A Bayesian hierarchical model to integrate dietary exposure and biomarker measurements for the risk of cancer

UNIGE's Data Science Day

Marta Pittavino

UNIVERSITÉ DE GENÈVE

GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT

September 16, 2021

Background

<u>**In nutritional epidemiology**</u>, **dietary exposure** is measured with:

1) <u>Dietary questionnaires **Q**</u>: with detailed questions on habitual diet
2) <u>24-hour dietary recalls **R**</u>: reference measurements based previous day

Errors:

- Self-reported assessments of dietary exposure are prone to <u>random and systematic measurement errors</u>
- <u>Estimates</u> of the association between dietary factors and risk of disease can be <u>biased</u>

To account for exposure misclassification:

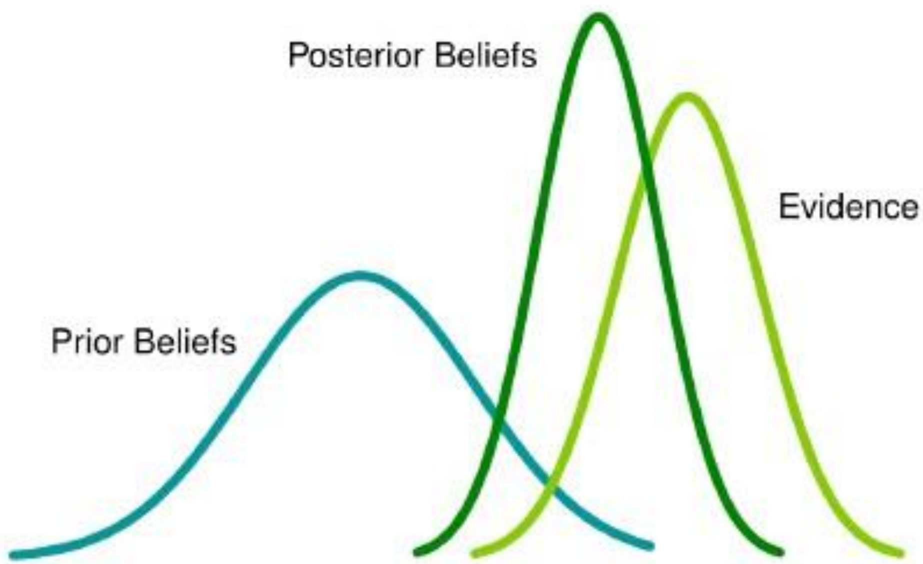
It has been suggested to complement self-reported dietary assessments with objective measurements, such as <u>dietary biomarkers **M**</u>

A holistic approach with ALL measurements still missing!

The Bayesian approach

. . . . The essence of the Bayesian approach is to provide a mathematical rule explaining how you change your existing beliefs in the light of new evidence. In other words, it allows scientists to combine new data with their existing knowledge or expertise . . .

The Economist, September 30th 2000



Research Question:

Is it possible to estimate the relationship between true T dietary exposure and the risk of cancer Y by modelling both self-reported Q, R and biomarker M measurements?

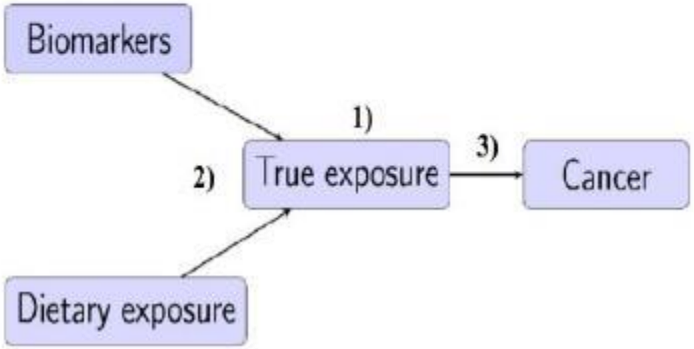
UNIVERSITÉ DE GENÈVE

GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT

Yes, with a Bayesian hierarchical model!

It enables to estimate the association between Diet and Cancer in order to:

- integrate self-reported (Q, R) and objective measurements (M)
- explore and correct for the measurement error structure of the data
- estimate the true (latent factor) dietary intakes (T)
- relate dietary exposures to risk of cancer (Y)



There are three structural components:

1) an exposure model (T), to define the distribution of unknown true exposure
2) a measurement model (Q,R,M), to disclose the relationship between observed measurements (Q,R,M) and the true exposure T

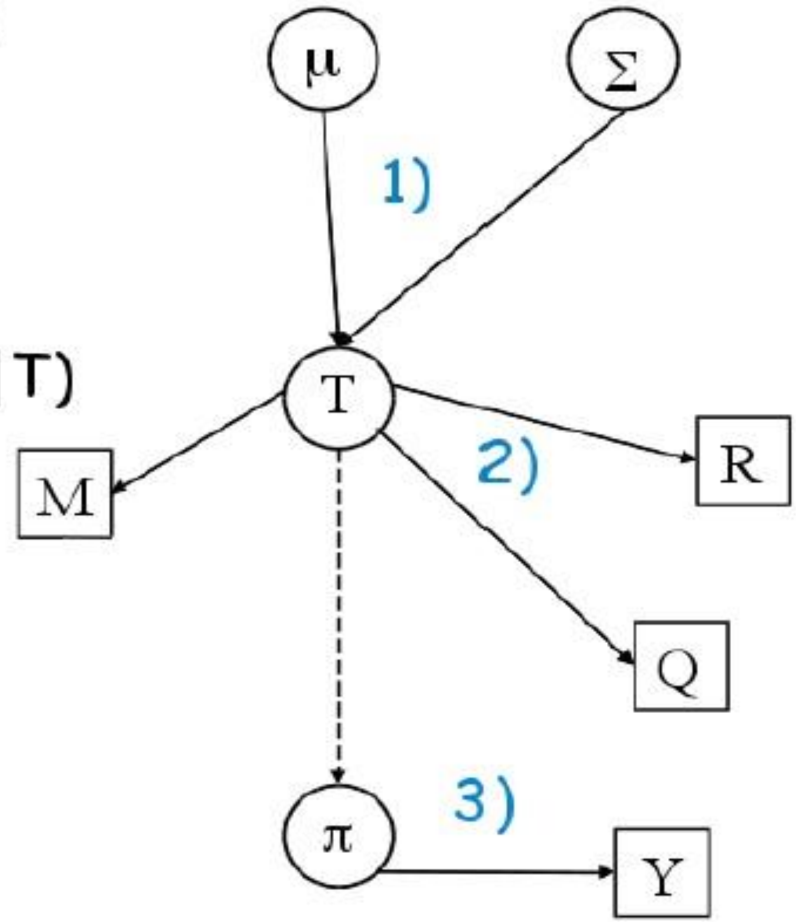
3) a disease model (Y), to estimate the relationship between dietary exposures and disease status Y

The Bayesian hierarchical model

1) Exposure Model $p(T|\mu, \Sigma)$

2) Measurement Model $p(Q,R,M|T)$

3) Disease Model $p(Y|T)$



Why using a Bayesian hierarchical model?

- Cornerstone of hierarchical modeling is **conditional thinking**
- Joint distribution can be written as product of conditional probabilities:

  P(T; (Q,R,M); Y) = P(T)*P(Q,R,M|T)*P(Y|T)
- Hierarchical models are used to build complex models through simpler conditional independent relationships
- Easier to express conditional models than full joint models

Bayesian hierarchical models are flexible tools, used to answer complex questions!

Application to the European Prospective Investigation into Cancer and Nutrition (EPIC) study

Data on:

- dietary and biomarker measurements of **vitamin-B6 and B9, folate**
- two (**kidney and lung**) cancers study

RESEARCH ARTICLE

For more information:

A Bayesian hierarchical framework to integrate dietary exposure and biomarker measurements into aetiological models

Marta Pittavino[1*] | Martyn Plummer[2†] | Mattias Johansson[3] | Pietro Ferrari[1]