Big Biobanks in the East and West

Zhengming Chen
Professor of Epidemiology
Nuffield Department of Population Health,
University of Oxford

Archie Cochrane Lecture, Green Templeton College,
Oxford, 14 March 2019
25 biggest countries: Trend in risk of death at ages 50-69, 1970-2010
CHINA, 2010: main causes of the 9M deaths per year at ALL AGES

0.4M communicable (half at ages 0-4)
0.8M injuries (accident/violence/suicide)
8M NCD (half before age 75)*

*2M cancer, 2M cardiac, 2M stroke, 1M COPD,
1M other NCD (eg renal, diabetes, cirrhosis)

Source: Global Burden of Disease for 2013; numbers are rounded to the nearest million
Big causes of adult mortality

Tobacco and alcohol
Chronic infection
Blood pressure
Blood lipids
Adiposity
Genetics
Diet
Air?
Large, unexplained cancer mortality variation

Oesophagus cancer

Nasopharynx cancer

Females only, so little effect of tobacco or alcohol
(Red = high mortality >10x green = low mortality)
Strengths of prospective biobank studies

- Survey is pre-disease onset; little reverse causality
- Appropriate controls, from the same population
- Stored DNA & plasma, for multi-omic studies
- Study many different disease outcomes

KEY NEEDS:
BIG studies in DIVERSE populations
SIZE matters: SBP vs IHD mortality, by age

5,000 adults

500,000 adults

Usual SBP (mmHg)

IHD mortality & 95% CI

Age

80-89
70-79
60-69
50-59
40-49

128
64
32
16
8
4
2
1

120 140 160 180

* Prospective Studies Collaboration (PSC), Lancet 2002
SIZE matters: Standing height in CKB

10,000 genotyped: only 2 hits passed threshold for reliable evidence (P<10^{-8})

100,000 genotyped: Many reliable associations
Need for big biobanks in EAST and WEST

- Wide range of exposures (eg, adiposity in China & UK)
- Genetic heterogeneity (eg, flushing gene limits alcohol)
- Very different disease rates (eg, stroke in north China)
- Contrasting associations (eg, BMI in China and UK)
Blood cholesterol in EAST and WEST

3.5 mmol/L  4.5 mmol/L  5.5 mmol/L

Rural China  Urban China  US / UK
Blood cholesterol vs IHD in EAST and WEST

USA (MRFIT): 5-7 mmol/L

China: 3.5-5 mmol/L *

Benefits of LDL-cholesterol lowering irrespective of the initial levels

Average LDL-cholesterol (mmol/L)

% with major vascular events

Statin-allocated
Placebo-allocated

Upper LDL third
Lower LDL third

2.0 2.5 3.0 3.5 4.0

Average LDL-cholesterol (mmol/L)
Declines in stroke mortality: not fully explained

Annual deaths at age 35-69 yrs per 1000
Stroke mortality in 65 counties across China: High rates, wide variation, NOT explained by BP

10% dead by age 70 →

UK
## Large biobank studies in EAST and WEST

<table>
<thead>
<tr>
<th>Cohort (&amp; location)</th>
<th>Participants (with samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico City</td>
<td>150,000</td>
</tr>
<tr>
<td>EPIC (Europe)</td>
<td>350,000</td>
</tr>
<tr>
<td>Million Veterans (US)</td>
<td>&gt;600,000</td>
</tr>
<tr>
<td>UK Biobank</td>
<td>500,000</td>
</tr>
<tr>
<td>Kadoorie Biobank (China)</td>
<td>512,000</td>
</tr>
</tbody>
</table>

Plus many new on-going biobanks in US (~1M), China (>1M) & elsewhere
China Kadoorie Biobank (CKB)

- 512K recruited from 10 localities in 2004-08
- Participants interviewed, measured, and had plasma and DNA taken for long-term storage
- All followed up indefinitely via electronic record linkage to deaths and ALL hospital episodes
- Periodic resurvey of 5% surviving participants (for enhancements and sources of variation)

www.ckbiobank.org
CKB: Ten study areas in China

- Urban
- Rural

Haikou
Harbin
Qingdao
Suzhou
Zhejiang

Gansu
Henan
Sichuan
Hunan
Liuzhou
Haikou
**CKB:** >512,000 people recruited during 2004-8

(Sample collection: ~100%)
CKB: Flow diagram of field work

Managed by ~100 bespoke IT systems
CKB: Centralised sample management system
**CKB: No. with selected diseases by 1.1.2017***

(>45K deaths, >1M hospital episodes, <1% lost)

<table>
<thead>
<tr>
<th>Condition</th>
<th>No. of Episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>56,641</td>
</tr>
<tr>
<td>Ischaemic Heart Disease</td>
<td>52,501</td>
</tr>
<tr>
<td>Diabetes</td>
<td>31,936</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>29,385</td>
</tr>
<tr>
<td>Cancer</td>
<td>28,250</td>
</tr>
<tr>
<td>Copd</td>
<td>19,688</td>
</tr>
<tr>
<td>Fracture</td>
<td>19,001</td>
</tr>
<tr>
<td>Cataract</td>
<td>14,203</td>
</tr>
<tr>
<td>Cardiac Arrhythmia</td>
<td>10,597</td>
</tr>
<tr>
<td>Coronary Revascularisation</td>
<td>9,097</td>
</tr>
<tr>
<td>Chronic Liver Disease</td>
<td>8,160</td>
</tr>
<tr>
<td>Kidney Disease</td>
<td>7,776</td>
</tr>
<tr>
<td>Pulmonary Heart Disease</td>
<td>5,444</td>
</tr>
<tr>
<td>Heart Failure</td>
<td>5,017</td>
</tr>
<tr>
<td>Asthma</td>
<td>5,188</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>3,068</td>
</tr>
<tr>
<td>Arthritis</td>
<td>2,839</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>2,701</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>1,294</td>
</tr>
<tr>
<td>Depression</td>
<td>1,189</td>
</tr>
<tr>
<td>Retinopathy</td>
<td>1,097</td>
</tr>
<tr>
<td>Parkinson's Disease</td>
<td>898</td>
</tr>
<tr>
<td>Dementia</td>
<td>538</td>
</tr>
<tr>
<td>Self-Harm</td>
<td>481</td>
</tr>
<tr>
<td>Snake Bite</td>
<td>417</td>
</tr>
<tr>
<td>Sepsis</td>
<td>313</td>
</tr>
<tr>
<td>Earthquake</td>
<td>54</td>
</tr>
</tbody>
</table>

*2017 & 2018 data are being processed

>80K episodes of stroke, IHD, renal disease, cancer already adjudicated
**CKB: Selected publications on CVD, 2015-19**

(>200 papers submitted/published, mainly after 2015)

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Key publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td><em>Lancet</em> 2015, <em>Lancet Public Health</em> 2018</td>
</tr>
<tr>
<td>LDL-cholesterol</td>
<td><em>Nature Med</em> 2019</td>
</tr>
<tr>
<td>HDL-cholesterol</td>
<td><em>JAMA Cardiol</em> 2018</td>
</tr>
<tr>
<td>Alcohol</td>
<td><em>Lancet</em> 2019</td>
</tr>
<tr>
<td>Physical activity</td>
<td><em>BMC Public Health</em> 2015, <em>JAMA Cardiol</em> 2017</td>
</tr>
<tr>
<td>Household air pollution</td>
<td><em>Eur Heart J</em> 2015, <em>JAMA</em> 2018</td>
</tr>
<tr>
<td>Adiposity</td>
<td><em>IJE</em> 2015, <em>Lancet Global Health</em> 2018</td>
</tr>
<tr>
<td>Metabolomics</td>
<td><em>JACC</em> 2018</td>
</tr>
</tbody>
</table>
Simple data can reveal history & effects of policy

No. of participants

Great famine

Birth year

Mean age at 1st birth (years)

One child policy

Urban

Rural

Int J Epidemiol 2011 & 2013
CKB: BMI, SBP & risk of stroke types

BMI vs. SBP

SBP vs. ischaemic stroke

SBP vs. haemorrhagic stroke

Chen Z, et al. Lancet Global Health 2018
**CKB: BMI & risk of stroke types**
(Black: no adjusting for SBP, white: given SBP)

**Ischaemic stroke**

**Haemorrhagic stroke**

Chen Z, et al. Lancet Global Health 2018
**CKB: LDL-cholesterol & risk of stroke types**

**Ischaemic stroke**
(N=5475)

RR per 1 mmol/L: 1.17 (1.10-1.25)

**Haemorrhagic stroke**
(N=4776)

RR per 1 mmol/L: 0.86 (0.80-0.92)

LDL-C vs. stroke risk in CKB and statin trials

(RR per 1 mmol/L reduction in LDL-C)

(I) Ischaemic stroke

<table>
<thead>
<tr>
<th>Number of Case / Control</th>
<th>RR (95% CI)</th>
<th>P for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational study (CKB)</td>
<td>5475 / 6290</td>
<td>0.85 (0.80 - 0.91)</td>
</tr>
<tr>
<td>Genetic study (CKB)</td>
<td>5567 / 9742</td>
<td>0.75 (0.60 - 0.95)</td>
</tr>
<tr>
<td>Randomised trials</td>
<td>2431 / 3045</td>
<td>0.80 (0.76 - 0.84)</td>
</tr>
</tbody>
</table>

(II) Intracerebral haemorrhage

<table>
<thead>
<tr>
<th>Number of Case / Control</th>
<th>RR (95% CI)</th>
<th>P for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational study (CKB)</td>
<td>4776 / 6290</td>
<td>1.16 (1.08 - 1.25)</td>
</tr>
<tr>
<td>Genetic study (CKB)</td>
<td>4911 / 9742</td>
<td>1.13 (0.91 - 1.40)</td>
</tr>
<tr>
<td>Randomised trials</td>
<td>494 / 404</td>
<td>1.17 (1.03 - 1.32)</td>
</tr>
</tbody>
</table>
NET effects of 1 mmol/L lower LDL-C in China

a) Low risk

b) High risk

Stroke patients, after hospital discharge in CKB:
Risk of a further stroke or heart attack is 20% in year 1, then 10% / year
Relevance of HDL cholesterol, given LDL-C

Ischaemic heart disease in Western population (PSC)

Ischaemic stroke in China (CKB)

Rate ratio (95% CI)

Usual HDL cholesterol (mmol/L)

Baseline LDL-C:
- ≥ 2.39 mmol/L
- < 2.39 mmol/L
CETP removes cholesterol from HDL particles, reducing HDL-Cholesterol levels in the blood. Would blocking it be good?
Mendelian randomisation (MR) studies to predict potential drug effects

Genetic studies

Randomise at conception: lifelong CETP defect vs not

<table>
<thead>
<tr>
<th>Defect*</th>
<th>No defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ HDL-C</td>
<td>10 mg/L</td>
</tr>
<tr>
<td>Disease outcome</td>
<td>Disease outcome</td>
</tr>
</tbody>
</table>

Randomised trials

Randomise drug treatment: CETP inhibitor vs placebo

<table>
<thead>
<tr>
<th>Anacetrapib</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ HDL-C</td>
<td>10 mg/L</td>
</tr>
<tr>
<td>Disease outcome</td>
<td>Disease outcome</td>
</tr>
</tbody>
</table>

*East Asian variant rs2303790 (A>G) reduces CETP activity
**CKB**: Relation of *CETP* genetic defect (rs2303790) with vascular disease, and with other diseases

Vascular disease - little or no effect  
(CETP defect vs not: RR 1.0, CI 0.9-1.1)

Eye disease - possible side effect  
(CETP defect vs not: RR 1.45, CI 1.13-1.80)

CKB findings are consistent with the trial results,  
(but did not cost $1 billion)
Genetic research can improve drug development

- Predict main effects and side effects of inhibitors
- Identify and validate inhibitor targets
- Re-purpose existing drugs
Cigarette consumption & lung cancer in US

- Cigarettes/day, (males + females)
- Male lung cancer, 35–69
- Female lung cancer, 35–69
- Annual mortality per 100 000 (mean of 7 age specific rates, ages 35–69)
Annual Chinese cigarette production, 1952-2011

First national cohort study

CKB

Tobacco Control 2014; 23: 167
CHINA, two studies: RRs for all-cause mortality, by age began smoking (black: urban, white: rural)

First study (1991-99)
- RR=2.5 for urban men who started before age 15 yrs

Second study (2006-15)

RR=2.5 for urban men who started before age 15 yrs
CKB: Male smoking patterns by year of birth

c) Age started smoking regularly

d) Smoked cigarettes when first started
**CKB: Smoking prevalence among women**

Similar mean age of starting

![Graph showing smoking prevalence among women](image)
Contrasting male and female trends in tobacco-attributed mortality in China: evidence from successive nationwide prospective cohort studies

Zhengming Chen, Richard Peto, Maigeng Zhou, Andri Iona, Margaret Smith, Ling Yang, Yu Guo, Yiping Chen, Zheng Bian, Garry Lancaster, Paul Sherliker, Shutao Pang, Hao Wang, Hua Su, Ming Wu, Xianping Wu, Junshi Chen, Rory Collins, Liming Li, for the China Kadoorie Biobank (CKB) collaborative group*

“As the population grows and the proportion of male deaths due to smoking increases, annual tobacco deaths will rise from about

1 million in 2010 to
2 million in 2030 and
3 million in 2050, unless there is widespread cessation.”

Lancet 2015; 386: 1447
Going to the pub is officially good for you, according to Oxford University researchers

Conversation is the key, according to a new study
Alcohol metabolism

Alcohol $\rightarrow$ Acetaldehyde $\rightarrow$ Acetate

Accumulation causes nausea and flushing

ADH accelerated by genetic variant (rs1229984 G>A) (70% East vs 1-2% West)

ALDH disabled by genetic variant (rs671 G>A) (20% East vs 0% West)
CKB: GWAS of alcohol intake in 42K men
Mean alcohol intake in men, according to 9 possible ALDH2-rs671 and ADH1B-rs 1229984 genotypes.
Mean alcohol intake, predicted by genotypes

Men (n=60,982)

- C1: 4
- C2: 18
- C3: 34
- C4: 78
- C5: 130
- C6: 256

Women (n=90,046)

- C1: <2
- C2: <2
- C3: <2
- C4: 4
- C5: 5
- C6: 8

Units: g/week
CKB: Alcohol drinking and stroke risk, in men

Total stroke
9189 events

1.38 (1.26, 1.51)
per 280 g/week
p<0.0001

RR (95% CI)
0.75
1.00
1.25
1.50
1.75
2.00

Mean alcohol (g/week) predicted from genotype and area

0 100 200 300 400 500
**CKB: Alcohol drinking and stroke risk, in men**

**Genetic Epidemiology**

- **a) Ischaemic stroke**
  - 5700 events
  - Relative Risk (RR) with 95% Confidence Interval (CI)
  - 1.27 (1.13, 1.43)
  - per 280 g/week
  - p=0.0001

- **b) Intracerebral haemorrhage**
  - 2760 events
  - Relative Risk (RR) with 95% Confidence Interval (CI)
  - 1.56 (1.30, 1.84)
  - per 280 g/week
  - p=0.0001

**Observational Epidemiology**

- **a) Ischaemic stroke**
  - 14,930 events
  - Relative Risk (RR) with 95% Confidence Interval (CI)
  - 1.26 (1.19, 1.39)
  - per 280 g/week
  - p=0.0001

- **b) Intracerebral haemorrhage**
  - 3466 events
  - Relative Risk (RR) with 95% Confidence Interval (CI)
  - 1.59 (1.37, 1.85)
  - per 280 g/week
  - p=0.0001
**CKB: Current and future research**
(For hypothesis testing and generating)

- Prospective analyses of lifestyle factors and disease risks
- Case-cohort studies of biomarkers and risks of major diseases
- GWAS of diseases, quantitative traits and lifestyle factors
- PheWAS of LOF variants for drug target discovery & validation
- MR studies to assess causal role of non-genetic factors
- Health care delivery and utilisation in urban and rural China
- Health effects of environmental exposures (eg, air pollution)

**KEY NEEDS: Turn samples into data**
(eg, genotyping, proteomics, metabolomics in all 0.5M)
China Kadoorie Biobank (CKB, 500,000) and UK Biobank (UKB, also 500,000)

- **LARGE** numbers, allowing **RELIABLE** assessment of risks
- **FULL** effects of each exposure (e.g., smoking) are assessed on **ALL** health outcomes (mortality, disability, dementia)
- **NOVEL** discoveries from detailed measurements at entry and, particularly, on blood, with careful disease classification
Workers of the world unite