NCRM National Centre for Research Methods

Biological Research: Biological Data Quality Issues

Tarani Chandola

NCRM, University of Manchester

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

UNIVERSITY OF Southampton

MANCHESTER 1824 The University of Manchester

THE UNIVERSITY OF EDINBURGH

Methodological considerations

- Biosocial research framework
- Biological data quality issues
- Missing data in biosocial research

Biological Data in Social Surveys

- Does the mode or collection condition influence levels of biomarkers?

- What Quality Control (QC) processes are important?

Investigating mode/ collection conditions

A number of biosocial studies tend to use two methods to collect biomarker data:

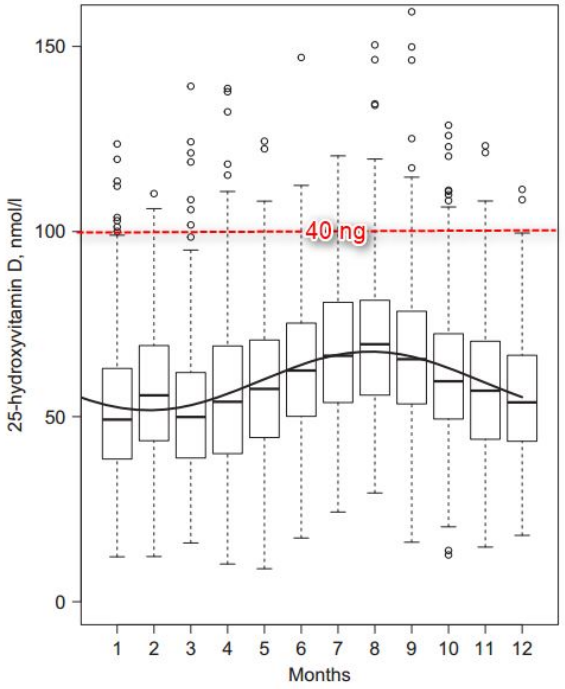
- Clinic collection: participants are invited to a clinic where blood samples are collected and stored/processed immediately
  - ALSPAC
  - National Study Health and Development (NSHD) @63
  - Hertfordshire Cohort Study
  - Whitehall II
- Home Nurse visit: participants are visited at home; blood samples are collected and posted to the laboratory
  - Understanding Society: UKHLS
  - Southampton Women's Study
  - National Child Development Study
  - NSHD@53
  - Health Survey for England
  - English Longitudinal Study of Ageing

Clinic vs home visit biological data

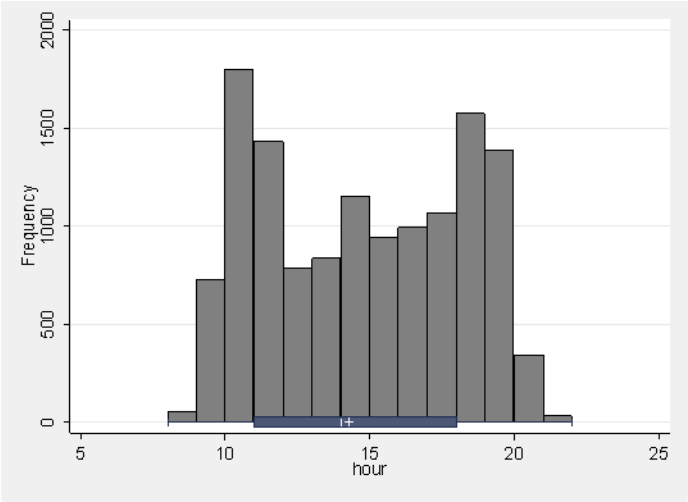
- Ideal sample collection conditions (usually in a clinical setting):
  - Venepuncture to collect blood sample
  - Immediate sample processing
  - Immediate analyte measurement or storage in a -80 freezer
- Typical sample collection conditions (usually in population surveys):
  - Venepuncture to collect blood sample
  - Delayed sample processing and storage

Are collection conditions important?

Clinic vs Home

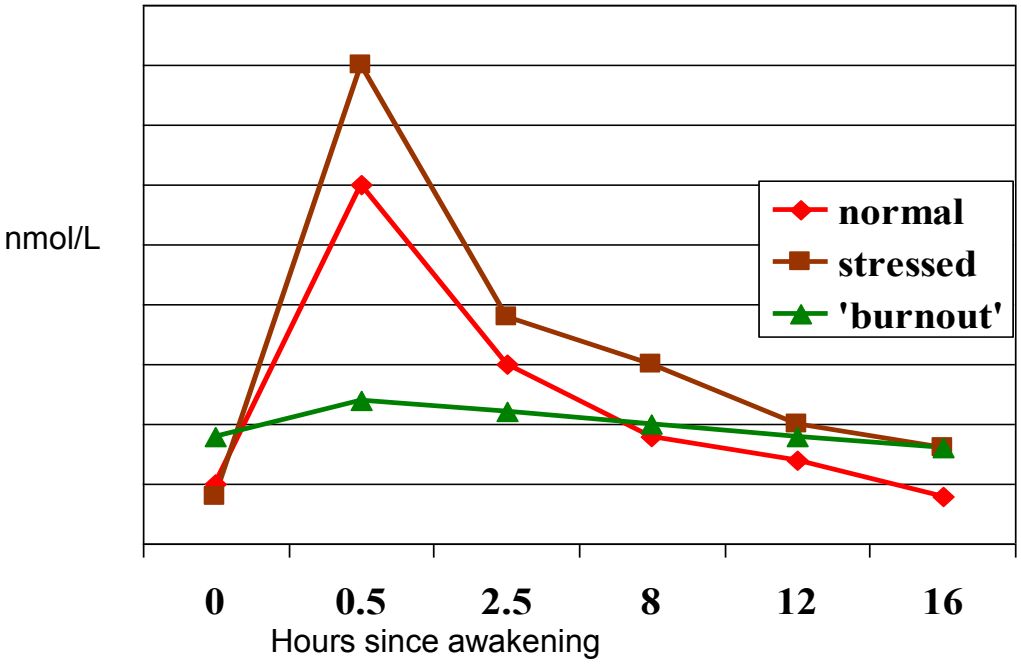


Daily variation: many biomarkers vary by time of day



Bi-modal distribution of nurse visits

Understanding Society: The UK Household Longitudinal Study
https://www.understandingsociety.ac.uk/



Diurnal variation in cortisol

Quality control Data on biomarkers

Internal and external Quality control (QC) processes

Internal:

- Some biomarkers have impossible values
- How well do measurements compare across time within a laboratory (ie does a sample measure the same if the measurement is made on day ‘a’ as it does on day ‘a’+1).
- Measured through intra-assay coefficient of variation
- Less than 5% is within acceptable limits.

Quality control Data on biomarkers

External quality control reflects how the laboratory compares to other laboratories measuring the same analyte.

-Measured through the standard deviation index (SDI) an index of total error, including components of inaccuracy and imprecision

- lower values of SDI suggest more accurate measures

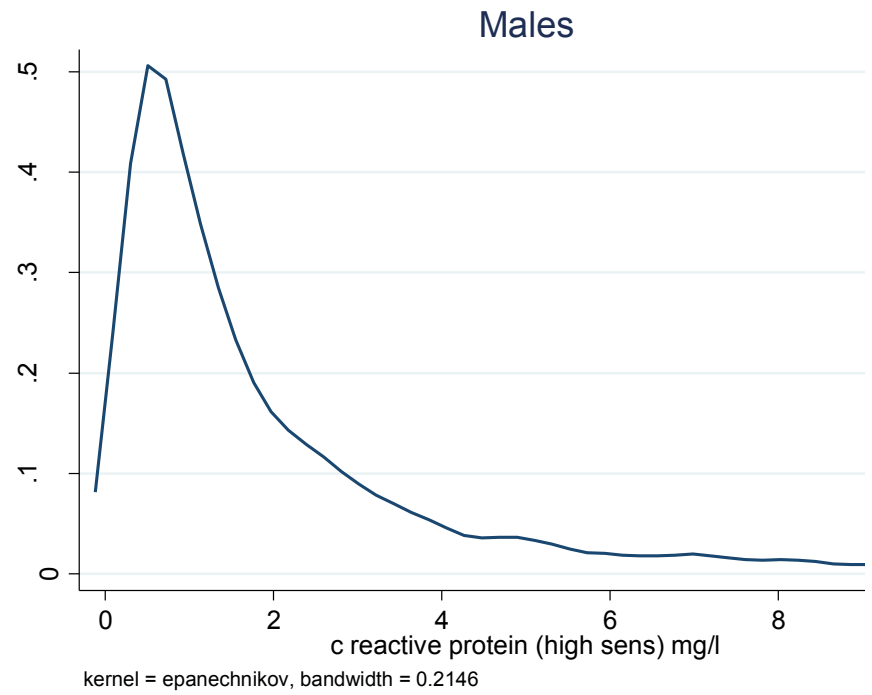
- score below 1 SDI is good, and between 1-2 SDI is acceptable

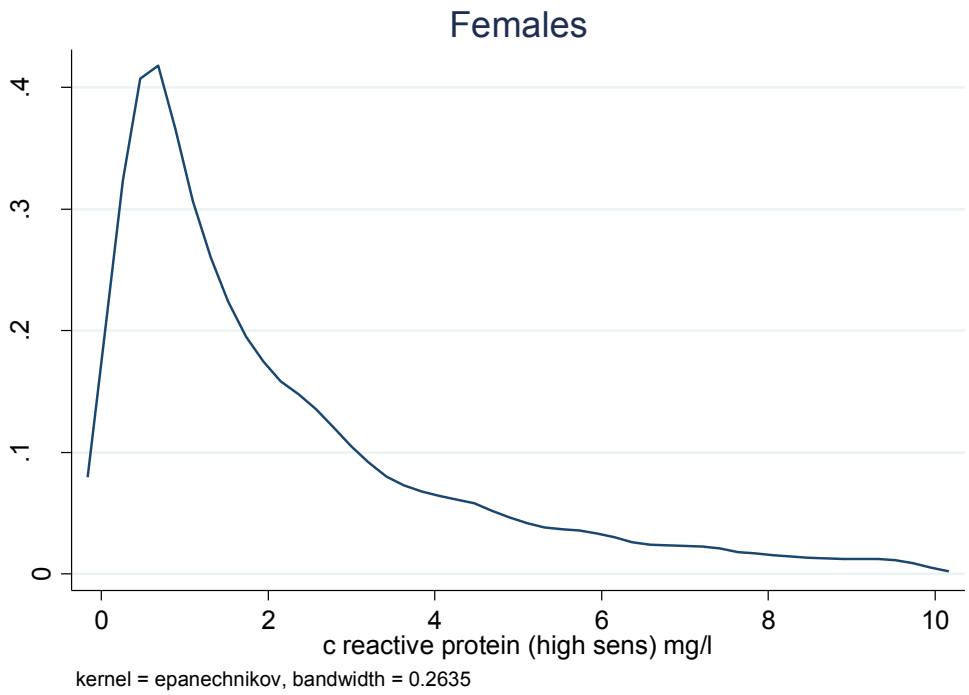
Specific biomarkers: methodological considerations

C-Reactive Protein (CRP)

- Systemic inflammation: 3-10mg/L
- Current/recent infections: >10mg/L (often excluded)
- In general, CRP levels of over 3mg/L are considered as levels that are high risk for Cardiovascular disease.
- CRP is influenced by medication: anti-inflammatory medications, statins and contraception and hormone replacement therapy

Distribution of C-Reactive Protein (CRP) by gender





Notes: Excluded cases with CRP>10mg/L

Methodological considerations when analysing biological data

- Consider:
  - normal ranges of biological variables (if available)
  - identify outliers
  - relevant medication use
  - transformations (for skewed biological dependent variables)
  - context of blood sampling like time of day, room temperature, recent operations, smoking, food & alcohol, etc
  - quality control processes in producing biological data