Session 1: Background and Concept

Why we need to measure individual exposomes

Paul Elliott
Professor of Epidemiology and Public Health
Imperial College London

NAS Workshop on Emerging Technologies for Measuring Individual Exposomes
Washington, December 8-9, 2011
Outline

- Genes & Environment: Disease/risk factor trends
- Exposome concept
- Measurement issues
- Omics: GWAS — MWAS
- Mendelian randomisation
- EWAS
- Cohort resources
Age standardised death rates from coronary heart diseases in men, all ages from the European Union, eastern European countries (Bulgaria, Czech Republic, Hungary, Poland, Romania, and Slovakia), USA, and Japan, 1965-1997

Obesity trends worldwide

Blood Pressure Rise with Age

Mean systolic blood pressure (mmHg), US population, NHANES III Phase I (1988-1991) by age, ethnic group and gender

INTERSALT Study


© Imperial College London
Lifestyle factors – especially diet and weight gain – are key in explaining the rise in BP with age and the consequent prevalence of high BP at older ages.

Adapted from Rose G “Sick individuals and sick populations” *Int J Epidemiol* 1985; 14: 32-38

Poulter et al. *BMJ* 1990; 300: 967-72
“Genetics loads the gun, but Environment pulls the trigger”

After Elliott Proctor Joslin MD,
A pioneering American diabetologist and founder of the Joslin Diabetes Center, 1869-1962

The main determinants of health

“Individual and Population Exposomes”

Cell (2008)
134: 714-717.
Exposome Concept

Exposure → Biomarkers of exposure (incl. –omics) → Intermediate –omics biomarkers of early effects → Disease

Epidemiological study

“Lifecourse”

Environmental Sensors & Modelling

© Imperial College London
Northern Finland Birth Cohorts

- Followed from antenatal period to present day
- Self-administered questionnaires
- Hospital records, maternity cards, national registers
- Clinical exam, blood samples: chemistry, extracted DNA (GWAS), metabolomics
- Stored serum, plasma, cells

© Imperial College London
N Finnish 1966 Birth Cohort

SBP and birth weight

Geometric mean CRP by BMI tertile at 14 y and tertile of BMI change 14-31 y

Including twins (solid line), singletons (dotted line). Adjusted for sex, gestational age, family SES, parity, maternal height & weight before pregnancy, maternal age, maternal smoking after 2nd month of pregnancy, alcohol & BMI at age 31 years.

Cut-points for BMI tertiles were 14 y: 18.26/19.89 kg/m² (males), 18.29/20.13 kg/m² (females). BMI change 14-31 y: 4.54/6.92 kg/m² (males), 2.86/5.45 kg/m² (females).


© Imperial College London
While *genetic data* are a (fixed) digital read-out...

Environmental *exposure data* vary over time, are continuously distributed, with wide dynamic range...

And *difficult to measure*....

Such that ... the quality of the exposure data has been regarded as the *Achilles heel* of environmental epidemiology.
Non-systematic exposure misclassification:
The “Regression Dilution” problem

Reliability = \frac{\sigma^2_B}{\sigma^2_B + \sigma^2_W}

e.g.,
Serum cholesterol = 0.7
24-hr urinary sodium = 0.3
## Risk Factors for Cervix Cancer

**Questionnaire data...**

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>No. of case subjects</th>
<th>No. of control subjects</th>
<th>RR #1*</th>
<th>RR #2†</th>
<th>RR #3 95% CI‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifetime No. of sex partners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>113</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0 (0.5-2.1)</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>57</td>
<td>1.6</td>
<td>1.6</td>
<td>1.0 (0.6-2.1)</td>
</tr>
<tr>
<td>3-5</td>
<td>127</td>
<td>116</td>
<td>2.6‡</td>
<td>2.4§</td>
<td>1.2 (1.0-3.4)</td>
</tr>
<tr>
<td>6-9</td>
<td>117</td>
<td>70</td>
<td>4.2§</td>
<td>4.1§</td>
<td>1.8 (1.0-3.3)</td>
</tr>
<tr>
<td>10+</td>
<td>117</td>
<td>73</td>
<td>4.4§</td>
<td>4.4§</td>
<td></td>
</tr>
</tbody>
</table>

**Biomarker data...**

<table>
<thead>
<tr>
<th>HPV test result†</th>
<th>Case subjects</th>
<th>Control subjects</th>
<th>RR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>89</td>
<td>373</td>
<td>1.0</td>
<td>—</td>
</tr>
<tr>
<td>Types 6, 11, 42, or other or unknown type</td>
<td>108</td>
<td>52</td>
<td>8.7</td>
<td>5.8-13.0</td>
</tr>
<tr>
<td>Types 31, 33, 35, 39, 45, 51, or 52</td>
<td>117</td>
<td>15</td>
<td>33.0</td>
<td>18.0-59.0</td>
</tr>
<tr>
<td>Types 16 or 18</td>
<td>158</td>
<td>13</td>
<td>51.0</td>
<td>28.0-94.0</td>
</tr>
<tr>
<td>Total</td>
<td>472</td>
<td>453</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*CIN included a cytologic diagnosis of condylomatous atypia or CIN 1-3.
†HPV typing results grouped hierarchically according to risk when multiple types were present: 16 group > 31 group > 6 group.

Schiffman MH et al. *JNCI* 1993;85:958-64.
Plasma vitamin C, fruit and vegetable intake & CHD risk

678 cases in EPIC Norfolk among 11,134 free of CHD at baseline

Bingham S et al. *Int. J. Epidemiol* 2008; 37:978-987
Arsenic (As) in soil, house dust and urine in the South West of England

<table>
<thead>
<tr>
<th></th>
<th>Cargreen</th>
<th>Gunnislake</th>
<th>Devon GC</th>
<th>Ratio across sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil (µg g⁻¹)</td>
<td>37</td>
<td>365</td>
<td>4500</td>
<td>1: 10: 122</td>
</tr>
<tr>
<td>House dust (µg g⁻¹)</td>
<td>49</td>
<td>217</td>
<td>1167</td>
<td>1: 4: 24</td>
</tr>
<tr>
<td>Urine Total As (µg g⁻¹ creatinine)</td>
<td>4.7</td>
<td>9.2</td>
<td>10.0</td>
<td>1: 2: 2</td>
</tr>
</tbody>
</table>

GC = Great Consols

Some exposures are not (well) captured by biomarkers...

(Peters et al *JECH* 2011)

Modelling of individual RF exposures at residential address for nationwide study of childhood cancers and proximity to mobile phone masts

Elliott P et al. *BMJ* 2010; 340: c3077

3D distribution of power density around the focus of the beam simulated using Gaussian functions
# Air Pollution: Oxford Street Study

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>Hyde Park</th>
<th>Oxford Street</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrafines (pt/cc)</td>
<td>19459</td>
<td>66627</td>
</tr>
<tr>
<td>PM2.5 (mg/m³)</td>
<td>16.9</td>
<td>32.7</td>
</tr>
<tr>
<td>E. carbon (µg/m³)</td>
<td>1.7</td>
<td>8.1</td>
</tr>
<tr>
<td>NO2 (mg/m³)</td>
<td>36.1</td>
<td>143.0</td>
</tr>
<tr>
<td>CO (mg/m³)</td>
<td>0.97</td>
<td>1.26</td>
</tr>
</tbody>
</table>

**Supematant Myeloperoxidase** (Sputum)

- E. carbon (µg/m³)
- NO2 (mg/m³)

**FEV₁ all**

- % change from baseline
- Hours after start of exposure

---

McCreanor *et al.* *NEJM* 2007;357:2348-58

© Imperial College London
Omics

• New genomic, metabolomic, proteomic and other omic techniques can measure thousands of markers in individual subjects
  – Some (many) will be associated with disease by chance alone
  – Results must therefore be replicated (cross-validation and external validation)
  – Example of GWAS which sets stringent $P$-values (Bonferroni) with replication
Genome-Wide Association Studies (GWAS)

- **Untargeted** analysis
- Has led to discovery of hundreds of new replicated genetic associations in past 4 years
- Typically examine 2M+ common variants (measured and imputed) across the genome
  - Problem of multiple testing → Bonferroni correction (genome-wide $5 \times 10^{-8}$)
- Large sample size (consortia, meta-analyses)
- Replication
Coronary Heart Disease GWAS

WTCCC1

9p21.3

N (cases) ~ 2000

N (cases) ~ 22000

Conclusion After adjustment for traditional cardiovascular risk factors, a genetic risk score comprising 101 single nucleotide polymorphisms was not significantly associated with the incidence of total cardiovascular disease.

JAMA. 2010;303(7):631-637
BMI GWAS (GIANT) N (discovery) = 123,000

Metabolome-Wide Association Study (MWAS)

Human metabolic phenotype diversity and its association with diet and blood pressure

Elaine Holmes¹*, Ruey Leng Loo¹,²*, Jeremiah Stamler³, Magda Bictash¹,², Ivan K. S. Yap¹,², Queenie Chan¹, Tim Ebbels¹, Maria De Iorio¹, Ian J. Brown², Kirill A. Veselkov¹, Martha L. Davilus³, Hugo Kesteloot⁴, Hirotsgu Ueshima⁵, Liancheng Zhao⁶, Jeremy K. Nicholson¹ & Paul Elliott²

Shanxi  N China

Guangxi  S China

© Imperial College London
Geographical Metabolic Phenotype Mapping (INTERMAP)
China, Japan, USA, UK - 17 sub-populations, male/female, n = 4630 (24-h urine)

“Metabolome-Wide Association Studies” for novel hypothesis generation…
e.g…..A possible new role for formate in human BP regulation?

MWAS of animal/vegetable protein intake

Holmes et al Nature 2008; 493:396-400
GWAS-metabolomics

Suhre et al. Nature 2011; 477: 54-62
Mendelian Randomisation

- CRP levels but **NOT** CRP lowering variants (CRP gene) are associated with CHD
- Suggests therapeutic strategies to lower CHD risk by targeting specific reductions in CRP levels are unlikely to be fruitful


© Imperial College London
LPA genetic variants associated with Lp(a) lipoprotein levels and CHD risk

Clarke R et al. *NEJM* 2009; 361:2518-28
Environment-Wide Association Study (EWAS) of Type 2 Diabetes

Large blood-based prospective studies of disease associations

*Collected:*
EPIC (Europe) 420,000
Kadoorie (China) 500,000
UK Biobank 500,000

*Piloting:*
German National Cohort 200,000
Biobank Qatar 80,000

Studies in different populations allow investigation of a wide range of exposures and diseases
UK Biobank

- Data and samples on 500,000 men and women ages 40-69
- Prospective follow up over many years

### Baseline samples

<table>
<thead>
<tr>
<th>Vacutainer tube</th>
<th>Fractions</th>
<th>Number of aliquots</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-80°C</td>
</tr>
<tr>
<td>EDTA (9 ml) × 2</td>
<td>Plasma</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Buffy coat</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Red cells</td>
<td>1</td>
</tr>
<tr>
<td>LH (PST)</td>
<td>Plasma</td>
<td>3</td>
</tr>
<tr>
<td>Clot activator (SST)</td>
<td>Serum</td>
<td>3</td>
</tr>
<tr>
<td>ACD</td>
<td>DMSO blood</td>
<td>–</td>
</tr>
<tr>
<td>EDTA (4 ml)</td>
<td>Haematology (immediate)</td>
<td>–</td>
</tr>
<tr>
<td>Urine</td>
<td>Urine</td>
<td>4</td>
</tr>
<tr>
<td>Total Aliquots</td>
<td></td>
<td>19</td>
</tr>
</tbody>
</table>

PST, plasma separation tube; SST, serum separation tube; LH, lithium heparin.

Summary

• Bringing together large prospective cohorts with high-throughput omics technologies offers unprecedented opportunity for discovery on the causes and mechanisms of disease

• This needs to be linked with improved methods of exposure assessment to optimize understanding of exposure→disease relationships

© Imperial College London