

# The modified diet history methodology of the Malmö Diet Cancer cohort

**Edited by Elisabet Wirfält, Emily Sonestedt**

**Resource persons:**

- Dietary data collection at baseline:

Ulrika Ericson (ulrika.ericson@med.lu.se)

Peter Wallström (especially supplements) (peter.wallstrom@med.lu.se)

- Consultation on dietary data handling and analysis:

Ulrika Ericson (ulrika.ericson@med.lu.se)

Isabel Drake (isabel.drake@med.lu.se)

Emily Sonestedt (emily.sonestedt@med.lu.se)

...

Research group of Nutritional Epidemiology;  
Diabetes and cardiovascular disease – genetic epidemiology,

Lund University  
Department of Clinical Sciences Malmö  
CRC building 60 level 13, Jan Waldenströms gata 35  
SE-205 02 Malmö, Sweden

## Contents

1. Introduction.....	3
Dietary data .....	3
2. The modified diet history method used in the Malmö Diet and Cancer study.....	3
Dietary data collection.....	4
Dietary variables .....	9
Basic food groups (G-variables) .....	10
Cooked foods (R-variables).....	11
Energy and nutrients from foods.....	11
Energy and nutrients from supplements.....	15
3. Validation study: Malmö Food Study .....	15
4. Considerations regarding statistical power using the MDC diet history.....	18
5. Other methodological papers from the MDC cohort.....	19
Change in interview routines in 1993 of the MDC diet history.....	19
Mis-reporting of energy intakes using the MDC diet history.....	20
6. Recommendations regarding dietary data handling and data analysis.....	22
7. Recommended literature – suggestions.....	25
8. References.....	25

## 1. Introduction

Nutritional epidemiology uses epidemiological methods to examine associations between diet and disease. The interpretation of observed associations in nutrition epidemiology is a challenge because of the complexities surrounding “diet” as an exposure.

- Food contains a vast number of bio-active components. We do not eat single nutrients or food items, but different amounts and frequencies of food combinations. Since many nutrients and dietary components are present in the same foods, co-variation between dietary variables is common, making the study of “independent effects” of single nutrients very difficult.
- Diet is part of the overall lifestyle and culture of an individual, causing dietary factors to co-vary with many lifestyle and socio-economic factors.
- Nutritional epidemiology typically uses self-reported data, sometimes collected with brief questionnaires that may be difficult to complete accurately by responders, making the presence of measurement errors inevitable.

### Dietary data

Self-reported dietary data is commonly collected through questionnaires, diaries or interviews. These methods vary in their ability to assess diet accurately. The perfect dietary assessment method does not exist. Therefore a thorough understanding of the methods, their strengths and their limitations, is essential when handling and analyzing dietary data and interpreting outcomes in nutritional epidemiology. Collaboration with experienced nutrition epidemiologists and the inclusion of a nutritionist in the study team is recommended for any research project seeking to measure diet or analyzing dietary data.

Dietary assessment methods can be classified as “current diet” methods and “usual diet” methods. The former collects information about what is actually eaten during one or several single days. Typically, a diary is kept and everything eaten during the day is recorded. Today, a “usual diet” method is most often a food frequency questionnaire. Respondents are typically presented with a list of foods or food groups, and asked how often they eat certain foods (“what do you usually eat?”). With a semi-quantitative questionnaire the individual is also asked to indicate the typical portion size chosen.

The information (what?, how much?, how often?) is summarized into the average amount of food consumed (in g) per day. Most studies also convert the food information (by using a food database) to the average intakes of energy (kcal, MJ) and nutrients (g, mg, µg) consumed each day.

## 2. The modified diet history method used in the Malmö Diet and Cancer study

The Malmö Diet and Cancer (MDC) study started in March 1991 (4). Until May 1995 the participants were recruited from the total cohort of people living in Malmö born between 1926 and 1945. After that date the cohort was extended to include men born between 1923 and 1945 and women born between 1923 and 1950. Eligible were those living in the City of

Malmö and with Swedish reading and writing skills. People with mental incapacity who were not able fulfil the baseline examinations were not included. When recruitment closed, more than 30 000 individuals had joined the study, and 28 098 persons had completed all baseline examinations (5) (6). The data collection included dietary habits, socio-economics, medical history, and lifestyle habits using questionnaires and interview. Anthropometrics, body composition, and blood pressure were collected through direct measurements. Blood samples were collected, frozen, and stored for biochemical analysis at a later stage. Participants visited the study centre twice. During the first visit, the study procedures and questionnaires were explained and distributed, participants gave their informed written consent, direct measurements were made and blood samples collected. Approximately two weeks later, the questionnaires completed at home were reviewed and the diet history interview conducted.

### Dietary data collection

The modified diet history methodology was especially developed for the MDC study. The choice of methodology was guided by the need to assess total diet in a middle-aged and older urban population where the daily eating habits regularly included traditional cooked meals and infrequently included fast foods and meals eaten out of home. The development of the method was led by Eva Callmer during the 1980's and by Irene Mattisson and Ulla Johansson during the early 1990's.

The method consists of three parts:

- a) a 7-day food diary ("Menybok")
- b) a 168-item semi quantitative diet history questionnaire ("Kosthistoria")
- c) a 45-60 min diet history interview.

This methodology combined a "current diet" method (food diary) with a "usual diet" method (diet history questionnaire). The combination of the two parts reflects total diet. The relative validity of this methodology, compared to similar methodologies in other populations, has proven to be very high (1-3).

In the **7-day food diary ("menu book")**, participants recorded "cooked/main meals" (such as lunch and dinner, i.e. meals that you eat on a daily basis but with day-to-day variation regarding content), cold beverages (i.e., milk, juice, soft drinks, water and alcoholic beverages), drugs, natural remedies, and dietary supplements during 7 consecutive days (Figure 1). Breakfast was not included.

## Exempel

<b>Dag 1</b>		Veckodag <i>tisdag</i>	Datum <i>22/10-91</i>
Måltid Plats	Livsmedel		Beskrivning
<i>middag</i>	<i>sillflundra</i>		<i>panerad, stekt i milda</i>
<i>hemma</i>	<i>sås</i>		<i>steksky + vispgrädde</i>
	<i>potatis</i>		<i>kokt</i>
	<i>grönsaksblandning m. en klick bregott</i>		<i>ärter, majs, paprika</i>
	<i>lingonsylt</i>		
	<i>jordgubbskräm m. mjölk</i>		<i>Bob, gammaldags</i>
<i>kvällsmat</i>	<i>pizza</i>		<i>m. skinka</i>
<i>restaurang</i>	<i>sallad</i>		<i>vitkål, paprika</i>
	<i>dressing</i>		<i>olja + vinäger</i>
Dryck	Mängd under dygnet	Dryck	Mängd under dygnet
Saft		Vin, starkvin	2,5 dl
Läsk	1 burk (33 cl)	Starksprit	
Läsk; light		Vatten	2 dl
Apelsinjuice	2 dl	Mineralvatten	33 cl
Lättöl kl I	33 cl	Lätt – minimjök	
Folköl kl II		Mellanjök	
Starköl, mellanöl		Standardmjök	
Cider		Gammaldagsmjök	
Annan dryck (ej kaffe, te och choklad) , vilken?		<i>Äppeljuice</i>	3 dl
Annan dryck (ej kaffe, te och choklad) , vilken?			
Vitaminer, mineraler, naturmedel och läkemedel			Mängd
Namn	Tillverkare		
<i>Multiplex comp N</i>	<i>Ferrosan</i>		1 tablett
<i>Alvedon</i>	<i>Astra</i>		3 tabl (500 mg/st)
<i>Laktulos</i>	<i>ACO</i>		30 ml (670 mg/ml)

Figure 1: Example from the 7-day food diary ("menu book")

In the **diet history questionnaire** the general meal pattern were first recorded (see Figure 2). Then, the frequency and portion-size information of foods consumed regularly (see Figure 3) and with low day-to-day variation (i.e., hot beverages (see example for coffee in Figure 3), sandwiches, edible fats, breakfast cereals, yoghurt, milk, fruits, cakes, candies and snacks) were recorded. The reference period of the questionnaire was the preceding year. At home

the participant estimated the usual portion-sizes of foods reported in the **questionnaire** from a booklet with 48 black and white photographs (Figure 4).

### Exempel

2

## Måltidsordning

Beskriv i stora drag de **måltider**, även mellanmål, som Du vanligen äter/dricker under en dag! Ange **måltidens namn**, **tid för måltiden**, **vad** den huvudsakligen består av och **hur många gånger per vecka** Du brukar äta den måltiden. Om **lördag** och **söndag** skiljer sig mycket från vardagarna, kan Du ange två måltidsordningar. Se **exempel**.

Måndag–fredag	Måltidens namn	Tid ca	Vad består den av?	Hur många gånger per vecka?
	<i>frukost</i>	<i>6.30</i>	<i>kaffe + 2 smörgåsar</i>	<i>5</i>
	<i>förmiddag</i>	<i>ca 10</i>	<i>frukt eller skorpa och te</i>	<i>5</i>
	<i>lunch</i>	<i>12</i>	<i>lagad mat, råkost, bröd, öl</i>	<i>5</i>
	<i>eftermiddagskaffe</i>	<i>14</i>	<i>kaffe + vetebulle</i>	<i>5</i>
	<i>kvällsmat</i>	<i>18</i>	<i>3–4 smörgåsar med olika pålägg</i>	<i>5</i>
	<i>kvällste</i>	<i>20</i>	<i>te + liten kaka (fredag kväll grogg, ost eller räkor)</i>	<i>4 + 1</i>

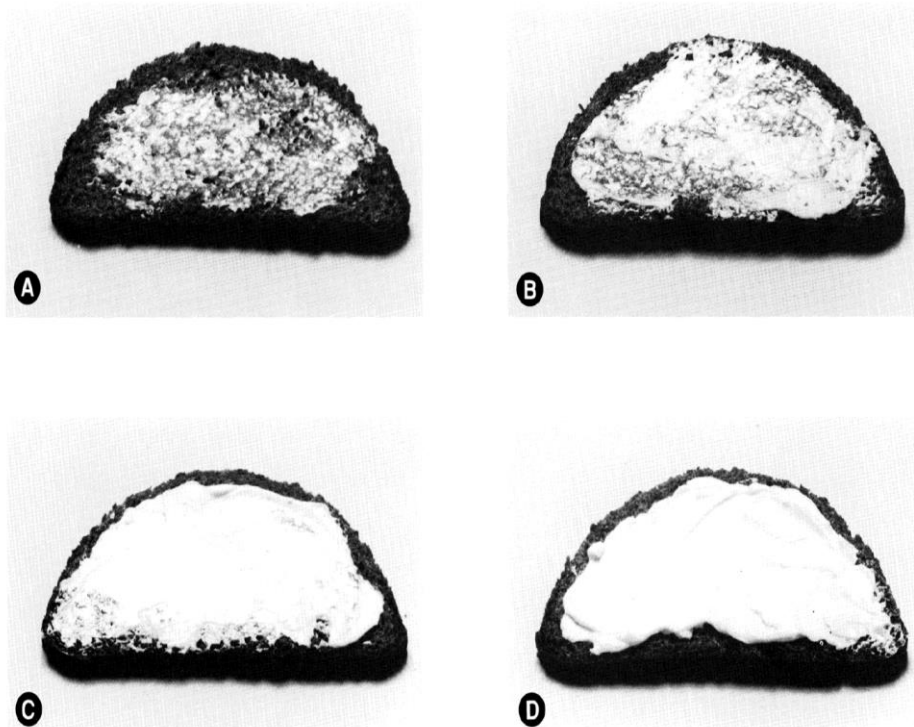
  

Lördag–söndag	Måltidens namn	Tid ca	Vad består den av?	Hur många gånger per vecka?
	<i>frukost</i>	<i>8</i>	<i>kaffe, gröt, smörgås, juice</i>	<i>2</i>
	<i>förmiddagskaffe</i>	<i>10.30</i>	<i>kaffe + kaka</i>	<i>2</i>
	<i>lunch</i>	<i>12</i>	<i>smörgåsar + öl</i>	<i>2</i>
	<i>eftermiddagskaffe</i>	<i>15</i>	<i>kaffe + wienerbröd</i>	<i>2</i>
	<i>middag</i>	<i>18</i>	<i>lagad mat, efterrätt, vin på lördag</i>	<i>2</i>
	<i>kvällste</i>	<i>20</i>	<i>te + skorpor</i>	<i>2</i>

Figure 2: Example of meal order recorded in the diet history questionnaire

	Dricker sällan eller aldrig	Antal koppar per dag	per vecka	Se bild	Mängd per kopp Ringa in
<b>Kaffe</b>					
Kaffe, svart _____		_____	_____	1	A B C D E F
Kaffe med mjölk eller grädde _____		_____	_____	1	A B C D E F
Lättmjölk, minimjölk i kaffe _____		_____	_____	2	A B C
Mellanmjölk i kaffe _____		_____	_____	2	A B C
Standardmjölk i kaffe _____		_____	_____	2	A B C
Gammaldagsmjölk i kaffe _____		_____	_____	2	A B C
Kaffegrädde i kaffe _____		_____	_____	2	A B C
Vispgrädde i kaffe _____		_____	_____	2	A B C
Socketer i kaffe _____		_____	_____	_____	bitar eller _____ tsk per kopp
Sötningemedel i kaffe _____		_____	_____	_____	bitar eller _____ tsk per kopp

Figure 3: Example from the diet history questionnaire (coffee)



**Figure 4: Example from the booklet of photographs “Bilderbok”**

During the **diet history interview** an extensive book of photographs was used to estimate usual portion-sizes of dishes and foods in the **food diary**. Typically, a set of 4 photos (A-D), displaying 4 different portion-sizes of the same dish, was shown to the participants. One set of photos was shown for each dish, or food, registered in the **food diary**. The participants were not limited to the amounts indicated by the photos, but were encouraged to describe their usual portion-sizes as exactly as possible. Thus, portion-sizes could be expressed in several ways i. e. “half the size of C”, “between B and C”, “D plus A” etc.

During the **diet history interview** participants were asked to describe in detail how foods and dishes recorded in the **food diary** were prepared (e.g. the type of fats used for cooking, specific ingredients in mixed dishes).

The food data was entered into the data system using the interactive computer software KOSTSVAR (AIVO AB). The interactive software guided the interviewer through a system of “recipe identifiers”. These specifically helped identifying preparation methods and ingredients in mixed dishes. A “recipe identifier” indicating the type of dish (e.g., casserole with meat) was first entered. The following menu on the screen listed potential codes, indicating the specific constituents of different casseroles. The interviewers chose the most appropriate code and made necessary adaptations of the recipe concurrently, depending on the information given by the participant.

The MDC diet history method included the option of exchanging a maximum of four ingredients in standard (default) recipes. The ingredient changes focused on the amount and quality of fat (type of dietary fat, liquid in sauces, casseroles, meat, and fish) and vegetables. In addition, the MDC method included the option of creating new individual recipes during the dietary interview. This procedure was used (by judgement of the dietary interviewer)

when standard recipes, with ingredient exchanges, did not cover the recipe described by the participant.

The **questionnaire** and the **food diary** were also checked, according to predefined rules (from extensive documentation/interview instructions for interviewers), so that reported food consumption did not overlap and were in concordance with the overall meal pattern reported by the participant.

The specific food information obtained from the **7-day food diary** and **diet history interview** was coded and entered during or shortly after the interview (at the second visit). The information from the **questionnaire** was entered by the dietary interviewer or secretarial staff shortly after the participants' second visit to the study centre.

Information from the **diet history questionnaire** on portion-sizes and frequencies was entered into the computer and converted into grams. Each food item and ingredient in the **food diary** was coded and summarized into the average amount of food consumed (g) per day. All food information from the two sources (i.e., the **food diary/diet history interview** and the **questionnaire**) was summarized into the average daily consumption of food groups (gram per day) for each individual. The food intake data was converted into nutrient intake data by the combined use of the interactive computer software, and the Swedish Food Database PC KOST2 -93 of the Swedish National Food Administration. PC KOST2-93 contains approximately 1600 basic foods; additional recipes and food codes were added specifically for the MDC study.

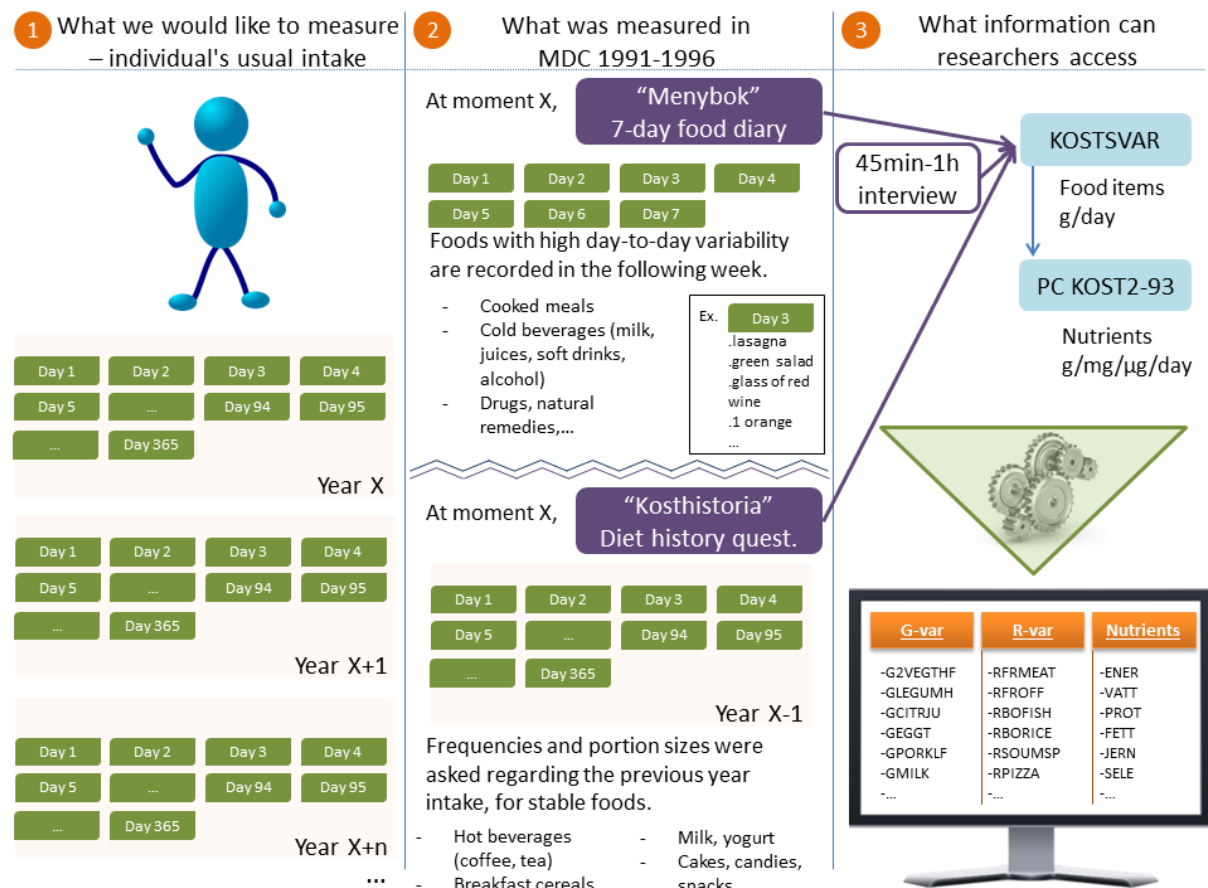
During the six-year baseline examination period, a total of fifteen interviewers conducted the dietary interviews. A variable indicating diet interviewer is available. Several measures were taken to facilitate standardized dietary data collection procedures across dietary interviewers: a continuous and extensive in-service training of interviewers; interactive computer software for standardized entering and coding of data; an extensive set of coding rules for food items and mixed dishes; and a quality control program of collected data. Weekly training sessions and bi-yearly workshops were conducted to discuss and solve problems related to coding and entering of dietary data. In addition, the two head nutritionists (Irene Mattisson (I.M.) and Ulla Johansson (U.J)) conducted weekly inspections of questionnaires and menu books (randomly selected from each dietary interviewer), and regularly listened in on dietary interviews. Extreme portion-sizes were identified through a monthly, computerised quality control routine, and were either verified or corrected if erroneous. In addition, the extreme and median values of total energy, all nutrients, and major food groups were regularly inspected, and erroneous values attended to. Finally, the age and gender specific ratios of total energy intake (EI) to basal metabolic rate (BMR) were computed, using the formula for BMR identified by a joint FAO/WHO/UNU expert consultation (7). Extreme and median values were identified, and the dietary reports of these individuals were checked for errors.

Due to reduction in grants, the dietary data collection method was changed to make the interviews faster (from 60 min to 45 min). The method used March 1991 to August 1994 is referred to the SLOW method, and the method used from 1<sup>st</sup> September 1994 until November 1996 is referred to as the FAST method. The changes involved the coding of



different dishes from food diary and the usage of standardized portion sizes. To estimate portion sizes in the food diary, the SLOW method used 180 photographs and the FAST method 75 photographs. See reference Wirfält 2002 for more detailed information.

A detailed description on what was paid attention to and what was not paid attention to can be found in a separate document.



**Figure 5. Overview of the diet assessment method in MDC**

## Dietary variables

Four dietary data files exist in the database:

1. Basic food groups (G-variables) ( file name: g1\_aggr.sav)
2. Cooked foods (R-variables) (file name: g2\_aggr.sav)
3. Energy and nutrient intake variables from the reported food intakes (file name: nvds\_kost\_91\_96.sav)
4. Energy and nutrient intake variables from the reported use of supplements (file name: nvds\_s\_91-96.sav)

## Basic food groups (G-variables)

The foods reported by the participant are in the G-food group variables delineated (i.e. converts mixed dishes) and described as basic food groups (i.e., vegetables, fruits, cereals, milk, meat, edible fats etc). A total of 124 food groups are available in the file (g1\_aggr.sav). Please see separate document (in Swedish) for names/labels and brief descriptions of G-food variables (i.e., Appendix: Livsmedelsgruppering 1 (g/livsmedelsgrupp)). Within each type of food the different varieties are aggregated according to botanical origin or animal species, and according to fat, fibre, vitamin C or beta-carotene content. Recipes are broken down into their ingredients and each food is brought to the correct group. Each variable in the file (g1.aggr.sav) indicate the average intake of each food group in grams per day. The G-food group variables in the variable list are aggregated from a detailed grouping scheme and takes into account the coding routines implemented in 1994. More detailed food-groups than the G-variables are not available in the MDC. However, a list of the underlying foods from KOSTSVAR (approximately 1600) is available and could be reviewed to better understand the categorization of food groups. Three different weights of foods were applied: weight as consumed (A), weight before cooking (B) and dry weight (C). The type of weight used is stated in the variable list for each food group. Note that these different weights influence how the food groups can be used, and it is very important to take this into account when aggregating foods. Always contact MDC nutritionists when aggregation of food groups is planned.

During 1991 the computer software did not allow any separation for fresh and heat-treated vegetables/fruits. Note that for vegetables and fruits only the summary variables (G2-variables) are available for the whole database (i.e., covers the whole baseline examination period, unlike specific variables that were not included in 1991 database). These summary variables cannot be used at the same time as the variables they are made from. Subsequently the variable "G2VEGTHF" (i.e., all vegetables) is the sum of nine vegetable variables (GVEGTH + GLEGUMH + GCARROT + GROOTVEF + GLEGUMF + GLEAFVEF + GCABBF + GTOMATOF + GOTHVEGF). Similarly, the variable G2NCITRT (i.e., non-citrus fruit) is the sum of three variables (i.e., GFRBETH (except citrusfruits) + GNCITRTF + GBERRTF), and the variable G2CITRT (i.e., citrus fruit) is the sum of two variables (GCITRF+ citruspart of GFRBETH). There are no limitations in the use of G-variables collected during the time period 1992-1996.

These "G2"-variables are always followed in the variable list by the variables they are constructed from. Thus use only the G2 variables for vegetables and fruits when data from data from the whole collection period (1991-1996) is analysed. The sum of the subgroups does not match g2vegthf for all individuals because 1) g2vegthf is converted to cooked/consumed weight, and 2) cooking losses and coding were handled differently for the 1991 database, compared with later. Thus researchers are also mostly advised to use the G2 variables instead of specific vegetables or fruits, as the summary variables contain much more information (they have no missing values, and have much less zero-consumers).

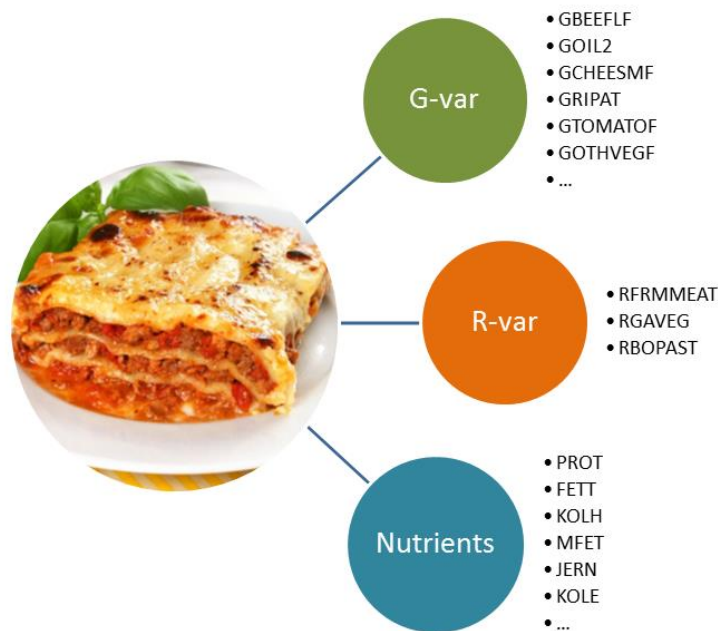
## Cooked foods (R-variables)

Although using the same original food information reported by the participant, the cooked food (R [recipe]-variables) partly handle and describe the information in a different way compared to the G-variables. The R-variables describe dishes (lunch/dinner) registered in the **7-day food diary**. Most of these food group variables describe both a specific food preparation procedure (i.e., frying, deep-frying, oven-baking, boiling and smoking) and the specific main ingredient (i.e., meat, fish, sausage, vegetables, potatoes etc). Other variables describe different types of dishes: casseroles (6 types), soups (3 types), pie, and pizza, salad as a main dish, egg-based dishes, and desserts. The weight is always the consumed weight. These variables are adapted to the change in coding routines implemented in September 1994 (when the coding of recipes was generalized) and can therefore be used on the whole material. Similarly to the G-variables the R-variables indicate the average consumption of each food group in grams per day.

PLEASE NOTE that the G- and R-variables are derived from the same food information. These variables are not independent, but basically describe the same thing from different angles. The G-variables are essentially raw foods and R-variables are cooked foods. Since the philosophies behind the G- and R-variables are different it is not possible to construct composites or summaries of them. Please see the separate table (in Swedish) for names/labels and brief descriptions of R-food variables (i.e., Appendix: Gruppering 2 (g/livsmedelsgrupp och näringsvärde/livsmedelsgrupp)).

## Energy and nutrients from foods

The energy and nutrient intake variables found in the SPSS-file 'nvds\_kost\_91\_96' describe the absolute intake amounts obtained from the food amounts reported by participants at baseline. The total sum of each nutrient variable was calculated as the mean daily intake from the **diet history questionnaire** plus the mean daily intake from the **7-day food diary**, thus the energy and nutrient intake variables for each individual is the sum of two un-weighted means.

**Figure 6. Overview of the diet variables****Table 1: The 50 energy and nutrient intake variables found in the MDC nutrient database**

Variable namne	Nutrient	Variable namne	Nutrient
enertot	Energy (kcal/day)	zink	Zinc (Zn) (mg/day)
vatt	Water (g/day)	sele	Selenium (Se) ( $\mu\text{g}/\text{day}$ )
prot	Protein (g/day)	alko	Alcohol (g/day)
fett	Fat (g/day)	mfet	Saturated fat (SFA) (g/day)
kolh	Carbohydrate (g/day)	mone	Monounsaturated fat (MUFA) (g/day)
fibe	Fibre, total (g/day)	pole	Polyunsaturated fat (PUFA) (g/day)
aska	Ash (g/day)	kole	Cholesterol (g/day)
reti	Retinol (mg/day)	msac	Monosaccharides (g/day)
karo	Beta-carotene (mg/day)	dsac	Disaccharides (g/day)
rekv	Retinol equivalents (mg/day)	sack	Sucrose (g/day)
dvit	Vitamin D ( $\mu\text{g}/\text{day}$ )	f410	Butyric acid, caproic acid, caprylic acid, capric acid; C4:0-C10:0 (g/day)
evit	Vitamin E (mg/day)	f120	Lauric acid; C12:0 (g/day)
toko	D- $\alpha$ -tocopherol (mg/day)	f140	Myristic acid; C14:0 (g/day)
asko	Ascorbic acid (mg/day)	f160	Palmitic acid; C16:0 (g/day)
tiam	Thiamin (mg/day)	f180	Stearic acid; C18:0 (g/day)
ribo	Riboflavin (mg/day)	f200	Arachidic acid; C20:0 (g/day)
fola	Folate ( $\mu\text{g}/\text{day}$ )	f161	Palmitoleic acid; C16:1 (g/day)
niac	Niacin (mg/day)	f181	Oleic acid; C18:1 (g/day)
niek	Niacin equivalents (mg/day)	f183	$\alpha$ -linolenic acid (n-3); $\gamma$ -linolenic acid (n-6) C18:3 (g/day)
b6	Pyridoxine (B <sub>6</sub> ) (mg/day)	f204	Arachidonic acid (n-6); C20:4 (g/day)

b12	Cobalamin (B <sub>12</sub> ) (µg/day)	f182	Linoleic acid (n-6); C18:2 (g/day)
kalc	Calcium (Ca) (mg/day)	f205	Eicosapentaenoic acid (n-3); C20:5 (g/day)
fosf	Phosphorus (P) (mg/day)	f225	Docosapentaenoic acid (n-3) C22:5 (g/day)
jern	Iron (Fe) (mg/day)	f226	Docosahexaenoic acid (n-3); C22:6 (g/day)
magn	Magnesium (Mg) (mg/day)		
natr	Sodium (Na) (g/day)		
kali	Potassium (K) (mg/day)		

Energy (kcal/day) (variable: enertot) is the sum of energy from protein (17kJ/4kcal per g), carbohydrates (not including fibre) (17kJ/4kcal per g), fat (37 kJ/9kcal per g), fiber (8kJ/2kcal) and alcohol (29 kJ/7 kcal per g). Fibre is also considered to provide energy (approximately half as much as other carbohydrates) because they are broken down in the colon; therefore, this new variable was constructed. This can also be done by adding the energy from fibre (8 kcal x fiber intake (in grams)) to the previous variable(ener). Energy can be expressed in kilojoule (kJ) or in kilocalories (kcal). 1 kcal=4.184 kJ. It is, however, recommended nowadays to express the energy intake in kJ or MJ (mega joule). When calculating the percent contribution of the different macro-nutrients to total energy (nutrient densities), factors in table 2 should be used. Please also see this table for how to calculate total energy and non-alcohol energy.

**Table 2: Energy provided by the macronutrients (per 1 gram)**

Macronutrient	Kcal	kJ
Protein	4	17
Carbohydrates (glycemic)	4	17
Fibre	2	8
Fat	9	37
Alcohol	7	29

Ash (variable aska) should not be confused with ascorbic acid (variable asko).

Fat. Values are given for total fat, saturated fat, monounsaturated fat, polyunsaturated fat, different fatty acids and cholesterol.

The variable total fat = the total sum of all fat components of the diet (i.e. fatty acids, glycerol, steroids etc)

Saturated fatty acids: FA 4:0-10:0 + 12:0 + 14:0 +16:0 + 18:0 + 20:0

Monounsaturated fatty acids = FA 16:1 + 18:1

Polyunsaturated fatty acids = FA 18:2 + 18:3 + 20:4 + 20:5 + 22:5 + 22:6.

Please pay attention to the fact that for example saturated fat is not the sum of all fatty acids because the backbone (glycerol) of the fat molecule is also included in the fat variables.

FA 18:3 includes both the n-3 ( $\alpha$ -linolenic acid) and n-6 fatty acids, 18:3 n-6 ( $\gamma$ -linolenic acid) is principally negligible in food. (NOTE: In **supplements** 18:3 is ONLY 18:3 n-6)  
FA 20:5, 22:5 and 22:6 are called "fish fatty acids" but they can also be found in meat, specially pork.

No information on trans-fatty acids is available in the food database. As the greatest part of trans-fatty acids is *trans* 18:1, they are included in the amount of monounsaturated fat and the fatty acid 18:1. In the spring 1995 the manufacturers of margarine changed their recipes to a lower amount of trans-fatty acids and a higher amount of saturated fatty acids. We did not change the content of fatty acids in different codes of margarine because the real change in trans-fatty acids would be impossible to estimate.

Carbohydrate. Data is available for total carbohydrates, mono-saccharides, disaccharides, and sucrose (a disaccharide). Carbohydrate content of foods is often calculated by difference (total weight of foods minus protein, fat, water, alcohol, and ash), and not analysed directly. Fibre is not included in the total carbohydrate variable.

Although questions were asked about the use of sugar, no specific emphasis was put during data collection on the usage of sweeteners in cakes, candies and some beverages, or how much sugar there was in jam, desserts etc. The ability to rank individuals on mono- and disaccharides and sucrose may be satisfactory, but the absolute intake estimates of these nutrients in individuals may be unreliable. In general, if research questions call for an examination of refined carbohydrates the corresponding food groups would be the preferred alternative.

Fibre. Total fibre includes all types of fibre. AOAC method has been used for analysing fibre content of foods. No data is available on sub-types.

Vitamin A. Vitamin A is expressed as retinol, beta-carotene and retinol equivalents. The latter is the sum of retinol and 1/6 of the amount of beta-carotene.

Vitamin E. Vitamin E is a group of tocopherols where alpha-tocopherol is the most active form. Data is available for both total vitamin E and alpha-tocopherol. Please note that in the supplement nutrient database values of vitamin E (variable evit) are not safe; the alpha-tocopherol variable is more conservative and therefore recommended.

Niacin. Niacin is expressed both as niacin (i.e., the content of preformed niacin in foods), and niacin equivalents which is the sum of preformed niacin plus the amount estimated to be formed from the amino acid tryptophan (60 mg tryptophan = 1 mg niacin).

Sodium. The amount of salt in different foods varies greatly and no special attention was paid to added sodium in recipes during data collection. Therefore, this variable should be used with caution or not at all.

Iron. In Sweden wheat flour was previously enriched with iron. However, in January 1995 the enrichment ceased. We did not change the iron values attached to the codes for wheat

flour, bread, cakes etc. because the real change in iron exposure would be impossible to estimate.

### Energy and nutrients from supplements

The energy and nutrient variables found in the SPSS-file 'nvds\_s\_91\_96' are the absolute amounts obtained from supplements, calculated as the mean daily intake from the registration in the 7-day food diary. The energy and nutrient variable names are identical to those of nutrients from foods (i.e., from the SPSS-file 'nvds\_kost\_91\_96') listed above. If the participant did not know the exact name of the supplement, default supplements were assigned. These were defined based on the common use of different brands.

Although, some supplements only contain a single nutrient or a few nutrients, all the nutrients that are listed in the food database are also found in the file for nutrients from supplements (i.e., if the supplement contains them). One notable exception is the fatty acid 18:3, which always is 18:3 n-6 (gamma-linolenic acid) in the supplement database. The same software was used for the supplement and the food database. This facilitates the calculation of the total daily intake (food + supplements) for all nutrients. However, in order facilitate the creation of summary variables (i.e., total intake variables) the supplement nutrients have a suffix added to them (i.e, ener\_s etc).

### 3. Validation study: Malmö Food Study

In the validation study (Malmö Food Study), nutrient and food intake were estimated by two different dietary methods and compared against a reference method of 18 days weighted food records in a random sample of the Malmö population in 1984-85. The validation study included 206 Malmö residents (101 men and 105 women) in the age range 50-69 years.

The two methods (A and B) were: A) an extensive food frequency questionnaire (250 food items) with portion size to be estimated from a booklet of 120 sets of photos and B) a method involving the combination of a shorter questionnaire (130 food items) and a two-week food record. Both methods produced fairly good correlations with the reference method and method B was the method chosen for the MDC study. However, the final dietary assessment method used in MDC, combining a 168-diet questionnaire, a 7-day food record and approximately 1 hour interview, is not completely the same as the validated method B. This is explained by the decisions made after the validation study (Malmö Food Study), depending on what was considered to work best and to be most feasible in the MDC study. The two-week food record was considered to be too laborious for the participants and was expected to lead to loss of participants. Further, a food record with a subsequent interview was considered to be best suitable for the age group of the study. The questionnaire worked relatively well for foods that were commonly used from day to day. Therefore, it was decided to use a less laborious food record (with less days than in the Malmö Food Study) together with a detailed interview and a more extensive questionnaire than what was used in the earlier version validated in the Malmö Food Study. It can be considered a strength that the validation of the dietary method was performed in the same population from where the MDC cohort was enrolled.

Validation correlation coefficients (i.e., Pearson's) of energy-adjusted nutrient intakes, comparing estimates from the modified diet history method with estimates from the 18 dietary records (i.e., the reference method) are shown in Table 3. The energy and nutrient correlation coefficients were amongst the highest compared to those found in validation studies of other "usual" diet instruments, performed in other populations (Willett).

**Table 3: Relative validity and reproducibility of the MDC dietary assessment method (nutrients)**

<b>Nutrient</b>	<b>Relative validity (men/women)</b>	<b>Reproducibility (men/women)</b>
Protein	0.54; 0.53	0.63; 0.54
Fat	0.64; 0.69	0.49; 0.52
Carbohydrates	0.66; 0.70	0.50; 0.49
Sucrose	0.60; 0.74	0.78; 0.46
Alcohol	0.80; 0.78	0.70; 0.82
Retinol	0.39; 0.72	0.58; 0.81
Beta-carotene	0.48; 0.70	0.61; 0.53
Tocopherol	0.65; 0.83	0.83; 0.74
Ascorbic acid	0.64; 0.71	0.67; 0.74
Folate	0.75; 0.75	0.33; 0.71
Calcium	0.70; 0.73	0.75; 0.37
Zink	0.58; 0.44	0.70; 0.49
Selenium	0.46; 0.44	0.69; 0.64
Fibre	0.74; 0.69	0.66; 0.70
Saturated fatty acids	0.56; 0.68	0.64; 0.62
Monounsaturated fatty acids	0.59; 0.66	0.46; 0.50
Polyunsaturated fatty acids	0.26; 0.64	0.68; 0.70

Similarly, validation correlation coefficients (energy-adjusted) of food group intakes, comparing estimates from the diet history with estimates from the 18 dietary records are shown in table 4.

**Table 4: Relative validity and reproducibility of the MDC dietary assessment method (foods)**

<b>Food</b>	<b>Relative validity (men/women)</b>	<b>Reproducibility (men/women)</b>
Vegetables	0.65; 0.53	0.71; 0.76
Fruits	0.60; 0.77	0.80; 0.81
Potatoes	0.69; 0.51	0.82; 0.43
All Milk	0.83; 0.84	0.82; 0.70
Milk low fat	0.92; 0.92	0.79; 0.65
Milk high fat	0.76; 0.75	0.73; 0.74



Cream	0.47; 0.52	0.48; 0.42
Cheese	0.47; 0.59	0.71; 0.71
Meat products	0.84; 0.92	0.96; 0.57
Meat low fat	0.43; 0.51	0.46; 0.31
Meat high fat	0.40; 0.80	0.73; 0.23
Fish	0.35; 0.70	0.87; 0.33
Eggs	0.57; 0.74	0.78; 0.22
Cereals	0.74; 0.73	0.48; 0.56
Bread	0.50; 0.58	0.45; 0.65
Rice, pasta	0.35; 0.24	0.41; 0.23
Added fats and oils	0.54; 0.66	0.62; 0.46
Fat 80%	0.72; 0.74	0.77; 0.62
Fat 40-60%	0.76; 0.85	0.85; 0.46
Beer	0.68; 0.74	0.58; 0.68
Wine	0.53; 0.63	0.83; 0.79
Spirits	0.57; 0.67	0.75; 0.83

Mean daily intake of energy (kcal/MJ) from the **MDC modified diet history** compared to the mean intakes from the 18-days weighted records were on average overestimated by 21% [However, there was no indication that nutrient densities were over- or under-estimated: protein (%energy) 0%; fat (%energy) 0%; carbohydrates (%energy) 1%]. Men overestimated on average alcohol consumption by 186%, while women underestimated alcohol by 21%.

In conclusion, the MDC methodology is a modified diet history methodology, giving “usual diet” data. As such these data are suitable to rank individuals on their usual dietary intakes, and to examine associations with other variables and disease end-points. As indicated by the energy-adjusted validity correlation coefficients the relative validity for most variables, both nutrients and foods, is very high (1) (2) (3). However, the methodology is not ideal for estimating absolute mean intakes, and to determine cut-offs in absolute terms.

#### References:

- I. Riboli E et al. The Malmö Food Study: validity of two dietary assessment methods for measuring nutrient intake. *Int J Epidemiol.* 1997;26 Suppl 1:S161-73.
- II. Elmståhl S et al. The Malmö Food Study: the relative validity of a modified diet history method and an extensive food frequency questionnaire for measuring food intake. *Eur J Clin Nutr.* 1996 Mar;50(3):143-51.
- III. Elmståhl S et al. The Malmö Food Study: the reproducibility of a novel diet history method and an extensive food frequency questionnaire. *Eur J Clin Nutr.* 1996 Mar;50(3):134-42.
- IV. Callmer E et al. Dietary assessment methods evaluated in the Malmö Food Study. *J Intern Med.* 1993 Jan;233(1):53-7.

## 4. Considerations regarding statistical power using the MDC diet history

The original research program of the Malmö Diet Cancer study (submitted to major Swedish funding agencies in December 1991) includes extensive computations and discusses issues related to statistical power when examining self-reported dietary data in this population. These considerations were based on the Malmö population distribution in 1991, and the cancer incidence from 1988 and 1989 in this population (as assessed from the national population registries).

The table below describes the Relative Risk (RR) estimates that are observable over quintiles of dietary intake, based on 3 controls per case, when hypothesising that the true risk gradient associated with a specific dietary exposure ranges from 1.00 in the 1<sup>st</sup> (lowest) quintile to 3.00 in the 5<sup>th</sup> (highest) quintile. The table illustrates that if the dietary validation coefficients have lower values the observed risk estimates will gradually be attenuated. Therefore, there is a clear danger in observing non-elevated RRs when the dietary validation coefficients have low values (below 0.5), even if the true risk is elevated.

Dietary validation coefficient	RRs associated with quintiles of dietary exposure				
	1 <sup>st</sup> quintile	2 <sup>nd</sup> quintile	3 <sup>rd</sup> quintile	4 <sup>th</sup> quintile	5 <sup>th</sup> quintile
1.00 ("truth")	1.00	1.50	2.00	2.50	3.00
0.7	1.00	1.28	1.49	1.71	1.99
0.6	1.00	1.23	1.40	1.57	1.79
0.5	1.00	1.18	1.31	1.43	1.61

The validity of the dietary assessment methodology influences the required sample size. Given a validity coefficient of the dietary variable of 0.6, and a "true" risk gradient of 1 to 3 over quintiles of nutrient intake, sufficient statistical power (i.e., 80% and  $\alpha=0.05$ ) to observe a risk gradient of 1 to 1.79 over quintiles was in this population reached when 283 cancer cases had accumulated. For instance, by the end of 1999 a sufficient number (the above 283) of incident breast cancer cases had accumulated among women of the full cohort ( $n=17\ 035$ ). By December 31, 2007, a total of 732 invasive breast cancer cases had accumulated among women free of cancer (but including those with cancer cervix in situ) at study entry ( $n=15\ 773$ ). Since, many dietary variables have energy-adjusted validation correlation coefficients above 0.5 or 0.6 in the MDC study, and a sufficient number of cases currently (i.e., August 2010) is available for (at least) breast and prostate cancer, the power to examine dietary hypotheses is satisfactory.

It is, however, important to recognize that between several dietary and lifestyle variables (especially nutrients) strong correlations are seen, and that co-linearity attenuates risk estimates (and can even change the direction of the association) if variables are examined in the same model. The observed relative risk therefore depends both on the validity of the diet assessment method, and on co-linearity between the examined variables. These issues have been examined by Elmståhl and Gullberg in the paper cited below (11). Similarly, it

should be recognized that low validity of co-variables (usually used for adjustments) will also attenuate risk estimates.

Elmståhl S, and Gullberg B. Bias in diet assessment methods--consequences of collinearity and measurement errors on power and observed relative risks. *Int J Epidemiol.* 1997 Oct;26(5):1071-9.

**BACKGROUND:** If several risk factors for disease are considered in a regression model and these factors are affected by measurement errors, the observed relative risk will be attenuated. In nutritional epidemiology, several nutrient variables show strong correlation, described as collinearity. The observed relative risk will then depend not only on the validity of the chosen diet assessment method but also on collinearity between variables in the model.

**METHODS:** The validity of different diet assessment methods are compared. The correlation coefficients between common nutrients and foods are given using data from the Malmö Food Study. Intake of nutrients and foods were assessed with a modified diet history method, combining a 2-week food record for beverages and lunch/dinner meals and a food frequency questionnaire for other foods. The study population comprised 165 men and women aged 50-65 years. A multivariate logistic regression model is used to illustrate the effect of co-linearity on observed relative risk (RRo).

**RESULTS:** A moderate to high correlation between risk factors will substantially influence RRo even when using diet assessment methods with high validity. Methods with low validity might even give inverse RRo.

**CONCLUSION:** It is stressed that caution must be exercised and only a selected number of variables should be included in the model, especially when they are highly inter-correlated, since RRo might be severely biased.

## 5. Other important methodological papers from the MDC cohort

### Change in interview routines in 1994 of the MDC diet history

- Wirfält E et al. A methodological report from the Malmö Diet and Cancer study: development and evaluation of altered routines in dietary data processing. *Nutr J.* 2002 Nov 19;1:3.

Although, a change in dietary data collection routines is not recommended during the active data collection phase, reality sometimes forces such undesired changes. In the spring of 1993 an unforeseen reduction of grants initiated measures to simplify dietary data collection routines and to make faster interviews possible. The change did not affect the instructions given to participants, but only involved the coding of food information. The change in coding routines was preceded by a phase evaluating possible options for simplifying the procedures, and a series of tests to examine the effect of different coding and portion-size alternatives on dietary intakes. The tests suggested that alterations in the handling of dietary data (when

examined under carefully controlled situations) were possible without substantial impact on the ranking ability or mean nutrient intake levels.

When implemented in full, however, the altered routines could possibly affect dietary interviewers and study-participants, and subsequently, influence observed intakes in unpredictable ways. Therefore mean dietary estimates collected right before and after the alteration of routines was also compared in a separate study (12). This comparison indicated that energy intakes were significantly lower for both gender groups after the change in routines, but energy-adjusted fat intakes were significantly higher with the second method version.

These results suggest that descriptive studies using the MDC data should preferably *select subsets of the population assessed with either method version to avoid distortion of observed intakes*. The impact of altered interview routines on the outcome of analytical studies (i.e. when dietary data is used to rank individuals) probably is small. Depending on the specific research question, *adjustment in analysis for year and season of data collection, and dietary interviewer and method version* may need to be considered. Regardless, method version differences between cases and non-cases need to be monitored. However, in nested case-control studies when matching cases and controls on the date of blood collection these influences are minimized.

### Mis-reporting of energy intakes using the MDC diet history

- Mattisson I et al. Misreporting of energy: prevalence, characteristics of misreporters and influence on observed risk estimates in the Malmö Diet and Cancer cohort. *Br J Nutr.* 2005 Nov;94(5):832-42.

The prevalence of misreporting of energy in the Malmö Diet and Cancer cohort, including the anthropometric, socio-economic and lifestyle characteristics of the misreporters, was examined in a separate study (13). This study also examined the influence of excluding misreporters on risk estimates of post-menopausal breast cancer. Information of reported energy intake (EI) was obtained from a modified diet history method. A structured questionnaire provided information on lifestyle and socio-economic characteristics. Individual physical activity level (PAL) was calculated from self-reported information on physical activity at work, leisure time physical activity and household work, and from estimates of hours of sleeping, self-care and passive time. Energy misreporting was defined as having a ratio of EI to BMR outside the 95 % CI limits of the calculated PAL. Logistic regression analyzed the risk of being a low-energy reporter or a high-energy reporter.

Almost 18% of the women and 12 % of the men were classified as low-energy reporters, 2.8 % of the women and 3.5 % of the men were classified as high-energy reporters. In both genders high BMI, large waist circumference, short education and being a blue-collar worker were significantly associated with low-energy reporting. High-energy reporting was significantly associated with low BMI, living alone and current smoking. When mis-reporters of energy were excluded from the analyses, the previously observed association with energy-adjusted intakes of fat and fibre remained largely unchanged. The results add

support to the practice of energy adjustment as a means to reduce the influence of errors in risk assessment.

### Past food habit change

- Sonestedt E, et al. Past food habit change is related to obesity, lifestyle and socio-economic factors in the Malmo Diet and Cancer Cohort. *Public Health Nutr.* 2005 Oct;8(7):876-85.

**OBJECTIVES:** To examine if obesity status and socio-economic and lifestyle factors are associated with self-reported past food habit change, and also whether the level of obesity depends on the reason for change (14).

**DESIGN:** Cross-sectional analysis within the Malmo Diet and Cancer (MDC) study using data from the baseline examination and the extensive socio-economic and lifestyle questionnaire including questions of past food habit change. The risk of having changed food habits in the past was examined using logistic regression. Mean differences in obesity status across categories of reasons for past food habit change were examined using analysis of variance.

**SETTING:** Malmö, the third largest city in Sweden.

**SUBJECTS:** A sub-sample (15 282 women and 9 867 men) from the MDC cohort recruited from 1992 to 1996.

**RESULTS:** Individuals with body mass index (BMI)  $>30 \text{ kg m}^{-2}$  had an increased risk of having reported past food habit change compared with individuals with BMI  $<25 \text{ kg m}^{-2}$  (odds ratio (OR) = 1.63, 95% confidence interval (CI) = 1.48-1.83 for women; OR = 1.53, 95% CI = 1.32-1.76 for men). The highest level of obesity was observed among individuals who had changed their diet due to reasons related to the metabolic syndrome. Changers were more likely to be highly educated and to live alone, be retired, ex-smokers and non-drinkers at baseline.

**CONCLUSIONS:** Because past food habit change is related to obesity and other lifestyle and socio-economic factors, a complex confounding situation may exist that could seriously influence observed relationships between diet and disease. Studies need to collect information on past food habit change and take this information into account in the analysis and when interpreting study outcomes.

- Sonestedt E et al. Both food habit change in the past and obesity status may influence the association between dietary factors and postmenopausal breast cancer. *Public Health Nutr.* 2007 Aug;10(8):769-79.

**OBJECTIVE:** Valid dietary data are essential when trying to identify whether or not one or more dietary exposures are responsible for disease. We examined diet composition in women who reported dietary change in the past compared with non-changers, and how the associations between dietary factors and postmenopausal breast cancer are influenced by dietary change, obesity status and misreporting of energy (15).

**DESIGN:** A population-based prospective cohort study. Data were obtained by a diet history method, anthropometrical measurements and an extensive lifestyle questionnaire including items on past food habit change.

**SETTING:** The Malmö Diet and Cancer (MDC) study, conducted in Malmö, Sweden.

**SUBJECTS:** A subsample of 12,781 women from the MDC cohort recruited from 1991 to

1996. A total of 428 postmenopausal women were diagnosed with incident breast cancer, during 9.2 years of follow-up.

RESULTS: Past food habit changers reported healthier food habits and lower energy intake compared with non-changers, a finding that raises issues regarding possible reporting biases. When excluding diet changers, the trend of increased breast cancer risk across omega-6 fatty acid quintiles was stronger, and a tendency of decreased risk emerged for 'fruit, berries and vegetables'. When excluding individuals with non-adequate reports of energy intake, risk estimates were similar to that of the whole sample. In women with body mass index < 27 kg m<sup>-2</sup>, significant trends of increased breast cancer risk were seen for total fat and omega-6 fatty acids, and of decreased risk for 'fruit, berries and vegetables'.

CONCLUSIONS: This study indicates that both obesity and self-reported past food habit change may be important confounders of diet-breast cancer relationships. The study demonstrates that sensitivity analysis, through stratification, may facilitate interpretation of risk relationships and study results.

## 6. Recommendations regarding dietary data handling and data analysis.

After 20 years of experience (starting around 1995) from dietary data analysis of the MDC cohort, the involved nutritional epidemiologist has gained substantial knowledge of the specific characteristics of the data set. Each researcher wishing to examine MDC dietary data would benefit greatly from incorporating this experiential knowledge into their projects.

Since the food information (i.e., portion-sizes and frequencies) has been converted into the absolute amounts consumed per day, the dietary variables are continuous and quantitative variables. However, if compared to other numerical variables (like body-weight or biochemical measures) the characteristics may be quite different. These are some typical features of dietary data:

- Food choices and nutrient intakes may vary greatly day by day. Therefore many daily records may be needed to accurately assess specific nutrients. Please note that foods specifically assessed through the “menu-book” may be affected by this day-to-day variation.
- Self-report dietary data are often skewed.
- Self-report dietary data are often associated with different types of mis-reports and measurement errors leading to mis-classification of exposures, which may cause biases in data analysis, or attenuation of associations between exposure and disease.
- Because food items contain many different nutrients and other bioactive components, projects may have difficulties separating the effects of specific nutrients.
- Because lifestyle and socio-economic factors commonly co-vary with food habits, projects may have difficulties separating the effects of diet and other lifestyle factors.
- Individual characteristics may influence the examined association. That is obesity and genetic factors may modulate the development of disease.
- Other factors, like change of food habits over time, dieting or social norms and attitudes may cause “reporting” errors. That is the reported dietary exposure may not be the “true” dietary exposure that influenced disease development.

Skewed distributions and zero-consumers. Many dietary variables have non-normal/skewed distributions. This is especially true for alcohol and many food variables where large numbers of zero-consumers are common. Both skewness and zero-consumers need to be handled prior to analysis; mainly data transformation and categorization of zero-consumers. Dietary variables are often categorized into quintiles, quartiles or tertiles (after zero-consumers are taken care of – i.e., separated into a different category), which serves two purposes: handling of skewness and avoidance of undue influence of extreme (unreliable) exposures.

Energy-adjustment. Nutrition epidemiology research has shown that self-report dietary intakes need to be examined as dietary proportions, because absolute intakes will not give interpretable results (2). This is well recognized and four energy-adjustment approaches have been described: the nutrient density, the residual, the standard multivariate, and the partitioning energy adjustment approaches (16).

The major reasons behind the need for energy-adjustment are:

- Energy needs varies greatly between individuals depending on body size, gender and physical activity, resulting in differences in energy intakes, therefore only iso-caloric nutrient intakes are of interest.
- In epidemiological studies different energy intake levels mirror energy needs and the degree of physical activity. Mostly epidemiological studies need to study the effect of macro-nutrients (eg. fat) separately from that of total energy intake or physical activity.
- Because big individuals consume large amounts of food, the intakes of most nutrients (also those that do not contribute energy) vary with the total energy intake, and causes variation in nutrient intakes that are not linked to the composition of diet.
- Energy intake may be the primary cause of disease.
- If energy-intake is associated with disease, but not the direct cause, the effect of specific nutrients may become influenced or distorted by total energy intake (17) (18).
- An additional benefit of energy-adjustment, which has been recognized and discussed extensively during the last decade, is that it removes or dilutes the negative influence caused by measurement errors in dietary assessment (18) (19).

Energy-adjusted models usually include both the energy-adjusted variable (i.e., the energy% (E%) or residual variable) and total energy. If physical activity also is included in the model there is some danger of over-adjustment because self-reported energy intake in epidemiology studies may not reflect the real intake, but rather expended energy. “Non-alcohol energy” (i.e. energy from protein, carbohydrates and fat) is often used when computing residuals, because a focus on the dietary composition is achieved. In these cases the multivariate model should, however, include total energy (including energy from alcohol).

**Note 1.** When the “nutrient density” approach is used in multivariate analysis it will be equal to the “residual” approach only if total energy-intake also is included in the model.

**Note 2.** If the “nutrient density”-model is used to energy-adjust non-energy giving nutrients an artificial association with total energy could be created. Therefore the “residual” model is preferred for non-energy giving nutrients.

Gender specific analysis. Men and women have different energy needs and subsequently different energy intakes. Also, the selection of foods may differ e.g. fruits and vegetable and cakes and pastries in the MDC are higher in women in spite of lower energy intake. In addition, the validity of dietary data may vary by gender. It is also possible that other confounding variables may differ by gender.

Method variables. Several methodological variables are available in the dataset: method, season and diet-assistant (KA). Because of the change in coding routines September 1<sup>th</sup>, 1994, cohort analysis should in the basic model (i.e., the “crude” data analysis) include season and method version. At least the distribution of season, method-version and diet assistant needs to be examined in relation to the selected exposure variables and end-point variables (case/control-status).

In “nested case-control” studies where cases and controls are matched on age and blood sample date, the study design will at least partly control for variation caused by season, method version, and diet-assistant. Still it is recommended to examine the distribution of these variables across cases and controls.

Descriptive analysis. Since the MDC modified diet history methodology is a “usual diet method” it gives a qualitative description of what the participants usually eat, and is suitable for ranking individuals. When the average dietary composition of groups of individuals is described, relative variables should be used. Because of the change in coding routines in September 1<sup>st</sup>, 1994, projects that wish to compare intakes between population groups should use the sub-set of individuals that completed the baseline examination after that date. Please also see the study described above; Wirfält et al. Nutr J. 2002 (12).

Use of sub-sets of the cohort, or sensitivity analysis. Research indicates that under- and mis-reporting of dietary intakes is common in overweight and obese individuals. Similarly, those dieting, or with previous disease (like heart-disease, dyslipidemia, hypertension or diabetes) are often found to be underreporting intake. Such individuals also tend to have changed their food habits in the past.

(A) The baseline database include one questionnaire item of **specific importance**: “Have you substantially changed your food habits in the past due to ill-health or other reasons?” (Sonestedt et al. 2005, 2007)

(B) Estimated energy expenditure compared to self-reported energy intake has been examined in order to identify those with reasonable energy intakes. That project (Mattisson et al. 2005) identified individuals with too low energy intakes, adequate energy intakes and too high energy intakes (i.e., identified energy mis-reporters).

**Note 1.** Only individuals reporting *no* change of food habits in the past, and only those reporting *adequate energy intakes* should be included in cross-sectional studies (i.e., when using baseline data only).

**Note 2.** In prospective studies individuals with and without food habit change and with or without mis-reports should be examined separately, at least in a *final sensitivity analysis*. Projects may also exclude these individuals from analysis altogether (i.e., identify a study-sample without them).



## 7. Recommended literature – suggestions.

- Willett W. Nutritional Epidemiology. New York Oxford: Oxford University Press; 2013
- Margetts, B. M. & Nelson, M. Design concepts in nutritional epidemiology. Oxford New York Tokyo: Oxford University Press; 1997.
- Food, nutrition, physical activity and the prevention of cancer: A global perspective. Washington, DC: World Cancer Research Fund / American Institute for Cancer Research (WCRF/AICR); 2007.
- Modern Epidemiology, 3<sup>rd</sup> edition. Eds. Kenneth Rothman, Sander Greenland, Timothy L. Lash. Philadelphia London Tokyo: Wolters Kluwer Lippincott Williams & Wilkins; 2008.

## 8. References.

1. Riboli E, Elmståhl S, Saracci R, Gullberg B, Lindgärde F. The Malmö Food Study: Validity of Two Dietary Assessment Methods for Measuring Nutrient Intake. *International Journal of Epidemiology* 1997;26(1(Suppl 1)):S161-S73.
2. Willett W. Nutritional Epidemiology. New York Oxford: Oxford University Press; 1998. p. 101-47.
3. Thiébaud AC, Kipnis V, Schatzkin A, Freedman LS. The role of dietary measurement error in investigating the hypothesized link between dietary fat intake and breast cancer--a story with twists and turns. *Cancer Invest* 2008;26(1):68-73.
4. Berglund G, Elmståhl S, Janzon L, Larsson SA. The Malmö Diet and Cancer Study: Design and feasibility. *Journal of Internal Medicine* 1993;233:45-51.
5. Manjer J, Carlsson S, Elmståhl S, Gullberg B, Janzon L, Lindström M, Mattisson I, Berglund G. The Malmö Diet and Cancer Study: Representativity, Cancer Incidence and Mortality in Participants and Non-participants. *European Journal of Cancer Prevention* 2001;10(6):489-99.
6. Manjer J, Elmståhl S, Janzon L, Berglund G. Invitation to a population based cohort study: differences between subjects recruited using various strategies. *Scandinavian Journal of Public Health* 2002;30:103-12.
7. FAO/WHO/UNU. Energy and Protein Requirements Report of a Joint FAO/WHO/UNU Expert consultation. Geneva: World Health Organization; 1985.
8. Callmer E, Riboli E, Saracci R, Åkesson B, Lindgärde F. Dietary assessment methods evaluated in the Malmö food study. *J Intern Med* 1993;233:53-7.
9. Elmståhl S, Riboli E, Lindgärde F, Gullberg B, Saracci R. The Malmö Food Study: the relative validity of a modified diet history method and an extensive food frequency questionnaire for measuring food intake. *European Journal of Clinical Nutrition* 1996;50(3):143-51.
10. Elmståhl S, Gullberg B, Riboli E, Saracci R, Lindgärde F. The Malmö food study: The reproducibility of a novel diet history method and an extensive food frequency questionnaire. *European Journal of Clinical Nutrition* 1996;50:134-42.

11. Elmståhl S, Gullberg B. Bias in diet assessment methods - consequences of collinearity and measurement errors on power and observed relative risks. *International Journal of Epidemiology*1997;26(5):1071-9.
12. Wirfält E, Mattisson I, Johansson U, Gullberg B, Wallström P, Berglund G. A methodological report from the Malmö Diet and Cancer study: Development and evaluation of altered routines in dietary data processing. *Nutrition Journal* [serial on the Internet]. 2002; 1(3): Available from: <http://www.nutritionj.com/content/1/1/3>.
13. Mattisson I, Wirfält E, Andrén-Aronsson C, Wallström P, Sonestedt E, Gullberg B, Berglund G. Misreporting of energy: prevalence, characteristics of misreporters and influences on observed risk estimates in the Malmö Diet and Cancer cohort. *British Journal of Nutrition*2005;94(5):832-42.
14. Sonestedt E, Wirfalt E, Gullberg B, Berglund G. Past food habit change is related to obesity, lifestyle and socio-economic factors in the Malmo Diet and Cancer Cohort. *Public Health Nutr*2005 Oct;8(7):876-85.
15. Sonestedt E, Gullberg B, Wirfalt E. Both food habit change in the past and obesity status may influence the association between dietary factors and postmenopausal breast cancer. *Public Health Nutr* [[Epub 2007 Mar 5] ]. 2007;10(8):769-79.
16. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *American Journal of Clinical Nutrition*1997;65(suppl):1220S-8S.
17. Kipnis V, Freedman LS, Brown CC, Hartman A, Schatzkin A, Wacholder S. Interpretation of Energy Adjustment Models for Nutritional Epidemiology. *Am J Epidemiol*1993;137(12):1376-80.
18. Kipnis V, Freedman LS, Brown CC, Hartman AM, Schatzkin A, Wacholder S. Effect of measurement error on energy-adjustment models in nutritional epidemiology. *American Journal of Epidemiology*1997;146(10):842-55.
19. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, Schatzkin S, Carroll RJ. Structure of Dietary Measurement Error: Results of the OPEN Biomarker Study. *Am J Epidemiol*2003;158:14-21.