The Diet, Cancer and Health study – a prospective cohort study

The importance of record linkage in Cancer Epidemiology

Anne Tjønneland
ENCR, SEP
September 28th, 2018
Danish Cancer Society Research Center

250 researchers and technicians from all parts of the world

- **Diet, Genes and Environment**
- Virus, Lifestyle and Genes
- Cell Stress and Survival
- Cell Death and Metabolism
- Genome Integrity
- Survivorship
- Translational Cancer Research
- Statistics, Bioinformatics and Registry
- Danish Centre for Translational Breast Cancer Research
Agenda

• Cohort description
• Record linkage examples
• Perspectives
Diet, Cancer and Health cohort

• Baseline data collection 1993–1997
• Follow up questionnaires 1999-2002

57,053 healthy participants, 50–64 y
• 27,178 men
• 29,875 women

Follow up for disease events (31/12 2016)
• 14,000 deaths, all causes
• 14,875 incident cancers
• 1,909 diagnosed colorectal cancer
• 2,495 diagnosed prostate cancer
• 2,311 diagnosed breast cancer
Diet, Cancer and Health cohort

- **Baseline data:**
  - Food frequency questionnaires, 24HDR (subset)
  - Lifestyle questionnaires
  - Biological specimens
    - Blood
    - Urine
    - Adipose tissue
    - Toenail clippings
  - Physical measurements
    - Weight, height, standing height, sitting height
    - Waist circumference, hip circumference
    - Blood pressure
Key data sources for follow up

- Clinical Cancer Databases
  + detailed, prospective clinical data
  + quality improvement research
  - relatively new

- Danish Cancer Registry
  + incident cancer
  - limited data on stage and treatment

- National Patient Registry
  + all hospital contacts since 1977
  - created for administration purposes

- Diet, Cancer and Health Cohort
  + prospective data on diet and lifestyle
  - updates based on postal letters/emails
Other registries:

- The Civil Registration System (from 1968)
- The National Diabetes Registry (from 2006-11)
- School Health Records Registry
- Birth Records
- The Danish Pension Fund Registry to individual employment history
- CPR and exposure modelling for air pollution, traffic noise etc.
- Statistics Denmark, incl. Danish Prescription Registry

Follow-up rate: 99.8 %
EPIC collaboration

- European collaboration, EPIC, 10 countries, 500,000 participants.
- Continuous collaboration for more than 25 years among 25 research institutions in Europe
- Common EPIC database at IARC and Imperial College, London
- Working groups on cancer end points, and other chronic diseases
- Very productive collaboration with app. 100 publications pr. year
- Monthly telephone meetings, and 1-2 yearly face-to-face meetings
- Extended collaboration through i.e. Cohort Consortium
25 years of research

>1000 scientific papers:

• Hormone therapy during menopause increases the risk of breast cancer, ovarian cancer and endometrial cancer

• Alcohol intake increases the risk of breast cancer

• Whole grain intake protects against colorectal cancer

• High pre-diagnostic blood levels of enterolactone improves survival after breast cancer

• Healthy Nordic diet reduce overall mortality

• Air pollution increases the risk of lung cancer
Diet, Cancer and Health – Next Generations

Rationale and objectives

To extend the existing Diet, Cancer and Health (DCH) cohort by recruiting “next generations”

Overall aims:

- Enable trans-generational studies of the pathogenesis of multiple cancers and other diseases
- Valuable in the search for biomarkers and omics technologies for early detection and exposure
Identification of “next generations”
- using The Danish Civil Registration System

Kost, kræft og helbred
Original Diet, Cancer and Health cohort

G1
n=57,053

G2
Child of G1
2 parents in G1
n= 20,470

G2A
Spouse to G2
/Parent to G3
n= 63,772

G3
Child of G2
n= 127,684

G2
Child of G1
1 parent in G1
n= 67,250
Recruitment process

Invitation letter → DCH-NG homepage → Webprofile (Web)
Questionnaires

**Food frequency questionnaire**
- 366 food items

**Background and lifestyle questionnaire**
- Socio-demographics
- Work conditions and environment
- Smoking habits and history
- Alcohol habits and history
- Physical activity
- Sleep pattern and quality
- Weight history
- Medication and medical history
- Family history of disease
- Female/male reproductive factors
- Quality of life (SF36)
- Family relations
Baseline assessment at the study center

**Physical measurements**

<table>
<thead>
<tr>
<th>Physical measurements</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthropometry</td>
<td>Height, weight, waist, hip</td>
</tr>
<tr>
<td>Bioimpedance</td>
<td>Body composition (e.g. fat and muscle mass)</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Pulse rate and blood pressure</td>
</tr>
<tr>
<td>App</td>
<td>Physical activity, steps</td>
</tr>
</tbody>
</table>

**Biological samples (non-fasting)**

<table>
<thead>
<tr>
<th>Biological samples (non-fasting)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine</td>
<td>Spot urine</td>
</tr>
<tr>
<td>Saliva</td>
<td>Pure saliva, saliva with added RNAlater</td>
</tr>
<tr>
<td>Feecal sample</td>
<td>Imidiately frozen and preserved for DNA extraction</td>
</tr>
<tr>
<td>Blood, storage</td>
<td>Plasma, serum, buffy coat, erythrocytes, RNA extraction tube</td>
</tr>
<tr>
<td>Blood, up-front</td>
<td>Triglycerides, total, HDL and LDL cholesterol, HbA1c, hs-CRP, creatinine</td>
</tr>
</tbody>
</table>
DCH-NG MAX  n=500

Data collection

- Height
- Weight
- Bioimpedance
- Blood pressure
- Waist circumference
- Hip circumference
- Food frequency questionnaire
- Lifestyle questionnaire
- 24-hour dietary recall
- Blood
- Urine
- Saliva
- Faeces

Analyses

HbA1c, triglyceriders, total cholesterol, LDL cholesterol, HDL cholesterol, creatinin, hs-CRP

Untargeted **metabolomics** using LC-QTOF-MS

**Gut microbiota** analysed by 16S rRNA and whole genome sequensing

**GWAS** incl. cardio metabolic traits (lipids, lipoproteins, BMI, waist) and diseases (cardiovascular and Type 2 diabetes)
Diet, Cancer and Health – Next generation status

>51,000 participants have registered (response rate ~26%)
>42,000 participants visited the study center (300-350 visits/week)

Data collection is almost complete: 99.5-100% for all measurements including anthropometry, blood pressure and blood, urine and saliva samples

~24,000 fecal samples will be available for future research

Data collection to end in 2018
Courses of Death Registry Statistics Denmark

Participation and mortality among 80,996 men and 79,729 women invited to the DCH study.
Overall Mortality (log MRR) participants/non-participants in DCH study

**Fig. 2** Log rate ratio of overall mortality (logMRR) between participants and non-participants in the prospective Danish “Diet, Cancer and Health” Study stratified by sex
Mortality rate ratios for participants and non-participants in DCH study (1993-2008), men

<table>
<thead>
<tr>
<th>Socioeconomic indicator</th>
<th>Men</th>
<th></th>
<th>Non-participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants</td>
<td>Non-participants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MRR</td>
<td>95 % CI</td>
<td>MRR</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1.00</td>
<td>(Reference)</td>
<td>2.06</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic/high school</td>
<td>1.89</td>
<td>(1.72–2.07)</td>
<td>3.68</td>
</tr>
<tr>
<td>Vocational training</td>
<td>1.40</td>
<td>(1.29–1.53)</td>
<td>2.76</td>
</tr>
<tr>
<td>Higher education</td>
<td>1.00</td>
<td>(Reference)</td>
<td>1.77</td>
</tr>
<tr>
<td><strong>Income (quartile)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>2.94</td>
<td>(2.66–3.24)</td>
<td>5.46</td>
</tr>
<tr>
<td>2nd</td>
<td>1.77</td>
<td>(1.60–1.95)</td>
<td>3.07</td>
</tr>
<tr>
<td>3rd</td>
<td>1.20</td>
<td>(1.08–1.33)</td>
<td>2.12</td>
</tr>
<tr>
<td>4th</td>
<td>1.00</td>
<td>(Reference)</td>
<td>1.61</td>
</tr>
</tbody>
</table>
Conclusion:

- Mortality differs within social strata

- Self selection is based both on health at enrolment and also on a lifestyle keeping you healthier throughout the course of the study

- Mortality rates differed, even after accounting for differences in SEP between participants and non-participants
# Diet, Nutrition, Physical Activity and Breast Cancer Survival (by Outcome)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>All Cause Mortality</th>
<th>Breast Cancer Mortality</th>
<th>Second Primary Breast Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decreased Risk</td>
<td>Increased Risk</td>
<td>Decreased Risk</td>
</tr>
</tbody>
</table>

## Strong Evidence
- Convincing
- Probable

## Limited Evidence
- Suggestive
  - Physical activity: Before diagnosis ≥12 months after diagnosis
  - Body fatness: Before diagnosis <12 months after diagnosis

## Strong Evidence
- Substantial effect on risk unlikely

### Foods Containing Fibre
- Before diagnosis ≥12 months after diagnosis

### Foods Containing Soy
- Before diagnosis ≥12 months after diagnosis

---

**Notes:**
- STRONG: Evidence strong enough to support a judgement of a convincing or probable causal relationship and generally justify making recommendations
- LIMITED: Evidence that is too limited to justified making specific recommendations
- Post menopause only
Lignans – prognosis after breast cancer

Published in final edited form in:

Dietary lignan intakes in relation to survival among women with breast cancer: the Western New York Exposures and Breast Cancer (WEB) Study

Susan E. McCann1, Lilian U. Thompson2, Jing Nie3, Joan Dom3, Maurizio Trevisan4, Peter G. Shields5, Christine B. Ambrosone6, Stephen B. Edge2, Hsin-Fang Li5, Christina Kasprzak1, and Jo L. Freudenheim3

Postmenopausal (n = 807)

<table>
<thead>
<tr>
<th>Intake</th>
<th>Cases</th>
<th>Death Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;155</td>
<td>33</td>
<td>1.00</td>
</tr>
<tr>
<td>155–227</td>
<td>33</td>
<td>0.91 (0.55–1.52)</td>
</tr>
<tr>
<td>227–318</td>
<td>29</td>
<td>0.78 (0.46–1.33)</td>
</tr>
<tr>
<td>&gt;318</td>
<td>21</td>
<td>0.49 (0.26–0.91)</td>
</tr>
<tr>
<td>P for trend</td>
<td></td>
<td>0.02</td>
</tr>
</tbody>
</table>

Breast cancer mortality

<table>
<thead>
<tr>
<th>Intake</th>
<th>Cases</th>
<th>Death Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;155</td>
<td>18</td>
<td>1.00</td>
</tr>
<tr>
<td>155–227</td>
<td>20</td>
<td>0.94 (0.48–1.87)</td>
</tr>
<tr>
<td>227–318</td>
<td>11</td>
<td>0.51 (0.23–1.15)</td>
</tr>
<tr>
<td>&gt;318</td>
<td>7</td>
<td>0.29 (0.11–0.76)</td>
</tr>
<tr>
<td>P for trend</td>
<td></td>
<td>0.01</td>
</tr>
</tbody>
</table>

Cox proportional hazards adjusting for age, race, total energy, stage at diagnosis, body mass index, and education.
Lignans is converted/fermented by the gut microbiota by to enterolactone.

Enterolactone is a weak estrogen.

Estrogen dependent mechanisms include agonist and antagonist effects on the estrogen receptor, depended on the level of estrogen exposures.

In vitro studies have found, enterolactone to inhibit metastasis and reduce cell proliferation.

Enterolactone may improve prognosis among post menopausal women with breast cancer.
Mortality and prediagnostic level of enterolactone

Enterolacton and breast cancer - metaanalyses

Figure 1. Meta-analysis: Association between lignan exposure and breast cancer prognosis in postmenopausal women. Results for all-cause mortality (a) and breast cancer-specific mortality (b) are pre-
Enterolactone and the Danish Prescription Registry

Flaxseed, whole grains, vegetables, berries etc.  Lignans Enterolactone og enterodiol

Use of antibiotics is associated with lower enterolactone plasma concentration

Anne K. Bolvig, Cecilie Kyø, Natalja P. Nørskov, Anne K. Eriksen, Jane Christensen, Anne Tjønneland, Knud E. Bach Knudsen and Anja Olsen

1 Department of Animal Science, Aarhus University, Tjele, Denmark
2 Unit of Diet, Genes and Environment, Danish Cancer Society Research Center, Copenhagen, Denmark

Scope: High enterolactone levels may have health benefits in relation to risk of noncommunicable diseases. Enterolactone is produced by the colonic microbiota after intake of lignans and treatment with antimicrobials may result in altered enterolactone production. This study investigates the association between antibiotic use and enterolactone concentration.

Methods and results: Using LC-MS/MS, enterolactone concentrations were quantified in plasma samples from 2237 participants from the Diet, Cancer and Health cohort. The participants were healthy at enrollment, but were later diagnosed with cancer. At enrollment, participants had blood drawn and completed a food frequency questionnaire and lifestyle questionnaire. Antibiotic use was assessed as reimbursed antibiotic prescriptions up to 12 months before enrollment. Antibiotic use <3 months before enrollment was associated with a 41% (Δmbs: −41; 95% CI: −52, −28) lower enterolactone concentration in women and 12% in men (Δmbs: −12; 95% CI: −31, 11), while antibiotic use >3-12 months before enrollment was associated with 26% lower enterolactone in women (Δmbs: −26; 95% CI: −37, −14) and 14% in men (Δmbs: −14; 95% CI: −28, 1).

Conclusion: Use of antibiotics up to 12 months before enrollment was associated with lower plasma enterolactone levels, especially among women.

Keywords: Antibiotics / Enterolactone / Epidemiology / Lignans / Microbiota

Additional supporting information may be found in the online version of this article at the publisher’s web-site
Antibiotic treatment/reimbursed prescription and levels of enterolactone, women

Table 2. Percentage difference in enterolactone plasma concentration and 95% CI by most recent antibiotic use among 2237 participants included in the Diet, Cancer and Health cohort

<table>
<thead>
<tr>
<th>R²</th>
<th>Female (n = 1106)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Crude model</td>
<td>Model-1&lt;sup&gt;a)&lt;/sup&gt;</td>
<td>Model-2&lt;sup&gt;b)&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>0.030</td>
<td>0.075</td>
<td>0.089</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ</td>
<td>95% CI</td>
<td>p-Value</td>
<td>95% CI</td>
<td>p-Value</td>
<td>95% CI</td>
<td>p-Value</td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>No antibiotic treatment</td>
<td>731</td>
<td>Ref.</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Antibiotic use 0–3 months</td>
<td>132</td>
<td>-41</td>
<td>-52</td>
<td>-28</td>
<td>&lt;0.0001</td>
<td>-39</td>
</tr>
<tr>
<td>Antibiotic use 3–12 months</td>
<td>243</td>
<td>-26</td>
<td>-37</td>
<td>-14</td>
<td>0.0002</td>
<td>-24</td>
</tr>
</tbody>
</table>

<sup>a</sup> Model-1 is adjusted for smoking, schooling, alcohol consumption, and BMI.
<sup>b</sup> Model-2 is adjusted for smoking, schooling, alcohol consumption, BMI, and whole-grain intake.

The percentage estimates were derived from regression with log-transformed values. The results presented are back-transformed log-values.

n, number of participants; Ref., reference (Δ = 0); R², fitness of model; Δ, estimates reported as percentage change in enterolactone concentration.
Conclusion:

- Level of enterolactone depends on the ‘recent’ intake of antibiotics.

- This should be taken into account, when analysing the association between enterolactone and disease endpoints.
Assessment of whole-grain intake
Assessment of whole-grain intake

• Most cohort studies have no information on whole-grain intake

• In some cohorts, the whole-grain intake is very low

• Dietary assessment of whole-grain intake from questionnaires can especially be difficult

• Biomarkers of intake could overcome some of these problems
Alkylresorcinols
- Biomarkers of whole-grain intake

- Phenolic lipids – found in the bran part of rye and wheat
- Unaffected by food processing
- Validated (measured in plasma) both in intervention studies and in cohort studies
  - Questionnaire vs. biomarker: r=0.25–0.57

<table>
<thead>
<tr>
<th>Alkylresorcinol</th>
<th>Abbreviation used</th>
<th>R</th>
<th>Molecular weight (g/mole)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-n-Heptadecylresorcinol</td>
<td>(C17:0)</td>
<td>C_{17}H_{35}</td>
<td>348</td>
</tr>
<tr>
<td>5-n-Nonadecylresorcinol</td>
<td>(C19:0)</td>
<td>C_{19}H_{39}</td>
<td>376</td>
</tr>
<tr>
<td>5-n-Hexadecylresorcinol</td>
<td>(C21:0)</td>
<td>C_{21}H_{43}</td>
<td>404</td>
</tr>
<tr>
<td>5-n-Tricosylresorcinol</td>
<td>(C23:0)</td>
<td>C_{23}H_{47}</td>
<td>432</td>
</tr>
<tr>
<td>5-n-Pentacosylresorcinol</td>
<td>(C25:0)</td>
<td>C_{25}H_{51}</td>
<td>460</td>
</tr>
</tbody>
</table>

*Figure 1. Structures of alkylresorcinols (ARs) commonly found in cereals.*
Research projects - methods

Questionnaire data/ Record Linkage
- Colorectal cancer (1100 cases)
- Myocardial infarction (2300 cases)
- Diabetes (7000 cases)
- Mortality (7800 deceased)

Biomarker
- Nested case-control design
- 1372 colorectal cases and 1372 matched controls
- The biomarker “Alkylresorcinols” analyzed using GC-MS
Plasma Alkylresorcinols, Biomarkers of Whole-Grain Wheat and Rye Intake, and Incidence of Colorectal Cancer


Manuscript received June 27, 2013; revised October 25, 2013; accepted October 29, 2013.

Correspondence to: Cecilie Kyre, MSc, PhD, Danish Cancer Society Research Center, Strandboulevarden 49, 2100 Copenhagen Ø, Denmark (e-mail: ceciliek@cancer.dk).

Background

Few studies have investigated the association between whole-grain intake and colorectal cancer. Because whole-grain intake estimation might be prone to measurement errors, more objective measures (e.g., biomarkers) could assist in investigating such associations.

Methods

The association between alkylresorcinols, biomarkers of whole-grain rye and wheat intake, and colorectal cancer incidence were investigated using prediagnostic plasma samples from colorectal cancer case patients and matched control subjects nested within the European Prospective Investigation into Cancer and Nutrition. We included 1372 incident colorectal cancer case patients and 1372 individual matched control subjects and calculated the incidence rate ratios (IRR) for overall and anatomical subsites of colorectal cancer using conditional logistic regression adjusted for potential confounders. Regional differences (Scandinavia, the Mediterranean, Central Europe) were also explored.

Results

High plasma total alkylresorcinol concentration was associated with lower incidence of distal colon cancer; the adjusted incidence rate ratio of distal colon cancer for the highest vs lowest quartile of plasma total alkylresorcinols was 0.48 (95% confidence interval [CI] = 0.28 to 0.83). An inverse association between plasma total alkylresorcinol concentrations and colon cancer was found for Scandinavian participants (IRR per doubling = 0.83; 95% CI = 0.70 to 0.98). However, plasma total alkylresorcinol concentrations were not associated with overall colorectal cancer, proximal colon cancer, or rectal cancer. Plasma alkylresorcinols concentrations were associated with colon and distal colon cancer only in Central Europe and Scandinavia (i.e., areas where alkylresorcinol levels were higher).

Conclusions

High concentrations of plasma alkylresorcinols were associated with a lower incidence of distal colon cancer but not with overall colorectal cancer, proximal colon cancer, and rectal cancer.

Mean plasma levels of alkylresorcinols

Wheat dominated diet  Rye dominated diet

France n=58  Italy n=374  Spain n=318  Greece n=64  UK n=428  NL n=298  Germany n=308  Sweden n=156  Denmark n=706  Norway n=32

Calculated using the ratio between two of the alkylresorcinol homologues
Studies on colorectal cancer - MEN

Hansen L et al. Int J Cancer 2012 (HELGA cohort)
Kyrø C et al. Cancer Causes Control 2013 (HELGA cohort)
Studies on colorectal cancer - WOMEN

Hansen L et al. Int J Cancer 2012 (HELGA cohort)
Kyrø C et al. Cancer Causes Control 2013 (HELGA cohort)
Studies on other diseases - MEN

Johnsen NF et al. Br J Nutr 2015 (HELGA cohort)
Helnaes et al., Am J Clin Nutr 2016 (Diet, Cancer and Health cohort)
Whole grains and type 2 diabetes – in preparation! (Diet, Cancer and Health cohort)
Conclusions and perspectives

- Whole grains associated with lower risk of colorectal cancer
- In new update of the WCRF/AICR report – recommendation for whole grains (2017)
- Also beneficial in relation to other non-communicable diseases and overall mortality!
<table>
<thead>
<tr>
<th><strong>2017</strong></th>
<th><strong>DIET, NUTRITION, PHYSICAL ACTIVITY AND COLORECTAL CANCER 2017</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STRENGTH OF EVIDENCE</strong></td>
<td><strong>DECREASES RISK</strong></td>
</tr>
<tr>
<td><strong>STRONG EVIDENCE</strong></td>
<td>Convincing</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Probable</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMITED EVIDENCE</td>
<td>Limited – suggestive</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limited – no conclusion</td>
</tr>
</tbody>
</table>
Conclusion – Use of Linkage data in cohort studies:

• More efficient data collection and lower participant burden, multiple outcome domains in the same cohort of individuals, at a low cost

• Collection of information that cannot be obtained by participants or biomarkers

• Increased information for correction of participant bias e.g. missing data, objective measures
Overall conclusion:

• Follow up in Cohort studies is not possible without linkage to national registries

• Better identification of high risk groups, and improvement of personal prevention and treatment

• Challenges in relation to data storage and handling

• Ethical aspects, balance protection of participants info vs nature and constraints of the research, GDPR

• Important and necessary contribution to public health research
Acknowledgements

**DCRC**
- Anja Olsen
- Cecilie Kyrø
- Louise Hansen
- Jytte Halkjær
- Nick Martinussen
- Katja Boll

**External partners**
- Rikard Landberg
  Chalmers, Sweden
- Bas Bueno-de-Mesquita
  RIVM and UMC Utrecht, The Netherlands
- Guri Skeie
  UiT, The Arctic University of Norway, Norway
Thank you for your attention