Socioeconomic inequalities in health expectancy with and without multiple morbidities: methodological challenges

Madhavi Bajekal, PhD, HonFIA
Mei Sum Chan
University College London
Overview

• Background and aims
• CPRD database
• Methodological challenges
  1. How representative is the CPRD of the English population?
  2. Which chronic diseases to include – and how many in all?
  3. Which modelling method/s to use?
    a) Cox survival model specification and provisional results
    b) Multi-state model (MSM) specification and provisional results
• Next challenges
Multimorbidity and socioeconomic disadvantage at older ages

• **What is Multimorbidity (MM)?**
  
  – “the co-occurrence of two or more chronic conditions within one person without specifying an index condition”
  
  – A chronic condition/disease is a “health problem that requires management over a period of years or decades” (WHO)

• **We know** that the level of deprivation affects:
  
  – The age of onset of MM, and the number of conditions
  
  – Disease combinations – physical and mental health more common in deprived than in affluent at ages <55

• **What we don’t know:**
  
  – Do older poor become morbid earlier in the life course and hence die younger? Or do they acquire more lethal diseases?
  
  – For similar disease combinations, is disease progression and survival different in deprived and advantaged groups?
Project Aims

- To quantify socioeconomic inequalities in survival; and in life expectancy spent ‘healthy’ (or disease-free) and with multiple morbidity (1, 2, 3+ diseases) - arising from differentials in age of onset, and rates of transition to multimorbidity and death

- To understand reasons for inequalities in disease progression and mortality rates: eg
  - do rich and poor get the same diseases, but either just get them later in life, or have better survival with multimorbidity; or
  - is it the differences in the type and combinations of diseases between social groups which explains most of the 4 year gap in LE @65?
Clinical Practice Research Datalink (CPRD) data extract
‘Big Data’: linked Electronic Health Records - CPRD

Notes: ECG = Electrocardiography, STEMI = ST-segment elevation Myocardial Infarction, ACEI = Angiotensin-converting-enzyme Inhibitor.

Cohort specification

Inclusion criteria

- **Open cohort design**, with patients becoming cohort members on the earliest date that *all three* of the following criteria are met:
  
  1. Registered in linked practices and with a valid postcode of residence *(to link to IMD)*
  2. UTS (up to (quality) standards) practice for at least 1 year
  3. All patients aged 45+ on 1 Jan 2001 and patients in participating practices who turn 45 between 1 Jan 2001 and 25 Mar 2010, irrespective of initial health status.

- **Follow up period** – *(from Jan 2001) to Mar 2010*

- Patients’ follow-up censored at the earliest date of death, deregistration from the practice, last data collection for the patient’s practice, or the overall study end date.

1.3 million patients with 12 million consultations relating to 30 chronic diseases
Challenge 1: How representative is the CPRD?
## CPRD – overall numbers (2007)

<table>
<thead>
<tr>
<th></th>
<th>GPRD (all practices) 2007</th>
<th>CPRD (linked practices) 2007</th>
<th>England 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population count (all ages)</td>
<td>3,187,261 (6.2%)</td>
<td>1,621,118 (3.2%)</td>
<td>51,261,945</td>
</tr>
<tr>
<td>Number of practices</td>
<td>c 500+</td>
<td>225</td>
<td>7,900</td>
</tr>
<tr>
<td>Percentage Female</td>
<td>50.5%</td>
<td>50.6%</td>
<td>50.8%</td>
</tr>
<tr>
<td>Median age group</td>
<td>40-44</td>
<td>35-39</td>
<td>35-39</td>
</tr>
<tr>
<td>Crude death rate (per 1000)</td>
<td>9.1</td>
<td>9.2</td>
<td>9.2</td>
</tr>
</tbody>
</table>
Population sampling fraction in CPRD-linked practices by age and deprivation quintile
Age-standardised mortality rates by deprivation quintile, 2007: England vs CPRD

### Males (age 25+)

<table>
<thead>
<tr>
<th>Deprivation Quintile</th>
<th>Rate per 100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Depr</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td></td>
</tr>
<tr>
<td>Most Depr</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
</tr>
</tbody>
</table>

### Females (age 25+)

<table>
<thead>
<tr>
<th>Deprivation Quintile</th>
<th>Rate per 100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Depr</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td></td>
</tr>
<tr>
<td>Most Depr</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
</tr>
</tbody>
</table>
Comparing CPRD Sample vs England population distribution by Region (2001-09)

Males

<table>
<thead>
<tr>
<th>Region</th>
<th>East of England</th>
<th>East Mids</th>
<th>London</th>
<th>N East</th>
<th>N West</th>
<th>S East</th>
<th>S West</th>
<th>West Mids</th>
<th>Yorks &amp; Humber</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRPD %</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Region %</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Females

<table>
<thead>
<tr>
<th>Region</th>
<th>East of England</th>
<th>East Mids</th>
<th>London</th>
<th>N East</th>
<th>N West</th>
<th>S East</th>
<th>S West</th>
<th>West Mids</th>
<th>Yorks &amp; Humber</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRPD %</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Region %</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>
Representativeness of the CPRD data – socioeconomic composition and mortality

- Deprived quintile (Q5) are significantly underrepresented in CPRD for both sexes
- Mortality rates (ASRs) were significantly lower than for England – overall, and for all quintile groups except Q2 males; for women quintile differences were smaller and not significant at 95% level.
- Adjustment for Regional imbalance in CPRD sample reduces, but doesn’t close, the life expectancy gap with England.
- Annual LE@65 trends using 2001 – 2009 data show similar pace of change in CPRD as in England (for all quintiles and both sexes)
- Hence, CPRD sample is both relatively ‘healthier’ and more affluent than the English population.
Validating single disease prevalence

- The prevalence by age band and sex for atrial fibrillation have been displayed here as an example.
- Age-sex prevalence rates from independent sources such as QOF, for most of the diseases in-scope, also matched up.

Age UK, Melzer et al, 2015
Challenge 2 - Which diseases to include?
Challenge – which diseases?

• No clear break point, no consensus

• Conflicting views on which long-term conditions to include
  – (e.g. pain; or syndromes e.g. sensory deficits)

• Always easier to justify adding a disease
  – Low prevalence but serious disease like MND

• Each disease requires 10-20 hours to produce a final, approved code list (Read code + ICD10 code)
Hypertension*  
Obesity*  
Diabetes  
COPD  
Asthma  
High cholesterol*  
Cancer (malignant)  
CHD (angina, heart attack)  
Depression  
Osteoarthritis  
Stroke  
Thyroid disorder  
Renal failure (CKD)  
Anxiety  
Osteoporosis  
Dementia  
Rheumatic Arthritis  
Heart Failure  
Chronic back pain  
Other arthritis

Listing long-term conditions based on prevalence/frequency

- Selecting TOP 5 – misses out Cancer
- Selecting Top 10 – misses out a MM patient with hypertension, kidney failure, heart failure, and osteoporosis
- Selecting Top 20 – misses out patients with diseases like Parkinson’s or liver disease

- Risk factors* or chronic diseases?
* hypertension, hypercholesterolemia

<table>
<thead>
<tr>
<th>Green</th>
<th>Blue</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma or COPD or bronchiectasis</td>
<td>Dementia or Alzheimer's</td>
<td>Osteoporosis</td>
</tr>
<tr>
<td>Atrial Fibrillation</td>
<td>Diverticulitis of intestine</td>
<td>Osteoarthritis (active Rx) or Chronic severe back pain</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Epilepsy</td>
<td>Peripheral Arterial Disease</td>
</tr>
<tr>
<td>Alcohol problems</td>
<td>Glaucoma</td>
<td>Parkinson's disease</td>
</tr>
<tr>
<td>Coronary Heart Disease</td>
<td>Hypothyroidism</td>
<td>Psychoactive substance misuse</td>
</tr>
<tr>
<td>Chronic Kidney Disease (CKD 4,5)</td>
<td>Heart Failure</td>
<td>Prostate disorders</td>
</tr>
<tr>
<td>Cancer (in last 5 years)</td>
<td>Inflammatory Bowel Disease</td>
<td>Psoriasis</td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>Learning Disabilities</td>
<td>Rheumatoid Arthritis</td>
</tr>
<tr>
<td>Depression (ever + SSRI last year)</td>
<td>Multiple Sclerosis</td>
<td>Stroke or TIA</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Motor Neurone Disease</td>
<td>Severe Mental Illness</td>
</tr>
</tbody>
</table>
Age-related increase in morbidity prevalence is steeper in Scotland than England.

This may be because Scottish study included more diseases (40 vs our 30); and/or country differences in disease onset/progression patterns.

Even in the 85+ age band, >10% of patients in England have none of 30 major chronic diseases; and just 70% would be considered MM (compared to 80% in Scotland).
Prevalence of multimorbidity (2+ diseases) by deprivation quintiles

- Clear social gradient in the prevalence of MM, with differentials narrowing with advancing age
- Age-related increase in MM steeper for men than for women
Challenge 3: Modelling methods
## Two survival models - based on disease-counts

<table>
<thead>
<tr>
<th><strong>Cox Proportional Hazards Model</strong></th>
<th><strong>Multistate Model</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Health state at entry</td>
<td>• Uses transitions between several health states to capture the temporal and stochastic processes underlying disease progression</td>
</tr>
<tr>
<td>• Time to death, with age as time-scale</td>
<td>• Except age, covariates fixed at entry</td>
</tr>
<tr>
<td>• ‘Ignorant’ of all health transitions between state-entry and death</td>
<td>• Prognostic value</td>
</tr>
<tr>
<td>• All covariates fixed at entry</td>
<td>• Complex model</td>
</tr>
<tr>
<td>• Predictive value</td>
<td></td>
</tr>
<tr>
<td>• Simple survival model, easy to understand</td>
<td></td>
</tr>
</tbody>
</table>
Cox survival model – model specification
Cox Model Specification

• Separate models by: **Sex**

• Model covariates: *all fixed at baseline*
  
  – **Age at study entry** – continuous (45+)
  
  – **Health state at baseline** – 4 categories, disease count based (healthy, 1 disease, 2 diseases, 3+ diseases)
  
  – **IMD quintiles** – 5 categories
  
  – **Smoking status** – 3 categories (never, ex, current smoker)
  
  – All interactions between age, health state and IMD have been incorporated
Modelling steps

1. Check proportionality assumption
2. Run Cox proportional hazards model
   - Hazard rates; calculate baseline hazard for reference group
3. Calculate and plot survival curves for sub-groups
4. Calculate area under the survival curve = estimated life expectancy
5. Plot LEs and read-off LE gap (or life years ‘lost’ relative to a ref group) for any age 45 and over.

• Reference group for Cox model:
  - Males
  - Q1 (least deprived)
  - Healthy at baseline
  - Non-smoker
Proportionality check - Kaplan-Meier curves

By IMD quintiles

By disease state

- No crossover in survival curves over time, by disease states, IMD quintiles and smoker status (not shown)
- Other proportionality checks done: log(-log) and Schoenfeld residuals
Cox survival model – provisional results (adjusted for smoking)
Within IMD, HRs by disease state - Males, Q3

Hazard ratios (HR)

- Social gradient in HRs, with partial convergence on the log scale at older ages
- (Females similar pattern)
Within IMD, Survival curves, by disease state - Males, Q3

- Consequently, survival probabilities lowest for males with 3+ diseases
Survival curves magnified – Males, Q3
Within IMD, LE and LE Gap, by disease state-Males, Q3

- Consistent LE pattern for all IMD groups: healthy>1d>2d>3+d
- Gaps are large: those with 3+d at age 45 can expect to live 8.5y years fewer for Q1; for Q5 the gap is even larger
Between IMD, 2 diseases at entry – Males, Q1 to Q5

Hazard ratios (HR)

- Social gradient in HRs until about age 90, convergence thereafter
- (Females similar pattern, but do not converge completely by age 100)
Between IMD, LE and LE lost, 2 diseases at entry - Males

Life expectancy

Life expectancy lost

- For men with 2 diseases at entry, differences in LE@45 between deprivation quintiles, were relatively small
- Ranging from 2.5 fewer years between Q5 and Q1 to 0.2 years fewer between Q2 and Q1 (for an expected LE@45 of about 35 years for those with 2d)
Key insights – Cox model

In a Cox model including 5 IMD quintiles, 4 disease states and 3 smoker categories, the gap in LE@45 (c38 years):
• was largest between disease states (c9 years; ‘healthy’ vs ‘3+d’)
• then between smoking status (c6 years; never vs current smokers)
• and smallest between deprivation (c3 years; Q1 vs Q5)

At age 65, equivalent differences were c7y, c5y and c2y, respectively.
Multi-state Model (MSM) – Model specification
Model specification

- Progressive model, i.e. no recovery; state 5 is absorbing (death)
- Intermediate state definition: **Diagnosis** of 1, 2, 3+ in-scope diseases
- Input data same as the disease count-based survival model
- Data split by: **Sex**, **IMD quintiles** and **Smoker status** (2 x 5 x 3 = 30 separate models)
- Smoker status is recorded at baseline – fixed, 3 categories (never, ex and current smokers)
- Model covariate: **Age** – time-dependent, continuous (65+)
‘Movers’ and ‘stayers’ in the dataset

- Significant proportions of the 1.3m patients have moved between each state, during the study period of up to 9¼ years.
- Approx. 60% are ‘stayers’, so MSM transition rates are largely informed by the remaining 40%.
- Patients in more deprived quintiles have more diseases at entry, and experience a larger degree of movement across states.
Model comparison and software check

Exponential: \( h_{rs}(t) = \exp(\beta_{rs}) \)

Gompertz: \( h_{rs}(t) = \exp(\beta_{rs} + \gamma_{rs}t) \)

Step: \( h_{rs}(t) = \exp(\beta_{rs} + \gamma_{rs,k}I(t \epsilon \Delta_k)) \)

Log-time: \( h_{rs}(t) = \exp(\beta_{rs} + \gamma_{rs} \log(t)) \)

<table>
<thead>
<tr>
<th></th>
<th>Exponential</th>
<th>Gompertz</th>
<th>Step</th>
<th>Log-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>6405</td>
<td>6118</td>
<td>6172</td>
<td>6117</td>
</tr>
</tbody>
</table>

- Models were fitted using the R package *msm* (Jackson, 2011)
- The fitting was checked by user-written R software: the likelihood function in *msm* is in accordance with our specification of known transition times
Multi-state Model (MSM) – Provisional results
Analysis of transition rates (model parameters) by age – Males Q3

- Mortality risk increases with age: and is lowest for those with no chronic disease and becomes progressively higher with each additional disease acquired
- Rate of transition from healthy to 1d (disease) are lowest for all groups; transitions from 1d to 2d and 2d to 3+d are higher (in this example very similar)
- Similar patterns are seen across IMD quintiles for males and females
Basic model – Males LE@65 by IMD

- Modelled LEs are systematically lower than period life table LE for CPRD by about 1.5y
- Social gradient in total LE is as expected – 3.1y gap between least-most deprived quintile
- When separated into years spent with MM (2+ diseases), most deprived males have the shortest ‘healthy’ life years; and, after becoming multimorbid, die earlier.
Males LE@65 – Time spent with/without MM by IMD and smoker status

- Years spent without MM: largest for never-smokers; similar for ex- and current smokers
- Years spent with MM: similar for never- and ex-smokers; and fewest for current smokers
Males, Q3 – Total LE by age and smoker status

- LE for never smokers > ex smokers > current smokers, and converges with age
- At age 65, the LE gap between never-smokers and those who had quit is 1.5 years and those who smoke is 4.3 years * (all in Q3)
- For both sexes and all SEC quintiles, the pattern across ages is broadly similar as above

*The difference in LE between never, ex and current smokers at age 50 are very close to estimates from longitudinal follow-up of the Whitehall study. Robert Clarke et al, 2009
Males Q3 – Time spent with/without MM by age and smoker status

Time spent *without* multimorbidity

- Never smokers spend most years of life without multimorbidity (or ‘healthy’) than either ex or current smokers
- Never and ex smokers spend equal number of years with multimorbidity, and more than do current smokers
- Hence, current smokers have lowest LE because they spend least time healthy and die quicker once they become multimorbid
- This pattern is similar for both sexes and across SEC quintiles

Time spent *with* multimorbidity:

- Never smokers
- Ex smokers
- Current smokers
Key insights – MSM model

• The ‘gap’ in LE between most and least deprived has 2 components:
  • The onset of MM is at an earlier age for the most deprived
  • Thereafter, progression to death is quicker for most than for least deprived

• Do IMD differences in smoker prevalence explain this?
  – No: For the same smoking status, people in deprived areas live shorter lives than those living in affluent areas – e.g. even for non-smokers, LE of most deprived was the lowest.
  – Yes: But the age of onset of MM is delayed for non-smokers; whereas it is earlier, and at similar ages, for ex and current smokers
  – Yes: Once MM sets in, never and ex smokers live the same number of years before death; current smokers with MM die sooner
Next Challenge - Disease clustering and analysis

A: Disease combinations analysis (dyads, triads) to:
   A1: identify common combinations, by broad age group and sex
   A2: identify lethal combinations of comorbidities (for index disease) using time to event Cox regression model

B: Cluster analysis
   B1: based on known clinical cluster typologies (eg cardiometabolic)
   B2: based on model-based clustering

C: Inequalities: relative mortality/survival ratios between and within disease clusters (by SEC)
Project essentials

Research Team:
1. **Dr Madhavi Bajekal (Principal Investigator)**, Honorary Senior Fellow, Dept of Applied Health Research (DAHR), UCL
2. **Mei Sum Chan**, Honorary Research Assistant, DAHR UCL
3. **Dr Mar Pujades**, University Academic Fellow, MRC Medical Bioinformatics Centre Leeds
4. **Dr Melvyn Jones**, Senior Lecturer, Dept of Primary Care UCL
5. **Dr Ardo van den Hout**, Lecturer, Dept of Statistical Science UCL
6. **Dr Mizan Khondoker**, Senior Lecturer, DAHR UCL
7. **Dr Spiros Denaxas**; Senior Lecturer, Farr UCL

Advisory Board: Prof Carol Jagger (Chair, Newcastle), Dr David Buck (King’s Fund), Prof José Iparraguirre (Age UK), Prof Fiona Matthews (Newcastle), Brian Ridsdale (IAA Mortality Working Group), Prof Chris Salisbury (Bristol)

**Funding**: CLAHRC matched funding by Legal & General Assurance Society. Lead sponsor – Joseph Lu

**Timeline**: Jan 2015 to Dec 2017 (3y)
Thank you for your attention

• Multimorbidity Project Contacts:
  – m.bajekal@ucl.ac.uk
  – mei.chan@ucl.ac.uk

• CLAHRC – North Thames contact sites
  – Website: http://www.clahrc-norththames.nihr.ac.uk/health-inequalities-multiple-morbidities/
  – Email: clahrc.norththames@ucl.ac.uk
  – Twitter: @CLAHRC_N_Thames
Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.
Project Funding & Disclaimer

• This presentation demonstrates independent research and was funded (in part) by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care (NIHR CLAHRC) North Thames at Bart’s Health NHS Trust.

• The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.
What is an NIHR CLAHRC?

NIHR Collaborations for Leadership in Applied Health Research and Care (CLAHRCs) are partnerships between universities, health and other organisations which have come together to conduct ground breaking applied health research that will have a direct impact on the health of patients with long term conditions and on the health of the public.
Stay in touch with NIHR CLAHRC North Thames

Website
www.clahrc-norththames.nihr.ac.uk

Email:
clahrc.norththames@ucl.ac.uk

Twitter:
@CLAHRC_N_Thames
References

• Salisbury C. Epidemiology & impact of multimorbidity in primary care. BJGP 2011
• Cooke G. Common general practice presentations and publication frequency. 42, 1, Jan 2013
• Jackson C. Multi-state models for panel data: The msm Package for R. Journal of Statistical Software 38, 2011
• van den Hout, A. ELECT: Estimation of life expectancies using continuous-time multi-state survival models. ELECT version 0.2. Vignette. 2016