sampling adequacy. As already in the first case, the scree plot of PCA2 suggested to retain only four components. However more comparable results to the first PCA were obtained if six components were kept. Therefore, a parallel analysis was performed to check for the maximum number of viable components (Figure 5). Some statisticians consider the parallel analysis as the best method to estimate the number of components which might be extracted in a PCA but this method is not part of the statistical package SPSS and is therefore not commonly used and was therefore not considered in the first PCA (Field, 2009 p641; Costello & Osborne 2005; Horn 1965). For the parallel analysis eigenvalues of the components of the PCA were compared with randomly generated eigenvalues that have the same characteristics as the data being analysed. Both eigenvalues were plotted against its component. PCA generated components which eigenvalues are greater than those from the randomly generated "counterparts" might be retained.

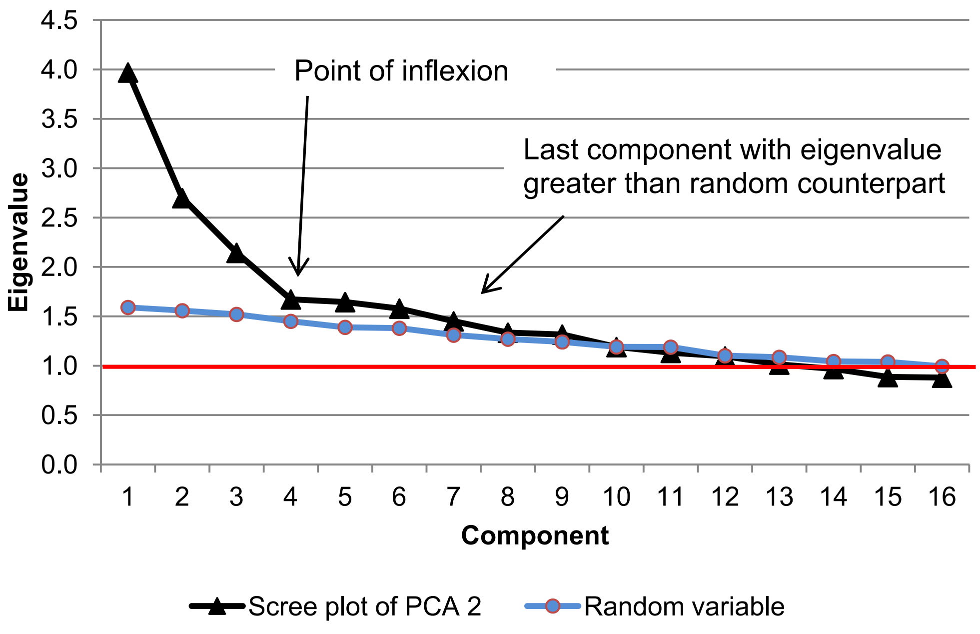


Figure 5: Scree plot of PCA and randomly generated data set

The results of the parallel analysis suggested to retain seven components but in this case three components had less than three food groups with a factor loading < -0.4 or >0.4 making the interpretation difficult. After all it was decided to keep six components which described 40.4% of the dietary variance (Table 11) and which showed comparable results to the first PCA.

Table 11: Summary of Principal Component Analysis with Varimax rotation retaining six dietary patterns based on 34 food groups (PCA2)

	Rotated Factor Loadings					
	Diet of the Rich (no alc)[1]	**Fruity Diet** (no alc)[1]	***Mchicha* Diet** (no alc)[1]	**Banana Diet** (no alc)[1]	**Starchy Diet** (no alc)[1]	**Fatty Diet** (no alc)[1]
Nuts	0.604	0.010	-0.007	0.090	0.120	-0.032
*Chapati**	0.593	0.039	0.138	0.017	-0.042	0.058
Soda drinks	0.574	0.206	-0.002	-0.185	-0.010	-0.098
Egg	0.571	-0.006	0.040	0.139	0.005	0.062
Pulses	0.491	-0.069	-0.162	0.009	0.274	0.054
Fish	0.086	0.612	0.178	0.068	-0.069	-0.103
Mango & Papaya	0.146	0.609	0.033	0.121	0.158	0.068
Avocado	-0.043	0.521	0.081	0.138	0.255	0.046
Watery fruits ♯	0.180	0.457	0.034	0.353	-0.140	-0.175
Onion	0.035	0.049	0.783	0.034	-0.004	0.046
Carrots & tomatoes	0.102	-0.005	0.703	-0.002	0.039	-0.077
Mchicha†	-0.108	0.184	0.630	0.033	0.247	0.030
Cucumber & okra	0.227	0.374	0.491	-0.077	0.044	0.021
Sugar	0.069	-0.074	-0.034	0.648	0.143	-0.113
Green banana	0.112	0.227	-0.040	0.588	-0.030	-0.152
Banana	0.187	0.334	0.005	0.584	-0.184	0.092
Starchy tubers	-0.317	0.089	-0.050	0.475	-0.056	0.100
Cabbage (white)	0.008	0.121	0.078	0.117	0.680	-0.025
Bread	0.368	-0.013	0.199	-0.134	0.596	-0.063
Mandazi§	-0.027	0.023	-0.026	0.009	0.561	0.080
Rice	0.496	-0.060	0.065	-0.049	0.551	0.002
Sunflower oil	0.243	-0.347	-0.045	0.062	0.084	-0.628
Milk	0.152	-0.107	-0.003	0.037	0.044	0.576
Mixed vegetable fats & oil	0.211	0.264	0.003	-0.220	0.090	0.557
Butter and lard	-0.328	-0.134	-0.235	0.155	0.009	0.490
Red meat	0.342	-0.165	0.279	0.080	0.105	0.432
Variance explained (%)	**8.8**	**7.1**	**6.7**	**6.5**	**6.1**	**5.2**

Food groups with factor loadings < 0.4and >-0.4, thus not presented: tea, chicken meat, juice, *uji* ∫, coffee, leguminous vegetables, potatoes. Rotation method Varimax with Kaiser Normalization. Rotation converged in 10 iterations.
[1] alcoholic beverages excluded, * unleavened East African flat wheat bread; ♯ oranges, watermelon and pineapple; † traditional Tanzanian food, synonymously used for a dish of amaranth leaves and e.g. onions, tomatoes or/ and carrots in various amounts; § East African donuts; ∫ thin millet or maize based porridge

Four of the six new components were very much comparable to the Diet of the Rich, *Mchicha*, Banana and Fatty Diet of the first PCA and were consequently called the same. In spite of the similarities there was one important change in the components "Diet of the Rich" and "Fatty Diet" from PCA1 to PCA2: the food group red meat, which loaded high in the Diet of the Rich of the first PCA now loaded high in the component called Fatty Diet of the second PCA.

The two "new" components were named "Fruity Diet" because of the high factor loadings of the fruit groups and "Starchy Diet" which were mainly characterized by foods rich in carbohydrates like rice, bread, and *Mandazi*: East African donuts. The Fruity Diet correlated negatively with property at a low level, r_s = -0.22, whereas the Starchy Diet correlated positively with property, r_s = 0.28 (all *P*s <0.001).

The food groups which factor loadings were too small to be considered were called "non contributors" and changed from PCA1 to PCA2. Potatoes, chicken meat, *uji*, juice, and coffee were non contributors in both PCAs. Other food groups were non contributors in PCA1 but had high loadings in PCA2 like mango & papaya (0.609), white cabbage (0.680), *Mandazi* (0.561) or vice versa like tea and leguminous vegetables (-0.410 and 0.537 respectively).

Associations between dietary patterns, socioeconomic parameters and BMI

Both Diets of the Rich correlated significantly with property ($r_{s\_property}$ = 0.37 and 0.35) and also at a medium level with education and occupation of the women but negatively with age ($r_{s\_education}$ = 0.45 and 0.51 and $r_{s\_occupation}$ = 0.43 and 0.46 respectively, r_{s_age} = -0.24 and -0.25, all *P*s <0.001). The Diet is also significantly and positively correlated with BMI measured at the interview, r_s = 0.35 and 0.36 respectively. This association is inverse to the Banana Diet of PCA 1 where $r_{s\_education}$ is -0.31, $r_{s\_occupation}$ is -0.25 and r_{s_age} is 0.25 (all *P*s <0.001). It is less pronounced for the Banana Diet of PCA (Table 12).

Table 12: Spearman correlation coefficients of dietary patterns, socioeconomic parameters and BMI

	PCA	Property level n = 345	Education n = 344	Occupation n = 344	Age n= 345	BMI at interview n = 341
Diet of the Rich	1	**0.372****	**0.448****	**0.426****	**-0.239****	**0.351****
	2	**0.351****	**0.514****	**0.458****	**-0.248****	**0.359****
***Mchicha* Diet**	1	-0.068	0.024	-0.066	**-0.144****	**0.232****
	2	-0.076	0.041	-0.043	-0.096	**0.225****
Banana Diet	1	**-0.155****	**-0.308****	**-0.252****	**0.246****	-0.060
	2	-0.058	**-0.262****	**-0.180****	0.247**	-0.031
Fatty Diet	1	**-0.176****	**-0.281****	**-0.248****	**0.113***	**-0.165****
	2	-0.060	**-0.150****	**-0.132***	0.053	-0.088
Fruity Diet	2	**-0.216****	**-0.112***	**-0.160****	-0.032	-0.007
Starchy Diet	2	**0.284****	**0.152****	**0.155****	-0.067	0.167**

*. Correlation is significant at the 0.05 level (2-tailed)
**. Correlation is significant at the 0.01 level (2-tailed)

Thus, the women with a high affiliation to the Diet of the Rich tend to be younger, to have a higher education and a better occupation than those women who tend to follow the Banana Diet which could therefore also be called “Traditional Diet”. Regarding the Fatty Diet the association of property, education and occupation is negatively and with age positively. Also BMI is negatively associated with the Fatty Diet but only significantly for the Fatty Diet of PCA 2. However, all these associations remain at a low level. Compared to the Diet of the Rich, women who follow the Fatty Diet tend to be older, have a lower education and occupation and are poorer.

3.3.5 Breast cancer risk: Basic model

Table 13 presents the odds ratios (OR) of a non-conditional multivariate logistic regression analysis examining the associations between early life events, reproductive behaviour and breast cancer. In this model, the basic model, a higher property level was associated with a reduced breast cancer risk. In relation to the reference the OR for medium property level was 0.34 and dropped to 0.22 for women with a high property level (95 % CI 0.19 – 0.61 and 95 % CI 0.09 – 0.55 respectively, all *P*s <0.01). A high BMI at 20 years was associated with a higher breast cancer risk (OR 1.31, 95 % CI 1.11 - 1.55, *P* <0.01). No significant risk association was found for BMI at the time of interview, the age at menarche, the age at first full-term pregnancy and no pregnancy. The odds ratio for lifelong lactation was just below one (OR 0.99, 95 % CI 0.98-1.00, *P* <0.01), indicating a decreased risk in the whole group.

Table 13: Logistic regression of basic breast cancer risk model

Variable	***P*-value**	**Odds Ratio**	**95% CI**	**n**
Property level				
Low	Reference			87
Medium	**<0.01**	**0.34**	**0.19 - 0.61**	198
High	**<0.01**	**0.22**	**0.09 - 0.55**	48
Body Mass Index (kg/m²)				
at 20 years	**<0.01**	**1.31**	**1.11 - 1.55**	333
at interview	0.15	0.94	0.87 - 1.02	333
Age at menarche (year)	0.07	0.84	0.69 - 1.01	333
Age at first full term pregnancy				
≤ 20years	Reference			193
> 20 years	0.15	1.52	0.86 - 2.69	122
No pregnancy	0.60	0.69	0.17 - 2.79	18
Menopausal status (pre-/postmenopausal)	0.76	1.15	0.48 - 2.74	149/ 184
Lifelong lactation (month)	**<0.01**	**0.99**	**0.98 - 1.00**	333

Adjusted for age, place of living;
Constant: p-value 0.56, OR 0.25; Cox & Snell R^2 = 0.15; Nagelkerke's R^2 = 0.21, Overall percentage correctly classified = 76 %

After dividing the group into quintiles of lactation (Table 14), the odds ratio decreased with prolonged lactation from 0.57 (95 % CI 0.25 – 1.31, *P* = 0.19) in the second quintile (55 – 90 months), to 0.16 (95 % CI 0.06 – 0.046, *P* <0.01) in the fourth quintile (114-131 month). In the fifth quintile (>131 months) the odds ratio increased again to 0.37 (95 % CI 0.14 - 0.97, *P* = 0.04).

Table 14: Logistic regression for five groups with increasing lifelong lactation

Variable	*P* value	Odds Ratio	95% CI	n
Lifelong lactation				
≤ 54 months	Reference			66
55-90 months	0.19	0.57	0.25 – 1.31	70
91-113 months	0.02	0.32	0.12 – 0.82	55
114-131 months	<0.01	0.16	0.06 – 0.46	69
> 131 months	0.04	0.37	0.14 – 0.97	73

Adjusted for age, place of living, property level, BMI at 20 years, BMI at interview, age at menarche, age at first full term pregnancy, no pregnancy and menopausal status
Constant: P = 0.58, OR 0.26; Cox & Snell R^2 = 0.17; Nagelkerke's R^2 = 0.23, Overall percentage correctly classified = 75 %

3.3.6 Breast cancer risk stratified for menopausal status

Since menopause is a known effect modifier (L. Yang *et al.* 2008) the study population was stratified for menopausal status. High property level showed a significant breast cancer risk reducing effect in both menopausal groups (OR 0.11, 95 % CI 0.02 – 0.75 and OR 0.17, 95 % CI 0.05 – 0.58 respectively, all *P*s <0.03) (Table 14). Medium property level was significantly associated with a decreased risk in postmenopausal women only (OR 0.18, 95 % CI 0.08 – 0.42, P <0.01). Late menarche was also associated with a decreased risk but for premenopausal women only. In both menopausal groups, a high BMI at the age of 20 was associated with a higher risk. Late first full-term pregnancy increased the risk among the postmenopausal women only, and the preventive effect of long lifetime lactation was more pronounced in premenopausal women. The quality measures of the logistic regression: the correlation coefficients of Cox & Snell and Nagelkerke improved slightly compared to the basic model which indicates an improvement in the risk estimations.

Table 15: Logistic regression of basic breast cancer risk model stratified for menopausal status

Menopausal status	Variable	*P* value	Odds Ratio	95% CI	n
Pre-Menopausal*	Property level				
	Low	Reference			41
	Medium	0.28	0.62	0.26 - 1.48	90
	High	**0.02**	**0.11**	**0.02 - 0.75**	18
	Body Mass Index (kg/m²)				
	at 20 years	**0.01**	**1.41**	**1.10 - 1.81**	149
	at interview	0.32	0.94	0.84 - 1.06	149
	Age at menarche (year)	**0.05**	**0.74**	**0.56 - 1.00**	149
	Age at first full term pregnancy				
	≤ 20 years	Reference			75
	> 20 years	0.92	1.05	0.44 - 2.49	68
	No pregnancy	0.23	4.29	0.41 - 45.19	6
	Lifelong lactation (month)	**<0.01**	**0.98**	**0.97 - 0.99**	149
Post-Menopausal**	Property level				
	Low	Reference			46
	Medium	**<0.01**	**0.18**	**0.08 - 0.42**	108
	High	**0.01**	**0.17**	**0.05 - 0.58**	30
	Body Mass Index (kg/m²)				
	at 20 years	**0.02**	**1.38**	**1.06 - 1.80**	184
	at interview	0.14	0.92	0.82 - 1.03	184
	Age at menarche (year)	0.68	0.95	0.72 - 1.24	184
	Age at first full term pregnancy				
	≤ 20 years	Reference			118
	> 20 years	**0.04**	**2.40**	**1.03 - 5.60**	54
	No pregnancy	0.38	0.40	0.05 - 3.11	12
	Lifelong lactation (month)	0.16	0.99	0.98 - 1.00	184

Adjusted for age and place of living;
*pre-menopausal: Constant: P = 0.47, OR 0.07; Cox& Snell R^2 = 0.21, Nagelkerke's R^2 = 0.3, percentage correct = 75 %;
**post-menopausal: Constant: P = 0.44, OR 0.05; Cox & Snell R^2 = 0.19, Nagelkerke's R^2 = 0.27, overall correct = 78 %

3.3.7 Breast cancer risk: Dietary patterns

Table 16 presents the results of the non-conditional multivariate and logistic regression examining the associations between dietary behaviour and breast cancer. According to these results three out of the four dietary patterns, the Mchicha, Banana and Fatty Diet are associated with an increased risk for breast cancer.

Table 16: Logistic regression: Dietary patterns and breast cancer

Variable	*P* value	Odds Ratio	95% CI	*n*
Dietary patterns (PCA 1)				
Diet of the Rich	0.95	1.01	0.79 - 1.30	345
Mchicha Diet	<0.01	1.47	1.14 - 1.88	345
Banana Diet	<0.01	1.94	1.43 - 2.63	345
Fatty Diet	<0.01	1.62	1.26 - 2.07	345

Adjusted for age.
Constant: *P* <0.01, OR 0.56; Cox & Snell R^2 = 0.13 ; Nagelkerke's R^2 = 0.18
Overall percentage correctly classified 74 %

3.3.8 Breast cancer risk: Dietary patterns and basic model

Dietary patterns of PCA1

After including socioeconomic parameters and reproductive variables of the basic model (page 41) to the dietary patterns of PCA1 in the logistic regression the odds ratio (OR) for the Mchicha Diet changed from a significant OR of 1.47 (95 % CI 1.14 - 1.88, *P* <0.01) to a non significant OR of 1.28 (95 % CI 0.97 - 1.7, *P* = 0.08). The Banana and the Fatty Diet were still associated with an increased breast cancer risk on a significant level (Table 16). The OR for the Fatty Diet increased to 3.04 (95 % CI: 1.34 - 6.91; *P* <0.01) among women with the highest consumption (4th Quartile).

No changes were found in the risk estimations of the parameters of the basic model. Hence, high property level and a long lifelong lactation were associated with a reduced risk, and a high BMI at 20 years of age was associated with an increased risk. The other parameters, BMI at interview, no pregnancy, age at menarche and menopausal status showed no risk association.

Table 17: Logistic regression: Dietary patterns (PCA1) and basic breast cancer risk model

Variable	*P* value	Odds Ratio	95% CI	*n*
Property level				
Low	Reference			87
Medium	**<0.01**	**0.39**	**0.21 - 0.72**	198
High	**0.01**	**0.26**	**0.09 - 0.75**	48
Body mass index (kg/m²)				
At 20 years	**0.01**	**1.26**	**1.05 - 1.51**	333
At interview	0.11	0.93	0.86 - 1.02	333
Age at first full term pregnancy				
≤ 20years	Reference			193
> 20 years	0.08	1.74	0.93 - 3.23	122
No pregnancy	0.79	0.82	0.18 - 3.70	18
Lifelong lactation	**0.02**	**0.99**	**0.98 - 1.00**	333
Dietary patterns (PCA 1)				
Diet of the Rich	0.40	1.15	0.83 - 1.59	333
Mchicha Diet	0.08	1.28	0.97 - 1.70	333
Banana Diet	**<0.01**	**1.75**	**1.20 - 2.57**	333
Fatty Diet	**0.01**	**1.50**	**1.12 - 1.99**	333

Adjusted for age, place of living, age at menarche, and menopausal status.
Constant: P = 0.81, OR 0.54; Cox & Snell R^2 = 0.20; Nagelkerke's R^2 = 0.29; Overall percentage correctly classified = 76%.

With increased affiliation to the Fatty Diet, total fat intake increased significantly (P = 0.04), whereas percentage of energy of total energy from fat did not change (P = 0.83). The ratio of polyunsaturated fatty acids to saturated fatty acids (P/S ratio) was inversely associated with breast cancer risk. The 4th quartile with lowest P/S ratio of 0.5 had the highest OR of 3.3 (95 % CI 1.5 – 7.2). In all subgroups the intake of saturated fatty acids was very similar whereas the intake of PUFA decreased with increased adherence (4th quartile) to the Fatty Diet (Figure 6). However, there was no risk association found between total fat intake (median 72 g/d) and breast cancer.

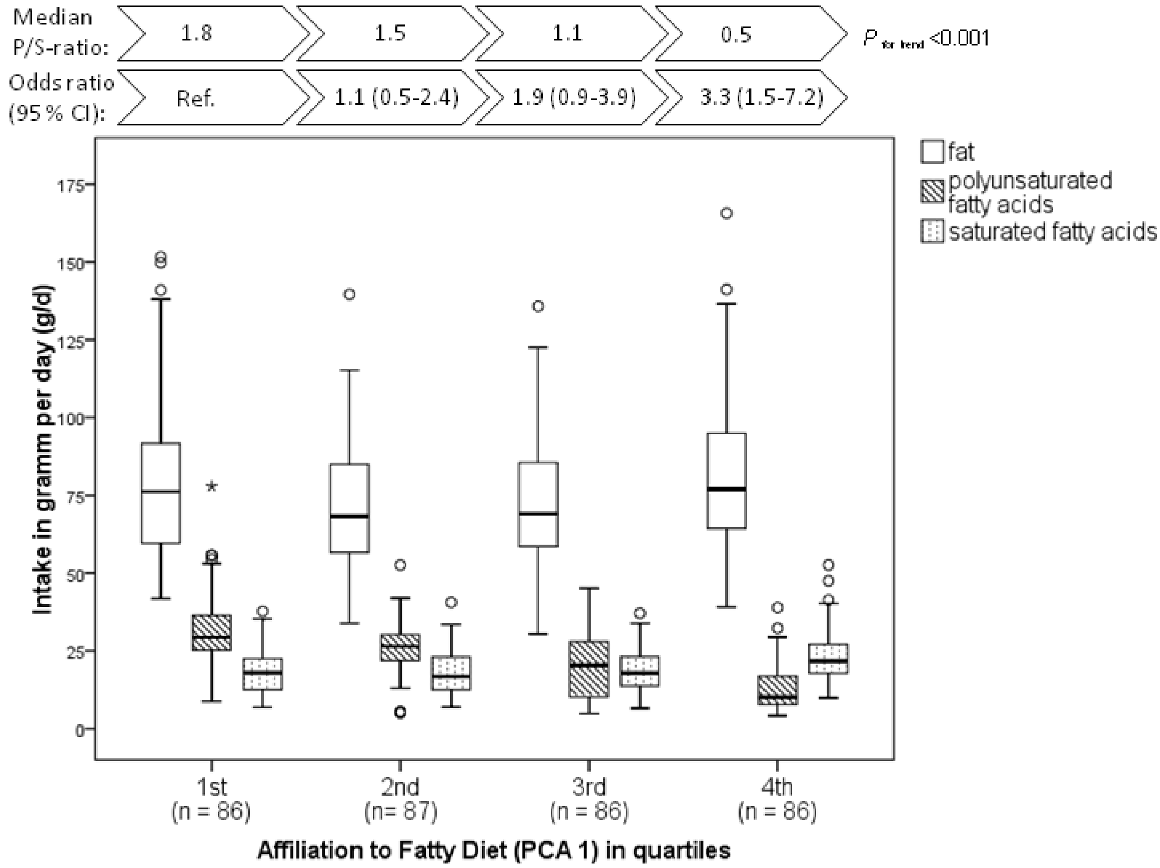


Figure 6: Intake of fat, polyunsaturated and saturated fatty acids per day and the related odds and P/S ratios in quartiles of the Fatty Diet (PCA1)

Dietary patterns of PCA2 and alcoholic beverages as distinct variables

Table 18 presents the results of the logistic regression including the dietary patterns retained in the second PCA with the alcoholic beverages included as distinct variable. The Mchicha Diet and the Banana Diet were no longer associated with breast cancer risk, but the new Fruity Diet and again a Fatty Diet very similar to the first Fatty Diet were associated with increased risk (OR 1.61 and 1.42, 95 % CI 1.14 - 2.28 and 1.08 - 1.87 respectively, both *Ps* = 0.01). No risk association was found for *Mbege* and beer & wine.

The risk estimations of the parameters of the basic risk model remained at the same level, like in the estimation including the dietary patterns of the first PCA (page 46).

Table 18: Logistic regression: Dietary patterns (PCA2), alcoholic beverages and basic breast cancer risk model

Variable	*P* value	Odds Ratio	95% CI	n
Property level				
Low	Reference			87
Medium	**0.00**	**0.37**	**0.20 - 0.71**	198
High	**0.01**	**0.27**	**0.09 - 0.77**	48
Body mass index (kg/m²)				
At 20 years	**0.01**	**1.27**	**1.06 - 1.53**	333
At interview	0.09	0.93	0.85 - 1.01	333
Age at first full term pregnancy				
≤ 20years	Reference			193
> 20 years	0.06	1.83	0.97 - 3.45	122
No pregnancy	0.80	0.82	0.18 - 3.84	18
Lifelong lactation	**0.02**	**0.99**	**0.98 - 1.00**	333
Dietary patterns (PCA2)				
Diet of the Rich (no alc)	0.17	1.28	0.90 - 1.59	333
Fruity Diet (no alc)	**0.01**	**1.61**	**1.14 - 2.28**	333
Mchicha Diet (no alc)	0.70	1.06	0.80 - 1.40	333
Banana Diet (no alc)	0.12	1.32	0.93 - 1.87	333
Starchy Diet (no alc)	0.86	1.02	0.78 - 1.34	333
Fatty Diet (no alc)	**0.01**	**1.42**	**1.08 - 1.87**	333
Mbege ‡	0.08	1.00	1.00 - 1.00	333
Beer & wine	0.87	1.00	1.00 - 1.00	333

Adjusted for age, place of living, age at menarche, and menopausal status.
Constant *P* = 0.64, OR 0.30; Cox & Snell R^2 = 0.21; Nagelkerke's R^2 = 0.29; Overall percentage correctly classified = 77%.
‡ often homemade opaque beer from bananas and millet

Like the Fatty Diet of PCA1, with increased affiliation to the Fruity Diet, total fat intake increased significantly ($P < 0.001$). The ratio of polyunsaturated fatty acids to saturated fatty acids (P/S ratio) was again inversely associated with breast cancer risk. The 4th quartile with lowest P/S ratio of 0.7 had the highest OR of 1.6 (95 % CI 0.7 – 3.3). In all subgroups the intake of saturated fatty acids was very similar whereas the intake of PUFA decreased with increased adherence (4th quartile) to the Fruity Diet (Figure 7).

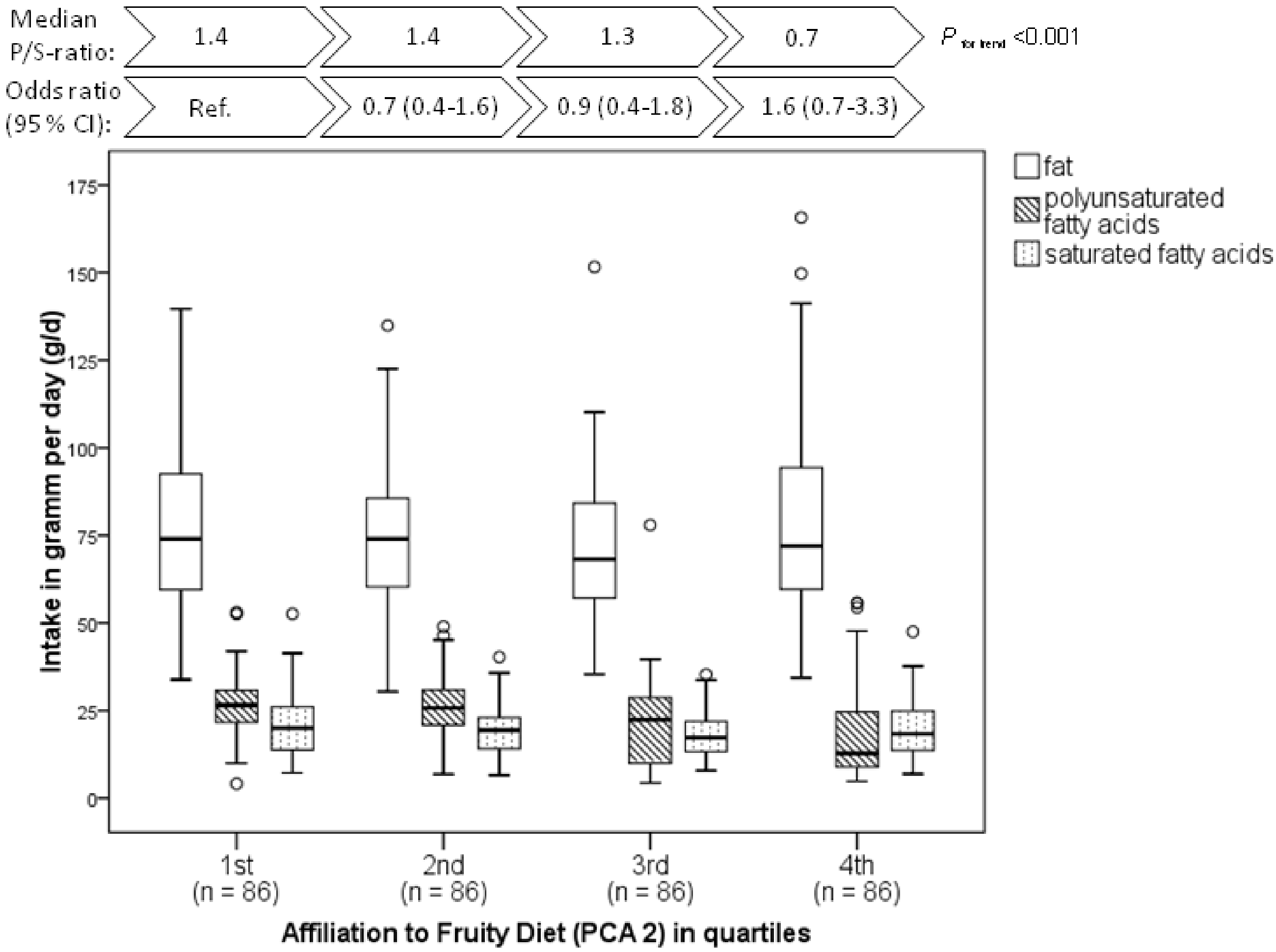


Figure 7: Intake of fat, polyunsaturated and saturated fatty acids per day and the related odds and P/S ratios in quartiles of the Fruity Diet (PCA2)

3.3.9 Breast cancer risk: Dietary patterns and basic risk model stratified for menopausal status

Premenopausal women

Table 19 shows the results of the logistic regression of the basic model plus the dietary patterns and alcoholic drinks stratified for premenopausal women. The Fruity and the Fatty Diet are significantly associated with increased breast cancer risk. Compared to the results of the basic risk model stratified for menopausal status (see chapter 3.3.5, p 41) changes in risk estimations were observed for property level, BMI at 20 years which changed to a non significant level. Also R^2 increased compared to the model basic model indicating an improvement in the risk estimation, the 95 % CIs increased as well indicating a lack of statistical power in the risk estimation.

Table 19: Logistic regression: Basic risk model, dietary patterns (PCA2) and alcoholic drinks in *premenopausal* women

Variable	*P* value	Odds Ratio	95% CI	n
Property level				
Low	Reference			41
Medium	0.82	1.14	0.38 - 3.40	90
High	0.09	0.14	0.02 - 1.34	18
Body Mass Index (kg/m²)				
at 20 years	0.12	1.27	0.94 - 1.73	149
at interview	0.45	0.95	0.82 - 1.09	149
Age at menarche (year)	0.29	0.84	0.60 - 1.17	149
Age at first full term pregnancy				
≤ 20 years	Reference			75
> 20 years	0.67	1.26	0.44 - 3.66	68
No pregnancy	0.13	8.00	0.56 - 115.23	6
Lifelong lactation (month)	**0.03**	**0.98**	**0.97 - 1.00**	149
Dietary patterns (PCA2)				
Diet of the Rich (no alc)	0.54	1.21	0.66 - 2.23	149
Fruity Diet (no alc)	**0.03**	**1.82**	**1.08 - 3.08**	149
Mchicha Diet (no alc)	0.13	1.38	0.91 - 2.11	149
Banana Diet (no alc)	0.11	1.62	0.89 - 2.94	149
Starchy Diet (no alc)	0.92	0.98	0.67 - 1.44	149
Fatty Diet (no alc)	**0.01**	**1.87**	**1.14 - 3.07**	149
Mbege‡	0.29	1.00	1.00 - 1.00	149
Beer & wine	0.45	1.00	0.99 - 1.00	149

Adjusted for age and place of living;
pre-menopausal: Constant: P = 0.43, OR 0.03; Cox& Snell R^2 = 0.33, Nagelkerke's R^2 = 0.46, percentage correct = 79 %;
‡ often homemade opaque beer from bananas and millet

Postmenopausal women

Among postmenopausal women neither a dietary pattern nor alcoholic beverages were associated with breast cancer as can be seen in Table 20 presenting the results of the logistic regression. Like in the basic model stratified for menopausal status, a high property level was associated with a decreased risk, OR = 0.15 (95 % CI 0.03-0.66), a high BMI at 20 years and a pregnancy at an older age (> 20 years) was associated with increased risk, OR = 1.43 and OR = 2.87, respectively (95 % CIs 1.08 – 1.89 and 1.12 – 7.35, respectively). Lifelong lactation was not associated with breast cancer among postmenopausal women.

Table 20: Logistic regression: Basic risk model, dietary patterns (PCA2) and alcoholic drinks in *postmenopausal* women

Variable	***P* value**	**Odds Ratio**	**95% CI**		**n**
Property level					
Low	Reference				46
Medium	**0.00**	**0.13**	**0.05 -**	**0.35**	108
High	**0.01**	**0.15**	**0.03 -**	**0.66**	30
Body Mass Index (kg/m²)					
at 20 years	**0.01**	**1.43**	**1.08 -**	**1.89**	184
at interview	0.08	0.88	0.77 -	1.02	184
Age at menarche (year)	0.86	0.97	0.73 -	1.30	184
Age at first full term pregnancy					
≤ 20 years	Reference				118
> 20 years	**0.03**	**2.87**	**1.12 -**	**7.35**	54
No pregnancy	0.47	0.43	0.04 -	4.28	12
Lifelong lactation (month)	0.30	0.99	0.98 -	1.01	184
Dietary patterns (PCA2)					
Diet of the Rich (no alc)	0.41	1.26	0.73 -	2.20	184
Fruity Diet (no alc)	0.16	1.53	0.85 -	2.75	184
Mchicha Diet (no alc)	0.57	0.84	0.46 -	1.54	184
Banana Diet (no alc)	0.47	1.21	0.72 -	2.01	184
Starchy Diet (no alc)	0.21	1.50	0.79 -	2.82	184
Fatty Diet (no alc)	0.32	1.23	0.83 -	1.82	184
Mbege‡	0.11	1.00	1.00 -	1.00	184
Beer & wine	0.75	1.00	1.00 -	1.00	184

Adjusted for age and place of living;
post-menopausal: Constant: P = 0.54, OR 0.08; Cox & Snell R^2 = 0.23, Nagelkerke's R^2 = 0.33, overall correct = 80 %
‡ often homemade opaque beer from bananas and millet

3.3.10 Breast cancer risk: Stratified for body mass index

Following the study of Sonestedt *et al.* (2007) who concluded that obesity and self-reported past food habits - as used in this study - may be important confounders of diet-breast cancer relationships, an analysis stratified into BMI groups was added. The study population was divided into three BMI groups. The median BMI of 25.8 kg/m² for all women was chosen as reference point for the middle group to allow reasonable sizes for further statistical analysis. Thus, the first group included all women with a BMI below 24 kg/m² (n = 89), the second all women with a BMI of 24 kg/m² up to 26 kg/m² (n = 91) and a third group included all women with a BMI above 26 kg/m² (n = 153). The median age of the BMI group <24 kg/m² was 50 years, the same as in the third group with a BMI >26 kg/m². The women in the BMI group 24-26 kg/m² were slightly older (median 53 years). The groups differed in BMI at age 20 years and BMI change over time (all *Ps* <0.001) (Figure 8). Lifelong lactation was equally distributed over all three categories.

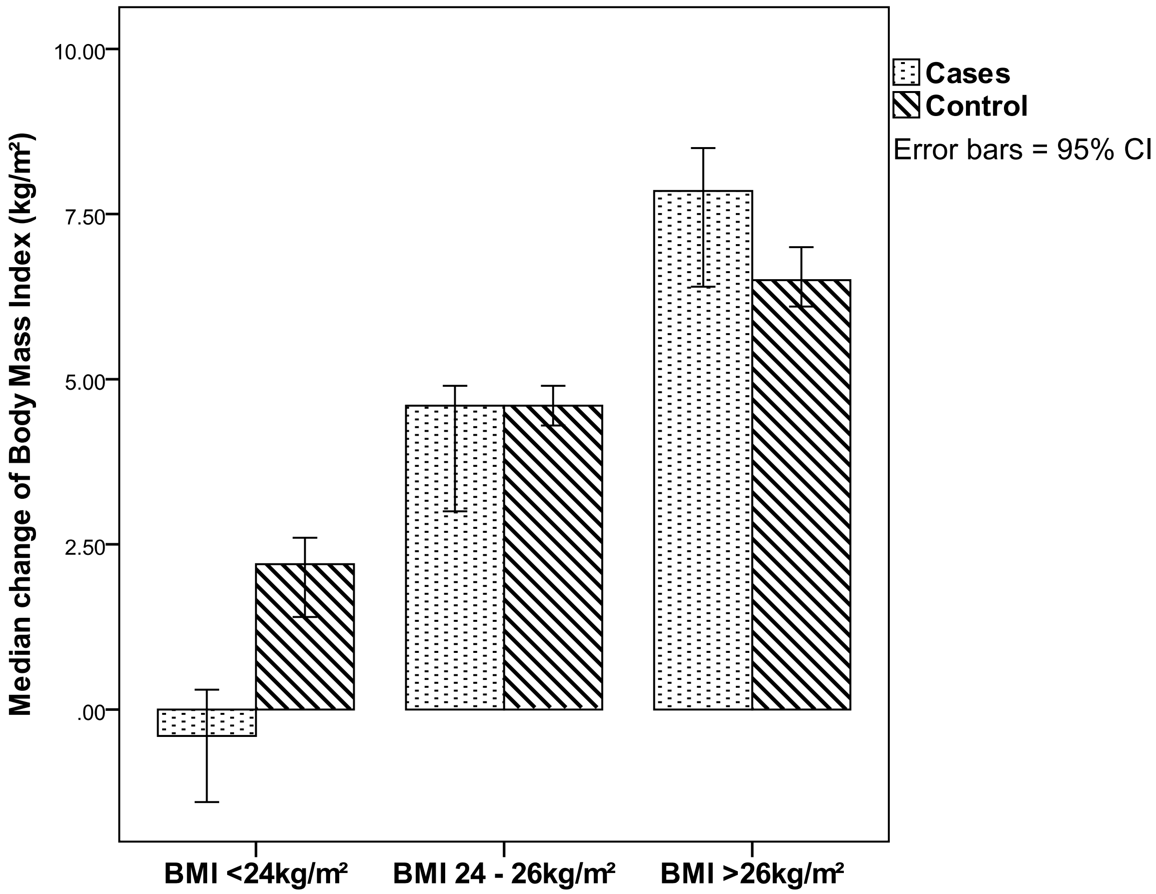


Figure 8: BMI change over time: BMI at interview minus BMI age 20 years

In the following a logistic regression of the basic model plus the dietary patterns including the alcoholic beverages as distinct variables stratified for the BMI group was performed on an explorative level.

Breast cancer risk among women with a BMI below 24 kg/m²

The results of the logistic regression of the first group, BMI <24 kg/m², are presented in Table 21. In this group the OR of BMI at 20 years remained above 1 indicating an increased risk with high BMI at a younger age, OR = 4.15 (95 % CI 1.41 – 12.27, *P* = 0.01) whereas a high BMI at interview was associated with a decreased risk, OR = 0.39 (95 % CI 0.21 – 0.7, *P* <0.01). Longer lactation remained a protective risk parameter, although not statistically significant anymore (OR = 0.89, 95 % CI 0.95 – 1.00, *P* = 0.08). A high affiliation to the Diet of the Rich and the Banana Diet were associated with increased breast cancer risk (OR 8.81, 95% CI 1.41 - 54.93, *P* = 0.02 and OR 3.87, 95% CI 1.21 – 12.40, *P* = 0.02). However, the 95 % confidence intervals for BMI at 20 years and both dietary patterns were very large and indicating a low precision of the risk estimation as with BMI at 20 years. Also the fact that the -2 Log likelihood estimation terminated at its maximum number of iterations and a final solution was not found indicated that the sample size for this kind of analysis was too small for a correct risk estimation.

Table 21: Logistic regression: Basic risk model and dietary patterns (PCA2) in women with a body mass index below 24kg/m²

Variable	*P* value	Odds Ratio	95% CI	n
Body mass index (kg/m²)				
At 20 years	**0.01**	**4.15**	**1.41 - 12.27**	89
At interview	**<0.01**	**0.39**	**0.21 - 0.70**	89
Lifelong lactation	0.08	0.98	0.95 - 1.00	89
Dietary patterns (PCA2)				
Diet of the rich (no alc)	**0.02**	**8.81**	**1.41 - 54.93**	89
Fruity diet (no alc)	0.06	2.73	0.97 - 7.71	89
Mchicha diet (no alc)	0.86	0.92	0.35 - 2.39	89
Banana diet (no alc)	**0.02**	**3.87**	**1.21 - 12.40**	89
Starchy diet (no alc)	0.49	1.32	0.60 - 2.90	89
Fatty diet (no alc)	0.18	1.78	0.77 - 4.08	89
Mbege‡	0.42	1.00	1.00 - 1.00	89
Beer & wine	0.37	1.00	1.00 - 1.00	89

Adjusted for age, place of living, property level, age at menarche, age at first full term pregnancy and menopausal status
Constant: P = 0.35, OR 0; Cox & Snell R^2 = 0.5; Nagelkerke's R^2 = 0.67; Overall percentage correctly classified = 84 %. -2 Log likelihood estimation terminated at iteration number 20 because maximum iterations had been reached. Final solution could not be found.
‡ often homemade opaque beer from bananas and millet

Breast cancer risk among women with a BMI between 24 and 26 kg/m²

The only parameter which could be considered as lifestyle or nutritional behaviour parameter which indicated a risk association in the BMI group 24-26 kg/m² was *Mbege*, the local brew (OR 1.003, 95% CI 1.000 – 1.006, P = 0.05). The only statistical significant risk association was found for menopausal status. Women who were postmenopausal had a much higher breast cancer risk, OR = 31.38 (P = 0.04), than premenopausal women, the reference group. Hence, the confidence interval for this risk estimation was very large (95% CI 1.22 – 804.24) indicating a low precision in this risk estimation. Previously estimated risk factors like BMI at 20 years and lifelong lactation were no longer statistically associated with breast cancer risk (Table 22).The -2 Log likelihood estimation terminated at its maximum iterations and a final solution was not found, thus again, the sample size might have been too small for this analysis.

Table 22: Logistic regression: Basic risk model and dietary patterns (PCA2) in women with a body mass index of 24-26kg/m²

Variable	*P* value	Odds Ratio	95% CI	n
Body mass index (kg/m²)				
At 20 years	0.44	1.28	0.69 - 2.37	91
At interview	0.55	1.41	0.46 - 4.34	91
Menopausal status				
premenopausal	Reference			39
postmenopausal	**0.04**	**31.38**	**1.22 - 804.24**	52
Lifelong lactation	0.33	0.99	0.96 - 1.01	91
Dietary pattern (PCA2)				
Diet of the rich (no alc)	0.53	0.64	0.16 - 2.58	91
Fruity diet (no alc)	0.23	0.44	0.12 - 1.66	91
Mchicha diet (no alc)	0.93	0.96	0.39 - 2.35	91
Banana diet (no alc)	0.40	0.58	0.16 - 2.10	91
Starchy diet (no alc)	0.20	2.28	0.65 - 7.91	91
Fatty diet (no alc)	0.69	0.83	0.34 - 2.04	91
Mbege‡	**0.05**	**1.003**	**1.00 - 1.006**	91
Beer & wine	0.90	1.00	0.99 - 1.00	91

Adjusted for age, place of living, property level age at menarche, age at first full term pregnancy
Constant *P* = 0.72, OR = 0; Cox & Snell R^2 = 0.25; Nagelkerke's R^2 = 0.39; Overall percentage correctly classified = 85 %. -2 Log likelihood estimation terminated at iteration number 20 because maximum iterations had been reached. Final solution could not be found.
‡ often homemade opaque beer from bananas and millet

Breast cancer risk among women with a BMI above 26 kg/m²

As presented in Table 23 the risk estimation of BMI at 20 years and at interview changed if estimations are done with women of the BMI group >26 kg/m² only. A high BMI at the age of 20 was not associated with a statistical significant increased risk anymore as reported previously for the whole group (see chapter 3.3.5). The odds ratio for BMI at interview changed from 0.94 in the basic model to 1.53 (95% CI 0.87 - 1.02, P = 0.15 and 1.20 – 1.96, P < 0.01, respectively). Among the dietary patterns only the Fatty Diet was associated with an increased risk (OR 1.66, 95% CI 1.0-2.78, P = 0.05).

Table 23: Logistic regression: Basic risk model and dietary patterns (PCA2) in women with a body mass index above 26kg/m²

Variable	*P* value	Odds Ratio	95% CI	n
Body mass index (kg/m²)				
At 20 years	0.06	1.32	0.98 - 1.76	153
At interview	**<0.01**	**1.53**	**1.20 - 1.96**	153
Lifelong lactation	0.34	0.99	0.98 - 1.01	153
Dietary pattern (PCA2)				
Diet of the rich (no alc)	0.60	0.86	0.50 - 1.50	153
Fruity diet (no alc)	0.21	1.42	0.82 - 2.46	153
Mchicha diet (no alc)	0.61	0.88	0.52 - 1.46	153
Banana diet (no alc)	0.79	1.09	0.59 - 2.01	153
Starchy diet (no alc)	0.53	1.22	0.65 - 2.29	153
Fatty diet (no alc)	**0.05**	**1.66**	**1.00 - 2.78**	153
Mbege‡	0.43	1.00	1.00 - 1.00	153
Beer & wine	0.86	1.00	1.00 - 1.01	153

Adjusted for age, place of living, property level, age at menarche, age at first full term pregnancy and menopausal status
Constant P = 0.01; OR = 0; Cox & Snell R^2 = 0.3; Nagelkerke's R^2 = 0.45; Overall percentage correctly classified = 85 %. -2 Log likelihood estimation terminated at iteration number 6 because parameters estimated changed by less than 0.001.
‡ often homemade opaque beer from bananas and millet

Differences in breast cancer estimations including dietary patterns

The logistic regression stratified for BMI group and showed diverse risk estimations regarding the possible association between breast cancer and nutritional behaviour. Only the Fatty Diet was repeatedly risk associated. The Fatty Diet of PCA1 and PCA2 were both associated with increased risk in the basic model. The latter was also found to increase risk in premenopausal women, and in women with a body mass index above 26kg/m². Although the Banana Diet was twice associated with increased risk, the Diets used were slightly different. The Banana Diet of PCA1 which showed an increased risk in women in the basic model did include *Mbege*, a local brew, unlike the Banana Diet of the PCA2 which showed a risk association in women with a BMI <24 kg/m². The Diet of the Rich showed only a risk association in women with a BMI <24 kg/m². A risk association with the Fruity Diet was found in postmenopausal women. The only dietary pattern which was not associated with breast cancer risk was the Starchy Diet.

3.3.11 Breast cancer risk estimation: Hunger periods

Out of the 15 women who reported a hunger threat before their menarche, 73 % remembered a moderate effect on body size and 7 % responded that their body size was affected a lot (Figure 9). The results were equally distributed between cases and controls ($P = 0.21$, Mann-Whitney U Test).

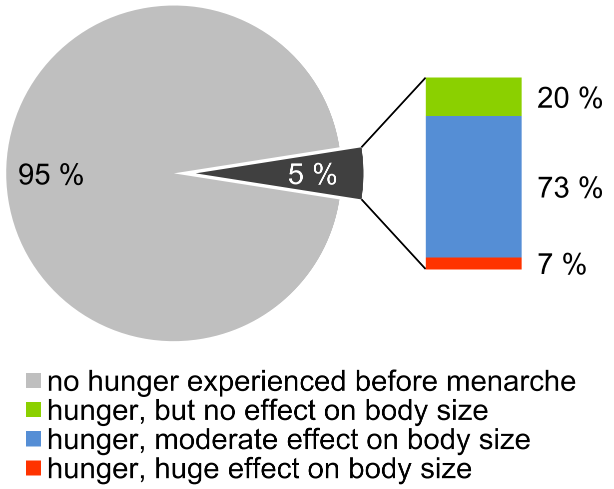


Figure 9: Prevalence of hunger before the menarche and effect on body size

Hunger threats after the menarche were reported by 87 women (31 %). Out of this group experienced 12 % a huge effect on body size, whereas 59 % reported a moderate effect and 30 % no effect (Figure 10). The results were also equally distributed between cases and controls ($P = 0.81$, Mann-Whitney U Test).

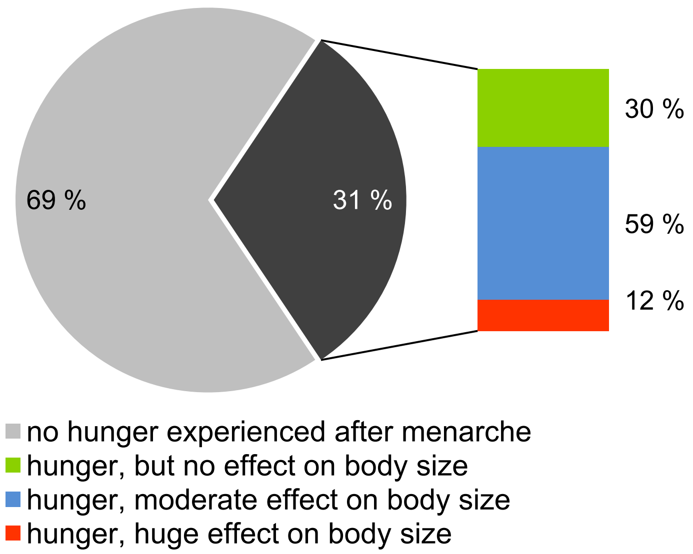


Figure 10: Prevalence of hunger after menarche and effect on body size

The number of women affected by hunger before their menarche with effect on their body size was too little to include into the logistic model. This was different for the parameter “hunger periods experienced after the menarche”. Hence, based on the reference parameter “no hunger or hunger, but no effect on body size”, the breast cancer and nutrition model estimated a decreased risk if the hunger experienced after the menarche had only a moderate effect on body size (OR 0.33, 95 % CI 0.13-0.85, P = 0.02). Hunger periods with huge effect on body size experienced after the menarche indicated an increased risk effect. However, the risk estimations were accompanied by large confidence intervals. The results were therefore considered as too explorative and therefore not reported.

3.3.12 Breast cancer risk estimation: Physical activity

The median physical activity level (PAL) was 1.37 for cases and controls and the distribution of the values was the same in both groups (P = 0.99, Mann-Whitney *U* Test). A PAL of <1.4 is considered to be equivalent to a sedentary lifestyle. Spearman correlation coefficients (r_s) showed a very low but significant association between PAL and energy intake (r_s = 1.3, P = 0.02). Other variables like education and occupation of the women as well as the property level were negatively associated with PAL (education r_s = -0.364, occupation r_s = -0.28, and property r_s = -0.20 respectively, all *P*s <0.01).

The inclusion of PAL as an additional variable into the logistic regression model on breast cancer and nutrition both stratified and not stratified for menopausal status did not change the findings already described. The odds ratio for PAL was estimated with a large confidence interval and thus finally neglected in the regression analysis.

3.4 Discussion

3.4.1 Lactation

Up to now estimations on an effect of prolonged lactation on breast cancer risk have been projected only. In a collaborative reanalysis of individual data from 47 epidemiological studies it was estimated that the incidence of breast cancer might decrease in high income countries from an observed cumulative incidence of 6.3 per 100 women to 2.7 per 100 women at the age of 70 years, if they had larger families and longer durations of breastfeeding (Beral *et al.* 2002). Thus would result in an average of 6.5 births instead of 2.5 and 24 months of breastfeeding each child instead of three. In this study the effect of prolonged lactation could be documented. The recorded median lifelong lactation of the parous women of 8 years (96 months, cases) and 9 years (108 months, controls) with a maximum of 20 years in both groups (240 months) is higher than elsewhere. Also the mean duration of 65 months reported in another case-control study from an African country was more than two years less than in this study (Okobia *et al.* 2006).

However, after an increasing protective effect on breast cancer due to prolonged lactation for 131 months (11 years), we observed a slight deceleration of this risk reducing effect (Table 14). The risk- reducing effect of breastfeeding may be fully utilised by then and other lifestyle factors may become more important, particularly in postmenopausal women.

The globally reported results of the effect of prolonged lactation on breast cancer risk in pre- and postmenopausal women are not consistent. According to our results, long lifetime lactation decreases the breast cancer risk in premenopausal women as also reported before (Galakshmi *et al.* 2009). One hypothesis was that the lack of effect in postmenopausal women might be due to a recall bias caused by the time span between past breastfeeding periods and the date of recall. Tanzania, as is common in sub-Saharan Africa, has a strong tradition of breastfeeding. According to the Tanzanian National Demographic Health Survey 2004, the median duration of breastfeeding per child is 21 months, congruent with the data of our study population (Tanzanian National Bureau of Statistics 2004). Findings of another study in the Kilimanjaro region conducted in 2004 also support our results (de Paoli 2004). Promislow *et al*. (2005) observed in their validation study on maternal recall of breastfeeding durations that there was a better recall of children from larger families. The family sizes in our study can be considered large, median number of children was 5. Therefore, we do not assume a reporting bias because of age or time span between event and recall. However, we cannot exclude nor estimate the bias due to rounding also observed in other studies (Beral *et al.* 2002). This may also be a non-reported problem in other studies.

Yang and Jacobsen concluded that there is no consensus about a relationship between breastfeeding and breast cancer (Yang *et al.* 2008). They suggest that three population groups should be considered in future analyses: nulliparous, parous women who have breastfed, and parous women who never breastfed. Here, the number of nulliparous women and parous women who never breastfed was too small for group comparisons. Hence we decided against group classification and chose lifetime lactation as a continuous variable starting with zero months. Studies in China and Japan found an inverse association between prolonged lactation and breast cancer risk for both pre- and postmenopausal women. Zheng *et al*. (2000) argue that the shorter lactation durations in Westernised countries result in inconsistent data. They did not report the median lifetime lactation, but it is probably lower than in Tanzania since their highest category started with 109 months, i.e. below the median lifetime lactation here. Due to lack of data we could not follow Minami *et al.* (2004) who postulate that different types of breastfeeding (exclusive, full, or partial) may be relevant.

Furthermore, the type of breast cancer might also be an important parameter in the discussion of the cancer-preventive effects of lactation. Beaber *et al.* (2008) observed different effects on postmenopausal women stratified for ductal, ductal-lobular, and lobular tumours. They suggested that the cell differentiation induced by breastfeeding may primarily exert its protective effect on the development of invasive ductal tumours rather than on other types. A risk-increasing effect of age at breastfeeding and ductal-lobular carcinoma could explain the different effects of lactation on breast cancer among postmenopausal women. This suggestion could be the cause behind our observation of a change in effect size with prolonged lactation, assuming a protective effect on one type of tumour and a risk- increasing effect of breastfeeding at an older age on another type.

3.4.2 Menarche

We found a low risk association between age at menarche and breast cancer among the premenopausal group. Comparing the results of age at menarche and breast cancer with other studies, the relevant risk factor might be age at menarche below 12 years. Women in our study population had a later median menarcheal age of 16 years and had their first child at a younger age compared to studies in Western countries (Chang-Claude *et al.*,2007; Ahlgren *et al*. 2004; McPherson *et al.* 2000; McCredie *et al.* 1998; La Vecchia *et al.* 1992). Age at first full-term pregnancy rather than menarcheal age was associated with breast cancer risk only among postmenopausal women. This association was also described in a multicentre, population-based US case-control study of white and African-American women (Li *et al.* 2007). They concluded that the interval between age at menarche and age at first delivery was important - a finding not supported by our data. This might be due to differences in menarcheal age. Thus, low age at menarche and late first full-term pregnancy may not be relevant as risk factors in a livelihood as in Tanzania. However an early menarche has been found to be associated with improved nutrition during childhood (Mesa *et al.* 2010; Bau *et al.* 2008; dos Santos Silva *et al.* 2002; Wadsworth *et al.* 2002). A study in Tanzania found the lowest age at menarche (13.2 years) among those girls whose fathers had a higher education and whose socioeconomic situation was very good or fairly good (Rebacz 2009). Thus, with improving socioeconomic status, age at menarche might decrease in low income countries as does the observed secular trend in Western countries (Ofuya 2008; Patton & Viner 2007; Gluckman & Hanson 2006; Parent *et al.* 2003), and this may become relevant for the risk of developing breast cancer.

3.4.3 Dietary pattern rich in fatty foods

No risk association was found with single nutrients but with the P/S-ratio (data not shown). As nutrients are ingested within diets, dietary patterns were obtained from two principal component analyses with Varimax rotation based on FFQ data and included into the basic logistic model. Several dietary patterns were associated with increased breast cancer risk. The most consistent results were found for two patterns, both called Fatty Diet. They were significantly associated with an increased risk in different logistical models. The Fatty Diets are basically characterized by a higher consumption of milk, mixed vegetable oils and fats, butter and lard, but a low consumption of sunflower oil. A diet rich in fat similar to our Fatty Diets was discussed by the EPIC-Potsdam study group using reduced rank regression, stating that specific fatty acids are less important in populations with a generally higher fat consumption (mean 8.3 to 10.4 g/MJ) (Schulz *et al.* 2008). However, this level of dietary fat intake as a proportion of energy intake was comparable to our study population (mean 10.2 g/MJ), but the mean total fat consumption in our population was 15 g per day lower because of the overall lower energy consumption than reported by Schulz and colleagues (2008). Here, the women's total fat intake was not associated with breast cancer risk (data not shown) although total fat intake increased significantly with increased affiliation to the fatty dietary patterns. Another prospective cohort study found a direct association between dietary fat intake including subtypes and post-menopausal invasive breast cancer (Thiébaut *et al.* 2007). They recorded at median 20.3 % energy intake from fat per day in the 1st

quintile and 40.1 % energy intake from fat per day in the last quintile. Only the latter energy intake level from dietary fat is comparable to our data. The wide range of fat intake observed in their study population may have resulted in an increased statistical power. This assumption was made on the hypotheses that a threshold effect may exist for dietary fat, such it would be difficult to detect an association between fat intake and breast cancer risk in Western populations (Thiébaut *et al.* 2007; Wynder *et al.* 1997). Thiébaut and colleagues (2007) referred to studies about Asian diets in which more people consume diets containing 20 % or less of energy from fat which have shown significant or borderline significant associations of fat intake and breast cancer risk. The median fat intake as percentage from energy intake in our study population was 39 % which is above this benchmark of 20 % and may explain why no association was found for our population. Regarding the fatty acid composition of the diet, the major fat sources reported by Thiébaut and colleagues (2007) were vegetable oils and fats, butter and mayonnaise. Except mayonnaise, these food items have also been identified by Schulz and colleagues (2008) and in our study as part of dietary patterns rich in fat which have been associated with a higher risk of developing breast cancer. Even if the fatty acid composition of foods varies intrinsically, this composition may be more important than the total fat intake. Our study population showed a negative association of the P/S ratio with breast cancer risk (Figure 6). This negative association was also observed in a case-control study among premenopausal women in Singapore but was attributed to PUFA intake only (Lee *et al.* 1991). However, results from the European Investigation into Cancer and Nutrition (EPIC) study and a case-control study in Connecticut (Sieri *et al.* 2008; Goodstine *et al.* 2003) supported the hypothesis that both saturated and polyunsaturated fatty acids influence inversely the oestrogen metabolism and mammary carcinogenesis (Rose *et al.* 1999; Key *et al.* 2002). In addition results from the Shanghai Women's Health study, a prospective cohort study, suggested that the relative amounts of n-6 PUFA to marine-derived n-3 PUFAs may be more important for the breast cancer risk than individual amounts of these fatty acids in the diet (Murff *et al.* 2011). They supported the hypothesis that the different PUFA compete as enzyme substrates inside membrane phospholipids (Bougnoux *et al.* 2006). This may also explain the contradictory results of other studies analyzing the effect of PUFAs on breast cancer risk (Saadatian-Elahi *et al.* 2004; MacLean *et al.* 2006; Holmes *et al.* 1999).

Investigators from the Black Women's Health Study, a prospective cohort study, identified a dietary pattern similar to our Fatty Diets called "Western diet" also based on a PCA with Varimax rotation and factor loadings for dairy products and meat at similar level (Agurs-Collins *et al.* 2009). However, they associated a lower risk for breast cancer only with another dietary pattern, the "prudent diet", characterized with a low consumption of meat and dairy products. Since both the Western and the prudent diet were more complex than in our study with each diet having more than 8 foods with factor loadings above 0.4, it is not known whether the non-relationship between the Western diet and breast cancer has been masked by a higher consumption of potentially preventive foods which in turn result in a high P/S ratio.

3.4.4 Effect of alcohol on dietary patterns risk association

The reported consumption of the local banana beer, *Mbege*, increased significantly with increased affiliation to the Fatty Diet (*P* for trend <0.001). The analysis of the basic logistic model including the dietary patterns of PCA1 estimated a higher breast cancer risk for women mainly following the Banana Diet. This pattern was also associated with a high consumption of *Mbege*, even though there is no risk association between alcohol intake and breast cancer risk in this study. According to the WRCF (2007) panel there is ample and generally consistent evidence from case-control and cohort studies that alcoholic drinks are a cause of pre- and post-menopausal breast cancer. In order to exclude a possible bias in the risk estimation of dietary behaviour and breast cancer risk we generated a second set of dietary patterns excluding the alcoholic beverages from the factor analysis. The risk-increasing effect of the Fatty Diet remained slightly less pronounced when alcoholic beverages were singularized and added separately into the risk estimation model. On the contrary, the *Mchicha* Diet and the Banana Diet were no longer associated with increased breast cancer risk. The latter is characterized by rapidly absorbable carbohydrates which have been found associated with an increased breast cancer risk (Lajous *et al.* 2008; Anderson & Badzioch 1993). If the alcoholic beverages did influence the risk estimation of the *Mchicha* and Banana Diet, the analysis keeping alcoholic beverages as separate food groups should visualize an increased risk association. However, the odds ratio of *Mbege* as well as bottled beer and wine was estimated to be 1.00 (all 95 % CIs 1.00 – 1.00, *P* = 0.08 and 0.87 respectively) indicating no change in risk. In our study the alcohol consumption was 8.2 g/d which is well below the recommended maximum intake of one drink per day in the

European code against cancer (Boyle *et al.* 2003). Thus, the alcohol intake in general was probably too low to show an effect.

3.4.5 Dietary pattern characterized by fruits

The Fruity Diet identified in the second PCA - keeping alcoholic beverages separate - was also associated with an increased breast cancer risk. This diet is characterized by a high consumption of fish, mango, papaya, avocados and watery fruits like oranges, watermelons and pineapples which are known for their high content of mono- and polyunsaturated fatty acids, vitamins, and micronutrients considered as potentially protective against cancer (Boffetta *et al.* 2010; Kim *et al.* 2009; Zhang *et al.* 2009; Malin *et al.* 2003; Longnecker *et al.*1997). Nevertheless, several other studies could not show an overall association between fruit and vegetable intake and breast cancer risk (Smith-Warner et al. 2001; van Gils et al. 2005).

The Fruity Diet was, like the Fatty Diet, inversely associated with the P/S ratio (P_{trend} <0.001). This was shown by dividing the study population into four groups according to the affiliation to a dietary pattern. Comparing the lowest to the highest quartile the change of the P/S ratio was caused by a reduced intake of polyunsaturated fatty acids (P_{trend} <0.001) but a stable saturated fatty acid intake (P_{trend} = 0.19). Thus, the accompanying dietary fat consumption might be the underlying cause for the risk effect associated with the Fruity Diet (Figure 7, page 49).

3.4.6 Body mass index

Estimated body mass index (BMI) data may be questioned, but the Spearman correlation between the calculated and estimated BMI at the time of the interview was in our study 0.75 (p <0.001 two-tailed), and 0.4 (p <0.001 two-tailed) for BMI at 20 years, which was considered acceptable. A recent validation study conducted by Keshtkar *et al.* (2010) in a cohort study in Iran supports this approach. They used the same method to estimate the BMI of 8,863 females showing that the body image pictogram is a valid instrument for discriminating normal and obese individuals.

Zhu *et al.* (2005) found the BMI at diagnosis (reference date) and the magnitude of change over time were associated with an increased risk of breast cancer among pre- and postmenopausal African- American women, but not the BMI at a young age (18 years). Renehan and colleagues (2008) identified a risk reduction effect of a high BMI among North American, European, and Australian premenopausal women but an increased risk among Asia-Pacific women. An increased risk from a high BMI was associated for postmenopausal women in all three regions. In our study a high BMI at the age of 20 increased the risk whereas the BMI at interview and magnitude of BMI - change over time had no effect (data not shown).

Some studies indicate that parameters behind the BMI might be relevant. Irigaray and colleagues (2007) indicated that adipose tissue acts as a reservoir for lipophilic environmental substances. These stored carcinogens can be released into the blood circulation inducing tumours and organochlorine exposure might be the woman's risk burden (Gatto et al. 2007). There is a high chance that the study population was exposed to pesticides without any protection since farming was their main income source, especially the production of coffee, sunflower seeds, and maize (Ngowi *et al.* 2007). Whether this is relevant to the increased risk of breast cancer and a high BMI at age of 20 deserves further research.

Another hypothesis is that being overweight or obese as measured by a high BMI induces higher blood oestrogen levels, promoting the proliferation of breast cancer cells (Key *et al*. 2001).

The risk estimations changed after stratifying for BMI (see chapter 3.3.10, page 52). In the BMI group 24-26 kg/m² all previously identified risk associations changed to no risk associations but menopausal status. This confirmed that age is an independent risk factor. Hence, the stratification for BMI also confirmed the risk potential of a Fatty Diet as identified in this study, especially for women with a BMI above 26 kg/m². In this group the risk estimations for BMI at age 20 changed from a significant increased risk effect to a non significant level. Yet BMI at interview became risk effective which is consistent with studies from Western countries (see above). Obese women are at higher risk to suffer from insulin resistance and possibly hyperinsulinemia which has been proposed as independent breast cancer risk factor (Gunter *et al.* 2008). However it was observed that the BMI change over time was much higher in the BMI group >26 kg/m² compared to the lower BMI groups. Whether this influenced the risk estimation on BMI deserves further investigations.

A change in the risk estimations after stratification for BMI categories were also found by Sonestedt and colleagues (2007) who primarily suggested the stratification on BMI. They argue that obesity status influences self-reported past food habits, especially its changes. Interpretation of risk relationships and study results will be facilitated if a sensitivity analysis through stratification is performed.

Although the quality criteria of the logistic model, Cox&Snell and Nagelkerke R^2, were highest if stratified for BMI groups, the risk estimations might be misleading. The high confidence interval and the statistical quality measures in the lower BMI groups indicated that these estimations remained at an explorative level only.

3.4.7 Property level

Reduced breast cancer risk was also associated with increasing property levels. At first sight this seems at odds with the statement that higher education and socioeconomic status are associated with an increased risk resulting from the lower number of parities and lactation. But in this study population, a "low property level" means the families own at maximum a bicycle looking at the property index used. If they would have had a radio in addition they would already belong to the group with a "medium property level". In low-income populations a small increase in property is primarily used to improve basic living conditions such as housing, nutrition, and access to health services (Kirkpatrick & Tarasuk 2003; Hoddinott & Yohannes 2002; Benus *et al.* 1976). Thus, it can be expected that the risk associations will change once the socio-economic situation improves to levels in high-income countries which is associated with younger age at menarche, low lifelong lactation, low number of parities and older childbearing age, thus, increasing breast cancer risk.

3.4.8 Hunger periods

Hunger threats are experienced very differently. While hunger is sometimes experienced as lack of commonly eaten staple foods, although there are other staple foods available, may hunger also be understood as inadequate intake of energy, macro- and micronutrients resulting in a low BMI. In order to distinguish between these diverse perceptions hunger was assessed as hunger resulting in reduced body size. A moderate change in body size due to hunger after menarche might be associated with a risk-decreasing effect but a huge change in body size might be associated with increased breast cancer risk. These findings are similar to other studies. Energy restriction in overweight and obese women may affect the hormonal and secretory profile of adipose tissue which may reduce cancer risk (Harvie *et al.* 2006). Whereas, a famine as experienced in the Netherlands in 1944/45 has been associated with increased risk in severely affected women (Elias *et al.* 2004). They proposed that the endocrine system may adapt to the severe energy restriction but responds inadequately to the period thereafter.

3.4.9 Physical activity

Industrialization and urbanisation are associated with increased breast cancer risk (Sitas *et al.* 2008; WCRF 2007; Okobia *et al.* 2006). Alongside this development people become more sedentary. Findings from the U.S. National Health and Nutrition Examination Survey 2003-2006 confirmed the statement of the WCRF 2007 report that sedentary ways of life are a cause of breast cancer (Lynch *et al.* 2011; WCRF 2007). This study could not confirm a relationship between physical activity and breast cancer. This might be due to the homogenous PAL values of cases and controls. Whether this is a true or false result cannot be analysed because a validation of the physical assessment like in other studies was not conducted; possibly resulting in an assessment bias (Irwin *et al*. 2011, 2008; Orsini *et al.* 2008).

3.4.10 Strength and limitations

The knowledge about breast cancer and breast self-examination is very poor in our study population, also observed by Okobia *et al.* (2006). Facilities for cancer diagnosis and treatment are rare in Tanzania. Even if breast cancer is diagnosed, many patients are not informed properly and do not know the full meaning of their diagnosis (Valentine 2010). Therefore, we had to exclude from the analysis the assessed data about breast cancer in the family history in order to avoid bias.

In the absence of health insurance, women with breast cancer needed some financial means to get access to health facilities. In order to minimize confounding errors due to different livelihood systems between cases and controls, we decided to select the controls also from within the hospital setting. Thereby, other selection biases cannot be excluded. Finally, most of the information was collected retrospectively, implying the risk of recall bias.

These data show the impact of reproductive and lifestyle factors on breast cancer aetiology of women in the Kilimanjaro Region. With regard to eating habits and dietary patterns, the diversity of the Kilimanjaro diet is low and it was less likely to miss important foods on the FFQ food list reducing the estimation bias for dietary behaviour. Due to low education levels and the poor infrastructure we do not expect socially desirable answers and participants are less likely to be informed about possible dietary impacts on health outcomes. The semi-quantitative FFQ allowed identifying non-consumers and frequent consumers on the basis of eating habits, nutrient and energy intake in the case and control groups. Also ready-to-use meals and eating out are uncommon in the Kilimanjaro region, which facilitated the identification of food groups based on single food items and less on complex meals.

A major limitation is the PCA method. There have been discussions that the PCA method is less suitable for risk estimations of dietary patterns, because of difficulties to find plausible linkages between dietary patterns and the observed disease (Hoffmann *et al.* 2004). Therefore, it was recommended using reduced rank regression based on response variables. However breast cancer develops over a long period of time. Thus, using response variables - such as biochemical parameters - is only possible in prospective studies.

Within a case-control study the possibilities for collecting retrospective data like birth weight are limited especially in a low income country like Tanzania where children are born at home and less likely to be registered (Ministry of Justice and Constitutional Affairs & The United Republic of Tanzania 2005). Thus, the present case-control study began to collect data on life experiences and lifestyle with age at menarche only missing out the possible linkages between diet of the mother during pregnancy and birth weight on breast cancer risk among premenopausal women (see page 24).

The total sample is relatively small compared to studies in high-income countries. Therefore we checked the sampling adequacy using the Kaiser-Meyer-Olkin measure (KMO) before running a Principal Component Analysis (PCA). The KMO which was 0.619 is considered as mediocre in our case (Field 2009; Kaiser 1970). A low KMO might result in a high unexplained variance. However, in this study we extracted 6 factors explaining 40.3 % variance which is a medium result compared to other studies e.g.: Hu *et al.* (1999) = 2 factors: 20 %, Arkkola *et al.* (2007) = 7 factors: 29.5 %, Shi *et al.* (2008)= 4 factors: 28.5 %, Lau *et al.* (2007) = 2 factors: 17.1 % . In addition, Bartlett's Test for Sphericity was 2599.25, $P < 0.001$ indicating that correlations between items were sufficiently large for a PCA.

Cox&Snell's and Nagelkerke's R^2 serve as quality criteria for logistic regression models. Whereas Cox&Snell's R^2 describes the contrast of likelihood estimations weighed by the sample size, estimates Nagelkerke's R^2 the proportion of the variance of the dependent variable through the independent variable. Values of ≥0.2 are considered as acceptable and values of ≥0.4 as good (Backhaus 2008, p270). An acceptable level was estimated for the basic model stratified for menopausal status and for the basic model plus dietary patterns. The values improved after stratification for BMI group.

The sample size required to achieve a high level of power in a logistic regression depends on the number of predictors and the size of the expected effect. Several studies showed that a sample size as used in this study (n = 345) allows to detect large and medium effects but might miss small effect. This study did not focus on small effects but aimed to understand whether globally acknowledged predictors are applicable to low income countries, thus, the sample size was considered acceptable.

3.4.11 Conclusion

Established risk factors for breast cancer in women of high-income populations are young age at menarche, late age at first child, short lactation, and being physically inactive. All of these factors seem to be valid in low-income countries as well. The traditional lifestyle of women in the studied Kilimanjaro Region of Tanzania is risk-reducing, nonetheless, breast cancer occurs.

Dietary patterns rich in fat and characterised by a low P/S ratio may be associated with a higher risk of breast cancer especially in women with a BMI above 26kg/m^2. The fatty acid composition is probably more important than total fat intake for the breast cancer risk.

With ongoing lifestyle changes, protective factors are decreasing, so an increase in breast cancer incidence is to be expected. Important public health measures to keep breast cancer incidence low include preventing women from becoming overweight in adolescence, and maintaining breastfeeding practices for up to two years per child. In addition dietary recommendations should consider the possible negative impact of a low P/S ratio. An important goal would be to provide access to adequate health-care services for early detection and management of breast cancer.

4 Summary

Breast cancer is the leading cause of death among women worldwide and the second most common cancer among women in the Kilimanjaro Region of Tanzania. Studies in industrialised countries identified age at menarche, age at first full-term pregnancy, and lactation as determining factors in the aetiology of breast cancer.

Apart for alcohol, no scientific agreement could be found on whether specific nutrients or dietary patterns have an impact on breast cancer risk.

Food frequency questionnaires are the primary method for measuring dietary intake in epidemiological studies to assess long-term dietary intake. Based on a food list the study population is asked to remember over a certain period of time how often they eat each listed food item. However, climate, culture, food availability, education and tradition influence individual food selection. Therefore it is necessary to adopt the food list to the local consumer habits.

The objective of this study was finally twofold: At first, a food frequency questionnaire had to be validated. Secondly, a case-control study was conducted with a focus on dietary patterns and breast cancer. This was done based on the hypotheses that nutrients are ingested within diets rather than as single nutrients. In addition it was assumed that in a community with a low risk lifestyle and reproductive behaviour like in Tanzania it is more likely to detect linkages between dietary intake and breast cancer than in communities of high income countries.

The food frequency questionnaire which had already been tested in a pilot study was validated between July 2005 and February 2006. In absence of a “gold standard” the 24h-recall method was chosen as validation tool. The 24h-recall is known for being less sensitive for recall bias but has a high day to day variability. Two data assessments were conducted in different agricultural seasons. The first data assessment included 78 women of whom 50 women completed the second round, i.e. 10-15 % of the expected study population of the breast cancer study which had been required for the validation study. The women were self selected by four interviewers based on the selection criteria of the breast cancer study at the Kilimanjaro Christian Medical Centre and in rural and urban Moshi. Their age ranged between 23 years and 70 years

After a positive test for interviewer effects the analysis of the FFQ data continued stratified for interviewer with a total of 46 data sets. Wilcoxon signed rank test were performed to test for seasonal effects as well as for differences between FFQ and 24h-recall.

In the case-control study 115 female breast cancer patients (cases) and 230 age- and district-matched women clinically free from breast cancer (controls) were interviewed about their reproductive history and socioeconomic condition. Semi-structured interviews including anthropometric measurements were conducted by trained enumerators. The validated semi-quantitative Food Frequency Questionnaire was used to assess the dietary intake. A logistic regression was performed to estimate breast cancer risk. Dietary patterns were obtained using principal component analysis with Varimax rotation and added as additional predictors to the basic logistic model.

Except for water there was no evidence for a seasonal effect if the FFQ is used. Differences in the intake of oils and fats between FFQ and 24h-recall had only been shown by one interviewer. However, the women had shown more difficulties in estimating their oil and fat consumption during the 24h-recall than during the FFQ. This can be seen looking at the Spearman correlation coefficients which were calculated to estimate the association between both instruments and to be able to compare the results with other studies. The coefficients were at a low and modest level and ranged from -0.2 (oils and fats) to 0.4 (cereals and fruits).

The median age of the women in the case-control study was 50 years (min/max 26 to 85 years). Estimated median BMI at age 20 was 21kg/m² in both cases and controls. Median lifelong lactation of the mothers was 96 months (cases) and 108 months (controls). A high BMI at 20 years was associated with an increased breast cancer risk (OR 1.31 95 % CI 1.11–1.55, *P* <0.01). The odds ratio for lifelong lactation was slightly below one (OR 0.99 95 % CI 0.98–1.00, *P* < 0.01). There was no significant association in risk for BMI at interview (median 25 kg/m² of cases and 26 kg/m² of controls), age at menarche (median 16 years), and age at first full-term pregnancy (median 20 years). The association of increased risk with higher BMI at age 20 years remained significant after stratification for menopause (premenopausal: OR 1.41 95 % CI 1.10–1.81, *P* = 0.01; postmenopausal: OR 1.38 95 % CI 1.06-1.80, *P* = 0.02). Late age at menarche and prolonged lifelong lactation were associated with a risk reduction among premenopausal women ($OR_{menarche}$ 0.74 95 % CI 0.56-1.00, *P* = 0.05; $OR_{lactation}$ 0.98 95 % CI 0.97-0.99, *P* <0.01). The adjusted logistic regression estimated an increased risk for a “Fatty Diet”, characterized by a higher consumption of milk, vegetable oils and fats, butter, lard, and red meat (OR = 1.42, 95 % CI 1.08-1.87; *P* = 0.01), and for a “Fruity Diet” characterized by a higher consumption of fish, mango, papaya, avocado, and watery fruits (OR = 1.61, 95 % CI 1.14-2.28; *P* = 0.01). Both diets showed an inverse association with the ratio between polyunsaturated and saturated fatty acids (P/S ratio).

The risk estimations changed after splitting the study population into three BMI groups (<24 kg/m², 24 – 26 kg/m² and >26 kg/m²). In the lowest BMI group risk estimations showed large confidence intervals and indicated that the group might be too small for reliable estimations. In the BMI group 24-26 kg/m² all previously identified risk associations changed to no risk associations but menopausal status. The increased risk association for Fatty Diet remained significantly in the BMI group >26 kg/m² only.

The FFQ was considered to be a reliable instrument for assessing the dietary intake in the Kilimanjaro Region. However, it was recommended to pay special attention to the training of the interviewers and there especially to the assessment of the fat and oil intake.

Long-standing lactation and reproductive behaviour are associated with a lower breast cancer risk in the region. As current changes in lifestyle affect age at menarche, reproductive behaviour, and nutritional status, an increased incidence of breast cancer is to be expected. Preventive efforts should include advice on reproductive and breastfeeding behaviour. A diet characterised by a low P/S ratio seems to be more important for the development of breast cancer than total fat intake.

5 Zusammenfassung

Brustkrebs ist einer der Haupttodesursachen weltweit und die zweithäufigste Todesursache bei Frauen in der Kilimanjaro Region Tansanias. Studien aus Industrieländern haben das Menarchealter, Alter am Ende der ersten ausgetragenen Schwangerschaft und lebenslange Laktationsdauer als wichtige Einflussfaktoren für die Entstehung von Brustkrebs identifiziert. Außer bei Alkohol gibt es kein Einvernehmen darüber, ob spezifische Nährstoffe oder Ernährungsmuster das Brustkrebsrisiko beeinflussen.

In epidemiologischen Studien werden am Häufigsten Ernährungshäufigkeitsfragebögen zur Erfassung der langfristigen Nahrungsaufnahme eingesetzt. Die Studienteilnehmerinnen werden anhand einer Liste befragt, wie häufig und in welchen Mengen sie welche Nahrungsmittel gewöhnlich essen. Die individuelle Nahrungsmittelauswahl wird jedoch von Klima, Kultur, Nahrungsverfügbarkeit, Bildung und Tradition beeinflusst. Die im Ernährungshäufigkeitsfragebogen verwendete Nahrungsmittelliste muss daher an die lokalen Konsumgewohnheiten angepasst werden.

Das Ziel dieser Studie war daher zweiteilig: Zunächst wurde ein validierter Ernährungshäufigkeitsfragebogen benötigt. Anschließend wurde eine Fall-Kontroll Studie mit dem Schwerpunkt Ernährungsmuster und Brustkrebs durchgeführt. Diese basierte auf der Hypothese, dass Nährstoffe nicht einzeln, sondern im Rahmen von Mahlzeiten aufgenommen werden. Außerdem wurde angenommen, dass ein mit geringen Brustkrebsrisiko verbundender Lebensstil und einem Reproduktionsverhalten wie in Tansania es eher ermöglicht mögliche Verbindungen zwischen Ernährungsverhalten und Brustkrebs aufzuzeigen als in Ländern mit hohem Einkommen.

Der bereits im Rahmen einer Pilotstudie getestete Ernährungshäufigkeitsfragebogen wurde zwischen Juli 2005 und Februar 2006 validiert. In Abwesenheit eines „Goldenen Standards" wurde der 24h Recall als Methode für die Validierung ausgewählt. Der 24h Recall ist bekannt dafür, dass er bezüglich Erinnerungslücken weniger anfällig ist als der Ernährungshäufigkeitsfragebogen, dafür aber mehr Tagesschwankungen unterliegt. Die Datenerhebung erfolgte in zwei unterschiedlichen landwirtschaftlichen Saisons. Die erste Erhebung schloss 78 Frauen ein von denen 50 die zweite Runde abschlossen. Dies entspricht der für die Validierung notwendigen 10-15 % der in der Brustkrebsstudie zu erwartenden Studienpopulation.

Die Frauen wurden von vier Interviewern auf Basis der Einschlusskriterien der Brustkrebsstudie rund ums Kilimanjaro Medical Center und in den beiden Landkreisen Moshi Land und Stadt selbst ausgewählt. Das Alter der Frauen schwankte zwischen 23 und 70 Jahren.

Nach einem positiven Test auf Interviewereffekte wurde die Analyse der Daten aus dem Ernährungshäufigkeitsfragebogen stratifiziert nach Interviewern mit 46 Fällen fortgesetzt. Mittels des *Wilcoxon signed rank* Tests wurde auf saisonale Effekte und Unterschiede zwischen dem Ernährungshäufigkeitsfragebogen und dem 24h Recall getestet.

In Rahmen der Fall-Kontroll Studie wurden 115 Brustkrebspatientinnen (Fälle) und 230 brustkrebsgesunde Frauen (Kontrollen), die über Alter und Wohnort mit den Fällen gepaart waren, über ihr Reproduktionsverhalten und sozioökonomische Situation befragt. Die halbstrukturierte Befragungen schlossen anthropometrischen Messungen mit ein und wurden von ausgebildeten Interviewern durchgeführt. Der validierte Ernährungshäufigkeitsfragebogen wurde zur Erfassung der Nahrungsaufnahme eingesetzt. Das Brustkrebsrisiko wurde mittels logistischer Regression geschätzt. Anhand einer Hauptkomponentenanalyse wurden Ernährungsmuster identifiziert, die im Weiteren als zusätzliche Risikofaktoren das logistische Grundmodell erweitert haben.

Unter Verwendungen des Ernährungshäufigkeitsfragebogens gab keinen Beleg für einen saisonalen Effekt außer für Wasser. Unterschiede in der Öl- und Fettaufnahme zwischen dem Ernährungshäufigkeitsfragebogen und dem 24h Recall waren nur bei einem Interviewer zu sehen. Jedoch hatten die Frauen unter Verwendung des 24h Recalls mehr Schwierigkeiten ihren Öl- und Fettverzehr abzuschätzen als beim Ernährungshäufigkeitsfragebogen. Dies wird bei der Betrachtung der Spearman Korrelation Koeffizienten deutlich, die berechnet wurden, um einerseits die Assoziation zwischen den beiden Instrumenten abzuschätzen, und um andererseits die Ergebnisse besser mit denen anderer Studien vergleichen zu können. Die Koeffizienten lagen auf niedrigem bis mittlerem Niveau und schwankten zwischen -0,2 (Öl und Fette) und 0,4 (Getreide und Obst).

Das Alter der Frauen in der Fall-Kontroll Studie war im Median 50 Jahre (min/max 26 bis 85 Jahre). Der geschätzte BMI im Alter von 20 Jahren war bei den Fälle und den Kontrollen im Median 21 kg/m². Bei den Fällen lag die lebenslange Laktationsdauer der Mütter im Median bei 96 Monaten und bei den Kontrollen bei 108 Monaten. Ein hoher BMI im Alter von 20 Jahren war mit einem erhöhten Brustkrebsrisiko verbunden (OR 1.31 95 % CI 1.11–1.55, P <0.01). Das Odds Ratio für lebenslange Laktationsdauer lag knapp unter 1 (OR 0.99 95 % CI 0.98–1.00, P < 0.01). Keine Risikoassoziation wurde für BMI zum Zeitpunkt des Interviews (Median 25 kg/m² bei den Fällen und 26 kg/m² bei den Kontrollen), Menarchealter (Median 16 Jahre) und Alter zum Ende der ersten ausgetragenen Schwangerschaft gefunden (Median 20 Jahre). Ein hoher BMI im Alter von 20 Jahren blieb nach der Stratifzierung nach Menopause signifikant mit einem erhöhten Brustkrebsrisiko assoziiert (prämenopausal: OR 1.41 95 % CI 1.10–1.81, P = 0.01; postmenopausal: OR 1.38 95 % CI 1.06-1.80, P = 0.02). Spätes Menarchealter und eine hohe Laktationsdauer wurden bei prämenopausalen Frauen mit einem verminderten Brustkrebsrisiko assoziiert ($OR_{menarche}$ 0.74 95 % CI 0.56-1.00, P = 0.05; $OR_{lactation}$ 0.98 95 % CI 0.97-0.99, P <0.01). Das adjustierte logistische Regressionsmodel schätzte ein erhöhtes Risiko für ein „fettiges Ernährungsmuster“ (OR = 1.42, 95 % CI 1.08-1.87; P = 0.01), gekennzeichnet durch einen hohen Verzehr von Milch, pflanzlichen Ölen und Fetten, Butter, Schmalz und rotes Fleisch und für ein „obstreiches Ernährungsmuster“ (OR = 1.61, 95 % CI 1.14-2.28; P = 0.01), gekennzeichnet durch einen hohen Verzehr an Fisch, Mango, Papaya, Avocado und wässrige Früchte. Beide Ernährungsmuster sind entgegengesetzt assoziiert mit dem Quotienten zwischen mehrfach ungesättigten und gesättigten Fettsäuren (P/S-Quotient).

Die Ergebnisse der Risikoschätzungen änderten sich nachdem die Studienpopulation in verschiedene BMI Gruppen aufgeteilt wurde (<24 kg/m², 24 – 26 kg/m² und >26 kg/m²). In der Gruppe mit einem BMI <24 kg/m² wiesen die Schätzungen hohe Konfidenzintervalle auf und deuteten darauf hin, dass die Gruppe zu klein gewesen sein könnte, um verlässliche Aussagen machen zu können. In der BMI Gruppe 24 – 26 kg/m² wechselten alle bisherigen signifikanten Risikoassoziationen zu nicht signifikant mit Ausnahme des Menopausenstatus. Das fettige Ernährungsmuster blieb nur in der BMI Gruppe >26 kg/m² signifikant mit einem steigenden Risiko assoziiert.

Der verwendete Ernährungshäufigkeitsfragebogen wird als verlässliches Instrument zur Erfassung der Ernährungsaufnahme in der Kilimanjaro Region angesehen. Besondere Aufmerksamkeit sollte jedoch auf die Ausbildung der Interviewer und dort insbesondere auf die Erfassung der Fett- und Ölaufnahme gelegt werden.

Die hohe lebenslange Laktationsdauer und das Reproduktionsverhalten in der Untersuchungsregion sind mit einem niedrigen Brustkrebsrisiko verknüpft. Der gegenwärtige Wandel im Lebensstil hat jedoch einen im Sinne des Brustkrebsrisikos negativen Effekt auf das Menarchealter, Reproduktionsverhalten und den Ernährungstatus; eine steigende Inzidenz ist zu erwarten. Präventionsmaßnahmen sollten eine Beratung hinsichtlich Reproduktions- und Stillverhalten einschließen. Eine Ernährung charakterisiert durch einen niedrigen P/S-Quotienten scheint für die Brustkrebsentwicklung bedeutender zu sein als die Gesamtfettaufnahmen.

6 Acknowledgements

This study would not have been possible without the women of the Kilimanjaro Region. Their willingness to participate without further incentive is greatly appreciated.

I wish to sincerely thank Professor Dr. Michael Krawinkel to accept me as a doctoral student in his working group and his continuous support during the research. His readiness to discuss the challenges during data analysis and reporting is very much appreciated. His letters requesting financial support were many times successful and allowed me to present the data in various international conferences. Special thanks go to Professor Dr. Ingrid Hoffmann for her willingness to support the research as referee requesting financial support for the validation study as well as being the second reviewer.

Special thanks also go to the research team itself: Dr. Antje Hebestreit who originally planned this study based on her results during her PhD and who did not stop supporting the finalization of this thesis; Britta Swai who carried out the Fine-Needle-Aspiration Cytology and the data entry but also contributed as enumerator as well as resource person about the livelihoods in Tanzania; Dr. Mark Swai for his contributions to the study design and implementation at the KCMC; Joel Julia Mfinanga for conducting interviews and physical examinations; Eunice Chibanda and Welima Bannister who assisted as enumerators in the validation study and Dani Swai who assisted as translator.

Many thanks go to the late Marion Mann and Prof. Dr. Heiner Boeing who assisted in planning the statistical analysis as well as to Dr. Gerrit Eichner and Dr. Johannes Herrmann who provided statistical guidance during the analysis of the dietary patterns. The latter was supported by Dr. Jürgen Erhardt who assisted in the application of the software package NutriSurvey® facilitating the data analysis of the food frequency questionnaire data which is greatly appreciated.

Furthermore I like to thank Timothy Bostick, Emma Sykes, Margaret Wightman and Richard Bliaut for the English editing.

In very good memory are the moral support and fruitful discussions with my colleagues, especially with Yujin Lee, Gudrun Keding, Sandra Habicht and Silvia Kaufmann. However, this thesis would not have been finished without the continuous support by my family and children: Dörte, Till and Kai, who many times took over my duties at home during my research activities.

Last but not least I would like to thank the former companies Henskes Ltd. in Laatzen and Plural, Sachse and Heinzelmann Ltd, as well as the German Academic Exchange Service (DAAD) for their financial support.

7 References

Agurs-Collins T, Rosenberg L, Makambi K, *et al.* (2009) Dietary patterns and breast cancer risk in women participating in the Black Women's Health Study. *American Journal of Clinical Nutrition*, 90(3), pp.621-628.

Ahlgren M, Melbye M, Wohlfahrt J, *et al.* (2004) Growth Patterns and the Risk of Breast Cancer in Women. *New England Journal of Medicine*, 351(16), pp.1619-1626.

Ahlgren M, Sørensen T, Wohlfahrt J, *et al.* (2003) Birth weight and risk of breast cancer in a cohort of 106,504 women. *International Journal of Cancer*, 107(6), pp.997-1000.

Ahrens W (2007) *Handbook of epidemiology*, corrected 2nd [ed.]. Berlin ; New York: Springer.

Ambrosini GL, de Klerk NH, Musk AW, *et al.* (2001) Agreement between a brief food frequency questionnaire and diet records using two statistical methods. *Public Health Nutrition*, 4(2), pp.255-264.

Ambrosini GL, Mackerras D, de Klerk NH, *et al.* (2003) Comparison of an Australian food-frequency questionnaire with diet records: implications for nutrition surveillance. *Public Health Nutrition*, 6(4), pp.415-422.

Anderson D & Badzioch M (1993) Familial breast cancer risks. Effects of prostate and other cancers. *Cancer*, 72(1), pp.114-119.

Anon Bonferroni Correction -- from Wolfram MathWorld. Available at: http://mathworld.wolfram.com/BonferroniCorrection.html [Accessed May 8, 2011].

Arkkola T, Uusitalo U, Kronberg-Kippilä C, *et al.* (2007) Seven distinct dietary patterns identified among pregnant Finnish women – associations with nutrient intake and sociodemographic factors. *Public Health Nutrition*, 11(02). Available at: http://www.journals.cambridge.org/abstract_S1368980007000420 [Accessed April 7, 2011].

Backhaus K (2008) *Multivariate Analysemethoden : eine anwendungsorientierte Einführung*, 12th ed. Berlin [u.a.]: Springer.

Bau AM, Ernert A, Schenk L, *et al.* (2008) Is there a further acceleration in the age at onset of menarche? A cross-sectional study in 1840 school children focusing on age and bodyweight at the onset of menarche. *European Journal of Endocrinology*, 160(1), pp.107-113.

Beaber EF, Holt VL, Malone KE, *et al.* (2008) Reproductive Factors, Age at Maximum Height, and Risk of Three Histologic Types of Breast Cancer. *Cancer Epidemiology Biomarkers & Prevention*, 17(12), pp.3427-3434.

Benus J, Kmenta J & Shapiro H (1976) The Dynamics of Household Budget Allocation to Food Expenditures. *The Review of Economics and Statistics*, 58(2), pp.129-138.

Beral V, et al. (2002) Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. *The Lancet*, 360(9328), pp.187-195.

Blackwood MA & Weber BL (1998) BRCA1 and BRCA2: from molecular genetics to clinical medicine. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 16(5), pp.1969-1977.

Boffetta P., Couto E, Wichmann J, *et al.* (2010) Fruit and Vegetable Intake and Overall Cancer Risk in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *JNCI Journal of the National Cancer Institute*, 102(8), pp.529-537.

Bohlscheid-Thomas S (1999) Überrprüfung der Reliabilität und Validität eines Fragebogens zu Ernährungsgewohnheiten für dessen Einsatz im deutschen Teil des EPIC-Projekts. Available at: https://vpn.uni-giessen.de/+CSCO+ch756767633A2F2F74726F2E6861762D747672666672612E7172+1555607686@16744448@1304934467@D532B92E30E453B93A9330AC8FD27FF4D23425B6+/geb/volltexte/2000/227/pdf/d000028.pdf [Accessed May 9, 2011].

Bougnoux P, Giraudeau B & Couet C (2006) Diet, cancer, and the lipidome. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 15(3), pp.416-421.

Boyle P, Autier P, Bartelink H, *et al.* (2003) European Code Against Cancer and scientific justification: third version (2003). *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO*, 14(7), pp.973-1005.

van den Brandt PA (2000) Pooled Analysis of Prospective Cohort Studies on Height, Weight, and Breast Cancer Risk. *American Journal of Epidemiology*, 152(6), pp.514-527.

Brennan SF, Cantwell MM, Cardwell CR, *et al.* (2010) Dietary patterns and breast cancer risk: a systematic review and meta-analysis. *American Journal of Clinical Nutrition*, 91(5), pp.1294-1302.

Brustad M, Skeie G, Braaten T, *et al.* (2003) Comparison of telephone vs face-to-face interviews in the assessment of dietary intake by the 24 h recall EPIC SOFT program—the Norwegian calibration study. *European Journal of Clinical Nutrition*, 57(1), pp.107-113.

Cade J, Thompson R, Burley V, *et al.* (2002) Development, validation and utilisation of food-frequency questionnaires – a review. *Public Health Nutrition*, 5(4), pp.567-587.

Chaiwun B & Thorner P (2007) Fine needle aspiration for evaluation of breast masses. *Current Opinion in Obstetrics & Gynecology*, 19(1), pp.48-55.

Chang-Claude J, Andrieu N, Rookus M, *et al.* (2007) Age at Menarche and Menopause and Breast Cancer Risk in the International BRCA1/2 Carrier Cohort Study. *Cancer Epidemiology Biomarkers & Prevention*, 16(4), pp.740-746.

Chen Yu, Ahsan H, Parvez F, *et al.* (2004) Validity of a food-frequency questionnaire for a large prospective cohort study in Bangladesh. *British Journal of Nutrition*, 92(05), p.851.

Cho E, Spiegelman D., Hunter DJ, *et al.* (2003) Premenopausal Fat Intake and Risk of Breast Cancer. *JNCI Journal of the National Cancer Institute*, 95(14), pp.1079-1085.

Costello AB & Osborne JW (2005) Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), pp.1–9.

Crispim SP, Ribeiro RCL, Silva MMS, *et al.* (2006) The influence of education in the validation process of a food frequency questionnaire for adults in Vicosa, Minas Gerais, Brazil. *Eur J Clin Nutr*, 60(11), pp.1311-1316.

Cureton E (1983) *Factor analysis : an applied approach.* [S.l.]: Lawrence Erlbaum.

DGEpi (2008) Guidelines and recommendations to assure good epidemiologic practice (GEP). Available at: http://www.dgepi.de/pdf/infoboard/stellungnahme/GEP_LL_english_f.pdf [Accessed June 4, 2011].

Ekbom A, Trichopoulos D, Adami HO, *et al.* (1992) Evidence of prenatal influences on breast cancer risk. *Lancet*, 340(8826), pp.1015-1018.

Elias SG, Peeters PHM, Grobbee DE, *et al.* (2004) Breast cancer risk after caloric restriction during the 1944-1945 Dutch famine. *Journal of the National Cancer Institute*, 96(7), pp.539-546.

Engeset D, Dyachenko A, Ciampi A, *et al.* (2009) Dietary patterns and risk of cancer of various sites in the Norwegian European Prospective Investigation into Cancer and Nutrition cohort: the Norwegian Women and Cancer study. *European Journal of Cancer Prevention*, 18(1), pp.69-75.

Erhardt JG (not dated) Nutrition surveys and assessment. Available at: http://www.nutrisurvey.de/ [Accessed May 11, 2011].

Evans DGR & Howell Anthony (2007) Breast cancer risk-assessment models. *Breast Cancer Research*, 9(5), p.213.

Faggiano F, Vineis P, Cravanzola D, *et al.* (1992) Validation of a method for the estimation of food portion size. , 3(4), pp.379-82.

Field A (2009) *Discovering statistics using SPSS : (and sex and drugs and rock "n" roll)*, 3rd ed. London: SAGE.

Fioretti F, Tavani A, Bosetti C, *et al.* (1999) Risk factors for breast cancer in nulliparous women. *British Journal of Cancer*, 79(11/12), pp.1923-1928.

Fung TT, Willett Walter C, Stampfer Meir J, *et al.* (2001) Dietary Patterns and the Risk of Coronary Heart Disease in Women. *Arch Intern Med*, 161(15), pp.1857-1862.

Gajalakshmi V, Mathew A, Brennan P, *et al.* (2009) Breastfeeding and breast cancer risk in India: A multicenter case-control study. *International Journal of Cancer*, 125(3), pp.662-665.

Gatto NM, Longnecker Matthew P., Press Michael F., *et al.* (2007) Serum organochlorines and breast cancer: a case–control study among African-American women. *Cancer Causes & Control*, 18(1), pp.29-39.

German Nutrition Society.;Austrian Nutrition Society.;Swiss Society for Nutrition Research.;Swiss Nutrition Association. (2002) *Reference values for nutrient intake*, 1st ed. Frankfurt/Main: Umschau Braus.

Gibney M, Margetts BM & Kearney JM (2004) *Public health nutrition* Oxford UK ;Ames Iowa: Blackwell Science.

Gibson R (2005) *Principles of nutritional assessment*, 2nd ed. Oxford: Oxford University Press.

Gibson RS & Ferguson E (2008) An interactive 24-hour recall for assessing the adequacy of iron and zinc intakes in developing countries. Available at: http://www.ifpri.org/sites/default/files/publications/tech08.pdf [Accessed November 12, 2011].

van Gils CH, Peeters PHM, Bueno-de-Mesquita H Bas, *et al.* (2005) Consumption of vegetables and fruits and risk of breast cancer. *JAMA: The Journal of the American Medical Association*, 293(2), pp.183-193.

Gluckman P & Hanson M (2006) Evolution, development and timing of puberty. *Trends in Endocrinology & Metabolism*, 17(1), pp.7-12.

Goodstine SL, Zheng Tongzhang, Holford TR, *et al.* (2003) Dietary (n-3)/(n-6) fatty acid ratio: possible relationship to premenopausal but not postmenopausal breast cancer risk in U.S. women. *The Journal of Nutrition*, 133(5), pp.1409-1414.

Gunter MJ, Hoover DR, Yu H, *et al.* (2008) Insulin, Insulin-Like Growth Factor-I, and Risk of Breast Cancer in Postmenopausal Women. *JNCI Journal of the National Cancer Institute*, 101(1), pp.48-60.

Hall JN, Moore S, Harper SB, *et al.* (2009) Global variability in fruit and vegetable consumption. *American Journal of Preventive Medicine*, 36(5), pp.402-409.e5.

Handa K & Kreiger N (2007) Diet patterns and the risk of renal cell carcinoma. *Public Health Nutrition*, 5(06). Available at: http://www.journals.cambridge.org/abstract_S1368980002001027 [Accessed May 1, 2011].

Harvie M & Howell A (2006) Energy balance adiposity and breast cancer - energy restriction strategies for breast cancer prevention. *Obesity Reviews: An Official*

Journal of the International Association for the Study of Obesity, 7(1), pp.33-47.

Hebestreit A (2004) Brustkrebs und Ernährung. Inaugural-Dissertation zur Erlangung des Grades eines Doktor oec.troph. im Fachbereich “Argarwissenschaften, Ökotrophologie und Umweltmanagement” der Justus-Liebig-Universität Gießen. Available at: http://geb.uni-giessen.de/geb/volltexte/2004/1544/pdf/HebestreitAntje-2004-05-17.pdf [Accessed January 12, 2011].

Hoddinott J & Yohannes Y (2002) *Dietary diversity as a food security indicator* Washington DC: IFPRI. Available at: http://ideas.repec.org/p/fpr/fcndbr/136.html#download [Accessed May 9, 2011].

Hoffmann K., Schulze M. B, Schienkiewitz A, *et al.* (2004) Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *American journal of epidemiology*, 159(10), p.935.

Holmes MD, Hunter DJ, Colditz GA, *et al.* (1999) Association of dietary intake of fat and fatty acids with risk of breast cancer. *JAMA: The Journal of the American Medical Association*, 281(10), pp.914-920.

Horn JL (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), pp.179–185.

Hu F B, Rimm E, Smith-Warner SA, *et al.* (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *The American Journal of Clinical Nutrition*, 69(2), pp.243-249.

Hu Frank B, Rimm EB, Stampfer Meir J, *et al.* (2000) Prospective study of major dietary patterns and risk of coronary heart disease in men. *The American Journal of Clinical Nutrition*, 72(4), pp.912-921.

Hunter DJ, Spiegelman D, Adami HO, *et al.* (1996) Cohort studies of fat intake and the risk of breast cancer--a pooled analysis. *The New England Journal of Medicine*, 334(6), pp.356-361.

IARC & WHO (2008) Global cancer statistics: Globocan 2008. Available at: http://www.iarc.fr/en/media-centre/iarcnews/2010/globocan2008.php [Accessed June 24, 2010].

Irigaray P, Newby J, Lacomme S, *et al.* (2007) Overweight/obesity and cancer genesis: More than a biological link. *Biomedicine & Pharmacotherapy*, 61(10), pp.665-678.

Irwin ML, McTiernan A, Manson Joann E, *et al.* (2011) Physical Activity and Survival in Postmenopausal Women with Breast Cancer: Results from the Women’s Health Initiative. *Cancer Prevention Research (Philadelphia, Pa.)*, 4(4), pp.522-529.

Irwin ML, Smith AW, McTiernan A, *et al.* (2008) Influence of pre- and postdiagnosis physical activity on mortality in breast cancer survivors: the health, eating, activity, and lifestyle study. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 26(24), pp.3958-3964.

Johannes CB, Crawford SL & McKinlay JB (1997) Interviewer effects in a cohort study. *American journal of epidemiology*, 146(5), p.429.

Jordan I, Hebestreit Antje, Swai B, *et al.* (2010) Breast cancer risk among women with long-standing lactation and reproductive parameters at low risk level: a case–control study in Northern Tanzania. *Breast Cancer Research and Treatment*. Available at: http://www.springerlink.com/index/10.1007/s10549-010-1255-7 [Accessed April 30, 2011].

Kabagambe EK, Baylin A, Allan DA, *et al.* (2001) Application of the method of triads to evaluate the performance of food frequency questionnaires and biomarkers as indicators of long-term dietary intake. *American Journal of Epidemiology*, 154(12), pp.1126-1135.

Kaiser HF (1970) A second generation little jiffy. *Psychometrika*, 35, pp.401-415.

Keshtkar AA, Semnani S, Pourshams A, *et al.* (2010) Pictogram use was validated for estimating individual's body mass index. *Journal of Clinical Epidemiology*, 63(6), pp.655-659.

Key T J, Verkasalo PK & Banks E (2001) Epidemiology of breast cancer. *The Lancet Oncology*, 2(3), pp.133-140.

Key T, Appleby P, Barnes I, *et al.* (2002) Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *Journal of the National Cancer Institute*, 94(8), pp.606-616.

Key T J, Appleby PN, Reeves GK, *et al.* (2003) Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. *Journal of the National Cancer Institute*, 95(16), pp.1218-1226.

Kim J, Lim S-Y, Shin A, *et al.* (2009) Fatty fish and fish omega-3 fatty acid intakes decrease the breast cancer risk: a case-control study. *BMC Cancer*, 9(1), p.216.

Kim MK, Sasaki S, Otani T, *et al.* (2005) Dietary patterns and subsequent colorectal cancer risk by subsite: A prospective cohort study. *International Journal of Cancer*, 115(5), pp.790-798.

Kim MK, Sasaki S, Sasazuki S, *et al.* (2004) Prospective study of three major dietary patterns and risk of gastric cancer in Japan. *International Journal of Cancer*, 110(3), pp.435-442.

Kirkpatrick S & Tarasuk V (2003) The relationship between low income and household food expenditure patterns in Canada. *Public Health Nutrition*, 6(6), pp.589-597.

Kroke A, Klipstein-Grobusch K, Voss S, *et al.* (1999) Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. *The American Journal of Clinical Nutrition*, 70(4), pp.439-447.

Lagiou P, Lagiou A, Samoli E, *et al.* (2006) Diet during pregnancy and levels of maternal pregnancy hormones in relation to the risk of breast cancer in the offspring. *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation (ECP)*, 15(1), pp.20-26.

Lajous M, Boutron-Ruault M-C, Fabre A, *et al.* (2008) Carbohydrate intake, glycemic index, glycemic load, and risk of postmenopausal breast cancer in a prospective study of French women. *The American Journal of Clinical Nutrition*, 87(5), pp.1384-1391.

Landig J, Erhardt JG, Bode JC, *et al.* (1998) Validation and comparison of two computerized methods of obtaining a diet history. *Clinical Nutrition (Edinburgh, Scotland)*, 17(3), pp.113-117.

Lau C, Glümer C, Toft U, *et al.* (2007) Identification and reproducibility of dietary patterns in a Danish cohort: the Inter99 study. *British Journal of Nutrition*, 99(05). Available at: http://www.journals.cambridge.org/abstract_S0007114507837494 [Accessed April 7, 2011].

Lee H, Gourley L, Duffy S, *et al.* (1991) Dietary effects on breast-cancer risk in Singapore. *Lancet*, 337(8751), pp.1197-1200.

Li CI, Malone KE, Daling JR, *et al.* (2007) Timing of Menarche and First Full-Term Birth in Relation to Breast Cancer Risk. *American Journal of Epidemiology*, 167(2), pp.230-239.

Longnecker M P, Newcomb PA, Mittendorf R, *et al.* (1997) Intake of carrots, spinach, and supplements containing vitamin A in relation to risk of breast cancer. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 6(11), pp.887-892.

Lynch BM, Friedenreich CM, Winkler EAH, *et al.* (2011) Associations of objectively assessed physical activity and sedentary time with biomarkers of breast cancer risk in postmenopausal women: findings from NHANES (2003-2006). *Breast Cancer Research and Treatment*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21553294 [Accessed May 16, 2011].

MacIntyre U, Venter C & Vorster H (2001) A culture-sensitive quantitative food frequency questionnaire used in an African population: 2. Relative validation by 7-day weighed records and biomarkers. *Public Health Nutrition*, 4(01). Available at: http://www.journals.cambridge.org/abstract_S136898000100009X [Accessed May 8, 2011].

MacLean CH, Newberry SJ, Mojica WA, *et al.* (2006) Effects of omega-3 fatty acids on cancer risk: a systematic review. *JAMA: The Journal of the American Medical Association*, 295(4), pp.403-415.

Mahabir S, Baer DJ, Giffen C, *et al.* (2006) Calorie intake misreporting by diet record and food frequency questionnaire compared to doubly labeled water among postmenopausal women. *European Journal of Clinical Nutrition*, 60(4), pp.561-565.

Malin AS, Qi D, Shu X-O, *et al.* (2003) Intake of fruits, vegetables and selected micronutrients in relation to the risk of breast cancer. *International Journal of Cancer*, 105(3), pp.413-418.

Marfell-Jones M, Olds T, Stewart A, *et al.* (2006) *International standards for anthropometric assessment* Potchefstroom, South Africa: ISAK.

Margetts BM & Nelson M (2006) *Design concepts in nutritional epidemiology*, 2nd ed. Oxford [u.a.]: Oxford Univ. Press.

Mattisson I, Wirfält E, Andrén C, *et al.* (2007) Dietary fat intake – food sources and dietary correlates in the Malmö Diet and Cancer cohort. *Public Health Nutrition*, 6(06). Available at: http://www.journals.cambridge.org/abstract_S1368980003000764 [Accessed April 7, 2011].

McCann SE, Weiner J, Graham Saxon, *et al.* (2007) Is principal components analysis necessary to characterise dietary behaviour in studies of diet and disease? *Public Health Nutrition*, 4(04). Available at: http://www.journals.cambridge.org/abstract_S1368980001000891 [Accessed April 7, 2011].

McCredie M, Paul C, Skegg DC, *et al.* (1998) Reproductive factors and breast cancer in New Zealand. *International Journal of Cancer. Journal International Du Cancer*, 76(2), pp.182-188.

McNaughton SA, Hughes MC & Marks GC (2007) Validation of a FFQ to estimate the intake of PUFA using plasma phospholipid fatty acids and weighed foods records. *The British Journal of Nutrition*, 97(3), pp.561-568.

McPherson K, Steel CM & Dixon JM (2000) ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ (Clinical Research Ed.)*, 321(7261), pp.624-628.

Mesa JM, Martínez J, Araújo C, *et al.* (2010) Growth patterns in early childhood and the onset of menarche before age twelve. *Revista De Saúde Pública*, 44(2), pp.249-260.

Mina K, Fritschi L & Knuiman M (2007) A valid semiquantitative food frequency questionnaire to measure fish consumption. *European Journal of Clinical Nutrition*, 61(8), pp.1023-1031.

Minami Y, Tsubono Y, Nishino Y, *et al.* (2004) The increase of female breast cancer incidence in Japan: emergence of birth cohort effect. *International Journal of Cancer. Journal International Du Cancer*, 108(6), pp.901-906.

Ministry of Justice and Constitutional Affairs & The United Republic of Tanzania (2005) Improving Services to Protect the Rights of the Most Vulnerable Chilcren. A country paper for presentation at the Eastern and Southern Africa

Confernce. Available at: http://unstats.un.org/unsd/vitalstatkb/Attachment320.aspx [Accessed May 16, 2011].

Murff HJ, Shu X-O, Li H, *et al.* (2011) Dietary polyunsaturated fatty acids and breast cancer risk in Chinese women: A prospective cohort study. *International Journal of Cancer*, 128(6), pp.1434-1441.

National Bureau of Statistics (2006) 2002 POPULATION AND HOUSING CENSUS- General Report. Available at: http://www.tanzania.go.tz/census/census/kilimanjaro.htm [Accessed February 9, 2006].

Nelson HD (2008) Menopause. *The Lancet*, 371(9614), pp.760-770.

Ngowi AVF, Mbise TJ, Ijani ASM, *et al.* (2007) Smallholder vegetable farmers in Northern Tanzania: Pesticides use practices, perceptions, cost and health effects. *Crop Protection*, 26(11), pp.1617-1624.

Ofuya ZM (2008) The age at menarche in Nigerian adolescents from two different socioeconomic classes. *Online Journal of Health and Allied Sciences*, 6(4).

Okobia M, Bunker C, Zmuda J, *et al.* (2006) Case-control study of risk factors for breast cancer in Nigerian women. *International Journal of Cancer. Journal International Du Cancer*, 119(9), pp.2179-2185.

Okobia MN, Bunker CH, Okonofua FE, *et al.* (2006) Knowledge, attitude and practice of Nigerian women towards breast cancer: a cross-sectional study. *World Journal of Surgical Oncology*, 4, p.11.

Okobia MN, Bunker CH, Okonofua FE, *et al.* (2006) Knowledge, attitude and practice of Nigerian women towards breast cancer: a cross-sectional study. *World Journal of Surgical Oncology*, 4, p.11.

Olopade OI, Fackenthal JD, Dunston G, *et al.* (2003) Breast cancer genetics in African Americans. *Cancer*, 97(1 Suppl), pp.236-245.

Orsini N, Mantzoros CS & Wolk A (2008) Association of physical activity with cancer incidence, mortality, and survival: a population-based study of men. *British Journal of Cancer*, 98(11), pp.1864-1869.

de Paoli M (2004) To breastfeed or not to breastfeed? Infant feeding dilemmas facing women with HIV in the Kilimanjaro Region, Tanzania.

Parent A-S, Teilmann G, Juul A, *et al.* (2003) The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocrine Reviews*, 24(5), pp.668-693.

Parkin DM, Sitas F, Chirenje M, *et al.* (2008) Part I: Cancer in Indigenous Africans--burden, distribution, and trends. *The Lancet Oncology*, 9(7), pp.683-692.

Parr CL, Barikmo I, Torheim LE, *et al.* (2002) Validation of the second version of a quantitative food-frequency questionnaire for use in Western Mali. *Public Health Nutrition*, 5(6), pp.769-781.

Patton GC & Viner R (2007) Pubertal transitions in health. *The Lancet*, 369(9567), pp.1130-1139.

Pellegrini N, Salvatore S, Valtueña S, *et al.* (2007) Development and validation of a food frequency questionnaire for the assessment of dietary total antioxidant capacity. *The Journal of Nutrition*, 137(1), pp.93-98.

Poslusna K, Ruprich J, de Vries JHM, *et al.* (2009) Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *British Journal of Nutrition*, 101, p.S73.

Prentice RL, Thomson C. A., Caan B, *et al.* (2007) Low-Fat Dietary Pattern and Cancer Incidence in the Women's Health Initiative Dietary Modification Randomized Controlled Trial. *JNCI Journal of the National Cancer Institute*, 99(20), pp.1534-1543.

Promislow JHE, Gladen BC & Sandler DP (2005) Maternal recall of breastfeeding duration by elderly women. *American Journal of Epidemiology*, 161(3), pp.289-296.

Rebacz E (2009) Age at menarche in schoolgirls from Tanzania in light of socioeconomic and sociodemographic conditioning. *Collegium Antropologicum*, 33(1), pp.23-29.

Rose DP & Connolly JM (1999) Omega-3 fatty acids as cancer chemopreventive agents. *Pharmacology & Therapeutics*, 83(3), pp.217-244.

Saadatian-Elahi M, Norat T, Goudable J, *et al.* (2004) Biomarkers of dietary fatty acid intake and the risk of breast cancer: A meta-analysis. *International Journal of Cancer*, 111(4), pp.584-591.

dos Santos Silva I, De Stavola BL, Mann V, *et al.* (2002) Prenatal factors, childhood growth trajectories and age at menarche. *International Journal of Epidemiology*, 31(2), pp.405-412.

Sawka MN, Cheuvront SN & Carter R (2005) Human Water Needs. *Nutrition Reviews*, 63(6), pp.30-39.

Schulz M, Hoffmann Kurt, Weikert C, *et al.* (2008) Identification of a dietary pattern characterized by high-fat food choices associated with increased risk of breast cancer: the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. *The British Journal of Nutrition*, 100(5), pp.942-946.

Shi Z, Hu X, Yuan B, *et al.* (2008) Vegetable-rich food pattern is related to obesity in China. *International Journal of Obesity (2005)*, 32(6), pp.975-984.

Sieri Sabina, Krogh Vittorio, Ferrari Pietro, *et al.* (2008) Dietary fat and breast cancer risk in the European Prospective Investigation into Cancer and Nutrition. *The American Journal of Clinical Nutrition*, 88(5), pp.1304-1312.

Singleton R (2010) *Approaches to social research*, 5th ed. New York: Oxford University Press.

Sitas F, Parkin DM, Chirenje M, *et al.* (2008) Part II: Cancer in Indigenous Africans--causes and control. *The Lancet Oncology*, 9(8), pp.786-795.

Smith-Warner SA, Spiegelman D, Yaun SS, *et al.* (2001) Intake of fruits and vegetables and risk of breast cancer: a pooled analysis of cohort studies. *JAMA: The Journal of the American Medical Association*, 285(6), pp.769-776.

Sonestedt E, Gullberg B & Wirfält E (2007) Both food habit change in the past and obesity status may influence the association between dietary factors and postmenopausal breast cancer. *Public Health Nutrition*, 10(08). Available at: http://www.journals.cambridge.org/abstract_S1368980007246646 [Accessed April 7, 2011].

Stunkard A, Sorensen T & Schulsinger F (1983) Use of a Danish adoption register for the study of obesity and thinness. In *The genetics of neurological and psychiatric disorders*. Research publications (Association for Research in Nervous and Mental Disease). New York: Raven Press, pp. p115-20.

Tanzanian National Bureau of Statistics (2004) Tanzania Demographic Health Survey. In Dar es Salaam, Tanzania, pp. pp177-202. Available at: http://www.measuredhs.com/countries/metadata.cfm?surv_id=253&ctry_id=39&SrvyTp=ctry&cn=Tanzania [Accessed May 9, 2011].

Thiébaut ACM, Kipnis V, Chang S-C, *et al.* (2007) Dietary fat and postmenopausal invasive breast cancer in the National Institutes of Health-AARP Diet and Health Study cohort. *Journal of the National Cancer Institute*, 99(6), pp.451-462.

Torheim LE, Barikmo I, Hatløy A, *et al.* (2001) Validation of a quantitative food-frequency questionnaire for use in Western Mali. *Public Health Nutrition*, 4(06). Available at: http://www.journals.cambridge.org/abstract_S1368980001001161 [Accessed May 8, 2011].

TPSF (2011) Tanzania Private Sector foundation. Available at: http://www.tpsftz.org/mapinfo.php?region=13 [Accessed June 4, 2011].

United Nations D of E and SA (2011) World Population Prospects, the 2010 Revision. Available at: http://esa.un.org/unpd/wpp/Excel-Data/population.htm [Accessed June 4, 2011].

Valentine G (2010) Knowledge and attitudes towards breast cancer among women in Rombo district, Kilimanjaro region. *Dar Es Salaam Medical Students' Journal*, 14(2), pp.44–47.

La Vecchia C, Negri E, Bruzzi P, *et al.* (1992) The role of age at menarche and at menopause on breast cancer risk: combined evidence from four case-control studies. *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO*, 3(8), pp.625-629.

Wadsworth M E J, Hardy RJ, Paul AA, *et al.* (2002) Leg and trunk length at 43 years in relation to childhood health, diet and family circumstances; evidence from the 1946 national birth cohort. *International Journal of Epidemiology*, 31(2), pp.383-390.

Wakai K, Dillon DS, Ohno Y, *et al.* (2000) Fat intake and breast cancer risk in an area where fat intake is low: a case-control study in Indonesia. *International Journal of Epidemiology*, 29(1), pp.20-28.

WCRF (2007) *Food, nutrition, physical activity, and the prevention of cancer : a global perspective* Washington DC: World Cancer Research Fund, American Institute for Cancer Research.

Weinberger K & Swai I (2006) Consumption of Traditional Vegetables in Central and Northeastern Tanzania. *Ecology of Food and Nutrition*, 45(2), pp.87-103.

WHO (2009a) Factsheet No 29 February 2009: Cancer. Available at: http://www.who.int/mediacentre/factsheets/fs297/en/print.html [Accessed July 15, 2009].

WHO (1995) *Physical status : the use and interpretation of anthropometry : report of a WHO Expert Committee.* Geneva: World Health Organization.

WHO (2009b) World Health Organization. Available at: http://www.who.int/whoisi/whostat/EN_WHS09_Full.pdf [Accessed September 19, 2009].

Willett W (1998) *Nutritional epidemiology*, 2nd ed. New York: Oxford University Press.

Wynder EL, Cohen LA, Muscat JE, *et al.* (1997) Breast cancer: weighing the evidence for a promoting role of dietary fat. *Journal of the National Cancer Institute*, 89(11), pp.766-775.

Yang L & Jacobsen KH (2008) A Systematic Review of the Association between Breastfeeding and Breast Cancer. *Journal of Women's Health*, 17(10), pp.1635-1645.

Zhang C-X, Ho SC, Chen Y-M, *et al.* (2009) Greater vegetable and fruit intake is associated with a lower risk of breast cancer among Chinese women. *International Journal of Cancer*, 125(1), pp.181-188.

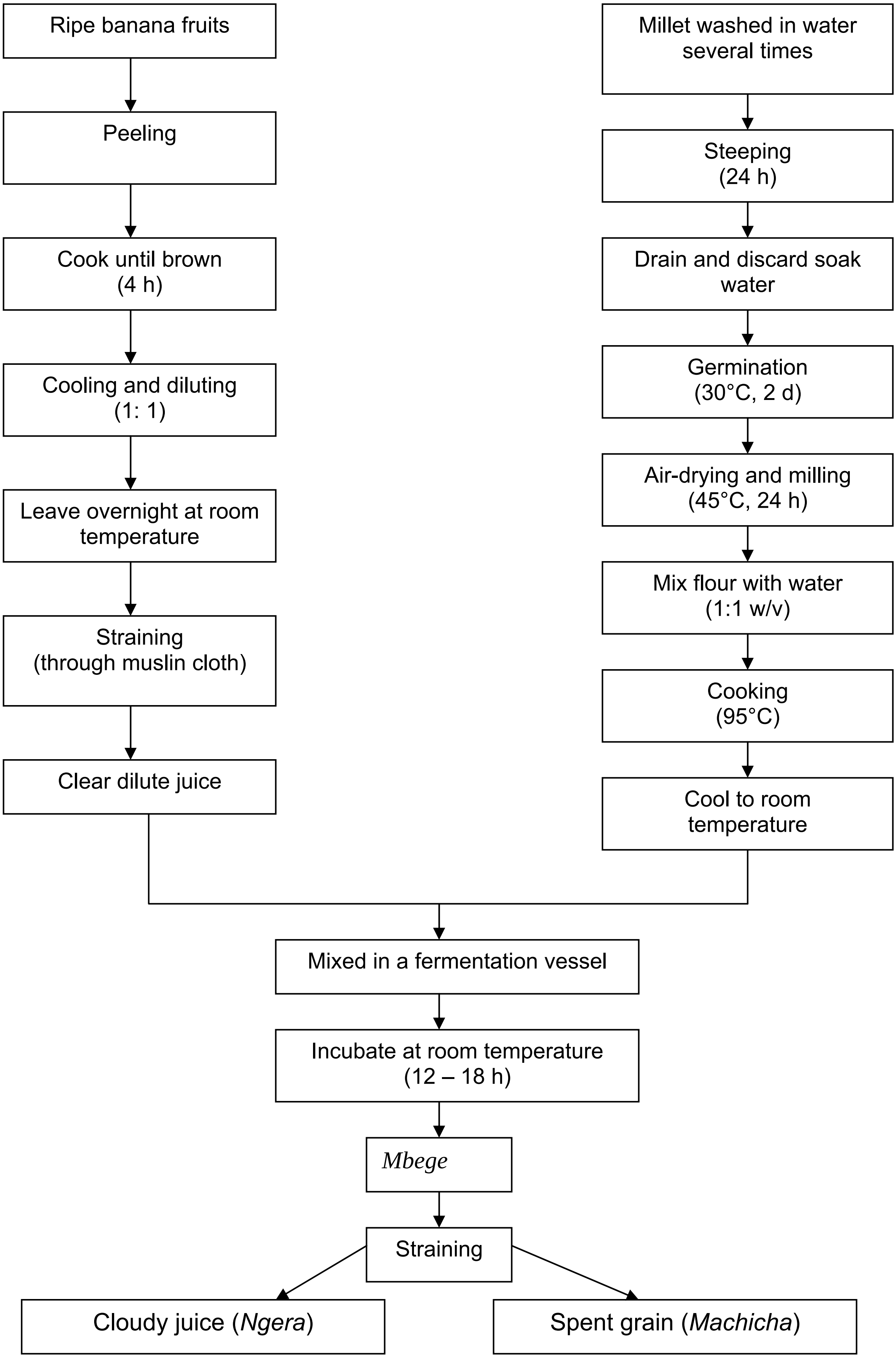
Zheng T, Duan L, Liu Y, *et al.* (2000) Lactation reduces breast cancer risk in Shandong Province, China. *American Journal of Epidemiology*, 152(12), pp.1129-1135.

Zhu K, Caulfield J, Hunter S, *et al.* (2005) Body mass index and breast cancer risk in African American women. *Annals of Epidemiology*, 15(2), pp.123-128.

8 Annex

8.1 Brewing process of *Mbege* 92
8.2 Questionnaire case-control study 93
8.3 Tables of Figures 99
8.4 Complete versions of the tables reported in the thesis 105
8.5 Publication 117
8.6 Publication under Review 118

8.1 Brewing process of *Mbege* (Hebestreit 2004)



8.2 Questionnaire of case-control study

Breast cancer case (BCa) Hospital control (HCo) Questionnaire Code
Visiting control (VCo) Population control (PCo) ________________

General questions

1 **Date of survey (day, month, year e.g. 1-1-2004)....[SURVDATE]** _____

2 **Hospital Reg. No..[Hospreg]................** _____

3 **Family Name..** _____

4 **Other Names..** _____

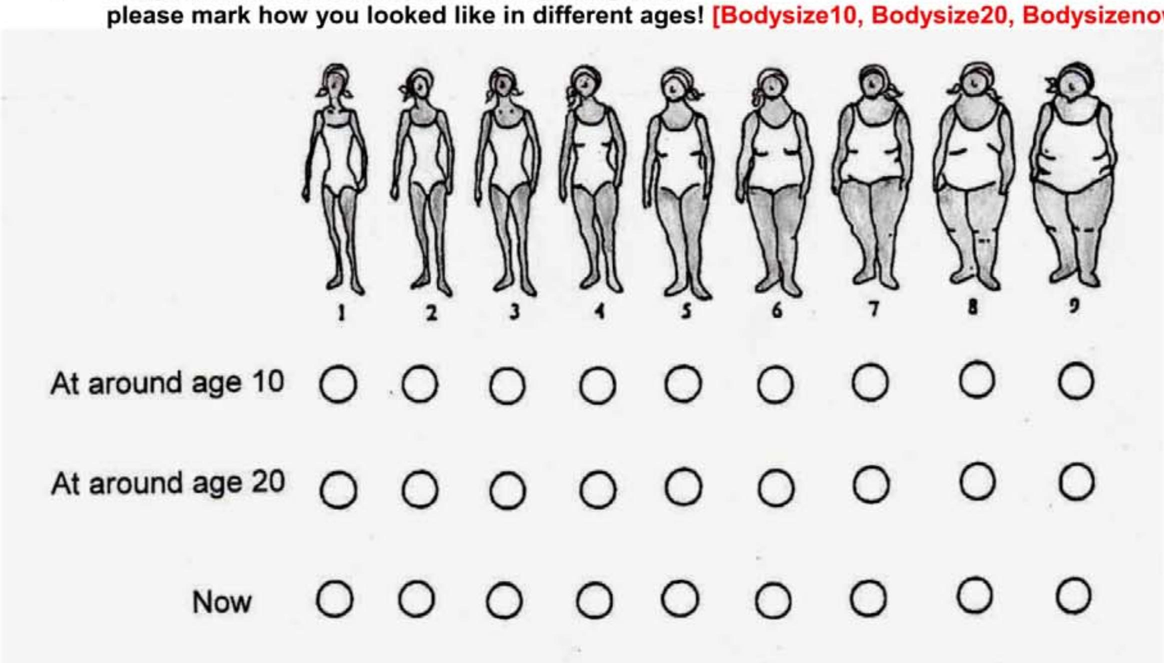
5 **How old are you?..[AGE].....** _____years

6 **Height in cm..[HEIGHT]** _____cm

7 **Body weight in kg...[Weight]...** _____kg

[BMIcalcul] _____ **BMI**

8 **Body size in different periods of life (pictogram):**
please mark how you looked like in different ages! [Bodysize10, Bodysize20, Bodysizenow]



9 **Tribe...[Tribe]...............................** _____

1] Chagga 2] Pare
3] Maasai 4] Meru
5] Arusha 77] Other

10 **Race...[Race].............................** _____

1] African 2] Asian 3] European 77] Other

1

Breast cancer case (BCa) Hospital control (HCo) Questionnaire Code

Visiting control (VCo) Population control (PCo) ________________

11 District...[District]......................... _____

1] Hai 2] Moshi urban

3] Moshi rural 4] Rombo

5] Arusha rural 6] Arusha urban 77] Other

12 How long do you live there?...................[Length]............................... _____

13 Did a close family member have cancer?......[Fammemb]......................... _____

1] yes 2] no 88] Don't know

14 What kind of cancer/ cause of death?..........[Kindcancer].......................... _____

1] Mamma 77] Other 88] Don't know

15 Have you ever had a breast cancer biopsy for lumps in the breast? .[Biopsy]... _____

1] yes 2] no 88] don't know

16 Major Signs

1] Weight loss over 10% of body weight
2] Unexplained chronic diarrhoea for more than one month
3] Unexplained prolonged fever for more than one month

In case we have more questions, can we contact you again?
Where will we find you?
Adress/ Balozi:

17 Minor Signs

1] Persistent cough 2] Generalized pruritic dermatitis
3] Recurrent zoster 4] Oropharyngeal candidiasis
5] Chronic progressive herpes virus infection
6] Generalized lymphadenopathy

18 WHO Conclusion [WHOconcl] _____

1] positive 2] negative

Reproductive History

19 Age at menarche?..........................[Agemenar].. _____

20 Do you still menstruate?................[Menopause].. _____

1] yes 2] no 3] pregnant now

21 When did your periods stop completely? (Age).....[Stopmeno]................ _____

22 Do you have children? ....[Children].. _____

1] yes - continue with the table

2] no - go to question 23

Breast cancer case (BCa) Hospital control (HCo) Questionnaire Code

Visiting control (VCo) Population control (PCo) ________________

Pregnancy	**What was your age when your first, second, third,...child was born?**	**How did the pregnancy end?** 1] baby was born alive → 2] abortion 3] baby was born dead	If baby was born alive: Did you breastfeed the baby? 1] yes 2] no	**If yes:** How many months did you breastfeed your baby? (66=still breastfeeding)
First pregnancy	**[Agechild1]Years**	**[pregnend1]**		**[Breastfeed1]**
Second pregnancy	**[Agechild2]Years**	**[pregnend2]**		**[Breastfeed2]**
Third pregnancy	**[Agechild3]Years**	**[pregnend3]**		**[Breastfeed3]**
Fourth pregnancy	**[Agechild4]Years**	**[pregnend4]**		**[Breastfeed4]**
Fifth pregnancy	**Years**			
Sixth pregnancy	**Years**			
Seventh pregnancy	**Years**			

23 Did / do you take contraceptives?.............[Contraceptives].................................... _____

1] yes 2] no 99] No answer (Tabelle: Zahl der verschiedenen Hormonellen Methoden)

24 Which contraceptives do you take? (note the months / years)........... _____

1] injection ....months / ....years [Typecontro_1 bis _3]

2] tablets ....months / ....years [howlong_1 bis _3]

3] implants or plaster ....months / ....years

4] Condoms

5] IUD

6] Natural (calendar method, interruption)

77] Other 99] No answer

Socioeconomic Questions

25 How many persons live in the household?........[HSHMEMO]............................. _____

Breast cancer case (BCa) Hospital control (HCo) Questionnaire Code

Visiting control (VCo) Population control (PCo) ________________

26 What schooling do you have?............[Educmoth]....................................... _____

1] Nil, illiterate 2] < 3 years

3] 3 years - finishing primary school 4] Secondary school education

5] University or other higher education

27 What is your main occupation?..........[Occup1 + 2]....................................... _____

1] No occupation 2] Housewife

3] Domestic servant 4] Industrial worker

5] Farmer 6] Sister/ Clergy

7] On daily wages 8] Dealer, trader or salesperson

9] Civil servant 10] Retired 77] Other

28 Which of these objects does your family has at home?...[Objecthouse_1 bis_5].. _____

1] Radio 2] TV 3] Bike 4] Motorcycle 5] Car

29 Try to remember the amount of time you spent in each activity on a <u>normal</u> day?

____

[Lightwork – Mediumwork – Heavywork]

____

____

____

____

____

30 How much do you smoke/do you use snuff? ......[tobacco]................................ _____

1] I don't smoke / use snuff 2] 1-5 cigarettes or snuff / day

3] 6-10 4] 11-15

5] 16-20 6] >20

31 Can you remember whether you were affected by hunger before your first menarche? ..___

1] yes - which year?________ for how long?_______ [hungbef – hyearbef - hdurbef]

did your body go visibly smaller? a) no b) moderate c) a lot [Hsizebef]

2] no - go to question 32

32 Were you affected by chronical hunger throughout your life?.[Hungchr]................. ______

1] yes – which period?________ how long was the period?_______ [Hungtime]

did your body go visibly smaller? a) no b) moderate c) a lot [hungsize]

2] no

4

Breast cancer case (BCa) Hospital control (HCo) Questionnaire Code

Visiting control (VCo) Population control (PCo) ________________

Food Fequency: What do you usually eat and drink in which amounts?

Item / amount ml	**daily**	**weekly**	**monthly**	**AMOUNT**
Tea (+ milk / + sugar) [m.M.597/ o.M.600]				
Coffee (+ milk / + sugar) [mM586/ oM 589]				
Milk **[469]**				
Soda **[578]**				
Water **[572]**				
Juice **[147]**				
Mbege **[1049]**				
Beer (bottle) **[601]**				
Konyagi / Spirits **[626]**				
Wine **[610]**				

	daily	**weekly**	**monthly**	**Piece**	**Hand**	**Amount**	**Spoon**
Meat							
chicken (kuku) [838]							
beef (n'gombe) [800]							
pork (nguruwe) [812]							
mutton (mbuzi) [818]							
Cereals							
maize on cob + Makande [911]							
Ugali [42]							
Uji from maize [42]							
millet + Uji [905]							
rice [46]							
chapati [18]							
bread [8]							
mandazi [123]							
Vegetables							
potato (sweet) [417]							
Potatoe (irish) [411]							
cassava + majimbi [416/931]							
Mtori (how much meat?) [mt]							
Mchicha [mc]							
Carrot [332]							
Cabbage [289]							
Onion [303]							
Cucumber [308]							
Tomatoes [314]							
Okra (bamia) [1224]							
Chinese [279]							

5

Breast cancer case (BCa) Hospital control (HCo) Questionnaire Code

Visiting control (VCo) Population control (PCo) ________________

	daily	weekly	monthly	Piece	Hand	Amount	Spoon
Pulses							
Beans [396]							
Lentils [393]							
Peas (njegere) [353]							
Soybeans [351]							
Mung beans (choroko) [941]							
Other							
Fruits							
bananas, ripe [s: 956; b: 954]							
bananas, unripe [955]							
Mangoes ☼ [232]							
Oranges ☼ [237]							
Melons ☼ [228]							
Avocado ☼ [218]							
Papaya ☼ [234]							
Pineapple ☼ [216]							
Other items							
Eggs [132]							
Dagaa [731]							
other type of fish (ocean/ lake?)							
Extra sugar + honey [682/1044]							
Margarine [644]							
Butter [648]							
coconut milk [372]							
cashew nuts (korosho)[373]							
ground nuts (karanga) [368]							
sunflower oil [638]							
mixed vegetable oil [631]							
Other oil – which?							
Kimbo [1038]							
Lard [652]							
Something we forgot?							
banana stew from hotel [1405]							
Please check whether the ingredients of the meals are mentioned above							

8.3 Tables of Figures

Table 1: Number of interviews per year of data assessment
Table to Figure 4, p32

Year	Total	Controls	Cases
1	20	12	8
2	137	90	47
3	188	128	60

Table 2: Initial Eigenvalues for scree plot of PCA1

Table to Figure 5, p36

	PCA1 Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	3.976	11.046	11.046
2	2.922	8.118	19.164
3	2.164	6.011	25.174
4	1.687	4.687	29.862
5	1.676	4.656	34.518
6	1.616	4.490	39.008
7	1.498	4.160	43.168
8	1.351	3.753	46.921
9	1.324	3.678	50.599
10	1.202	3.339	53.937
11	1.177	3.269	57.206
12	1.150	3.195	60.401
13	1.068	2.966	63.367
14	1.021	2.837	66.204
15	0.942	2.615	68.819
16	0.898	2.493	71.313
17	0.854	2.373	73.686
18	0.823	2.285	75.971
19	0.731	2.031	78.003
20	0.696	1.933	79.936
21	0.668	1.855	81.791

Table 3: Initial Eigenvalues of PCA 2 and Random Variable

Table to Figure 6, p39

	PCA 2 initial Eigenvalue			Random Variable initial Eigenvalue		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.970	11.678	11.678	1.591	4.680	4.680
2	2.698	7.934	19.612	1.558	4.582	9.262
3	2.144	6.307	25.919	1.521	4.472	13.734
4	1.672	4.918	30.837	1.450	4.266	18.000
5	1.646	4.840	35.677	1.390	4.087	22.087
6	1.579	4.645	40.322	1.381	4.060	26.148
7	1.450	4.264	44.586	1.311	3.855	30.002
8	1.336	3.929	48.515	1.270	3.736	33.738
9	1.317	3.872	52.388	1.242	3.652	37.390
10	1.190	3.501	55.888	1.191	3.504	40.893
11	1.132	3.330	59.218	1.189	3.498	44.391
12	1.095	3.221	62.439	1.102	3.241	47.632
13	1.014	2.981	65.420	1.086	3.195	50.827
14	0.968	2.847	68.267	1.044	3.071	53.898
15	0.886	2.607	70.874	1.039	3.057	56.954
16	0.880	2.588	73.462	0.994	2.925	59.879

Table 4: Intake of fat, polyunsaturated and saturated fatty acids per day in quartiles of Fatty Diet (PCA1), Table to Figure 7, p48

Quartile of the Fatty Diet (PCA1)			Total fat	PUFA	Saturated fatty acids
1st	N	Valid	86	86	86
		Missing	0	0	0
	Mean		78.94	31.56	18.16
	Std. Error of Mean		2.64	1.09	0.74
	Median		76.24	29.30	18.03
	Std. Deviation		24.49	10.07	6.83
	Variance		599.83	101.39	46.65
	Range		109.84	69.18	30.85
	Minimum		41.75	8.77	6.85
	Maximum		151.59	77.95	37.70
2nd	N	Valid	87	87	87
		Missing	0	0	0
	Mean		71.76	26.21	17.88
	Std. Error of Mean		2.29	0.82	0.71
	Median		68.25	26.42	16.76
	Std. Deviation		21.36	7.66	6.65
	Variance		456.20	58.67	44.23
	Range		105.79	47.57	33.67
	Minimum		33.85	5.01	6.90
	Maximum		139.64	52.58	40.57
3rd	N	Valid	86	86	86
		Missing	0	0	0
	Mean		72.17	19.61	18.84
	Std. Error of Mean		2.26	1.06	0.74
	Median		69.07	20.37	17.89
	Std. Deviation		20.99	9.81	6.90
	Variance		440.52	96.19	47.55
	Range		105.45	40.31	30.53
	Minimum		30.41	4.86	6.54
	Maximum		135.86	45.17	37.07
4th	N	Valid	86	86	86
		Missing	0	0	0
	Mean		80.47	13.43	23.05
	Std. Error of Mean		2.59	0.78	0.90
	Median		77.01	10.14	21.75
	Std. Deviation		24.00	7.25	8.36
	Variance		575.86	52.49	69.93
	Range		126.63	34.80	42.73
	Minimum		39.09	4.13	9.87
	Maximum		165.72	38.93	52.60

Table 5: Intake of fat, polyunsaturated and saturated fatty acids per day in quartiles of Fruity Diet (PCA2), Table to Figure 8, p50

Quartile of the Fruity Diet (PCA2)			Total fat	PUFA	Saturated fatty acids
1st	N	Valid	86	86	86
		Missing	0	0	0
	Mean		77.57	26.91	20.88
	Median		74.06	26.56	20.06
	Range		105.79	48.93	45.37
	Minimum		33.85	4.13	7.23
	Maximum		139.64	53.06	52.60
	Std. Error of Mean		2.53	0.91	0.93
	Std. Deviation		23.50	8.44	8.61
	Variance		552.46	71.15	74.10
2nd	N	Valid	86	86	86
		Missing	0	0	0
	Mean		75.29	25.35	19.29
	Median		74.04	25.77	19.43
	Range		104.47	42.17	33.72
	Minimum		30.41	6.82	6.54
	Maximum		134.88	48.99	40.26
	Std. Error of Mean		2.05	1.04	0.70
	Std. Deviation		19.00	9.65	6.53
	Variance		361.09	93.20	42.69
3rd	N	Valid	86	86	86
		Missing	0	0	0
	Mean		71.40	21.58	18.12
	Median		68.20	22.39	17.30
	Range		116.21	73.65	27.42
	Minimum		35.38	4.30	7.87
	Maximum		151.59	77.95	35.29
	Std. Error of Mean		2.15	1.26	0.68
	Std. Deviation		19.98	11.68	6.29
	Variance		399.33	136.54	39.58
4th	N	Valid	86	86	86
		Missing	0	0	0
	Mean		79.30	17.07	19.70
	Median		71.95	12.80	18.35
	Range		131.34	50.89	40.62
	Minimum		34.38	4.86	6.90
	Maximum		165.72	55.75	47.52
	Std. Error of Mean		3.02	1.27	0.87
	Std. Deviation		28.02	11.78	8.11
	Variance		785.16	138.71	65.82

Table 6: BMI change over time: BMI at interview minus BMI age 20 years

Table to Figure 9, p53

BMI group	N Valid	N Missing	BMI change (kg/m²) Mean	Median	SD	Variance
<24kg/m²	92	2	0.478	0.950	2.6906	7.239
24 - 26kg/m²	92	0	4.392	4.600	1.4140	1.999
>26kg/m²	154	1	7.147	6.950	2.8634	8.199

Table 7: Prevalence of hunger before the menarche and effect on body size

Table to Figure 10, p58

Hunger before menarche		Frequency	Percent	Valid Percent
Valid	Yes	15	4.3	5.4
	No	265	76.8	94.6
	Total	280	81.2	100
Missing	System	65	18.8	
Total		345	100	

Effect on Body size		Frequency	Percent	Valid Percent
Valid	No	3	0.9	20
	Moderate	11	3.2	73.3
	Huge	1	0.3	6.7
	Total	15	4.3	100
Missing	System	330	95.7	
Total		345	100	

Table 8: Prevalence of hunger after menarche and effect on body size

Table to Figure 11, p58

Hunger after menarche		Frequency	Percent	Valid Percent
Valid	Yes	87	25.2	31.2
	No	192	55.7	68.8
	Total	279	80.9	100
Missing	System	66	19.1	
Total		345	100	

Effect on body size		Frequency	Percent	Valid Percent
Valid	No	26	7.5	29.9
	Moderate	51	14.8	58.6
	Huge	10	2.9	11.5
	Total	87	25.2	100
Missing	System	258	74.8	
Total		345	100	

8.4 Complete versions of selected tables reported in the thesis

Add. Table 4: Differences between food frequency questionnaire and 24-recall: results of the Wilcoxon signed rank test (*p*-value) * and median g/day, standard deviation (SD) and inter-quartile range (IQ) stratified by interviewer

		Interviewer 1 (n=13)	Interviewer 2 (n=23)	Interviewer 3 (n=10)
Cereals	*P*-value	0.34	0.83	0.01
	median	30	-17	225
	SD	295.6	481.5	248.6
	IQ	298	669	231
Vegetables	*P*-value	0.07	0.197	0.03
	median	-149	124	-282.5
	SD	285.2	392.3	334.2
	IQ	292	407	641
Animal products	*P*-value	0.698	0.28	0.57
	median	-14	61	18
	SD	209.2	189.1	263.98
	IQ	265	266	317
Beverages	*P*-value	0.64	0.11	0.85
	median	-2	155	-149
	SD	660.2	645	533.9
	IQ	697	540	626
Fruits	*P*-value	0.89	0.02	0.77
	median	-17	299	-47
	SD	239.98	459.7	453.6
	IQ	262	625	456
Oils and fats	*P*-value	**0.0007**	0.22	0.79
	median	-36	1	0
	SD	23.3	27.8	16.9
	IQ	23	18	20

* P-value >0.003 = no statistical difference of the results of FFQ and 24h-recall

Add. Table 9: Energy and nutrient intake per day based on a food frequency

			Energy	Protein	Fat	Carbo-hydrates	% Energy of protein	% Energy of fat	% Energy of carbohydrates	PUFA	saturated fatty acids	alcohol
			Kcal/day	g/day			%			g/day		
Cases	N	Valid	115	115	115	115	115	115	115	115	115	115
		Missing	0	0	0	0	0	0	0	0	0	0
	Mean		1913.5	58.1	78.4	211.8	12.5	37.5	46.1	20.6	19.8	11.4
	Std. Error of Mean		55.4	2.6	2.5	7.0	0.3	0.7	0.7	1.1	0.8	1.4
	Median		1764.3	50.3	71.9	193.6	12.0	37.0	46.0	18.7	18.4	8.2
	Std. Deviation		593.6	27.6	26.6	75.1	3.4	7.7	7.1	11.8	8.2	15.2
	Range		3141.8	166.3	131.3	402.0	26.0	45.0	41.0	50.7	45.7	99.7
	Minimum		786.1	16.6	34.4	103.8	6.0	14.0	26.0	5.0	6.9	0.0
	Maximum		3927.9	182.9	165.7	505.7	32.0	59.0	67.0	55.8	52.6	99.7
Controls	N	Valid	230	230	230	230	230	230	230	230	230	230
		Missing	0	0	0	0	0	0	0	0	0	0
	Mean		1721.2	48.0	74.5	190.8	11.6	39.1	45.9	23.7	19.3	8.3
	Std. Error of Mean		25.7	0.9	1.4	3.5	0.1	0.4	0.4	0.7	0.5	0.6
	Median		1687.3	46.3	71.8	184.8	12.0	39.0	46.0	25.1	18.8	6.8
	Std. Deviation		390.5	13.6	20.9	52.4	1.8	6.6	6.1	10.6	7.1	8.4
	Range		2625.6	85.7	121.2	452.9	13.0	44.0	38.0	73.8	34.8	79.2
	Minimum		888.4	18.9	30.4	84.5	6.0	23.0	27.0	4.1	6.5	0.0
	Maximum		3514.0	104.6	151.6	537.4	19.0	67.0	65.0	78.0	41.4	79.2

Add. Table 10: Results of rotated PCA 1, retaining four dietary patterns based on 36 food groups

Food item	Diet of the Rich	*Mchicha* Diet	Banana Diet	Fatty Diet
Variance explained (%)	**9.1**	**7.9**	**7.6**	**5.3**
Rice	0.618	0.205	-0.143	-0.170
Nuts	0.587	-0.006	0.124	-0.089
Egg	0.557	-0.039	0.162	0.043
Chapati	0.556	0.062	0.055	0.009
Leguminous vegetables	0.537	-0.093	0.006	-0.026
Bread	0.503	0.362	-0.220	-0.190
Soda drinks	0.471	0.108	-0.028	-0.155
Red meat	0.453	0.103	-0.037	0.367
Mchicha†	-0.017	0.645	0.029	0.110
Cucumber & okra	0.209	0.581	0.032	0.038
Onion	0.089	0.579	-0.042	0.138
Carrots & tomatoes	0.145	0.516	-0.096	-0.007
Maize	-0.180	0.461	0.135	-0.085
Fish	-0.018	0.434	0.337	-0.085
Avocado	-0.016	0.413	0.347	0.067
Banana	0.145	0.030	0.667	0.073
Green Banana	0.086	0.008	0.616	-0.176
Sugar	0.153	-0.103	0.491	-0.166
Fruits♯	0.085	0.189	0.478	-0.218
Starchy tubers	-0.275	-0.063	0.461	0.136
Mbege‡	-0.295	0.050	0.442	0.246
Pulses	-0.070	0.281	0.415	0.134
Sunflower oil	0.203	-0.207	-0.071	-0.623
Milk	0.264	-0.079	-0.042	0.521
Butter and lard	-0.213	-0.254	0.055	0.457
Mixed vegetable fats and oil	0.263	0.191	-0.115	0.454
Tea	0.055	0.013	0.366	-0.410
Potatoes	0.349	0.228	0.244	0.077
Juice	0.339	0.030	-0.027	0.127
Chicken meat	0.254	0.133	0.083	-0.030
Mango & Papaya	0.119	0.389	0.388	0.059
Cabbage (white)	0.185	0.354	0.083	-0.097
Mandazi§	0.128	0.211	-0.060	-0.049
Uji ∫	0.167	-0.199	0.067	-0.050
Coffee	0.023	-0.049	0.271	0.055
Bottled beer & wine	0.107	-0.006	-0.087	-0.335

Varimax with Kaiser Normalization. Rotation converged in 7 iterations.
* unleavened East African flat wheat bread; † traditional Tanzanian food, synonymously used for a dish of amaranth leaves and e.g. onions, tomatoes or/ and carrots in various amounts; ‡ often homemade opaque beer from bananas and millet; § East African donuts; ∫ thin millet or maize based porridge; ♯ oranges, watermelon and pineapple

Add. Table 11: Results of the rotated PCA 2 based on 34 food groups, alcoholic beverages excluded, retaining six dietary patterns

	Diet of the Rich (non alc)	Fruity Diet (non alc)	Mchicha Diet (non alc)	Banana Diet (non alc)	Starchy Diet (non alc)	Fatty Diet (non alc)
Variance explained (%)	8.8	7.1	6.7	6.5	6.1	5.2
Nuts	0.604	0.010	-0.007	0.090	0.120	-0.032
Chapati*	0.593	0.039	0.138	0.017	-0.042	0.058
Soda drinks	0.574	0.206	-0.002	-0.185	-0.010	-0.098
Egg	0.571	-0.006	0.040	0.139	0.005	0.062
Pulses	0.491	-0.069	-0.162	0.009	0.274	0.054
Chicken meat	0.311	0.222	0.070	-0.015	-0.071	0.001
Juice	0.291	-0.211	0.264	0.095	-0.009	0.124
Fish	0.086	0.612	0.178	0.068	-0.069	-0.103
Mango & Papaya	0.146	0.609	0.033	0.121	0.158	0.068
Avocado	-0.043	0.521	0.081	0.138	0.255	0.046
Fruits[#]	0.180	0.457	0.034	0.353	-0.140	-0.175
Uji [∫]	0.109	-0.385	0.108	0.274	-0.026	-0.067
Onion	0.035	0.049	0.783	0.034	-0.004	0.046
Carrots & tomatoes	0.102	-0.005	0.703	-0.002	0.039	-0.077
Mchicha[†]	-0.108	0.184	0.630	0.033	0.247	0.030
Cucumber & okra	0.227	0.374	0.491	-0.077	0.044	0.021
Sugar	0.069	-0.074	-0.034	0.648	0.143	-0.113
Green Banana	0.112	0.227	-0.040	0.588	-0.030	-0.152
Banana	0.187	0.334	0.005	0.584	-0.184	0.092
Starchy tubers	-0.317	0.089	-0.050	0.475	-0.056	0.100
Coffee	-0.061	-0.098	0.058	0.371	0.075	0.040
Leguminous vegetables	-0.080	0.287	0.249	0.338	-0.027	0.094
Potatoes	0.240	0.059	0.220	0.304	0.279	0.097
Cabbage (white)	0.008	0.121	0.078	0.117	0.680	-0.025
Bread	0.368	-0.013	0.199	-0.134	0.596	-0.063
Mandazi§	-0.027	0.023	-0.026	0.009	0.561	0.080
Rice	0.496	-0.060	0.065	-0.049	0.551	0.002
Maize	-0.262	0.370	0.125	0.071	0.393	-0.045
Sunflower oil	0.243	-0.347	-0.045	0.062	0.084	-0.628
Milk	0.152	-0.107	-0.003	0.037	0.044	0.576
Mixed vegetable fats and oil	0.211	0.264	0.003	-0.220	0.090	0.557
Butter and lard	-0.328	-0.134	-0.235	0.155	0.009	0.490
Red meat	0.342	-0.165	0.279	0.080	0.105	0.432
Tea	0.092	0.204	-0.183	0.326	0.152	-0.345

Rotation method Varimax with Kaiser Normalization. Rotation converged in 10 iterations.
* unleavened East African flat wheat bread; ♯ oranges, watermelon and pineapple; † traditional Tanzanian food, synonymously used for a dish of amaranth leaves and e.g. onions, tomatoes or/ and carrots in various amounts; § East African donuts; ∫ thin millet or maize based porridge

Add. Table 13: Results of the logistic regression

Variable	Coefficient (β)	Standard Error	Wald χ^2	*P* - value	Odds Ratio	95% CI	N
Age	0·02	0·02	0·88	0·35	1·02	0·98 -1·05	333
Place of living	0·09	0·27	0·12	0·73	1·10	0·65 -1·86	333
Property level							
Low	Reference						87
Medium	-1·07	0·29	13·50	0·00	0·34	0·19 -0·61	198
High	-1·53	0·48	10·32	0·00	0·22	0·09 -0·55	48
BMI at 20 years	0·27	0·09	10·36	0·00	1·31	1·11 -1·55	333
BMI at interview	-0·06	0·04	2·12	0·15	0·94	0·87 -1·02	333
Age at menarche	-0·18	0·10	3·37	0·07	0·84	0·69 -1·01	333
Age at first full term pregnancy							
≤ 20years	Reference						193
> 20 years	0·42	0·29	2·03	0·15	1·52	0·86 -2·69	122
No pregnancy	-0·38	0·72	0·28	0·60	0·69	0·17 -2·79	18
Menopausal status	0·14	0·44	0·10	0·76	1·15	0·48 -2·74	149/ 184
Lifelong lactation	-0·01	0·00	8·33	0·00	0·99	0·98 -1·00	333
Constant	-1·40	2·43	0·33	0·56	0·25		

Cox & Snell R^2 = 0.15; Nagelkerke's R^2 = 0.21, Overall percentage correctly classified = 76 %

Add. Table 15: Results of the logistic regression stratified for menopausal status

Menopausal status of the women	Variable	Coefficient (β)	Standard Error	Wald χ^2	*P*-Value	Odds Ratio	95% Confidence Interval		N
Pre-menopausal*	Age	0·07	0·03	4·28	0·04	1·07	1·00 -	1·15	149
	Place of living	-0·37	0·42	0·78	0·38	0·69	0·31 -	1·57	149
	Property level								
	Low	Reference							41
	Medium	-0·49	0·45	1·18	0·28	0·62	0·26 -	1·48	90
	High	-2·25	1·00	5·06	0·02	0·11	0·02 -	0·75	18
	BMI at 20 years	0·35	0·13	7·45	0·01	1·41	1·10 -	1·81	149
	BMI at interview	-0·06	0·06	0·98	0·32	0·94	0·84 -	1·06	149
	Age at menarche	-0·30	0·15	3·90	0·05	0·74	0·56 -	1·00	149
	Age at first full term pregnancy								
	≤ 20years	Reference							75
	> 20 years	0·05	0·44	0·01	0·92	1·05	0·44 -	2·49	68
	No pregnancy	1·46	1·20	1·47	0·23	4·29	0·41 -	45·19	6
	Lifelong lactation	-0·02	0·01	8·19	0·00	0·98	0·97 -	0·99	149
	Constant	-2·68	3·72	0·52	0·47	0·07			
Post-menopausal**	Age	0	0·02	0	0·98	1·00	0·96 -	1·04	184
	Place of living	0·47	0·40	1·38	0·24	1·60	0·73 -	3·48	184
	Property level								
	Low	Reference							46
	Medium	-1·72	0·43	16·09	0·00	0·18	0·08 -	0·42	108
	High	-1·76	0·62	8·05	0·01	0·17	0·05 -	0·58	30
	BMI at 20 years	0·32	0·13	5·79	0·02	1·38	1·06 -	1·80	184
	BMI at interview	-0·09	0·06	2·20	0·14	0·92	0·82 -	1·03	184
	Age at menarche	-0·06	0·14	0·17	0·68	0·95	0·72 -	1·24	184
	Age at first full term pregnancy								
	≤ 20years	Reference							118
	> 20 years	0·88	0·43	4·12	0·04	2·40	1·03 -	5·60	54
	No pregnancy	-0·91	1·04	0·76	0·38	0·40	0·05 -	3·11	12
	Lifelong lactation	-0·01	0·01	1·98	0·16	0·99	0·98 -	1·00	184
	Constant	-2·95	3·79	0·61	0·44	0·05			

*pre-menopausal: Constant: *P* = 0.47, OR 0.07; Cox& Snell R^2 = 0.21, Nagelkerke's R^2 = 0.3, percentage correct = 75 %;
**post-menopausal: Constant: *P* = 0.44, OR 0.05; Cox & Snell R^2 = 0.19, Nagelkerke's R^2 = 0.27, overall correct = 78 %

Add. Table 16: Results of the logistic regression: dietary patterns and breast cancer

Variable	Coefficient (β)	Standard Error	Wald χ²	*P*-value	Odds Ratio	95% CI
Age	-0.00	0.01	0.13	0.72	1.00	0.98 - 1.02
Dietary patterns (PCA 1)						
Diet of the rich	0.01	0.13	0.00	0.95	1.01	0.79 - 1.30
Mchicha diet	0.38	0.13	9.11	0.00	1.47	1.14 - 1.88
Banana diet	0.66	0.16	17.93	0.00	1.94	1.43 - 2.63
Fatty diet	0.48	0.13	14.39	0.00	1.62	1.26 - 2.07
Constant	-0.58	0.50	1.35	0.25	0.56	

OR 0.56; Cox & Snell R^2 = 0.13 ; Nagelkerke's R^2 = 0.18; Overall percentage correctly classified 74 %

Add. Table 17: Results of the logistic regression: dietary patterns (PCA 1) and basic breast cancer risk model

Variable	Coefficient (β)	Standard Error	Wald χ^2	*P*-value	Odds Ratio	95% CI	N
Age	0.00	0.02	0.04	0.84	1.00	0.97 -1.04	333
Place of living	0.22	0.29	0.56	0.45	1.24	0.71 -2.18	333
Property level							
Low	Reference						87
Medium	-1.00	0.32	8.83	0.00	0.39	0.21 -0.72	198
High	-1.34	0.54	6.23	0.01	0.26	0.09 -0.75	48
Body mass index (kg/m²)							
At 20 years	0.28	0.09	6.12	0.01	1.26	1.05 -1.51	333
At interview	-0.07	0.04	2.63	0.11	0.93	0.86 -1.02	333
Age at menarche	-0.14	0.10	1.75	0.19	0.87	0.71 -1.07	333
Age at first full term pregnancy							
≤ 20years	Reference						193
> 20 years	0.55	0.32	3.03	0.08	1.74	0.93 -3.23	122
No pregnancy	-0.21	0.77	0.07	0.79	0.82	0.18 -3.70	18
Menopausal status	0.36	0.47	0.58	0.45	1.43	0.57 -3.59	149/ 184
Lifelong lactation	-0.01	0.00	5.93	0.02	0.99	0.98 -1.00	333
Dietary patterns (PCA 1)							
Diet of the rich	0.14	0.17	0.72	0.40	1.15	0.83 -1.59	333
Mchicha diet	0.25	0.14	3.01	0.08	1.28	0.97 -1.70	333
Banana diet	0.56	0.20	8.13	0.00	1.75	1.20 -2.57	333
Fatty diet	0.40	0.15	7.50	0.01	1.50	1.12 -1.99	333
Constant	-0.61	2.56	0.06	0.81	0.54		

Cox & Snell R^2 = 0.20; Nagelkerke's R^2 = 0.29; Overall percentage correctly classified = 76%.

Add. Table 18: Results of the logistic regression: dietary patterns (PCA 2), alcoholic beverages and basic breast cancer risk model

Variable	Coefficient (β)	Standard Error	Wald χ^2	*P*-value	Odds Ratio	95% CI	N
Age	0.00	0.02	0.04	0.85	1.00	0.97 -1.04	333
Place of living	0.20	0.29	0.48	0.49	1.22	0.70 -2.16	333
Property level							
Low	Reference						87
Medium	-0.98	0.33	9.12	0.00	0.37	0.20 -0.71	198
High	-1.32	0.54	5.98	0.01	0.27	0.09 -0.77	48
BMI at 20 years	0.24	0.09	6.41	0.01	1.27	1.06 -1.53	333
BMI at interview	-0.07	0.04	2.85	0.09	0.93	0.85 -1.01	333
Age at menarche	-0.12	0.11	1.25	0.26	0.89	0.72 -1.09	333
Age at first full term pregnancy							
≤ 20years	Reference						193
> 20 years	0.60	0.33	3.44	0.06	1.83	0.97 -3.45	122
No pregnancy	-0.20	0.79	0.06	0.80	0.82	0.18 -3.84	18
Menopausal status	0.33	0.47	0.48	0.49	1.39	0.55 -3.51	149/ 184
Lifelong lactation	-0.01	0.01	5.07	0.02	0.99	0.98 -1.00	333
Dietary patterns (PCA 2)							
Diet of the rich (non alc)	0.24	0.18	1.91	0.17	1.28	0.90 -1.59	333
Fruity diet (non alc)	0.48	0.18	7.44	0.01	1.61	1.14 2.28	333
Mchicha diet (non alc)	0.05	0.14	0.15	0.70	1.06	0.80 -1.40	333
Banana diet (non alc)	0.28	0.18	2.44	0.12	1.32	0.93 -1.87	333
Starchy diet (non alc)	0.02	0.14	0.03	0.86	1.02	0.78 1.34	333
Fatty diet (non alc)	0.35	0.14	6.10	0.01	1.42	1.08 -1.87	333
Mbege‡	0.00	0.00	3.07	0.08	1.00	1.00 1.00	333
Beer & wine	0.00	0.00	0.03	0.87	1.00	1.00 1.00	333
Constant	-1.22	2.60	0.22	0.64	0.30		

Cox & Snell R^2 = 0.21; Nagelkerke's R^2 = 0.29; Overall percentage correctly classified = 77%;
‡ often homemade opaque beer from bananas and millet

Add. Table 21: Results of the logistic regression in women with a BMI below 24 kg/m²

Body mass index <24 kg/m²							
Variable	**Coefficient (β)**	**Standard Error**	**Wald χ²**	***P*-value**	**Odds Ratio**	**95 % CI**	**N**
Age	0.044	0.045	0.949	0.330	1.045	0.956 - 1.142	89
Place of living	1.419	1.037	1.870	0.171	4.133	0.541 - 31.574	89
Property level							
Low			2.041	0.360			35
Medium	-1.158	0.811	2.041	0.153	0.314	0.064 - 1.539	52
High	-21.554	24526.463	0.000	0.999	0.000	0.000 -0.	2
Body mass index (kg/m²)							
At 20 years	1.423	0.553	6.629	0.010	4.151	1.405 - 12.268	89
At interview	-0.955	0.304	9.855	0.002	0.385	0.212 -0 .698	89
Age at menarche	0.117	0.405	0.084	0.772	1.124	0.509 - 2.487	89
Age at first full term pregnancy							
≤ 20years			3.583	0.167			56
> 20 years	0.396	1.047	0.143	0.706	1.485	0.191 - 11.570	27
No pregnancy	-3.588	2.193	2.677	0.102	0.028	0.000 - 2.035	6
Menopausal status	0.014	1.320	0.000	0.992	1.014	0.076 - 13.467	44/ 45
Lifelong lactation	-0.023	0.013	3.113	0.078	0.977	0.953 - 1.003	89
Dietary patterns (PCA 2)							
Diet of the rich (non alc)	2.176	0.934	5.428	0.020	8.808	1.413 - 54.927	89
Fruity diet (non alc)	1.006	0.529	3.610	0.057	2.733	0.969 - 7.713	89
Mchicha diet (non alc)	-0.084	0.488	0.029	0.864	0.920	0.353 - 2.394	89
Banana diet (non alc)	1.353	0.594	5.191	0.023	3.870	1.208 - 12.397	89
Starchy diet (non alc)	0.277	0.402	0.474	0.491	1.319	0.600 - 2.898	89
Fatty diet (non alc)	0.575	0.424	1.839	0.175	1.777	0.774 - 4.078	89
Mbege‡	0.001	0.001	0.638	0.424	1.001	0.998 - 1.004	89
Beer & wine	0.001	0.001	0.817	0.366	1.001	0.999 - 1.004	89
Constant	-9.733	10.482	0.862	0.353	0.000		

Cox & Snell R^2 = 0.5; Nagelkerke's R^2 = 0.67; Overall percentage correctly classified = 84 %. -2 Log likelihood estimation terminated at iteration number 20 because maximum iterations had been reached. Final solution could not be found.;
‡ often homemade opaque beer from bananas and millet

Add. Table 22: Results of the logistic regression in women with a BMI of 24 - 26 kg/m²

Body mass index 24 – 26 kg/m²							
Variable	**Coefficient (β)**	**Standard Error**	**Wald χ²**	***P*-value**	**Odds Ratio**	**95 % CI**	**N**
Age	-0.094	0.064	2.163	0.141	0.910	0.803 - 1.032	91
Place of living	0.004	0.793	0.000	0.996	1.004	0.212 - 4.748	91
Property level							
Low			6.972	0.031			21
Medium	-2.654	1.009	6.916	0.009	0.070	0.010 - .509	63
High	-2.491	1.748	2.032	0.154	0.083	0.003 - 2.545	7
Body mass index (kg/m²)							
At 20 years	0.246	0.315	0.610	0.435	1.279	0.690 - 2.374	91
At interview	0.340	0.575	0.351	0.554	1.406	0.456 - 4.335	91
Age at menarche	-0.344	0.305	1.272	0.259	0.709	0.390 - 1.289	91
Age at first full term pregnancy							
≤ 20years			1.783	0.410			55
> 20 years	1.166	0.873	1.783	0.182	3.211	0.580 - 17.785	33
No pregnancy	-21.265	23062.766	0.000	0.999	0.000	0.000 - 0	3
Menopausal status	3.446	1.655	4.335	0.037	31.377	1.224 - 804.237	39/ 52
Lifelong lactation	-0.014	0.014	0.941	0.332	0.987	0.960 - 1.014	91
Dietary patterns (PCA 2)							
Diet of the rich (non alc)	-0.443	0.709	0.390	0.532	0.642	0.160 - 2.577	91
Fruity diet (non alc)	-0.824	0.680	1.470	0.225	0.438	0.116 - 1.662	91
Mchicha diet (non alc)	-0.039	0.455	0.007	0.932	0.962	0.394 - 2.348	91
Banana diet (non alc)	-0.552	0.661	0.699	0.403	0.576	0.158 - 2.101	91
Starchy diet (non alc)	0.822	0.636	1.672	0.196	2.275	0.654 - 7.906	91
Fatty diet (non alc)	-0.185	0.458	0.163	0.687	0.831	0.339 - 2.041	91
Mbege	0.003	0.002	3.936	0.047	1.003	0.803 - 1.032	91
Beer & wine	0.000	0.003	0.017	0.896	1.000	0.212 - 4.748	91
Constant	-5.693	15.955	0.127	0.721	0.003		

Cox & Snell R^2 = 0.25; Nagelkerke's R^2 = 0.39; Overall percentage correctly classified = 85 %. -2 Log likelihood estimation terminated at iteration number 20 because maximum iterations had been reached. Final solution could not be found.
‡ often homemade opaque beer from bananas and millet

Add. Table 23: Results of the logistic regression in women with a BMI of >26 kg/m²

Body mass index > 26 kg/m²							
Variable	**Coefficient (β)**	**Standard Error**	**Wald χ²**	***P*-value**	**Odds Ratio**	**95 % CI**	**N**
Age	-0.005	0.033	0.026	0.872	0.995	0.932 - 1.062	153
Place of living	0.241	0.501	0.231	0.631	1.272	0.476 - 3.395	153
Property level							
Low			3.670	0.160		-	31
Medium	-1.079	0.653	2.731	0.098	0.340	0.094 - 1.222	83
High	-1.536	0.845	3.299	0.069	0.215	0.041 - 1.129	29
Body mass index (kg/m²)							
At 20 years	0.275	0.148	3.430	0.064	1.317	0.984 - 1.761	153
At interview	0.425	0.125	11.544	0.001	1.530	1.197 - 1.956	153
Age at menarche	-0.188	0.173	1.169	0.280	0.829	0.590 - 1.165	153
Age at first full term pregnancy							
≤ 20years			0.538	0.764		-	82
> 20 years	0.262	0.569	0.211	0.646	1.299	0.426 - 3.964	62
No pregnancy	0.973	1.363	0.509	0.475	2.646	0.183 - 38.279	9
Menopausal status	-0.203	0.788	0.067	0.796	0.816	0.174 - 3.821	66/ 87
Lifelong lactation	-0.008	0.008	0.902	0.342	0.992	0.977 - 1.008	153
Dietary patterns (PCA 2)							
Diet of the rich (non alc)	-0.148	0.283	0.272	0.602	0.863	0.495 - 1.503	153
Fruity diet (non alc)	0.349	0.281	1.549	0.213	1.418	0.818 - 2.457	153
Mchicha diet (non alc)	-0.133	0.262	0.257	0.612	0.875	0.524 - 1.464	153
Banana diet (non alc)	0.085	0.313	0.074	0.786	1.089	0.589 - 2.012	153
Starchy diet (non alc)	0.202	0.320	0.396	0.529	1.223	0.653 - 2.292	153
Fatty diet (non alc)	0.507	0.259	3.842	0.050	1.661	1.000 - 2.757	153
Mbege	0.001	0.001	0.637	0.425	1.001	0.999 - 1.002	153
Beer & wine	0.000	0.002	0.031	0.861	1.000	0.996 - 1.003	153
Constant	-14.813	5.999	6.098	0.014	0.000		

Cox & Snell R^2 = 0.3; Nagelkerke's R^2 = 0.45; Overall percentage correctly classified = 85 %. -2 Log likelihood estimation terminated at iteration number 6 because parameters estimated changed by less than 0.001.
‡ often homemade opaque beer from bananas and millet

8.5 Publication in the Journal: “Breast Cancer Research and Treatment”

Breast Cancer Res Treat
DOI 10.1007/s10549-010-1255-7

EPIDEMIOLOGY

Breast cancer risk among women with long-standing lactation and reproductive parameters at low risk level: a case–control study in Northern Tanzania

Irmgard Jordan · Antje Hebestreit · Britta Swai · Michael B. Krawinkel

Received: 15 July 2010 / Accepted: 7 November 2010

Abstract Breast cancer is the leading cause of death among women worldwide. Studies in industrialised countries identified age at menarche, age at first full-term pregnancy, and lactation as determining factors in the aetiology of breast cancer. 115 female breast cancer patients (cases) and 230 age- and district-matched women clinically free from breast cancer (controls) were interviewed about their reproductive history and socioeconomic condition. Semi-structured interviews including anthropometric measurements were conducted by trained enumerators. The median age was 50 years (min/max 26 to 85 years). Estimated median BMI at age 20 was 21 kg/m^2 in both cases and controls. Median lifelong lactation of the mothers was 96 months (cases) and 108 months (controls). A high BMI at 20 years was associated with an increased breast cancer risk (OR 1.31 95% CI 1.11–1.55, $P < 0.01$). The odds ratio for lifelong lactation was slightly below one (OR 0.99 95% CI 0.98–1.00, $P < 0.01$). There was no significant association in risk for BMI at interview (median 25 kg/m^2 of cases and 26 kg/m^2 of controls), age at menarche (median 16 years), and age at first full-term pregnancy (median 20 years). The association of increased risk with higher BMI at age 20 years remained significant after stratification for menopause (premenopausal: OR 1.41 95% CI 1.10–1.81, $P = 0.01$; postmenopausal: OR 1.38 95% CI 1.06–1.80, $P = 0.02$). Late age at menarche and prolonged lifelong lactation were associated with a risk reduction among premenopausal women ($OR_{menarche}$ 0.74 95% CI 0.56–1.00, $P = 0.05$; $OR_{lactation}$ 0.98 95% CI 0.97–0.99, $P < 0.01$). In conclusion, long-standing lactation and reproductive behaviour are associated with a lower breast cancer risk in the region. As current changes in lifestyle affect age at menarche, reproductive behaviour, and nutritional status, an increased incidence of breast cancer is to be expected. Preventive efforts should include advice on reproductive and breastfeeding behaviour.

Keywords Breast cancer · Tanzania · Menarche · Breastfeeding · Lactation · BMI · Lifestyle

I. Jordan · M. B. Krawinkel (✉)
Institute of Nutritional Sciences, Justus-Liebig-University Giessen, Giessen, Germany
e-mail: Michael.Krawinkel@uni-giessen.de

A. Hebestreit
Bremen Institute for Prevention Research and Social Medicine, Bremen, Germany

B. Swai
Kilimanjaro Christian Medical Center, Moshi, Tanzania

Introduction

Breast cancer is the leading cause of death among women worldwide [1, 2]. Less than 10% of women’s breast cancer cases are caused by inherited mutations of either BRCA1 or BRCA2 genes [3, 4]. Although distinct mutations in these genes have been identified in black people in the US [5], little is known of the prevalence of these mutations as well as of lifestyle factors associated with breast cancer in African populations [6]. In Tanzania, breast cancer is currently regarded as the second most common cancer in women after tumours of the cervix uteri. In 2008 the national age-standardised breast cancer rate was estimated to be 10.1 per 100,000 [7]. Between 1998 and 2006 the WHO supported cancer registry based at the Kilimanjaro Christian Medical Centre counted annually an average of 36 new breast cancer cases in the region. The real number is probably higher, because exclusive consultation with

Published online: 20 November 2010

8.6 Publication submitted to "European Journal of Nutrition", under review

European Journal of Nutrition

Dietary patterns and breast cancer risk among women in Northern Tanzania: a case-control study

--Manuscript Draft--

Manuscript Number:	EJN-D-11-00226R1
Full Title:	Dietary patterns and breast cancer risk among women in Northern Tanzania: a case-control study
Article Type:	Original Contribution
Keywords:	breast cancer, dietary pattern, PCA,PUFA, Tanzania
Corresponding Author:	Irmgard Jordan Justus-Liebig-University Giessen Giessen, GERMANY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Justus-Liebig-University Giessen
Corresponding Author's Secondary Institution:	
First Author:	Irmgard Jordan
First Author Secondary Information:	
All Authors:	Irmgard Jordan
	Antje Hebestreit, Dr
	Britta Swai
	Michael B Krawinkel, Professor Dr.
All Authors Secondary Information:	
Abstract:	Background: Breast cancer is the second most common cancer among women in the Kilimanjaro Region of Tanzania. It was tested within a case-control study in this region whether a specific dietary pattern impacts on the breast cancer risk. Methods: A validated semi-quantitative Food Frequency Questionnaire was used to assess the dietary intake of 115 female breast cancer patients and 230 healthy age-matched women living in the same districts. A logistic regression was performed to estimate breast cancer risk. Dietary patterns were obtained using principal component analysis with Varimax rotation. Results: The adjusted logistic regression estimated an increased risk for a "Fatty Diet", characterized by a higher consumption of milk, vegetable oils and fats, butter, lard, and red meat (OR = 1.42, 95% CI 1.08-1.87; P = 0.01), and for a "Fruity Diet" characterized by a higher consumption of fish, mango, papaya, avocado, and watery fruits (OR = 1.61, 95% CI 1.14-2.28; P = 0.01). Both diets showed an inverse association with the ratio between polyunsaturated and saturated fatty acids (P/S ratio). Conclusion: A diet characterised by a low P/S ratio seems to be more important for the development of breast cancer than total fat intake.
Suggested Reviewers:	

Dietary patterns and breast cancer risk among women in Northern Tanzania: a case-control study

Irmgard Jordan[1], Antje Hebestreit[2], Britta Swai[3], Michael B. Krawinkel[1]

[1] Institute of Nutritional Sciences of Justus-Liebig-University Giessen, Germany; [2] Bremen Institute for Prevention Research and Social Medicine, Germany; [3] Kilimanjaro Christian Medical Center in Moshi, Tanzania

Corresponding author:

Irmgard Jordan, Justus-Liebig University Giessen, Institute of Nutritional Sciences, Wilhelmstr. 20, 35392 Giessen, Germany

Tel: +49-641-9939052, email: Irmgard.Jordan@ernaehrung.uni-giessen.de

Abstract

Background: Breast cancer is the second most common cancer among women in the Kilimanjaro Region of Tanzania. It was tested within a case-control study in this region whether a specific dietary pattern impacts on the breast cancer risk.

Methods: A validated semi-quantitative Food Frequency Questionnaire was used to assess the dietary intake of 115 female breast cancer patients and 230 healthy age-matched women living in the same districts. A logistic regression was performed to estimate breast cancer risk. Dietary patterns were obtained using principal component analysis with Varimax rotation.

Results: The adjusted logistic regression estimated an increased risk for a “Fatty Diet”, characterized by a higher consumption of milk, vegetable oils and fats, butter, lard, and red meat (OR = 1.42, 95% CI 1.08-1.87; *P* = 0.01), and for a “Fruity Diet” characterized by a higher consumption of fish, mango, papaya, avocado, and watery fruits (OR = 1.61, 95% CI 1.14-2.28; *P* = 0.01). Both diets showed an inverse association with the ratio between polyunsaturated and saturated fatty acids (P/S ratio).

Conclusion: A diet characterised by a low P/S ratio seems to be more important for the development of breast cancer than total fat intake.

Keywords: breast cancer, dietary pattern, PUFA, Tanzania

Introduction

Established factors for breast cancer development are age at menarche, age at menopause, age at first full-term pregnancy, breastfeeding, and alcohol consumption at all ages [1-7]. A high percentage of total body fat, and tall height at adulthood in postmenopausal women are associated with an increased risk [8-11]. Several studies have looked at possible linkages between single nutrient intake as well as foods or dietary patterns and breast cancer [12-18]. However, there has been limited evidence suggesting only that consumption of total dietary fat and special dietary patterns influence breast cancer risk, but no internationally accepted conclusion was reached up to now [7, 19-20].

In Tanzania, a low income country where breast cancer is currently the second most common cancer in women, the lifestyle characterized by long-standing lactation or late age at menarche has been associated with a lower breast cancer risk [21]. However, breast cancer occurs and a pilot case-control study in the Kilimanjaro Region of Northern Tanzania estimated an increased association between alcohol consumption and breast cancer [22]. A new case-control study looked at dietary patterns rather than single nutrients as nutrients are ingested within diets. A case-control design was chosen because of a lack of demographic data and infrastructural deficits for identification of all women affected to allow for a prospective study approach.

The study was carried out in collaboration between the Kilimanjaro Christian Medical Centre (KCMC) in Moshi, Tanzania, and the Institute of Nutritional Sciences of the University of Giessen, Germany.

Methods

Breast cancer patients and controls were recruited in the Kilimanjaro Region between 2004 and 2007. The detailed study methodology has been described previously [21]. In summary, cases were identified using fine needle aspiration cytology (FNAC) confirming primary breast cancer diagnosis. The hospital and visitor-based controls were matched according to age (± 1.5 years) and lived in the same district for at least five years during the past ten years. The controls were interviewed for their medical history and underwent a physical examination to exclude palpable breast cancer. After informed consent 115 cases and 230 controls were interviewed either by a trained nurse or a medical doctor in Swahili, based on a standardized questionnaire in English about their socio-economic situation, current and former lifestyle. At first the variables were tested for normal distribution, followed by their respective tests for statistically significant differences between cases and controls. The two control groups were analysed for differences in their socioeconomic status using the Mann-Whitney *U* Test [21].

The present analyses focus on the dietary patterns of both cases and controls using the data of a semi-quantitative food frequency questionnaire (FFQ). The FFQ food list was prepared based on market surveys at different seasons and completed after a pre-test. This list was finally validated in 2005 and 2006 based on two non-consecutive 24-hour recalls of 50 randomly selected women with a mean age of 40 years (23-70 years), who did not participate in the case-control study but lived in the same study region. The validation study covered two seasons with different food availability: dry and rainy season. Data collection was done by four trained enumerators. The training included estimation of quantities using common household measurements, e.g. cups, spoons, customary packing size, and solid foods in pieces or slices. Foods were prepared according to local standard recipes and weighed using household kitchen scales by the research staff. Countable foods such as onions, eggs or bananas were classified according to their size into small, medium and large. Samples of food pieces were obtained from the local market and mean weights were taken of each size. The matter of size was intensively discussed in the interviewer trainings to assure a common comprehension. A raw/ cooked coefficient was applied when large deviations between cooked and raw foods were expected after preparation, e.g. for dried cereals (pasta, rice) and dried legumes. The coefficients were calculated by cooking experiments done by the nutritionist but without calculating any loss of vitamins and minerals. Seasonal food availability on individual level was assessed within the interview especially for fruits and a seasonal factor was applied accordingly.

The FFQ data from both, the validation and the case-control study, were entered into NutriSurvey®, a nutrition software package, which generated tables of the individual food and nutrient intake per day, latter based on food composition tables from Tanzania, Kenya, Senegal, Mali and Germany [23, 24]. All data were converted to gramm intake per day for each food item.

For the validation study the data sets were merged into six food groups to describe invidual food intake: 1) cereals: bread, rolls, cereal products, grains, egg-free pasta; 2) vegetables: vegetables, pulses, potatoes, mushrooms; 3) animal products: eggs, dairy and cheese, meat, fish, poultry, sausages and other meat products; 4) beverages: non-alcoholic beverages, coffee, tea, water, alcoholic beverages; 5) fruits; 6) fats: oil, fats, butter. Since the values of most variables were not normally distributed non parametric tests were carried out in the subsequent analysis. The studied population had a low educational level and considering the relative high number of interviewers in relation to the study population the validation data were tested for interviewer effects before any statistical analysis was performed. The Kruskal Wallis Test chosen to test for homogenity between the interviewers showed interviewer effects in 100 % of the food groups confirmed by the Median One Way Test at a level of 83 %. Therefore, further analysis was carried out stratified by interviewer. The Wilcoxon signed rank test was used to test the 24h-recall and the FFQ for seasonal variabiltiy. It is a non parametric test equivalent to the paired t-test. In addition, the Wilcoxon signed rank test was used to test for differences in the results of the 24h-recall and FFQ. There was no evidence for a seasonal effect in breast cancer relevant food groups if the FFQ is used. Differences in the intake of oils and fats assessed by the validated FFQ and its reference, the 24h recall, could only be shown by one interviewer. This might be due to low quantification capacities of the studied population especially in this respective food group and especially during the 24h-recall.. Furthermore Spearman correlation was calculated with all interviewers grouped together for comparison with other studies which did not report whether they checked for interviewer bias. The correlation coefficient (r_s) was highest in the food group “fruits” (r_s = 0.39, p = 0.01) followed by “cereals” (r_s = 0.38, p = 0.01), “beverages” (r_s = 0.33, p = 0.01) and the food groups “animal products” and “vegetables” (r_s =0.27 and r_s = 0.14 respectively, both values not significant). A negative correlation was found in the food group “oils and fats” (r_s = -0.22, p = 0.13). This was, however, not statistically significant. There had been no consistent statistical differences between FFQs and the 24h-recalls, and the correlations were low to modest but comparable to other studies [25-27], thus, the tested FFQ was considered a reliable instrument to assess dietary intake in the Kilimanjaro Region.

However, it was recommended paying special attention to the training of interviewers and especially to the assessment of the oil and fat intake. In addition, a calculation for seasonal variability in fruit and vegetable intake was recommended to be used where applicable. The FFQ finally contained in total 65 food items.

From data on individual food intake of the case-control study population, dietary patterns were created using principal component analysis (PCA). Although there is a certain disagreement among statistical theorists about it [28-30] PCA was chosen for keeping the results comparable to other studies looking at dietary patterns and disease [20, 31-35]. The sampling adequacy of the food group variables for factor analysis was confirmed using the Kaiser-Meyer-Olkin measure. The food items listed in the FFQ were merged into 36 food groups for obtaining factors from the PCA defined as dietary patterns. Scree plots and parallel analysis were used to quantify the number of factors wanted [30]. Food groups with factor loadings between -0.4 and 0.4 were disregarded for defining the dietary patterns. The final dietary patterns were included in a non-conditional logistic regression model. At first the dietary patterns were analysed only adjusted for age. Secondly the dietary patterns were included into the basic model described elsewhere [21]. This model includes the matching variables of age, and place of living, and the acknowledged predictors in the aetiology of breast cancer from high-income countries. If possible the variables were entered as continuous variables. The variables “age at first full-term pregnancy” was categorised into three groups, “first full-term pregnancy ‘≤ 20 years’, ‘> 20 years’, and ‘no pregnancy’ “.

Descriptive statistics, principal component analysis and logistic regression were performed using the statistical package of SPSS version 18 (SPSS Inc.).

Ethical clearance was obtained from the Research and Ethics Committee of the KCMC, Moshi, Tanzania, and the Ethics Committee of the Faculty of Medicine of the University of Giessen, Germany.

Results

Selected socio-economic and reproductive characteristics of the study participants are presented in table 1 [21]. Mean age of all women was 50 years, 94 % of them had children. Mean age at menarche was 16 years, and 21 years at delivery of the first child. Mean lifelong lactation time was 88 months. Breast cancer patients had a significantly lower lifelong lactation time compared to controls. The basic logistic model estimated an increased risk for women with a higher BMI at 20 years but a reduced risk for women with a high property level and prolonged lactation.

The energy and nutrients per day of cases and controls based on the results of the FFQ are presented in table 2. Median energy consumption in all women was 1,714 kcal per day (min 786 kcal; max 3,928 kcal), median protein intake was 47 g/d (min 17 g/d; max 183 g/d), median fat intake was 72 g/d (min 30 g/d; max 166 g/d) and median carbohydrate intake was 188 g/d (min 85 g/d – max 537 g/d). Median percentage of food energy from protein was 12 %, from fat 39 % and from carbohydrates 46 %. Median alcohol intake from alcoholic drinks was 8.2 g/d (min 0 g/d; max 100 g/d). Main alcoholic drinks were *mbege* (often homemade, locally brewed beer), bottled beer and wine (median intake 57 g/d, min 0 g/d; max 298 g/d and 0 g/d, min 0 g/d; max 77 g/d respectively).

A principal component analysis (PCA) was conducted primarily on 36 food groups with Varimax rotation. The Kaiser-Meyer-Olkin Measure verified the sampling adequacy for the PCA, KMO = 0.621, which is considered as mediocre [36, 37]. Following Kaiser's criterion retaining all components with eigenvalues greater than one, 14 components would have been useful for further analysis. However, the number of food groups with factor loadings < -0.4 or >0.4 varied between 0 to 11, thus, the results were not interpretable. Consequently, it was decided to retain four components as suggested by the scree plot. These four components or dietary patterns describe 29.9 % of the variance in food intake. The first pattern is characterized by rice, nuts, eggs, *chapati* (unleavened East African flat wheat bread), leguminous vegetables, bread, soda and red meat. Since most of these food items are usually purchased we called it the "Diet of the Rich". Pattern two is characterized by *mchicha*, cucumber, okra, onions, carrots, tomatoes, maize, fish and avocado. *Mchicha* is the Swahili name for amaranth leaf, a traditional food in Tanzania and often synonymously used for a dish consisting of amaranth leaves and e.g. onions, tomatoes or/ and carrots in various amounts. The pattern was therefore named "*Mchicha* Diet". The third pattern is characterized by ripe and green banana, sugar, different fruits, tubers, pulses and *mbege*. The mountainous area of the Kilimanjaro Region is known for its various banana plants. Therefore pattern three was called "Banana Diet". Pattern four is characterized by a high consumption of milk, butter, lard, vegetable oils and fats and a low consumption of sunflower oil and tea. All of the positively loading food items relate to fat, thus we called this pattern "Fatty Diet". With increased affiliation to this Fatty Diet bread consumption decreased (1^{st} Quartile median= 17 g bread/d, 4^{st} Quartile median = 9 g bread/d; *P* for trend <0.001) and red meat consumption increased (1^{st} Quartile: median 44 g/d, 4^{th} Quartile: median = 52 g/d; *P* for trend = 0.09).

The non-conditional multivariate and logistic regression examining the associations between dietary behaviour and breast cancer showed an increased risk association with three out of the four dietary patterns: the *Mchicha*, Banana and Fatty Diet After including socio-economic parameters and reproductive variables in the logistic model the odds ratio (OR) for the *Mchicha* Diet changed from a significant OR of 1.47 (95 % CI 1.14 - 1.88, *P* <0.01) to a non significant OR of 1.28 (95 % CI 0.97 - 1.7, *P* = 0.08). The Banana and the Fatty Diet were still associated with an increased breast cancer risk on a significant level. The OR for the Fatty Diet increased to 3.04 (95%CI: 1.34 - 6.91; *P* <0.01) among women with the highest consumption (4^{th} Quartile). With increased affiliation to the Fatty Diet, total fat intake increased significantly (*P* = 0.04), whereas percentage of energy from fat did not change (*P* = 0.83) and whereas the ratio of polyunsaturated fatty acids to saturated fatty acids (P/S ratio) was inversely associated with breast cancer risk (Figure 1). However, there was no risk association found between total fat intake (median 72 g/d) and breast cancer. In addition there was no changed found in risk associations if energy was included into the risk model described above (OR energy = 1.00, 95%CI 1.00-1.00, *P* = 0.51).

The Banana Diet includes *Mbege* - a local, often homemade opaque beer from bananas and millet. Acknowledging that alcohol is an accepted risk factor for breast cancer, the factor analysis was repeated excluding the alcoholic beverages from the food group list. In order to get comparable results to the first PCA we generated six dietary patterns that described 40.3% of the dietary variance, and four of them were comparable to the Diet of the Rich, *Mchicha*, Banana and Fatty Diet of the first PCA (Table 3). Table 4 presents the results of the logistic regression including the second set of dietary patterns with the alcoholic beverages included as separated variable. The *Mchicha* Diet and the Banana Diet were no longer associated with breast cancer risk, but the new Fruity Diet and again a Fatty Diet very similar to the first Fatty Diet were associated with increased risk (OR 1.61, 95 % CI 1.14 - 2.28, *P* = 0.01 and OR 1.42, 95 % CI 1.08 - 1.87, *P* = 0.01 respectively).

Discussion

Several dietary patterns from two principal component analyses with Varimax rotation based on a FFQ were associated with increased breast cancer risk. Two patterns, both called Fatty Diet are basically characterized by a higher consumption of milk, mixed vegetable oils and fats, butter and lard, but a low consumption of sunflower oil (Tables 4 and 7). Both Fatty Diets were associated with an increased risk in different logistical models. A diet rich in fat similar to our Fatty Diets was discussed by Schulz *et al.* [38] using reduced rank regression, stating that specific fatty acids are less important in populations with a generally higher fat consumption (mean 8.3 to 10.4 g/MJ). However, this level of dietary fat intake as a proportion of energy intake was comparable to our study population (mean 10.2 g/MJ), but the mean total fat intake in our population was 15 g per day lower because of the overall lower energy consumption than reported by Schulz *et al.* [38]. Here, the women's total fat intake was not associated with breast cancer risk (data not shown) although total fat intake increased significantly with increased affiliation to the fatty dietary patterns. Another prospective cohort study found a direct association between dietary fat intake including subtypes and post-menopausal invasive breast cancer [39]. However, they recorded at median 20.3 % energy intake from fat per day in the 1^{st} quintile and 40.1 % energy intake from fat per day in the last quintile. Only the latter energy intake level from dietary fat is comparable to our data. The wide range of fat intake observed in their study population may have resulted in an increased statistical power. This assumption was made by Thiébaut *et al.* [39] based on the hypotheses from Wynder *et al.* [40] that a threshold effect may exist for dietary fat, such it would be difficult to detect an association between fat intake and breast cancer risk in Western populations. They referred to studies about Asian diets in which more people consume diets containing 20 % or less of energy from fat which have shown significant or borderline significant associations of fat intake and breast cancer risk [39]. The median fat intake as percentage from energy intake in our study population was 39 % which is above this benchmark of 20 % and may explain why no association was found for our population. Regarding the fatty acid composition of the diet, the major PUFA sources reported by Thiébaut *et al.* were vegetable oils and fats, butter and mayonnaise [39]. Except mayonnaise, these food items have also been identified by Schulz *et al.* [38] and in our study as part of dietary patterns rich in fat which have been associated with a higher risk of developing breast cancer. Even if the fatty acid composition of foods varies intrinsically, this composition may be more important than the total fat intake. Our study population showed a negative association of the P/S ratio with breast cancer risk (Figure 1). This negative association was also observed in a case-control study among premenopausal women in Singapore but was attributed to PUFA intake only [12]. However, results from the European Investigation into Cancer and Nutrition (EPIC) study [41] and a case-control study in Connecticut [42] supported the hypothesis of Rose *et al*. [43] and Key *et al.* [9] that both saturated and polyunsaturated fatty acids influence inversely the oestrogen metabolism and mammary carcinogenesis. In addition results from the Shanghai Women's Health study, a prospective cohort study, suggested that the relative amounts of n-6 PUFA to marine-derived n-3 PUFAs may be more important for the breast cancer risk than individual amounts of these fatty acids in the diet [44]. They supported the hypothesis that the different PUFA compete as enzyme substrates inside membrane phospholipids [45]: This may also explain the contradictory results of other studies analyzing the effect of PUFAs on breast cancer risk [46-48].

Investigators from the Black Women's Health Study, a prospective cohort study, identified a dietary pattern similar to our Fatty Diets called "Western diet" also based on a PCA with Varimax rotation and factor loadings for dairy products and meat at similar level [49]. However, they associated a lower risk for breast cancer only with another dietary pattern, the "prudent diet", characterized with a low consumption of meat and dairy products. Since both the Western and the prudent diet were more complex than in our study with each diet having more than 8 foods with factor loadings above 0.4, it is not known whether the non-relationship between the Western diet and breast cancer has been masked by a higher consumption of potentially preventive foods which in turn result in a high P/S ratio.

Effect of alcohol on dietary patterns risk association

The reported consumption of the local banana beer, *Mbege*, increased significantly with increased affiliation to the Fatty Diet (*P* for trend <0.001). Our data show a higher breast cancer risk for women mainly following the Banana Diet which was also associated with a high consumption of *Mbege*, even though there is no risk association between alcohol intake and breast cancer risk in this study. According to the WRCF panel there is ample and generally consistent evidence from case-control and cohort studies that alcoholic drinks are a cause of pre- and post-menopausal breast cancer [7]. In order to exclude a possible bias in the risk estimation of dietary behaviour and breast cancer risk we generated a second set of dietary patterns excluding the alcoholic beverages from the factor analysis. The risk-increasing effect of the Fatty Diet remained slightly less pronounced when alcoholic beverages were singularized and added separately into the risk estimation model. On the contrary, the *Mchicha* Diet and the Banana Diet were no longer associated with breast cancer risk although the latter is characterized by rapidly absorbable carbohydrates. Such carbohydrates have been associated with increased breast cancer risk [50, 20]. One would assume that if the alcoholic beverages did influence the risk estimation of the *Michicha* and Banana Diet, the analysis keeping alcoholic beverages as separate food groups should visualize an increased risk association. However, the odds ratio of *Mbege* as well as bottled beer and wine was estimated to be 1.00 (95 % CI 1.00 – 1.00, P = 0.08 and 0.87 respectively) indicating no risk association. In our study the alcohol consumption was 8.2 g/d which is well below the recommended maximum intake of one drink per day in the European code against cancer [51]. Thus, the alcohol intake in general was probably too low to show an effect.

Fruity Diet

The Fruity Diet identified in the second PCA - keeping alcoholic beverages separate - was also associated with increased breast cancer risk. This diet is characterized by a high consumption of fish, mango, papaya, avocados and watery fruits like oranges, watermelons and pineapples which are known for their high content of valuable fatty acids, vitamins, and micronutrients considered as potentially protective against cancer [52-56]. Nevertheless, several other studies could not show an overall association between fruit and vegetable intake and breast cancer risk [57, 58]. In our study the Fruity Diet is, like the Fatty Diet, inversely associated with the P/S ratio (P_{trend} <0.001) which is caused by a reduced intake of polyunsaturated fatty acids mainly from sunflower oil (P_{trend} <0.001) but a stable saturated fatty acid intake (P_{trend} = 0.19). Thus we concluded that it is not the fish and fruit intake but the accompanying dietary fat consumption which is associated with breast cancer (Figure 2).

Property level

Socioeconomic status (SES) is an international acknowledged indicator in epidemiological, economic and sociological studies. However there is no international consensus on assessing SES; income and household expenditures being the most commonly used measures of SES [59]. In low income countries like Tanzania these indicators are difficult to assess. Often poverty or possession scores are used instead. Studies from low-income countries looking at socioeconomic status and health have shown that a possession score might be even a better indicator of SES as this score allows greater discrimination in identifying health risks than a poverty index [59]. In this study a possession score called property which was used as proxy for SES showed an inverse association with breast cancer risk. This seems at odds with the statement that higher education and socioeconomic status are associated with an increased risk resulting from the lower number of parities and lactations. However parity and lactation was correlated with educational level only. Furthermore educational level was negatively correlated with age indicating a trend towards higher education among the younger women. Education might have an impact on breast cancer risk estimation in the future following the expectation that lactation and parity will reduce over time with increasing educational level. In addition, it is expected that the trend towards higher education and fewer children, thus reduced lifelong lactation will continue especially with all the efforts towards the MDGs.

We did not find a correlation between lactation, parity and property level. In the context of this study "low property level" means people can call a bicycle or a radio their own. If they own both they already belong into the group at "medium property level". Thus, any extra income is used first to improve basic living conditions like nutrition, sanitation and health before it is used for education. Alderman [60] points out that the relationship between possessions to nutrition provides only an indirect answer looking at social transfer programmes in low-income countries aiming at improvement in nutrition and health care seeking behaviour. However according to Hou *et al.* [61] it may not be surprising to observe an inverse association between SES and breast cancer risk as studies have shown that people with low SES develop triple-negative subtypes which accounts for a substantial proportion of breast cancer in Africa. Nevertheless they required confirmation by larger population-based studies.

Strengths and limitations

Our data show the impact of reproductive and lifestyle factors on breast cancer aetiology of women in the Kilimanjaro Region. With regard to eating habits and dietary patterns, the diversity of the Kilimanjaro diet is low and it was less likely to miss important foods on the FFQ food list reducing the estimation bias for dietary behaviour. Due to low education levels and the poor infrastructure we do not expect socially desirable answers and participants are less likely to be informed about possible dietary impacts on health outcomes. The semi-quantitative FFQ allowed us to identify non-consumers and frequent consumers on the basis of eating habits, nutrient and energy intake in the case and control groups. Also ready-to-use meals and eating out are uncommon in the Kilimanjaro region, which facilitated the identification of food groups based on single food items and less on complex meals.

The sample is relatively small compared to studies in Westernized countries. Therefore we controlled for sampling adequacy using the Kaiser-Meyer-Olkin measure (KMO) before running a Principal Component Analysis (PCA). The KMO which was 0.62 is considered as mediocre in our case [62, 63]. A low KMO might result in a high unexplained variance. However, in this study we extracted 6 factors explaining 40.3 % variance which is a medium result compared to other studies e.g.: Hu et al = 2 factors: 20%, Arkkola et al = 7 factors: 29.5%, Shi et al = 4 factors: 28.5%, Lau et al = 2 factors: 17.1% [63-67]. In addition, Bartlett's Test for Sphericity was 2599.25, $p < 0.001$ indicating that correlations between items were sufficiently large for a PCA. The sample size required to achieve a high level of power in a logistic regression depends on the number of predictors and the size of the expected effect. Several studies showed that a sample size like in our studies allows detecting large and medium effects but might miss small effects. Thus, the results of our study are moderately powered and need to be confirmed by studies with a larger study population.

A major limitation is the PCA method. There have been discussions that the PCA method is less suitable for risk estimations of dietary patterns, because of difficulties to find plausible linkages between dietary patterns and the observed disease [29]. Therefore, it was recommend to use reduced rank regression based on response variables. However breast cancer developes over a long period of time. Thus, using response variables - such as biochemical parameters - is only possible in prospective studies.

The knowledge about breast cancer and breast self-examination was very poor in our study population and facilities for cancer diagnosis and treatment are still rare in countries like Tanzania [68]. In order to avoid a bias, we excluded the family history data for cancer from the analysis.

In the absence of a general health insurance, patients have had to pay for getting access to the health facilities. With the aim to minimize confounding errors due to different livelihood systems between cases and controls, we decided to select the controls also from within the hospital setting. But thereby, other selection biases cannot be excluded.

In conclusion a dietary pattern rich in fat and characterised by a low P/S ratio may be associated with a higher risk of breast cancer. The fatty acid composition is probably more important than total fat intake for the breast cancer risk.

Acknowledgements:

We thank the participating women of the Kilimanjaro Region and the members of the research team. Greatly appreciated is the input of: Mark Swai for his contribution to the study design and implementation, Joel Julia Mfinanga for conducting the interviews and physical examinations, Johannes Herrmann for statistical guidance and data analysis, and Emma Sykes and Margaret Wightman for the English editing.

The contributions of the authors were as follows:

Irmgard Jordan assessed, analyzed and interpreted the data and drafted the manuscript; Antje Hebestreit and Michael Krawinkel were principal investigators and contributed to the manuscript; Britta Swai conducted the study implementation and data entry in Tanzania.

Conflict of interest notification/ financial support:

The study was partially funded by Henskes Ltd., Laatzen, Germany, Plural, Sachse and Heinzelmann Ltd., Germany, as well as the German Academic Exchange Service (DAAD). The authors have no conflict of interest

References:

1. Smith-Warner SA, Spiegelmann D, Yaun SS *et al.* (1998) Alcohol and breast cancer in women. JAMA 279, 7: 535-40
2. Fioretti F, Tavani A, Bosetti C, *et al.* (1999) Risk factors for breast cancer in nulliparous women. Br J Cancer 79 (11/12): 1923-28
3. Anonymous (2001) Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. The Lancet 358: 1389-99
4. Beral V, *et al.* (2002) Alcohol, tobacco and breast cancer – collaborative reanalysis of individual data from 53 epidemiological studies, including 58 515 women with breast cancer and 95 067 women without the disease. BJ Cancer 87, 11: 1234-45
5. Bofetta P, Hashibe M (2006) Alcohol and cancer. Lancet Oncol 7: 149-56
6. Evans DGR, Howell A (2007) Breast Cancer risk-assessment models. Breast Cancer Res 9 (5): 213
7. World Cancer Research Fund, American Insitute for Cancer Research (2007) Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR
8. van den Brandt PA, Spiegelmann D, Yaun SS, *et al*. (2000) Pooled Analysis of prospective cohort studies on height, weight, and breast cancer risk in: Am J Epidemiol 152 (6): 514-27
9. Key T, Appleby P, Barnes I, Reeves G (2002) Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 94: 606-16
10. Key T, Appleby PN, Reeves GK, *et al*. (2003) Body Mass Index, Serum Sex Hormones, and Breast Cancer Risk in Postmenopausal Women. J Natl Cancer Inst 95 (16): 1218-26
11. Ahlgren M, Melbye M, Wohlfahrt J, Thorkild IA, Sørensen MD (2004) Growth Patterns and the Risk of Breast Cancer in Women. N Engl J Med 351:1619-6
12. Lee HP, Gourley L, Duffy SW, Estève J, Lee J, Day NE (1991) Dietary effects on breast cancer risk in Singapore. The Lancet 337: 1197-200
13. Hunter DJ, Spiegelmann D, Adami HO, *et al.* (1996) Cohort studies of fat intake and the risk of breast cancer – a pooled analysis. N Engl J Med 334:356-61
14. Wakai K, Drupadi SD, Ohno Y, *et al.* (2000) Fat intake and breast cancer risk in an area where fat intake is low: a case-control study in Indonesia. Int J Epidemiol 29: 20-28
15. Mattisson I, Wirfält E, Andrén C, Gullberg B, Berlund G (2003) Dietary fat intake – food sources and dietary correlates in the Malmö Diet and Cancer cohort. Public Health Nutr 6,6: 559-69
16. Cho E, Spiegelmann D, Hunter DJ, *et al.* (2003) Premenopausal fat intake and risk of breast cancer. J Natl Cancer Inst 95: 1079-85

17. Prentice RL, Thomson CA, Caan B, *et al.* (2007) Low-fat dietary pattern and cancer incidence in the women's health initiative dietary modification randomized controlled trial. J Natl Cancer Inst 99, 20: 1534-43
18. Engeset D, Dyachenko A, Ciampi A, *et al.* (2009) Dietary patterns and risk of cancers of various sites in the Norwegian European Prospective Investigation into Cancer and Nutrition cohort: the Norwegian women and cancer study. Eur J Cancer Prev 18: 69-75
19. Bosetti C, Pelucchi C, La Vecchia C (2009) Diet and cancer in Mediterranean countries: carbohydrates and fats. Public Health Nutr 12(9A): 1595-1600
20. Brennan SF, Cantwell MM, Cardwell CR, *et al.* (2010) Dietary patterns and breast cancer risk: a systematic review and meta-analysis. Am J Clin Nutr 91: 1294-302
21. Jordan I, Hebestreit A, Swai B, Krawinkel MB (2010) Breast cancer risk among women with long-standing lactation and reproductive parameters at low risk level: a case-control study in Northern Tanzania. Breast Cancer Res Treat Published online 20 November 2010, DOI: 10.1007/s10549-010-1255-7
22. Hebestreit A (2004) Brustkrebs und Ernährung. Inaugural-Dissertation zur Erlangung des Grades eines Doktor oec.troph. im Fachbereich „Agrarwissenschaften, Ökotrophologie, Umweltmanagement" der Justus-Liebig Universität Gießen http://geb.uni-giessen.de/geb/volltexte/2004/1544/pdf/HebestreitAntje-2004-05-17.pdf latest access January 12, 2011
23. Landig J, Erhardt JG, Bode JC, Bode C (1998) Validation and comparison of two computerized methods of obtaining a diet history. Clin Nutr 17 (3): 113-7
24. Erhardt J (2010) http://www.nutrisurvey.de/ latest access January 14, 2011
25. Chen Y, Ahsan H, Parvez F, *et al.* (2004) Validity of a food-frequency questionnaire for a large prospective cohort study in Bangladesh. British Journal of Nutrition, 92(05), 851
26. MacIntyre U, Venter C, Vorster H (2001) A culture-sensitive quantitative food frequency questionnaire used in an African population: 2. Relative validation by 7-day weighed records and biomarkers. PHN, 4(01). Available at: http://www.journals.cambridge.org/abstract_S136898000100009X latest access May 8, 2011
27. Torheim LE, Barikmo I, Hatløy A, *et al.* (2001) Validation of a quantitative food-frequency questionnaire for use in Western Mali. PHN 4(06). Available at: http://www.journals.cambridge.org/abstract_S1368980001001161 latest acess May 8, 2011

28. McCann SE, Weiner J, Graham S, Freudenheim JL (2001) Is principal components analysis necessary to characterize dietary behavior in studies of diet and disease? Public Health Nutr 4,4: 903-8
29. Hoffmann K, Schulze MB, Schienkiewitz A, Nöthlings U, Boeing H (2003) Application of a new statistical method to derive dietary patterns in nutritional epidemiology. Am J Epidemiol 159: 935-44
30. Costello AB, Osborne J (2005) Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, 10(7) http://pareonline.net/getvn.asp?v=10&n=7 latest access March 24, 2011
318. Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelmann D, Willet WC (2000) Prospective study of major dietary patterns and risk of coronary heart disease in men. Am J Clin Nutr 72: 912-21
32. Fung TT, Willet WC, Stampfer MJ, Manson JE, Hu FB (2001) Dietary patterns and the risk of coronary heart disease in women. Arch Intern Med 161: 1857-62
33. Handa K, Keiger N (2002) Diet patterns and the risk of renal cell carcinoma. Public Health Nutr 5: 757-67
34. Kim MK, Sasaki S, Otani T, Tsugane S (2004) Prospective of three major dietary patterns and risk of gastric cancer in Japan. Int J Cancer 110: 435-42
35. Kim MK, Sasaki S, Otani T, Tsugane S (2005) Dietary patterns and subsequent colorectal cancer risk by subsite: a prospective cohort study. Int J Cancer 115: 790-8
36. Cureton, E. E./ D'Agostino, R. B. (1983) Factor analysis: an applied approach. Hillside, New Jersey. Lea Lawrence Erlbaum Associates, Publishers. p389-90
37. Backhaus K (2008) Multivariate Analysemethoden: eine anwendungsorientierte Einfu□hrung. 12th ed. Berlin [u.a.]: Springer; 575 p.
38. Schulz M, Hoffmann K, Weikert C, Nöthlings U, Schulze MB, Boing H (2008) Identification of a dietary pattern characterized by high-fat food choices associated with increased risk of breast cancer: the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Br J Nutr 100: 942-6
39. Thiébaut ACM, Kipnis V, Chang SC, *et al.* (2007) Dietary fat and postmenopausal invasive breast cancer in the National Institutes of Health-AARP diet and health study cohort. J Natl Cancer Inst 99: 451-62
40. Wynder EL, Cohen LA, Muscat JE, Winters B, Dwyer JT, Balckburn G (1997) Breast cancer: weighing the evidence for a promoting role of dietary fat. J Natl Cancer Inst 89: 766-75
41. Sieri S, Krogh V, Ferrari P, *et al.* (2008) Dietary fat and breast cancer risk in the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr 88:1304-12

42. Goodstine SL, Zheng T, Holford TR, *et al.* (2003) Dietary (n-3)/(n-6) fatty acid ratio: possible relationship to premenopausal but not postmenopausal breast cancer risk in US Women. J Nutr 133: 1409-14

43. Rose DP, Conolly JM (1999) Omega-3 fatty acids as cancer chemopreventive agents. Pharmacol Ther 83: 217-44

44. Murff HJ, Shu X-O, Li H, *et al.* (2010) Dietary polyunsaturated fatty acids and breast cancer risk in Chinese women: a prospective cohort study. Int J Cancer: first published online: 23 Nov 2010, DOI: 10.1002/ijc.25703

45. Bougnoux P, Giradeau B, Couet C (2006) Diet, cancer, and the lipidome. Cancer Epidemiol Biomarkers Prev 15: 416-21

46. Holmes MD, Hunter DJ, Colditz GA, *et al.* (1999) Association of dietary intake of fat and fatty acids with risk of breast cancer. JAMA 281 (10): 914-20

47. Saadatoan-Elahi M, Norat T, Goudable J, Riboli E (2004) Biomarkers of dietary fatty acid intake and the risk of breast cancer: a meta-analysis. Int J Cancer 111 (4): 584-91

48. MacLean CH, Newberry SJ, Mojica WA, *et al.* (2006) Effects of omega-3 fatty acids on cancer risk: a systematic review. JAMA 295 (4): 403-415

49. Argurs-Collins T, Rosenberg L, Makambi K, Palmer JR, Adams-Campbell L (2009) Dietary patterns and breast cancer risk in women participating in the Black Women's Health Study. Am J Clin Nutr 90: 621-8

50. Lajous M, Boutron-Ruault MC, Fabre A, Clavel-Chapelon, Romieu I (2007) Carbohydrate intake, glycemic index, glycemic load, and risk of postmenopausal breast cancer in a prospective study of French women. Am J Clin Nutr 87:1384-91

51. Boyle P, Autier P, Bartelink H, *et al.* (2003) European code against cancer and scientific justification: third version. Ann Oncol 14: 973-1005

52. Longnecker MP, Newcom PA, Mittendorf R, *et al.* (1997) Intake of carrots, spinach, and supplements containing vitamin A in relation to risk of breast cancer. Cancer Epidemiol Biomarkers Prev 6 (11): 887-92

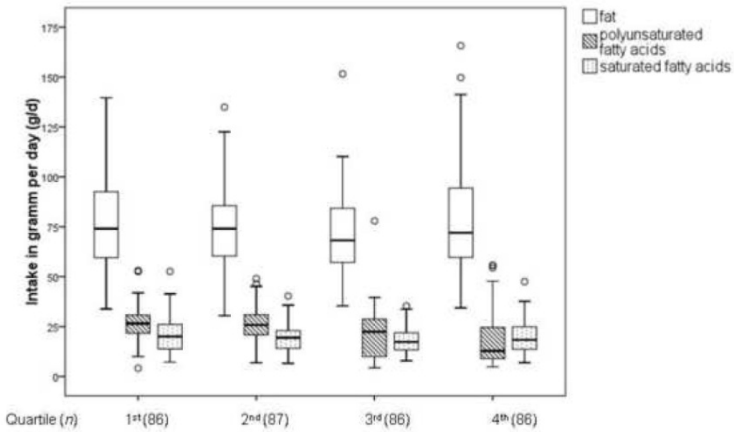
53. Malin AS, Qi D, Shu XO, *et al.* (2003) Intake of fruits, vegetables and selected micronutrients in relation to the risk of breast cancer. In J Cancer 105 (3): 413-8

54. Zhang CX, Ho SC, Chen YM, *et al.* (2009) Greater vegetable and fruit intake is associated with lower risk of breast cancer among Chinese women. Int J Cancer 124 (12): 2929-37

55. Kim J, Lim SY, Shin A, *et al.* (2009) Fatty fish and fish omega-3 fatty acid intakes decrease the breast cancer risk: a case-control study. BMC Cancer 9: 216

56. Bofetta P, Couto E, Wichmann J, *et al.* (2010) Fruit and vegetable intake and overall cancer risk in the European Prospective Investigation into Cancer and nutrition (EPIC). J Natl Cancer Inst 102: 529-37
57. Smith-Warner SA, Spiegelmann D, Yaun SS, *et al.* (2001) Intake of fruits and vegetables and risk of breast cancer: a pooled analysis of cohort studies. JAMA 285 (6): 769-76
58. van Gils CH, Peeters PH, Bueno-de-Mesquita HB, *et al.* (2005) Consumption of vegetables and fruits and risk of breast cancer. JAMA 293 (2): 183-93
59. Mohsena M, Mascie-Taylor CN, Goto R (2010) Association between socio-economic status and childhood undernutrition in Bangladesh; a comparison of possession score and poverty index. PHN 13, 1498-1504
60. Alderman H (2010) What do we want to know about the relationship of wealth or income to nutrition? PHN 13(10): 1485-1486.
61. Hou N, Huo D, Olopade OI (2011) Protective effect of longstanding lactation and reproductive factors: a case–control study in North Tanzania. Breast Cancer Research and Treatment. Available from: http://www.springerlink.com/index/10.1007/s10549-011-1391-8 latest access November 30, 2011
62. Kaiser, HF (1970) A second generation little jiffy. Psychometrika **35**: 401-415
63. Field A (2009) Discovering statistics using SPSS : (and sex and drugs and rock “n” roll) SAGE: London
64. Hu FB, Rimm E, Smith-Warner SA, *et al*. (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am. J. Clin. Nutr. Feb;69(2):243–9
65. Arkkola T, Uusitalo U, Kronberg-Kippilä C, *et al.* (2007) Seven distinct dietary patterns identified among pregnant Finnish women – associations with nutrient intake and sociodemographic factors. PHN. Available from:http://www.journals.cambridge.org/abstract_S1368980007000420 latest access April 7, 2011
66. Shi Z, Hu X, Yuan B, Hu G, *et al.* (2008) Vegetable-rich food pattern is related to obesity in China. Int J Obes (Lond). Jun;32(6):975–84.
67. Lau C, Glümer C, Toft U, et al. (2007) Identification and reproducibility of dietary patterns in a Danish cohort: the Inter99 study. BJN [Internet]. 99(05). Available from: http://www.journals.cambridge.org/abstract_S0007114507837494 latest access April 7, 2011
68. Okobia MN, Bunker CH, Okonofua FE, Osime U (2006) Knowledge, attitude and practice of Nigerian women towards breast cancer: A cross-sectional study. World J Surg Oncol 4: 11

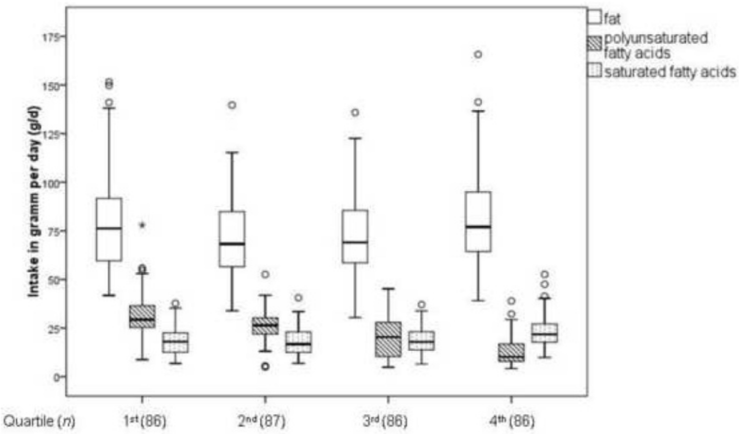
Figure
Click here to download high resolution image



Odds ratio	Reference	0.7 (0.4-1.6)	0.9 (0.4-1.8)	1.6 (0.7-3.3)	
P/S-ratio	1.4	1.4	1.3	0.7	P_{trend} <0.001

Figure 1: **Intake of fat, polyunsaturated and saturated fatty acids per day and its related odds and P/S ratios in quartiles of the Fruity Diet**

Figure
Click here to download high resolution image



Odds ratio	Reference	1.1 (0.5-2.4)	1.9 (0.9-3.9)	3.3* (1.5-7.2)	* $P < 0.05$
P/S-ratio	1.8	1.5	1.1	0.5	P_{trend} <0.001

Figure 1: **Intake of fat, polyunsaturated and saturated fatty acids per day and its related odds and P/S ratios in quartiles of the Fatty Diet**

Table

List of tables:

Table 1: Selected socio-economic and reproductive indicators

beginning of results – manuscript page 6

Table 2: Results of Principal component analysis with Varimax rotation (PCA 1)

close to description of first PCA – manuscript page 7

Table 3: Results of logistic regression, unadjusted

close to description of logistic regression model – manuscript page 8

Table 4: Basic breast cancer risk model and dietary patterns

close to description of logistic regression model – manuscript page 8

Table 1: Selected socio-economic and reproductive indicators [21]

Variable	Cases			Controls			*P* value*
	Median (min-max)		n	Median (min-max)		n	
Age (years)	50	(28-85)	115	50	(26-83)	230	0.620
Age at Menarche (years)	16	(11-20)	111	16	(13-20	230	0.267
Age at first full term pregnancy (years)	20	(14-35)	106	20	(13-41)	217	0.571
Number of children ⁂	5	(1-10)	106	5	(1-9)	217	0.219
Lifelong lactation (months)	90	(0-240)	114	108	(0-240)	230	0.045
Schooling (%)							0.119
Less than 3 years		27			18		
Finished primary school		54			59		
Finished secondary school		19			23		
Property level (%)							<0.0001
Low		47			18		
Medium		45			65		
High		8			17		
Women with children (%)		92			94		0.515

⁂ only parous women, * Mann-Whitney *U* Test: differences between cases and controls

Table 2: Results of rotated Principal component analysis (PCA 1)

Food item	Diet of the Rich	*Mchicha* Diet	Banana Diet	Fatty Diet
Variance explained (%)	**9.1**	**7.9**	**7.6**	**5.3**
Rice	0.618	0.205	-0.143	-0.170
Nuts	0.587	-0.006	0.124	-0.089
Egg	0.557	-0.039	0.162	0.043
*Chapati**	0.556	0.062	0.055	0.009
Leguminous vegetables	0.537	-0.093	0.006	-0.026
Bread	0.503	0.362	-0.220	-0.190
Soda drinks	0.471	0.108	-0.028	-0.155
Red meat	0.453	0.103	-0.037	0.367
Mchicha†	-0.017	0.645	0.029	0.110
Cucumber & okra	0.209	0.581	0.032	0.038
Onion	0.089	0.579	-0.042	0.138
Carrots & tomatoes	0.145	0.516	-0.096	-0.007
Maize	-0.180	0.461	0.135	-0.085
Fish	-0.018	0.434	0.337	-0.085
Avocado	-0.016	0.413	0.347	0.067
Banana	0.145	0.030	0.667	0.073
Green (cooking) banana	0.086	0.008	0.616	-0.176
Sugar	0.153	-0.103	0.491	-0.166
Watery fruits♯	0.085	0.189	0.478	-0.218
Starchy tubers	-0.275	-0.063	0.461	0.136
Mbege‡	-0.295	0.050	0.442	0.246
Pulses	-0.070	0.281	0.415	0.134
Sunflower oil	0.203	-0.207	-0.071	-0.623
Milk	0.264	-0.079	-0.042	0.521
Butter and lard	-0.213	-0.254	0.055	0.457
Mixed vegetable fats and oil	0.263	0.191	-0.115	0.454
Tea	0.055	0.013	0.366	-0.410

Food groups with factor loadings < 0.4and >-0.4: potatoes, juice, chicken meat, mango & papaya, cabbage (white), *mandazi*§, *uji*∫, coffee, bottled beer and wine. Rotation method Varimax with Kaiser Normalization. Rotation converged in 7 iterations.

* unleavened East African flat wheat bread; † traditional Tanzanian food, synonymously used for a dish of amaranth leaves and e.g. onions, tomatoes or/ and carrots in various amounts; ‡ often homemade opaque beer from bananas and millet; § East African donuts; ∫ thin millet or maize based porridge; ♯ oranges, watermelon and pineapple

Table 3: Results of the logistic regression: dietary patterns only

Variable	*P* value	Odds Ratio	95% CI	*n*
Dietary patterns (PCA 1)				
Diet of the Rich	0.95	1.01	0.79 -1.30	345
Mchicha Diet	0.00	1.47	1.14 -1.88	345
Banana Diet	0.00	1.94	1.43 -2.63	345
Fatty Diet	0.00	1.62	1.26 -2.07	345

Adjusted for age.
Constant: *P* value <0.01, OR 0.56; Cox & Snell R^2 = 0.13 ; Nagerkerke R^2 = 0.18
Overall percentage correctly classified 74%

Table 4: Basic breast cancer risk model and dietary patterns

Variable	*P* value	Odds Ratio	95% CI	*n*
Property level				
Low	Reference			87
Medium	0.00	0.39	0.21 -0.72	198
High	0.01	0.26	0.09 -0.75	48
Body mass index (kg/m²)				
At 20 years	0.01	1.26	1.05 -1.51	333
At interview	0.11	0.93	0.86 -1.02	333
Age at first full term pregnancy				
≤ 20years	Reference			193
> 20 years	0.08	1.74	0.93 -3.23	122
No pregnancy	0.79	0.82	0.18 -3.70	18
Lifelong lactation	0.02	0.99	0.98 -1.00	333
Dietary patterns (PCA)				
Diet of the Rich	0.40	1.15	0.83 -1.59	333
Mchicha Diet	0.08	1.28	0.97 -1.70	333
Banana Diet	0.00	1.75	1.20 -2.57	333
Fatty Diet	0.01	1.50	1.12 -1.99	333

Adjusted for age, place of living, age at menarche, and menopausal status.

Constant: *P* value = 0.81, OR 0.54; Cox & Snell R^2 = 0.20; Nagelkerke R^2 = 0.29; Overall percentage correctly classified = 76%.