Composite scores, social embodiment and risk of CVD: Evidence from the UK Biobank cohort

SoT – Risk Assessment Specialty Section webinar – 10th February, 2021

Marc Chadeau-Hyam
m.chadeau@imperial.ac.uk
LIFEPATH: Aims and overall approach

- **Aim:** Investigate mechanisms involved in the quality of ageing and health risk
- **Data types:** Social factors, biological markers, and health outcomes
- **Overarching framework:**

![Diagram showing the relationship between Social Factors & Experiences, Biology & Embodiment, and Health & Aging.](Diagram)

- **Social Factors & Experiences**
  - Heterogenous Data: capturing lifestyle, environmental, and cultural aspect of experiences
  - Dynamic effect:
    - Existence of critical life stages
    - Role of social mobility
  - Towards multi-level & life course modelling

- **Biology & Embodiment**
  - Complex biological response: account for the gradient of granularity
  - Three levels of investigation:
    - Synthetic scores
    - Prioritised pathways
    - OMICs profiling & integration
  - Complexity reduction to improve interpretability

- **Health & Aging**
  - Different outcome resolution:
    - Mortality
    - Functional outcomes
    - Specific diseases
  - Temporality:
    - Exposure effect
    - Incidence
  - Lifecourse and causal modelling
Stream 1: Linking Social Factors and health outcomes

- Robust Evidence that social factors affect health risk

Socioeconomic status and the 25 x 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women

Silke Struyf*, Cristian Carnot*, Mohsin Jakle*, Mouhia-Ameriene*, Peter Mouris, Alexandra Guda, Fritzie Rizal, Angelo PÉrez,

...
Stream 1: Linking Social Factors and health outcomes

- 46 cohorts support higher mortality in lower SES in men, HR=1.4
- 44 of the cohorts provide consistent results in women, HR=1.3
Stream 1: Linking Social Factors and Functional outcomes

- By age 85, low vs high SES is associated with a loss of 7-11 years of functioning.
- These cannot be attributed to health risk factors.
- Most risk factors (except physical activity) are related to smaller losses of functioning.
Stream 2: Investigating biomarkers of SEP

- Heterogenous Data: capturing lifestyle, environmental, and cultural aspects of experiences
- Dynamic effect:
  * Existence of critical life stages
  * Role of social mobility

Towards multi-level & life course modelling

- Complex biological response: account for the gradient of granularity
- Three levels of investigation:
  * Synthetic scores
  * Prioritised pathways
  * OMICs profiling & integration

Complexity reduction to improve interpretability

- Different outcome resolution:
  * Mortality
  * Functional outcomes
  * Specific diseases
- Temporality:
  * Exposure effect
  * Incidence

Lifecourse and causal modelling

- 3 approaches corresponding to a gradient of granularity
  1. Focus on prioritized pathways: Inflammation
  2. Synthetic scores
  3. OMICs approaches
Focus on Inflammation

- Data: 6 LIFEPATH cohorts (N=23,008), with SEP factors in the life-course & CRP measurement in blood as a proxy for inflammatory status.

- Aim: Explore the CRP-SEP association across country; in the life-course; evaluate the role of lifestyle factors and behaviors

- Results: overall higher inflammatory burden in lower SEP group
  - Consistent gradient irrespective of the SEP metric
  - Stronger associations in women
  - Lifestyle factors marginally attenuate the associations

Multi-cohort study identifies social determinants of systemic inflammation over the life course

Eloïse Berger, Raphaëlle Castagné, Marc Chadeau-Hyam, Murielle Bochud, Angelo d’Enrico, Martina Gandini, Maryam Karimi, Mika Kivimäki, Vittorio Krogh, Michael Marmot, Salvatore Panico, Martin Preisig, Fulvio Ricceri, Carlotta Sacerdote, Andrew Steptoe, Silvia Stringhini, Rosario Tumino, Paolo Vines, Lynne Delpiere & Michelle Kelly-Irving

https://doi.org/10.1038/s41467-019-08722-x
Complexity Reduction approach: Allostastic Load

- **Definition:** multi-system synthetic score capturing physiological wear-and-tear (6 systems included)
- **Data:** SKIPOGH study (N=1,128), with SEP & 14 blood-derived biomarkers
- **Results:** SEP-AL associations by gender
  - Main trend: higher AL for lower SEP categories
  - Stronger associations for education, and in women
  - Lifestyle factors marginally attenuate the associations

⇒ stronger gradient for early life SEP
The BHS an extension over the allostatic load

Early-life inequalities and biological ageing: a multisystem Biological Health Score approach in *Understanding Society*

Maryam Karimi, Raphaële Castagné, Cyrille Delpierre, Gaëlle Albertus, Eloïse Berger, Paolo Vineis, Meena Kumari, Michelle Kelly-Irving, Marc Chadeau-Hyam

Data: Understanding Society (N=9,088), with educational attainment & 16 blood-derived biomarkers capturing 6 physiological systems (including liver and kidney functions)

Aims: Define BHS as an extension from the AL

1. Explore BHS gradient across SEP groups and age classes
2. Quantify the relative contribution of each system
Investigating social gradients in composite scores

Results: systematic SEP-related gradient (higher scores in disadvantaged pop)

- Consistent results in men (A) and women (B)
- Gradient is observed in all age groups
- Gradient is not affected by adjustment on lifestyle factors
  \[ \Rightarrow \text{effects are detected in early adulthood and persist} \]
Step 3: From biology to Health Outcomes

- Heterogenous Data: capturing lifestyle, environmental, and cultural aspect of experiences
- Dynamic effect:
  - Existence of critical life stages
  - Role of social mobility

Complex biological response: account for the gradient of granularity
- Three levels of investigation:
  - Synthetic scores
  - Prioritised pathways
  - OMICs profiling & integration

Complexity reduction to improve interpretability

Different outcome resolution:
- Mortality
- Functional outcomes
- Specific diseases
- Inciden

Towards multi-level & life course modelling

Gradient of resolution:
1. Low resolution biological factors
2. Pathways
3. Composite Scores
4. Full-resolution OMICs profiles
Allostatic Load and mortality

https://doi.org/10.1007/s10654-018-0364-1

Allostatic load and subsequent all-cause mortality: which biological markers drive the relationship? Findings from a UK birth cohort

Raphaële Castagné¹ · Valérie Garès¹ · Maryam Karimi² · Marc Chadeau-Hyam² · Paolo Vineis²,³ · Cyrille Delpierre¹ · Michelle Kelly-Irving¹ · for the Lifepath Consortium

• Data: 1958 British Birth cohort (N=8,113) 14 blood-derived biomarkers. 132 deaths

• Aims
  1. Evaluate the effect of AL and its constituents on mortality
  2. Investigate the role of SEP (education) and behaviors in these associations.
**Allostatic Load and mortality**

Results: Hazard Ratio by system and biomarker and for AL

- Positive contribution of all markers/system to mortality (except NE)
- Attenuation upon adjustment on adulthood confounders
- Effect of AL, and most system remain significant after adjustment for behaviors and SEP (not shown).

⇒ AL at 44 predicts mortality irrespective of subsequent SE experiences
⇒ the multi-system AL predicts better than each system separately
BHS, mortality and incident pathologies: UKBiobank

- **Study Overview:** 502,536 volunteers from the UK aged 37-73 years at entry between 2006 and 2010.

- **Questionnaire data:** computer-based questionnaire on life-course exposures, medical history and treatments.

- **Anthropometric/clinical data:** from clinical assessment centres computer-based questionnaire on life-course exposures, medical history and treatments.

- **Mortality Outcomes:** linkage to death registers

- **Health status Follow-up:** range from 0.2 to 12.04 years

- **Incident pathologies** identified through linkage to NHS central registers, cancer and hospital registers, and/or nurse-administered questionnaire.

- **Biosampling:** participants donated one blood sample at baseline
  - Genome-wide scans were measured (N=672,345 genotyped SNPs in 488,377 participants)
  - Panel of 30 prioritised biomarkers
**Health Outcomes of interest in UK Biobank**

- **Cancer Outcomes:** all sites
- **CVD Outcomes:** coronary arterial disease, angina, stroke, and related outcomes
- **External cause mortality:** suicide and accident
- **Mortality Outcomes:** all-cause, cancer, CVD and external cause

<table>
<thead>
<tr>
<th></th>
<th>All-cause</th>
<th>Cancer</th>
<th>CVD</th>
<th>External causes</th>
<th>Other causes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>8,735</td>
<td>4,316</td>
<td>1,547</td>
<td>328</td>
<td>2,544</td>
</tr>
<tr>
<td>Females</td>
<td>5,661</td>
<td>3,698</td>
<td>352</td>
<td>160</td>
<td>1,451</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>14,396</strong></td>
<td><strong>8,014</strong></td>
<td><strong>1,899</strong></td>
<td><strong>488</strong></td>
<td><strong>3,995</strong></td>
</tr>
</tbody>
</table>

- **Incident pathologies:** cancer and CVD incidence

<table>
<thead>
<tr>
<th></th>
<th>Cancer</th>
<th>CVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>26,123</td>
<td>10,114</td>
</tr>
<tr>
<td>Females</td>
<td>26,320</td>
<td>5,539</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>52,443</strong></td>
<td><strong>15,653</strong></td>
</tr>
</tbody>
</table>
Biomarkers selection

• **UK Biobank Biomarkers**: 13 measuring 5 systems:
  1. **Metabolic system (N=4)**: Glycated haemoglobin (HbA1c), High-density lipoprotein cholesterol (HDL), Low-density lipoprotein cholesterol (LDL), Triglycerides (Tri);
  2. **Cardiovascular system (N=3)**: Systolic and diastolic blood pressure (SBP, and DBP, respectively), pulse (Pulse);
  3. **Inflammatory / immune system (N=2)**: C-Reactive Protein (CRP), and Insulin-like growth factor 1 (IGF-1);
  4. **Liver function (N=3)**: Alanine transaminase (ALT), Aspartate transaminase (AST), Gamma glutamyltransferase (GGT);
  5. **Kidney function (N=1)**: Creatinine (Cre);
Complexity Reduction: Composite score

- **Biomarkers dichotomisation:** we define the ’at-risk’ quartile as the lowest quartile for HDL and IGF-1, and the highest quartile for all remaining 11 biomarkers. Quartiles were defined for each gender and age group (<50, 50-64, and >64 years old) separately.

- **Scores derivation:** For a given system \( s \), and individual \( i \)

  \[
  \text{sub-BHS}^i_s = \sum_{k=1}^{13} \frac{I^i_k}{n_s}, \text{ where}
  \]

  \( I^i_k \) is the binary score for biomarker \( k \), and \( n_s \): # of biomarkers in system \( s \)

- **BHS definition:**

  \[
  BHS^i = \frac{1}{n_s} \sum_{s=1}^{5} \text{sub-BHS}^i_s \]

  where \( n_s \) is the number of systems in the BHS

  \( \Rightarrow \) the BHS and sub scores are all on the same scale (\( \in [0, 1] \))
Main Analytical Plan

- **Descriptive Analyses: Investigate social gradients in BHS**
  - Compare BHS levels by education level (low, intermediate, high)
  - Investigate the role of socially-patterned exposures and behaviours in these gradients (medical status, smoking, physical activity, alcohol, and BMI)

- **Survival Analyses: proportional hazards Cox models**
  - Investigate the role of the BHS (and sub-scores) in mortality and incident pathologies: setting the BHS (or sub-scores as predictor)
  - Attenuation analyses: sequential adjustment for (i) Education, (ii) Behaviours, (iii) BMI, (iv) Medical History

- **Investigate the role of Education:**
  - Similar survival analyses, setting the Education level as predictor
  - Attenuation analyses: sequential adjustment for (i) Behaviours, (ii) BMI, (iii) Medical History, and (iv) BHS
We selected a total of 366,748 participants (171,193 men and 195,555 women) who were free of cancer and CVD at baseline.
• Unambiguous social gradient in the BHS from UK Biobank
• Slightly right shifted BHS distributions in incident CVD cases

⇒ has the BHS an effect on mortality and incidence, independent of education?
## Survival analyses: Univariate Models

### Mortality

<table>
<thead>
<tr>
<th></th>
<th>All-cause</th>
<th>Cancer</th>
<th>CVD</th>
<th>External cause</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=4,428</td>
<td>N=2,225</td>
<td>N=681</td>
<td>N=220</td>
</tr>
<tr>
<td><strong>HR [95% CI]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BHS</strong></td>
<td>1.14 [1.12-1.16]</td>
<td>1.11 [1.09-1.14]</td>
<td>1.25 [1.20-1.31]</td>
<td>0.99 [0.91-1.08]</td>
</tr>
<tr>
<td></td>
<td>7.63x10^-44</td>
<td>1.00x10^-16</td>
<td>2.70x10^-24</td>
<td>8.49x10^-01</td>
</tr>
</tbody>
</table>

**System-specific sub-score**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metabolic</strong></td>
<td>1.05 [1.04-1.06]</td>
<td>1.04 [1.02-1.06]</td>
<td>1.14 [1.11-1.18]</td>
<td>0.97 [0.92-1.03]</td>
</tr>
<tr>
<td></td>
<td>2.66x10^-14</td>
<td>1.27x10^-05</td>
<td>5.80x10^-20</td>
<td>3.45x10^-01</td>
</tr>
<tr>
<td><strong>Cardiovascular</strong></td>
<td>1.05 [1.04-1.06]</td>
<td>1.04 [1.02-1.05]</td>
<td>1.10 [1.07-1.12]</td>
<td>1.04 [1.00-1.08]</td>
</tr>
<tr>
<td></td>
<td>2.99x10^-26</td>
<td>1.32x10^-07</td>
<td>1.06x10^-15</td>
<td>6.13x10^-02</td>
</tr>
<tr>
<td><strong>Inflammatory</strong></td>
<td>1.07 [1.06-1.08]</td>
<td>1.06 [1.05-1.08]</td>
<td>1.09 [1.07-1.11]</td>
<td>1.01 [0.97-1.05]</td>
</tr>
<tr>
<td></td>
<td>3.20x10^-57</td>
<td>7.09x10^-22</td>
<td>3.51x10^-15</td>
<td>7.00x10^-01</td>
</tr>
<tr>
<td><strong>Liver</strong></td>
<td>1.03 [1.02-1.04]</td>
<td>1.03 [1.01-1.04]</td>
<td>1.04 [1.02-1.07]</td>
<td>1.02 [0.98-1.06]</td>
</tr>
<tr>
<td></td>
<td>1.89x10^-11</td>
<td>2.06x10^-04</td>
<td>3.61x10^-04</td>
<td>3.73x10^-01</td>
</tr>
<tr>
<td><strong>Kidney</strong></td>
<td>0.99 [0.98-0.99]</td>
<td>1.00 [0.98-1.01]</td>
<td>0.99 [0.97-1.01]</td>
<td>0.95 [0.91-0.99]</td>
</tr>
<tr>
<td></td>
<td>1.24x10^-03</td>
<td>4.20x10^-01</td>
<td>5.75x10^-01</td>
<td>2.05x10^-02</td>
</tr>
</tbody>
</table>

### Incidence

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=20,962</td>
<td>N=7,925</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HR [95% CI]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BHS</strong></td>
<td>1.02 [1.01-1.03]</td>
<td>1.01x10^-04</td>
<td></td>
<td>1.15 [1.13-1.16]</td>
</tr>
<tr>
<td></td>
<td>1.28x10^-93</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**System-specific sub-score**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metabolic</strong></td>
<td>1.01 [1.00-1.01]</td>
<td>1.00 [1.00-1.01]</td>
<td>1.01 [0.99-1.00]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.20x10^-03</td>
<td></td>
<td></td>
<td>1.00 [0.99-1.00]</td>
</tr>
<tr>
<td><strong>Cardiovascular</strong></td>
<td>1.00 [1.00-1.01]</td>
<td>1.00 [1.00-1.01]</td>
<td>1.00 [1.00-1.01]</td>
<td>1.00 [1.00-1.01]</td>
</tr>
<tr>
<td></td>
<td>1.05 [1.05-1.06]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inflammatory</strong></td>
<td>1.01 [1.00-1.02]</td>
<td>1.00 [1.00-1.01]</td>
<td>1.00 [1.00-1.01]</td>
<td>1.00 [1.00-1.01]</td>
</tr>
<tr>
<td></td>
<td>1.31x10^-51</td>
<td></td>
<td>6.72x10^-32</td>
<td></td>
</tr>
<tr>
<td><strong>Liver</strong></td>
<td>1.00 [1.00-1.01]</td>
<td>1.00 [1.00-1.01]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.42x10^-10</td>
<td></td>
<td>1.42x10^-10</td>
<td></td>
</tr>
<tr>
<td><strong>Kidney</strong></td>
<td>1.00 [1.00-1.01]</td>
<td>1.00 [1.00-1.01]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.59x10^-01</td>
<td></td>
<td>8.59x10^-01</td>
<td></td>
</tr>
</tbody>
</table>
## Survival analyses: Univariate Models

### Mortality

<table>
<thead>
<tr>
<th></th>
<th>All-cause</th>
<th>Cancer</th>
<th>CVD</th>
<th>External cause</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td>4,428</td>
<td>2,225</td>
<td>681</td>
<td>220</td>
</tr>
<tr>
<td><strong>HR [95% CI]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BHS</strong></td>
<td>1.14 [1.12-1.16]</td>
<td>1.11 [1.09-1.14]</td>
<td>1.25 [1.20-1.31]</td>
<td>0.99 [0.91-1.08]</td>
</tr>
<tr>
<td></td>
<td>7.63x10^{-44}</td>
<td>1.00x10^{-16}</td>
<td>2.70x10^{-24}</td>
<td>8.49x10^{-01}</td>
</tr>
</tbody>
</table>

### Incidence

<table>
<thead>
<tr>
<th></th>
<th>Cancer</th>
<th>CVD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td>20,962</td>
<td>7,925</td>
</tr>
<tr>
<td><strong>HR [95% CI]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BHS</strong></td>
<td>1.02 [1.01-1.03]</td>
<td>1.15 [1.13-1.16]</td>
</tr>
<tr>
<td></td>
<td>1.01x10^{-04}</td>
<td>1.28x10^{-93}</td>
</tr>
</tbody>
</table>

### System-specific sub-score

<table>
<thead>
<tr>
<th></th>
<th>Metabolic</th>
<th>Cardiovascular</th>
<th>Inflammatory</th>
<th>Liver</th>
<th>Kidney</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HR [95% CI]</strong></td>
<td>1.05 [1.04-1.06]</td>
<td>1.04 [1.02-1.06]</td>
<td>1.07 [1.06-1.08]</td>
<td>1.03 [1.02-1.04]</td>
<td>0.99 [0.98-0.99]</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>2.66x10^{-14}</td>
<td>1.27x10^{-05}</td>
<td>3.20x10^{-57}</td>
<td>1.89x10^{-11}</td>
<td>1.24x10^{-03}</td>
</tr>
</tbody>
</table>

### Observations

- **BHS** is associated to increased mortality, from all-cause, cancer and **CVD**: HR range 1.11 to 1.25, $p < 10^{-16}$ in men.
- None of the scores are related to external cause mortality.
- Unlike other systems, kidney weakly contributes to mortality.
### Survival analyses: Univariate Models

#### Men

<table>
<thead>
<tr>
<th></th>
<th>All-cause Mortality</th>
<th>Cancer Incidence</th>
<th>CVD Incidence</th>
<th>External cause Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td>4,428</td>
<td>2,225</td>
<td>681</td>
<td>220</td>
</tr>
<tr>
<td><strong>HR [95% CI]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BHS</strong></td>
<td>1.14 [1.12-1.16]</td>
<td>1.11 [1.09-1.14]</td>
<td>1.25 [1.20-1.31]</td>
<td>0.99 [0.91-1.08]</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>7.63x10^{-44}</td>
<td>1.00x10^{-16}</td>
<td>2.70x10^{-24}</td>
<td>8.49x10^{-01}</td>
</tr>
</tbody>
</table>

**System-specific sub-score**

<table>
<thead>
<tr>
<th></th>
<th>Metabolic</th>
<th>Cardiovascular</th>
<th>Inflammatory</th>
<th>Liver</th>
<th>Kidney</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HR [95% CI]</strong></td>
<td>1.05 [1.04-1.06]</td>
<td>1.04 [1.02-1.05]</td>
<td>1.07 [1.06-1.08]</td>
<td>1.03 [1.02-1.04]</td>
<td>0.99 [0.98-0.99]</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>2.66x10^{-14}</td>
<td>1.27x10^{-05}</td>
<td>3.20x10^{-57}</td>
<td>1.89x10^{-11}</td>
<td>1.24x10^{-03}</td>
</tr>
</tbody>
</table>

#### Incidence

<table>
<thead>
<tr>
<th></th>
<th>Cancer</th>
<th>CVD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td>20,962</td>
<td>7,925</td>
</tr>
<tr>
<td><strong>HR [95% CI]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **BHS** is associated to increased cancer and CVD incidence
- **Weaker effects for cancer incidence**: HR 1.02 ($p<10^{-4}$)
- **All systems but kidney are associated to CVD incidence**
- **Only Metab and Inflamm systems are associated to cancer incidence**
## Survival analyses: Univariate Models

<table>
<thead>
<tr>
<th></th>
<th>Mortality</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All-cause</td>
<td>Cancer</td>
</tr>
<tr>
<td></td>
<td>N=2,716</td>
<td>N=1,689</td>
</tr>
<tr>
<td><strong>BHS</strong></td>
<td>HR [95% CI]</td>
<td>p-value</td>
</tr>
<tr>
<td></td>
<td>1.09 [1.07-1.12]</td>
<td>8.38x10^{-16}</td>
</tr>
<tr>
<td><strong>System-specific sub-score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolic</td>
<td>1.04 [1.03-1.06]</td>
<td>3.91x10^{-08}</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>1.03 [1.02-1.04]</td>
<td>1.29x10^{-06}</td>
</tr>
<tr>
<td>Inflammatory</td>
<td>1.04 [1.03-1.05]</td>
<td>2.81x10^{-12}</td>
</tr>
<tr>
<td>Liver</td>
<td>1.03 [1.02-1.04]</td>
<td>1.51x10^{-07}</td>
</tr>
<tr>
<td>Kidney</td>
<td>1.00 [0.99-1.01]</td>
<td>9.93x10^{-01}</td>
</tr>
</tbody>
</table>

- Similar conclusions in Women
- Weaker effects mortality than in men
- Stronger effect size estimated for incidence
- Weaker p-values in women
None of the BHS, or system-specific sub-scores were associated to external cause mortality, irrespective of the adjustment & gender

⇒ External cause mortality serves as a negative control outcome
BHS and cancer & CVD incidence

- Contribution of BHS & all systems except kidney to cancer incidence
- Modest effect attenuation by education; stronger attenuation by BMI
- Fully adjusted HR for CVD incidence: \( HR \ 1.11 \ (p<10^{-46}) \)
- None of the scores remain associated to cancer incidence in the fully adjusted model
Explore the Role of Education

Education is associated with:

- Increased CVD Incidence
- All-cause and Cancer (men only)
- Cancer incidence & CVD Mortality (men only for low education)
Explore the Role of Education

Education is associated with:

- Modest effect attenuation by BHS; stronger attenuation by behaviours
- Education is only associated with CVD incidence in both gender and all education groups the fully adjusted model
Sensitivity analyses: unsupervised score

Approach: consider the first PC from a PCA for the 13 biomarkers or all biomarkers in each system

<table>
<thead>
<tr>
<th></th>
<th>A– All-cause</th>
<th>B– Cancer</th>
<th>C– CVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
</tr>
<tr>
<td>+ Education</td>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td>+ Behaviours</td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
<td><img src="image9" alt="Graph" /></td>
</tr>
<tr>
<td>+ BMI</td>
<td><img src="image10" alt="Graph" /></td>
<td><img src="image11" alt="Graph" /></td>
<td><img src="image12" alt="Graph" /></td>
</tr>
<tr>
<td>+ Medical</td>
<td><img src="image13" alt="Graph" /></td>
<td><img src="image14" alt="Graph" /></td>
<td><img src="image15" alt="Graph" /></td>
</tr>
</tbody>
</table>

- similar conclusions for mortality than when using the BHS
Sensitivity analyses: unsupervised score

**A – Cancer**

- Unadjusted
- + Education
- + Behaviours
- + BMI
- + Medical

**B – CVD**

- Similar conclusions for mortality than when using the BHS
- Similar conclusions for incidence
- However, much smaller effect size estimates
  \[\Rightarrow\] possible scale effect
Sensitivity analyses: Revisiting CVD definition

Approach: including in CVD systemic and circulatory diseases

- Weaker associations for mortality
- Liver and Inflammatory scores are no longer associated to CVD mortality
Sensitivity analyses: Revisiting CVD definition

As for mortality, effect on CVD mortality and incidence are weaker than when using the BHS.

Weaker results especially in women

⇒ BHS and subscores seem to predict better CAD, than systemic and circulatory diseases.
BHS and mortality & morbidity: causal assessment

- **Approach:** one sample, two-step least square Mendelian randomisation
  1. Identify the genetic instrument for the BHS (N=172 SNPs)
  2. Infer the instrumentally-explained BHS (2% of BHS explained)
  3. Infer the causal effect using Cox models
  4. Adjust the model for education
  5. Adopt a multivariable MR approach (including the instrumentally-explained education)

- **Data:** N=672,345 SNPs assayed in all 366,748 participants

<table>
<thead>
<tr>
<th></th>
<th>Base model</th>
<th>Base model + Education</th>
<th>Base model + Education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>p-value</td>
<td>HR</td>
</tr>
<tr>
<td>All-cause mortality</td>
<td>1.03</td>
<td>6.09 × 10^{-1}</td>
<td>1.00</td>
</tr>
<tr>
<td>Cancer mortality</td>
<td>0.99</td>
<td>8.91 × 10^{-1}</td>
<td>0.96</td>
</tr>
<tr>
<td>CVD mortality</td>
<td>1.12</td>
<td>4.43 × 10^{-1}</td>
<td>1.11</td>
</tr>
<tr>
<td>Cancer incidence</td>
<td>1.01</td>
<td>6.29 × 10^{-1}</td>
<td>1.01</td>
</tr>
<tr>
<td>CVD incidence</td>
<td>1.31</td>
<td>3.32 × 10^{-11}</td>
<td>1.30</td>
</tr>
</tbody>
</table>

⇒ Results are suggestive of a causal link between BHS and CVD incidence only that is independent of education
### BHS and cancer & CVD incidence

**Harrell's C-statistic**

<table>
<thead>
<tr>
<th>C−statistic</th>
<th>Education</th>
<th>BHS</th>
<th>Health Behaviours</th>
<th>Education + Health Behaviours</th>
<th>BHS + Health Behaviours</th>
<th>BHS + Education</th>
<th>BHS + Education + Health Behaviours</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age−scale</strong> (adjusted for age)</td>
<td>0.54 (0.53–0.54)</td>
<td>0.57 (0.56–0.57)</td>
<td>0.57 (0.57–0.58)</td>
<td>0.58 (0.57–0.58)</td>
<td>0.59 (0.58–0.59)</td>
<td>0.57 (0.57–0.58)</td>
<td>0.59 (0.58–0.60)</td>
</tr>
<tr>
<td><strong>Time−scale</strong> (including age)</td>
<td>0.67 (0.66–0.67)</td>
<td>0.65 (0.65–0.66)</td>
<td>0.66 (0.65–0.67)</td>
<td>0.67 (0.66–0.68)</td>
<td>0.67 (0.67–0.68)</td>
<td>0.68 (0.67–0.68)</td>
<td>0.67 (0.66–0.68)</td>
</tr>
</tbody>
</table>

- BHS is as predictive as Behaviours
- Complementarity of BHS, Education and Behaviours

⇒ Can we find biological support for this complementarity
BHS in UK Biobank: Conclusions

- Main results:
  - BHS is strongly associated to increased mortality (all-cause, cancer and CVD)
  - Effect attenuation mainly through behaviours
  - Strong effect of BHS and all sub-scores on CVD incidence surviving adjustment for socially-patterned exposures and behaviours
  - All scores except kidney are contributing to this association
  - MR is suggestive of a causal link between BHS and CVD incidence, independent of education
  - for both mortality and incidence analyses: limited role of education
  - Stronger effect of BHS than that of all subscores

⇒ BHS captures complementary physiological features that are disease relevant
⇒ such features are independent/complementary to education
BHS in UK Biobank: Conclusions

Strength & Limitations

- UK Biobank data: unique resource
- First analysis of Biological ageing in relation to mortality and morbidity
- Thorough investigation of the (lack of) role of education
- Representativity: UK Biobank suffers from Healthy Volunteer Bias
- Limited number of biomarkers (in particular for some systems)
- Outcome definitions are wide

⇒ investigate site-specific cancer and investigate different CVD outcomes
⇒ Explore common effects across outcomes
BHS in UK Biobank: Conclusions

- Interpretation:
  - An 10% increase in the BHS will increase the risk of incident CVD by 10 (over 4 years follow-up)
  - Of the (N=21,311) UK Biobank participants with 2 serial biomarker measurements, 25% (N=5,126) are exposed to that excess risk

⇒ Explore the mechanisms that are independent/complementary to education and disease-relevant
OMICS and composite scores: the NFBC cohort

NFBC is a birth cohort (1966) including N=12,000 mother-child pairs:

- Individual characteristics including lifestyle and social factors, and comorbidities
- Biomarker measurements (including HDL, LDL and total cholesterol, triglycerides)
- Other measurements (blood pressure, pulse, spirometry)

⇒ 5 physiological systems assayed

- **Metabolic** system: HDL cholesterol, total cholesterol, triglycerides
- **Cardiovascular** system: systolic and diastolic blood pressure, pulse
- **Inflammatory/immune** system: C-reactive protein, protein acetyllys
- **Kidney** function: creatinine
- **Liver** function: albumin
The BHS in the NFBC cohort

- As expected, lower score in healthy participants
- As observed in UK Biobank, higher BHS in CVD cases
Annotated NMR data in the NFBC cohort

After filtration 93 NMR Variables: 9 families and one pathway

- Amino acids (N=9)
- Apolipoproteins (N=2)
- Cholines (N=3)
- Fatty acids (N=8)
- **Glycolysis and gluconeogenesis** (N=5)
- Ketone bodies (N=3)
- Phosphoglycerides (N=1)
- Very low density lipoproteins (N=26)
- Low and intermediate density lipoproteins (N=18)
- High density lipoproteins (N=18)
Annotated NMR data in the NFBC cohort

⇒ Some correlations within groups (cholines, fatty acids, LDL, VLDL)
An sgPLS model for the BHS

Research Questions: can we identify metabolic markers of the BHS?

- can we use prior information on the metabolites to improve interpretability \(\rightarrow\) grouping
- can we select features within the groups \(\rightarrow\) sparsity through penalisation

\[\Rightarrow\] Some irrelevant features are discarded within (N=6) selected groups (VLDL, Apo AI, degree of unsaturation)
• 43 selected metabolites (from 7/10 groups)
• The group of cholesterol measurements appears redundant

⇒ is there a sub score differential?
OMICs and composite scores: the NFBC cohort

Question: can we identify sub-score specific patterns (sgPLS for subscores)

- Clear and functionally-relevant system specific relationships
  ⇒ current extensions: model scores as multivariate outcomes & multi-OMICs
Scores, risk factors and CVD prediction: UK Biobank

Background: Polygenic Genetic Scores incrementally improves CVD prediction (Elliott J et al., 2020; JAMA)

- The Pooled Cohort Equation (PCE) is an established score for risk of CVD (C-statistics: 0.76)
- Including a bespoke and re-calibrated PRS only increases the C-statistics by 0.02

Question: can other factors (including biochemistry) improve prediction over established scores? (Elliott J et al.; Submitted)

- Include biochemistry biomarkers in the prediction model
- Using stability selection and Random Forrest, identify and evaluate the relative importance of the selected predictors among:
  1. Variables included in the PCE or QRISK3 algorithms (N=21)
  2. Genetic information summarised by the PRS
  3. Biochemistry biomarkers (N=26)
  4. Haematology data (N=23)
Scores, risk factors and CVD prediction: UK Biobank

⇒ Identification of N=12 variables with selection proportion ≥ 0.8 in Men
⇒ Consistent results with the Random Forest
A series of models sequentially including variables (ranked importance) are fitted in the training set and C statistics in the test set.

⇒ Very limited increase in C statistic when including more than the calibrated number of variables
⇒ However: biochemistry adds to the prediction
Scores, risk factors and CVD prediction: UK Biobank

Evaluation of prediction performances: The C statistics in the test set

<table>
<thead>
<tr>
<th></th>
<th>Men (1,885/34,694)</th>
<th></th>
<th>Women (1,163/53,712)</th>
<th></th>
<th>Full population (3,048/88,406)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>C stat (95% CI)</td>
<td>#</td>
<td>C stat (95% CI)</td>
<td>C stat (95% CI)</td>
<td></td>
</tr>
<tr>
<td>PCE</td>
<td>0.732 [0.721-0.742]</td>
<td></td>
<td>0.684 [0.671-0.698]</td>
<td></td>
<td>0.713 [0.696-0.730]</td>
<td></td>
</tr>
<tr>
<td>Stability selection</td>
<td>12</td>
<td>0.726 [0.713-0.740]</td>
<td>11</td>
<td>0.745 [0.728-0.762]</td>
<td>0.762 [0.752-0.773]</td>
<td></td>
</tr>
<tr>
<td>Random Forest</td>
<td>16</td>
<td>0.727 [0.714-0.740]</td>
<td>13</td>
<td>0.739 [0.722-0.756]</td>
<td>0.760 [0.749-0.770]</td>
<td></td>
</tr>
</tbody>
</table>

⇒ Increase in C statistic of 0.05 for the full population when including other covariates from the selected groups
⇒ of this only up to 2% can be attributed to PRS
⇒ need to include other factors including environmental and other molecular data
⇒ identify other data sources with (multi-) OMICs data ⇒ need to focus on finer outcomes
Ongoing Questions:

Summary from the BHS analyses; Composite scores are

- capturing (biological, social) gradients in the population
- explanatory of incident conditions and mortality
- complementary to established factors

Ensuing research questions:

- Can we refine the scores such that they include Exposome features?
  ⇒ Clustering approaches
- Are and how much are (constituents of) scores complementing established risk factors?
  ⇒ variable selection and prediction models including Expotyes &/or their constituents
- Can we elucidate causal links between scores, their drivers and health outcomes?
  ⇒ Mediation & Causal models
Acknowledgments

Imperial College London

- B Bodinier
- D Vuckovic
- V Zuber

Imperial Team: P Elliott, P Vineis, A Berlanga, S Bowden, S Dagnino, A Pengelly, M Whittaker, M Hedges, J Elliott, D Petrovic, N De Toro Eadie, T Wright, K Asamoah

EU Financial Support: