The reliability and validity of global questions on disability: results from a cross-survey comparison of disability estimates in Great Britain

Madhavi Bajekal

REVES, May 2006
Aims of the Research

Two broad aims:
• Existing surveys: why did estimates differ (prevalence and absolute numbers)
• Consultation (disability organisations, experts): sources used, gaps and improving/tailoring dissemination
Three types of estimates based on global questions evaluated

- Limiting long-standing illness or disability
  (all ages 16+)

- Work-limiting disability
  (to state pensionable age or SPA)

- DDA (Disability Discrimination Act) defined disabled
  (all ages 16+)
Limiting long-standing illness (LLSI) : harmonised version

• Do you have any long-standing illness, disability or infirmity? By long-standing I mean anything that has troubled you over a period of time or is likely to affect you over a period of time?

   If yes,

   – Does this illness or disability (do any of these illnesses or disabilities) limit your activities in any way? Yes/No
Census 1991 and 2001: limiting long-term illness (LLTI)

Census 1991

• Do you have any long-term illness, health problem or *handicap* which limits your daily activities or the work you can do?

Include problems which are due to old age

Yes/No

Census 2001

• Do you have any long-term illness, health problem or *disability* which limits your daily activities or the work you can do?

• Include problems which are due to old age.

Yes/No
Work Limiting Disability

• **LFS** (1 filter + 2 questions)
  – long-term health problem or disability that you expect will last more than a year (*if yes*)
  – restricted in the kind and/or amount of paid work you **might do**

• **FRS** (1 question, 3 responses)
  – restricted in amount and/or type of work **can do** because of an injury, illness or disability (unable, restricted, not restricted) (FRS-1)
  – As above + long-term health problem filter (derived FRS-2)
DDA Definition (LFS version)

- Do you have any health problems or disabilities that you would expect to last for more than a year? (If yes)
- Do these health problems or disabilities, when taken singly or together, substantially limit your ability to carry out normal day-to-day activities?

If you are receiving treatment, please consider what the situation would be without medication or treatment

- plus progressive conditions (e.g. cancer)
- plus people with a DDA disability in the past
Methods: Reliability and Validity

• Reliability: consistent results from same subjects over time (test-retest), by different interviewers (inter-rater), in different contexts (portability)

• Validity:
  – Criterion: no ‘gold’ standard measure
  – Construct: association between indicator and conceptually related measure
  – Content: (or face) validity

• Secondary analysis of existing data places severe limits on reliability/validity testing
Empirical Comparisons – Q1 (a)

- Do estimates based on the **same question**, covering the **same population** and for the **same time period** differ between surveys? (**reliability/portability**)
  - overall and age/sex specific rates
  - survey context, mode, question order, proxy responses

- **Analysis approach:**
  - data-sets for 2001, or closest year
  - cross-sectional (e.g. LFS first wave only), adults 16+
  - except for selection probability weights, no grossing, non-response weights
  - age-standardised rate (especially when comparing with earlier/later estimates)
LLSI 2001, by age (FRS, GHS, HSE, Cen01)
DDAc, by age, 2001, (LFS vs Omnibus)
Summary: Same question, same year, same population, different surveys – Q1 (b)

- **LLSI** (GHS, FRS, Omnibus, HSE, Cen01)
  - range 23% - 24%, (21% Cen01, 26% HSE)
  - more similar pre-retirement (<65)
  - sensitive to q wording (handicap vs disability – 5 pp higher in Cen01 v Cen91)
  - HSE - context effect, higher rates of reporting

- **DDAc** (LFS, OMN)
  - Question order effects (LFS higher: 6 pp@ 25%)

- **WLD** (LFS, FRS-2)
  - significant difference (LFS higher: 3 pp @18%)
  - why? Different question, context effects
Empirical Comparisons – Q2 (a)

- Are questions measuring similar underlying concepts associated? (convergent validity)
  - Examine patterns of overlap between responses by the same individual to different questions which essentially tap into the same underlying construct (eg. DDAc vs WLD) in the same survey

- Analysis approach:
  - Cross-tabulations ( % overlaps)
  - measures of association (kappa)
Overlaps: same person, different measure
LFS - WLD & DDAc to SPA

- No LSI: 70%
- LSI only: 8%
- WLD only: 14%
- WLD+DDAc: 21%
- DDAc only: 3%

Legend:
- No LSI
- LSI only
- WLD only
- WLD+DDAc
- DDAc only
Overlaps: same person, different measure (convergent validity - summary) – Q2 (b)

- Overall % agreement, kappa (association)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Agreement</th>
<th>Kappa</th>
<th>Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLSI v DDA (all)</td>
<td>90%</td>
<td>0.7</td>
<td>3/26=12%</td>
</tr>
<tr>
<td>LLSI v WLD (SPA)</td>
<td>95%</td>
<td>0.8</td>
<td>1/18=1%</td>
</tr>
<tr>
<td>DDA v WLD (SPA)</td>
<td>93%</td>
<td>0.7</td>
<td>5/21=24%</td>
</tr>
</tbody>
</table>

- good overall agreement / overlap
- some response variation (eg. people say yes to WLD, but ‘no’ to DDA (5% of 21%), comprehension issues, labour market?)
Empirical Comparisons – Q3 (a)

• Do single-item questions discriminate by level of disability? *(discriminant validity)*
  – Limited data due to lack of severity indicator, therefore check if implicit gradient in disability severity holds (e.g. DDA>LLSI>LSI>not disabled)

• Analysis approach: compare mean EQ5D scores by age and sex of not disabled, LSI only, LLSI only, DDA current disabled.
Mean EQ5D scores by degree of (implied) severity – Q3 (b)
Different estimates, by age group
Sources of variation between estimates -1

• **Who is included?** Population coverage
  – non-coverage (age group, communal estb, hard to reach, geography)
  – estimates relating to specific age groups (all, adults, SPA)

• **What is being measured?** Definitions
  – global estimates similar to SPA, by age & overall (e.g. LLSI=15.7%, DDAc=15.4%), but not older
  – global estimates higher than disability survey estimates (e.g. DS96/7: 12.4% to SPA)
How it is being measured?

• Differences because of
  – question wording (~33% Cen01 v Cen91)
  – question order (~33% LFS v Omnibus)
  – filters/screen (false negatives, ~30% HSE01)
  – self-reports (false positives, ~33% LFS)
  – proxies (more proxy, lower estimates, ratio 3:1 LFS, FRS, GHS)
  – context/survey sponsorship effects (~6% HSE v FRS)
  – interview mode (self-completion v tel v f-to-f)
Technical review - suggestions for future (1)

- Primary collection
  - distinguish clearly between ill-health and impairment/disability (no filter on health)
  - incl indicator of severity in global question
  - q testing, harmonisation of global q’s
  - specialist survey at regular intervals, with global question/s to cross-walk, validate, calibrate
  - if ‘can do’ measure then info on use of devices (affects long-term trends)
  - longitudinal panel to understand underlying processes (incidence, recovery, mortality), causes & risk factors
Publication and Contact details

• “Review of disability estimates and definitions”
  DWP In-house report 128 (2004)
  Bajekal M, Harris T, Breman R, Woodfield K

• Contact details:
  madhavi.bajekal@ons.gov.uk