# Appendices

## Appendices Table of Contents

[Appendices Table of Contents 337](#_Toc298356206)

[Chapter 1 Appendices 338](#_Toc298356207)

[Appendix 1A: WLD questions in the LFS 338](#_Toc298356208)

[Chapter 2 Appendices 339](#_Toc298356209)

[Appendix 2A: Detail of the systematic data review 339](#_Toc298356210)

[Appendix 2B: Construction of the demands/control scales 358](#_Toc298356211)

[Chapter 3 Appendices 363](#_Toc298356212)

[Appendix 3A: Construction of WLD scale 363](#_Toc298356213)

[Appendix 3B: Description of control variables 364](#_Toc298356214)

[Appendix 3C: Multiple Imputation 372](#_Toc298356215)

[Appendix 3D: Results of sensitivity analyses 373](#_Toc298356216)

[Chapter 4 Appendices 374](#_Toc298356217)

[Appendix 4A: Results of sensitivity analyses 374](#_Toc298356218)

[Chapter 5 Appendices 376](#_Toc298356219)

[Appendix 5A: Construction of the demands/control scales 376](#_Toc298356220)

[Appendix 5B: Empirical Bayes estimates 379](#_Toc298356221)

[Chapter 6 Appendices 381](#_Toc298356222)

[Appendix 6A: Control variables used in BHPS 381](#_Toc298356223)

[Appendix 6B: Physical demands 387](#_Toc298356224)

[Appendix 6C: Sensitivity analyses 388](#_Toc298356225)

[Chapter 7 Appendices 391](#_Toc298356226)

[Appendix 7A: Sensitivity analyses 391](#_Toc298356227)

[Appendix Bibliography 393](#_Toc298356228)

## Chapter 1 Appendices

### Appendix 1A: WLD questions in the LFS

The combined sample size for the LFS data is 746,000 observations, consisting of around 10,000 observations for each year of 1992-2006, 60,000 observations for each year of 1985-1991, and 180,000 observations for 1984. This appendix details the changes over time in question wording, question filtering, response rates and proxy variables.

#### Question wording

Before 1997, there was a single WLD question asking about limitations on ‘the kind of work you can do’. From 1997 this was supplemented with a further question on ‘the amount of paid work’. I therefore combined the two separate post-1997 WLD questions into a single variable, with individuals classified as having a WLD if they report *either* type of limitation.

There are also very minor changes in question wording in 1995 and 1996 (from ‘*limit* the kind of work’ 1994 to ‘*affect* the kind of work’ 1995, and from ‘*which affect* the kind of work that you *can do*’ 1995 to ‘*which would affect* the kind of work that you *might do*’ 1996). I have assumed that these minor changes did not affect the reported prevalence of WLD, but if these years are excluded then the overall rise in WLD in the chained series is from 10.1% to 14.2% rather than to 16.0% (the apparent rise is 1.3 percentage points 1994-5 and 0.5 percentage points 1995-6).

#### Question filtering

There are a number of changes in question filtering:

* From 1984-8, individuals were only asked the WLD question if they reported having a ‘health problem or disability’ from a list given on a showcard (including an ‘other health problems or disabilities’ option).
* From 1989-1996, the WLD question was asked to all working-age respondents.
* From 1997-, the question was again filtered, now based on whether the individual reported a longstanding illness (a ‘health problem or disability’ expected to last for more than a year).

The discontinuity from the 1988/89 change is likely to be smaller than the 1996/7 change – the screening question is similar to WLD but broader, whereas the 1997 screening question refocuses WLD on *longstanding* health problems.

#### Response rates and proxies

There has been a decline in response rates in the LFS, but this took place since 1997 so cannot explain the rise 1984-1997.[[1]](#footnote-1) Proxy responses in the LFS are much higher than other surveys ([Bajekal et al 2004:135](#_ENREF_2)), but the level of proxy responses has been constant over time (authors’ calculations). Hence there is no reason to think that methodological factors are behind the rise in WLD – especially as methodological changes are more likely to produce one-off steps in WLD rather than the consistent rise we see over the 1990s.

## Chapter 2 Appendices

### Appendix 2A: Detail of the systematic data review

This appendix provides the main body of the systematic data review, formed of the details of each pairwise comparison of the levels of a given variable in two different years. This is presented in a series of tables, one for each of the two dimensions of demands and the four dimensions of control.

Each pairwise comparison includes my assessment of the comparability (and representativeness) of the data. Many of these quality assessments are based on a smaller number of general issues with particular surveys, which are described here to preserve readability in the tables:

* EWCS: in the absence of any additional issues, the comparability of all the EWCS surveys is set to ‘moderate’. This is because they use ‘random walk’ sampling methods, where there is no known list of people/addresses, but instead the interviewer calls at every xth house on a predefined route. Random walk methods are theoretically an acceptable way of generating a random sample, but in practice there are various ways in which this fails in practice (non-random walks, interviewer biases in selecting participants) ([Lynn et al 2004](#_ENREF_22)) and response rates are typically much lower. Random walk methodologies therefore result in more biased samples ([Hoffmeyer-Zlotnik 2003](#_ENREF_16)).
* WiB 2000: WiB appears to have some comparability issues compared to EiB/the Skills Surveys for unknown reasons. Superficially it should be comparable based on question wording and sampling methodologies, but two pieces of evidence suggest problems: (i) WiB shows different levels of D1 (working very hard) and other variables to SS01 that was undertaken at a similar time; (ii) an occupation-level ecological regression finds lower correlations of WiB with all of EiB/SS97/SS01/SS06 than between the same questions in the other surveys. Given that the SS01/WiB difference is not statistically significant when comparable weights are constructed, this may be partly due to sampling error.
* BSA: The standard sample frame changed over the period considered here from the Electoral Register to the Postal Address File (on the basis that the Electoral Register became too partial to be valid for this purpose). BSA 1991 used a split sample to compare the characteristics of samples achieved from these different sources; my comparison of work-related variables from BSA 1991 suggests that there are no noticeable differences between them, but the power of this comparison is low.
* WERS: There are multiple significant levels in WERS at which refusal can take place, from workplaces that refuse to take part in the main survey, participating workplaces that then refuse to hand them on to staff, and even among those workplaces that hand them on, non-responses from employees – all of which is likely to result in biases by the relationship of employees with employer ([see Forth et al 2010:584-6](#_ENREF_14)). Aside from considerable problems of bias, the overall response rate declined substantially (43% 1998, 33% 2004), an unexplained problem that is unlikely to be solved by the non-response weights. One author who has looked at WERS trends has noted that the comparability of these surveys is weaker than for the Skills Surveys (Web Appendix 1a).

Details on the individual surveys used in the data review is available in Web Appendix 2a. Surveys used the most comparable weights available and were restricted to a common subsample that were asked the questions; these details for each pairwise comparison can be inferred from the table in Web Appendix 2a.

Table A1: Trends in job demands (working hard/fast)

| Question (code) | Did demands rise or fall 1 | Sig 1 | Comparability | Years | *Further detail* | |
| --- | --- | --- | --- | --- | --- | --- |
| Greatest change in response categories | Sources |
| *Early 1990s to late 1990s* |  |  |  |  |  |  |
| My job requires that I work very hard (D1) a | ↑↑ | \*\*\* | Moderate  Slight changes to question wording (different preamble, no showcard in SS97) | 1992-1997 | Strongly agree:  31.7% to 39.9% | EiB (n=3828)  SS97 (n=2461) |
| Does your job involve working at very high speed? (D4\_B) | ↑↑↑ | \*\* | Low  Sampling change from 1991 (random sample from Electoral Register) to 1995 (random walk). Non-response weights also differ 1991-1995 (occupation only included 1995), and it seems EWCS91 excluded non-UK-nationals | 1991-1995 | ¾ of the time or more:  19.7% to 29.9% | EWCS91 (n=940)  EWCS95 (n=960) |
| Do you have to work very fast? (D6) b | ≈↑ | ns | Moderate  Based on the 1st half of 1993 compared to the 2nd half of 1994 (full-year samples show no trend). Change in screening question to more explicitly include the self-employed in 1994 | 1993-1994 | Often:  38.1% to 39.3% | HSE93 (n=4344)  HSE94 (n=4805) |
| Do you have to work very intensively? (D10) b | ≈↑ | ns | Moderate  Based on the 1st half of 1993 compared to the 2nd half of 1994 (full-year samples show half the increase on the latent scale). Change in screening question to more explicitly include the self-employed in 1994 | 1993-1994 | Often:  41.0% to 42.3% | HSE93 (n=4347)  HSE94 (n=4797) |
| *Late 1990s to early 2000s* |  |  |  |  |  |  |
| My job requires that I work very hard (D1) | ≈↑ | ns | Moderate  See general WiB note. Also slight changes to question wording (different preamble, no showcard in SS97) | 1997-2000 | Strongly agree:  39.9% to 41.4% | SS97 (n=2461)  WiB (n=2457) |
| My job requires that I work very hard (D1) | **↓** | ns | Moderate  See general WiB note. | 2000-2001 | Strongly agree:  41.4% to 38.3% | WiB (n=2457)  SS01(n=4470) |
| Does your job involve working at very high speed? (D4\_B) | ↓↓ | † | Moderate  See general EWCS note. No information on response rate provided (only the 'cooperation rate' among those confirmed as eligible, which stays constant). Unclear whether design weight used in 1995. | 1995-2000 | Never:  31.7% to 37.9% | EWCS95 (n=960)  EWCS00(n=1372) |
| *Early 2000s to mid 2000s* |  |  |  |  |  |  |
| My job requires that I work very hard (D1) | ↑ | \* | High | 2001-2006 | Strongly agree:  38.3% to 42.1% | SS01(n=4470)  SS06 (n=6929) |
| How often does work involve working at very high speed? (D4) | ≈≈ | ns | High | 2001-2006 | Never /almost never:  23.4% to 23.9% | SS01(n=4459)  SS06 (n=6912) |
| Does your job involve working at very high speed? (D4\_B) | ≈≈ | ns | Moderate  See general EWCS note. No information on response rate provided (only the 'cooperation rate' among those confirmed as eligible, which falls by 10% 2000-2005). | 2000-2005 | Almost all time +:  20.0% to 24.4%  (Scale shows negligible *decline*) | EWCS00(n=1372)  EWCS05 (n=859) |
| I work under a great deal of tension (D12) | ≈≈ | ns | High | 2001-2006 | Strongly agree: 20.9% to 19.9%  (Scale shows negligible trend) | SS01(n=4465)  SS06 (n=6923) |
| How often does work involve working to tight deadlines? (D15) | ↑ | \* | High | 2001-2006 | ½ time or more:  66.8% to 69.7% | SS01(n=4464)  SS06 (n=6922) |
| *Mid 2000s to late 2000s* |  |  |  |  |  |  |
| I have to work very fast at work (D8) | ≈↓ | ns | High | 2004-2008 | Always:  12.2% to 9.6% | PWCS04-4(n=751)  PWCS08-4(n=516) |
| I have to work very intensively at work (D11) | ≈≈ | ns | High | 2004-2008 | Always:  17.0% to 15.9%  (Scale shows negligible trend) | PWCS04-4(n=752)  PWCS08-4(n=515) |
| *--COMBINING PERIODS--* |  |  |  |  |  |  |
| My job requires that I work very hard (D1) | ↑↑↑ | \*\*\* | High  Only very minor changes in question wording | 1992-2006 | Strongly agree: 31.7% to 42.1% | EiB (n=3828)  SS06 (n=6929) |
| How often does work involve working at very high speed? (D4) | ↑↑↑ | \*\*\* | High | 1992-2006 | Never /almost never:  49.6% to 23.9% | EiB (n=3827)  SS06 (n=6912) |
| Does your job involve working at very high speed? (D4\_B) | ↑ | \* | Low  Sampling change from 1991 (random sample from Electoral Register) to 1995 (random walk). Non-response weights also differ 1991-1995 (occupation only included 1995), and it seems EWCS91 excluded non-UK-nationals | 1991-2005 | All of the time:  6.5% to 9.2% | EWCS91 (n=940)  EWCS05 (n=859) |
| I work under a great deal of tension (D12) | ↑↑ | \*\*\* | High | 1992-2006 | Disagree or strongly disagree:  51.6% to 41.6% | EiB (n=3813)  SS06 (n=6923) |

**Key:**

**↑↑↑/↓↓↓** Very large rise/fall in demands (>10 percentage points in at least one category); **↑↑** Large rise/fall in demands (5-10%); **↑** Moderate rise/fall in demands (2-5%); **≈**↑ Small rise/fall in demands (1-2%); **≈≈** no change (<1%).

† p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001. For binary outcomes this is from a logistic regression model, for ordinal outcomes this is from an ordered logistic regression model.

Table A2: Trends in job demands (having enough time)

| Question (code) | Did demands rise or fall 1 | Sig 1 | Comparability | Years | *Further detail* | |
| --- | --- | --- | --- | --- | --- | --- |
| Greatest change in response categories | Sources |
| *Early 1990s* |  |  |  |  |  |  |
| Do you have enough time to do everything? (D19) | ↑ | † | Moderate  Based on the 1st half of 1993 compared to the 2nd half of 1994 (full-year sample shows similar results). Change in screening question to more explicitly include the self-employed in 1994 | 1993-1994 | Often:  30.4% to 28.3% | HSE93 (n=4339)  HSE94 (n=4796) |
| *Late 1990s to early 2000s* |  |  |  |  |  |  |
| Never have enough time to get everything done on my job (D20\_B) | ↓↓↓ | \*\* | Low  EB56-1 was conducted immediately following 9/11, and therefore has a very low response rate (21% vs. 53% in 1996) | 1996-2001 | Disagree or strongly disagree:  31.2% to 42.3% | EB44-3 (n=544)  EB56-1 (n=490) |
| You have enough time to get the job done (D21\_B) | ↓ | \* | Moderate  See general EWCS note. No information on response rate provided (only the 'cooperation rate' among those confirmed as eligible, which stays constant). Unclear whether design weight used in 1995. | 1995-2000 | Yes:  69.6% to 75.2% | EWCS95 (n=985)  EWCS00 (n=1404) |
| *Early 1990s to early 2000s* |  |  |  |  |  |  |
| Never have enough time to get everything done on my job (D20) | ↑↑ | \*\*\* | Moderate  See general WiB note. | 1992-2000 | Strongly agree:  20.5% to 27.1% | EiB (n=3806)  WiB (n=2447) |
| *Early 2000s to mid 2000s* |  |  |  |  |  |  |
| Never have enough time to get everything done on my job (D20\_B) | ↑↑↑ | \* | Low  EB56-1 was conducted immediately following 9/11, and therefore has a very low response rate (21% vs. 53% in 1996) | 2001-2004 | Disagree or strongly disagree:  42.3% to 29.9% | EB56-1 (n=490)  ESS04 (n=751) |
| *Mid/late 1990s to mid 2000s* |  |  |  |  |  |  |
| Never have enough time to get everything done on my job (D20\_B) | ≈↓ | ns | Low  Different sampling (ESS04 is a random sample based on the PAF; EB44-3 is a random walk sample), different inclusion criteria (non-EU nationals excluded from EB44-3, different self-employment questions), and different weights (EB44-3 has non-response weights, ESS04 has design weights only) | 1996-2004 | Strongly agree:  21.0% to 15.6%  (Scale difference is much smaller) | EB44-3 (n=544)  ESS04 (n=751) |
| Never have enough time to get everything done on my job (D20\_BS) | ≈≈ | ns | Moderate  See general WERS note. | 1998-2004 | Disagree or strongly disagree:  26.9% to 27.9% | WERS98 (n=25393)  WERS04 (n=19430) |
| *Mid 2000s to late 2000s* |  |  |  |  |  |  |
| I have unrealistic time pressures at work (D24) b | ↓↓ | \* | Low  In general the PWCS series provides highly comparable trends. However, non-response weights are not available for the 2004-8 trend, and using these weights reduces the size of the trend 2005-7. Caution should therefore be used with this trend. | 2004-2008 | Never or seldom:  51.9% to 57.6% | PWCS04 (n=1629)  PWCS08 (n=547) |

**Key:**

**↑↑↑/↓↓↓** Very large rise/fall in demands (>10 percentage points in at least one category); **↑↑** Large rise/fall in demands (5-10%); **↑** Moderate rise/fall in demands (2-5%); **≈**↑ Small rise/fall in demands (1-2%); **≈≈** no change (<1%).

† p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001. For binary outcomes this is from a logistic regression model, for ordinal outcomes this is from an ordered logistic regression model.

Table A3: Trends in job control (control over work pace/effort)

| Question (code) | Did control rise or fall 1 | Sig 1 | Comparability | Years | *Further detail* | |
| --- | --- | --- | --- | --- | --- | --- |
| Greatest change in response categories | Sources |
| *Mid 1980s to early 1990s* |  |  |  |  |  |  |
| Which are important in determining how hard you work? Own discretion (C2\_B) | ↑↑ | \* | Low  SCELI and EiB have very different sampling methodologies (EiB is a random sample, SCELI is a stratified sample of only 8 areas). [Sig level here is after adjustment for clustering of SCELI]. The question wording in SCELI was also slightly different (no option of 'targets you are set' alongside 'own discretion') | 1986-1992 | Yes:  61.3% to 66.7% | SCELI (n=3633)  EiB (n=3372) |
| Can you...slow down your pace of work for a day when you want to? (C10) | ↑↑↑ | \*\*\* | Low  Design weights are not available for SCMB, which makes establishing the representativeness of the sample difficult | 1984-1996 | Yes:  48.0% to 60.1% | SCMB (n=779)  Omni96-4 (n=608) |
| *Early 1990s to late 1990s* |  |  |  |  |  |  |
| How much influence do you have on how hard you work? (C1\_B) | ↓↓ | \*\*\* | Moderate  Question wording is identical but different preamble ("How much influence does (i) your work group; (ii) you personally; (iii) your supervisor or superior have on..." , RATHER THAN "How much influence do you personally have on...") | 1992-1997 | A great deal:  70.8% to 64.4% | EiB (n=3392)  SS97 (n=2193) |
| Which are important in how hard you work? Own discretion (C2\_B) z | ≈≈ | ns | Moderate  Slight changes in question wording (EiB doesn't have the 'none of these' option, but it does have 'targets you are set' which is lacking in SS97) | 1992-1997 | Yes:  66.6% to 67.6% | EiB (n=3372)  SS97 (n=2195) |
| Which are important in how hard you work? Own discretion (C2\_B) z | ↓↓↓ | \*\*\* | Low  Different sampling methodology (EiB is random sample, EB44-3 is random walk), and v slight difference in wording (EB44-3 has 'pay incentives such as bonuses etc' rather than just 'pay incentives') | 1992-1996 | Yes:  67.0% to 41.0% | EiB (n=3372)  EB44-3 (n=416) |
| Which are important in how hard you work? Own discretion (C2\_B) z | ↑↑↑ | \*\*\* | Low  Different sampling methodology (SS97 is random sample, EB44-3 is random walk), wording is slightly different (EB44-3 includes the category 'targets you are set' but not the 'none of these' option, and has 'pay incentives such as bonuses etc' rather than just 'pay incentives') | 1996-1997 | Yes:  41.0% to 67.6% | EB44-3 (n=416)  SS97 (n=2195) |
| Are you able, or not, to choose or change...your speed or rate of work (C7) | ↑↑ | \*\*\* | Low  Sampling change from 1991 (random sample from Electoral Register) to 1995 (random walk). Non-response weights also differ 1991-1995 (occupation only included 1995), and it seems EWCS91 excluded non-UK-nationals | 1991-1995 | Yes:  72.6% to 80.4% | EWCS91 (n=938)  EWCS95 (n=962) |
| *Late 1990s to early 2000s* |  |  |  |  |  |  |
| How much influence do you have on how hard you work? (C1) | ↓↓↓ | \*\*\* | High | 1997-2001 | A great deal:  66.6% to 54.0% | SS97 (n=2465)  SS01 (n=4467) |
| Which are important in how hard you work? Own discretion (C2) | ↓↓ | \*\*\* | High | 1997-2001 | Yes:  67.4% to 62.4% | SS97 (n=2467)  SS01 (n=4468) |
| Which are important in determining how hard you work? Own discretion (C2\_B) | ↓↓↓ | \*\*\* | Moderate  Very slight changes in ordering of question responses. See also general WiB note. | 1997-2000 | Yes:  67.6% to 53.8% | SS97 (n=2195)  WiB (n=2101) |
| Which are important in determining how hard you work? Own discretion (C2\_B) | ↑↑ | \*\*\* | Moderate  See general WiB note. | 2000-2001 | Yes:  53.8% to 61.9% | WiB (n=2101)  SS01 (n=4003) |
| How much influence do you have about...the pace at which you work (C5) | ↑ | \*\*\* | Moderate  See general WERS note. | 1998-2004 | A lot:  34.1% to 36.9% | WERS98 (n=25563)  WERS04 (n=19541) |
| Are you able, or not, to choose or change...your speed or rate of work (C7) | ↓↓ | \*\* | Moderate  See general EWCS note. No information on response rate provided (only the 'cooperation rate' among those confirmed as eligible, which stays constant). Unclear whether design weight used in 1995. | 1995-2000 | Yes:  80.4% to 73.3% | EWCS95 (n=962)  EWCS00 (n=1362) |
| *Early 2000s to mid 2000s* |  |  |  |  |  |  |
| How much influence do you have on how hard you work? (C1) | ≈↑ | ns | High | 2001-2006 | A great deal:  54.0% to 55.2% | SS01 (n=4467)  SS06 (n=6922) |
| Which are important in how hard you work? Own discretion (C2) | ↓ | \*\*\* | High | 2001-2006 | Yes:  62.4% to 57.5% | SS01 (n=4468)  SS06 (n=6928) |
| Are you able, or not, to choose or change...your speed or rate of work (C7) | ↓ | ns | Moderate  See general EWCS note. No information on response rate provided (only the 'cooperation rate' among those confirmed as eligible, which falls by 10% 2000-2005). | 2000-2005 | Yes:  73.3% to 70.0% | EWCS00 (n=1362)  EWCS05 (n=852) |
| *Mid 2000s to late 2000s* |  |  |  |  |  |  |
| I have a say in my own work speed (C6) | ↓ | ns | High | 2004-2008 | Always:  28.1% to 23.9% | PWCS04-4 (n=751)  PWCS08-4 (n=515) |
| *\*\* MULTIPLE PERIODS \*\** |  |  |  |  |  |  |
| How much influence do you have on how hard you work? (C1\_B) | ↓↓↓ | \*\*\* | Moderate  Question wording is identical but different preamble ("How much influence does (i) your work group; (ii) you personally; (iii) your supervisor or superior have on..." , RATHER THAN "How much influence do you personally have on...") | 1992-2006 | A great deal:  70.8% to 51.8% | EiB (n=3392)  SS06 (n=6107) |
| Which are important in determining how hard you work? Own discretion (C2\_B) z | ↓↓↓ | \*\*\* | Moderate  Slight changes in question wording (EiB doesn't have the 'none of these' option in SS06, but unlike SS06 it has the option of 'targets you are set') | 1992-2006 | Yes:  66.6% to 56.3% | EiB (n=3372)  SS06 (n=6113) |
| Which are important in determining how hard you work? Own discretion (C2\_B) z | ↓ | \* | Low  SCELI and SS06 have very different sampling methodologies (SS06 is a random sample, SCELI is a stratified sample of only 8 areas). [Sig level here is after adjustment for clustering of SCELI]. The question wording in SCELI was also slightly different (no option of 'none of these' alongside 'own discretion') | 1986-2006 | Yes:  61.3% to 56.8% | SCELI (n=3633)  SS06 (n=6113) |
| Are you able, or not, to choose or change...your speed or rate of work (C7) | ↓ | ns | Low  Sampling change from 1991 (random sample from Electoral Register) to 2005 (random walk). Non-response weights also differ 1991-2005 (occupation only included 2005), and it seems EWCS91 excluded non-UK-nationals | 1991-2005 | Yes:  72.6% to 70.0% | EWCS91 (n=938)  EWCS05 (n=852) |
| Does someone else decide how much work you do or how fast you work? (C8) d | ↑↑ | \*\*\* | Low  Design weights are not available for SCMB, which makes establishing the representativeness of the sample difficult. See also general WiB note. | 1984-2000 | Yes:  32.4% to 22.5% | SCMB (n=929)  WiB (n=2043) |

**Key:**

**↑↑↑/↓↓↓** Very large rise/fall in demands (>10 percentage points in at least one category); **↑↑** Large rise/fall in demands (5-10%); **↑** Moderate rise/fall in demands (2-5%); **≈**↑ Small rise/fall in demands (1-2%); **≈≈** no change (<1%).

† p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001. For binary outcomes this is from a logistic regression model, for ordinal outcomes this is from an ordered logistic regression model.

z C2 correlates relatively poorly with other questions on job control, and is therefore excluded from the job control scales used elsewhere in this thesis. This suggests that it is measuring something different from the other questions in this section (perhaps unsurprisingly given that it is part of a larger question on sources of effort pressure).

Table A4: Trends in job control (control over how work is done)

| Question (code) | Did control rise or fall 1 | Sig 1 | Comparability | Years | *Further detail* | |
| --- | --- | --- | --- | --- | --- | --- |
| Greatest change in response categories | Sources |
| *1980s to 1990s* |  |  |  |  |  |  |
| My job allows me to design or plan most of my daily work (C21\_S) | ↑ | † | Moderate  Possible non-response biases (C21\_S is in the self-completion part of the BSA questionnaire, which a number of participants refuse after completing the face-to-face interview. No non-response weights are supplied). See also general BSA note. | 1989-1993 | Most of my work:  39.3% to 43.0% | BSA89 (n=699)  BSA93 (n=1364) |
| How much choice do you have over the way in which you do your job (C23) | ↓↓ | \*\* | Low  SCELI and SS97 have very different sampling methodologies (EiB is a random sample, SCELI is a stratified sample of only 8 areas) | 1986-1997 | A great deal:  51.7% to 43.9% | SCELI (n=3594)  SS97 (n=2195) |
| *Early to late1990s* |  |  |  |  |  |  |
| Do you have a choice in deciding HOW you do your work? (C29) | ≈≈ | ns | Moderate  Change in screening question to more explicitly include the self-employed in 1994 | 1993-1994 | Never/almost never or seldom:  15.5% to 14.6% | HSE93 (n=4357)  HSE94 (n=4810) |
| My job allows me to design or plan most of my daily work (C21\_S) | ↑ | ns  (\* 89-95) | Moderate  Possible non-response biases (C21\_S is in the self-completion part of the BSA questionnaire, which a number of participants refuse after completing the face-to-face interview. No non-response weights are supplied). See also general BSA note. | 1993-1995 | Most of my work:  43.0% to 46.2% | BSA93 (n=699)  BSA95 (n=1364) |
| How much influence do you have on deciding how to do the task? (C25\_B) | ↓↓ | \*\*\* | Moderate  Question wording is identical but different preamble ("How much influence does (i) your work group; (ii) you personally; (iii) your supervisor or superior have on..." , RATHER THAN "How much influence do you personally have on...") | 1992-1997 | A great deal:  56.9% to 49.7% | EiB (n=3390)  SS97 (n=2194) |
| How much influence on deciding quality standards to work to? (C49\_B) | ↓↓↓ | \*\*\* | Moderate  Question wording is identical but different preamble ("How much influence does (i) your work group; (ii) you personally; (iii) your supervisor or superior have on..." , RATHER THAN "How much influence do you personally have on...") | 1992-1997 | A great deal:  69.7% to 51.1% | EiB (n=3389)  SS97 (n=2191) |
| *Late 1990s to early/mid 2000s* |  |  |  |  |  |  |
| How much choice do you have over the way in which you do your job (C23) | ↓↓ | \*\* | High | 1997-2001 | A great deal:  44.3% to 38.6% | SS97 (n=2195)  SS01 (n=4004) |
| How much influence do you have on deciding how to do the task? (C25) | ↓↓ | \*\*\* | High | 1997-2001 | A great deal:  52.7% to 46.2% | SS97 (n=2466)  SS01 (n=4470) |
| How much influence do you have about...how you do your work? (C26) | ≈↑ | \* | Moderate  See general WERS note. | 1998-2004 | None:  5.4% to 4.3% | WERS98 (n=25557)  WERS04 (n=19596) |
| How much influence on...deciding quality standards to work to? (C49) | ↑ | † | High | 1997-2001 | Not much or  None at all;  19.0% to 15.0% | SS97 (n=2463)  SS01 (n=4465) |
| *Early 2000s to mid 2000s* |  |  |  |  |  |  |
| How much choice do you have over the way in which you do your job (C23) | ↓ | \* | High | 2001-2006 | A great deal:  38.6% to 36.0% | SS01 (n=4004)  SS06 (n=6115) |
| How much influence do you have on deciding how to do the task? (C25) | ≈≈ | ns | High | 2001-2006 | Not much /  none at all  15.5% to 16.4% | SS01 (n=4470)  SS06 (n=6931) |
| How much influence on deciding quality standards to work to? (C49) | ≈≈ | ns | High | 2001-2006 | Not much /  None at all;  15.0% to 16.3%  (Scale change is negligible) | SS01 (n=4465)  SS06 (n=6928) |
| *Mid 2000s to late 2000s* |  |  |  |  |  |  |
| I have some say over the way I work (C24) | ↓ | ns | High | 2004-2008 | Strongly agree:  30.8% to 26.9% | PWCS04-4 (n=793)  PWCS08-4 (n=547) |
| I have a choice in deciding how I do my work (C30) | ↓ | ns | High | 2004-2008 | Never or Seldom:  16.3% to 18.9% | PWCS04-3/4 (n=1631)  PWCS08-4 (n=546) |
| *--WHOLE PERIOD--* |  |  |  |  |  |  |
| Can design/plan important aspects of own work or largely defined for you? (C18) | ↑ | \* | Low  Design weights are not available for SCMB, which makes establishing the representativeness of the sample difficult. See also general WiB note. | 1984-2000 | Can design/plan:  38.3% to 43.2% | SCMB (n=931)  WiB (n=2044) |
| How much choice do you have over the way in which you do your job (C23) | ↓↓↓ | \*\*\* | Low  SCELI and SS97 have very different sampling methodologies (EiB is a random sample, SCELI is a stratified sample of only 8 areas) | 1986-2006 | A great deal:  51.7% to 36.3% | SCELI (n=3594)  SS06 (n=6115) |
| How much influence do you have on deciding how to do the task? (C25\_B) | ↓↓↓ | \*\*\* | Moderate  Question wording is identical but different preamble ("How much influence does (i) your work group; (ii) you personally; (iii) your supervisor or superior have on..." , RATHER THAN "How much influence do you personally have on...") | 1992-2006 | A great deal:  56.9% to 42.4% | EiB (n=3390)  SS06 (n=6116) |
| How much influence on deciding quality standards to work to? (C49\_B) | ↓↓↓ | \*\*\* | Moderate  Question wording is identical but different preamble ("How much influence does (i) your work group; (ii) you personally; (iii) your supervisor or superior have on..." , RATHER THAN "How much influence do you personally have on...") | 1992-2006 | A great deal:  69.7% to 51.2% | EiB (n=3389)  SS06 (n=6113) |

**Key:**

**↑↑↑/↓↓↓** Very large rise/fall in demands (>10 percentage points in at least one category); **↑↑** Large rise/fall in demands (5-10%); **↑** Moderate rise/fall in demands (2-5%); **≈**↑ Small rise/fall in demands (1-2%); **≈≈** no change (<1%).

† p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001. For binary outcomes this is from a logistic regression model, for ordinal outcomes this is from an ordered logistic regression model.

Table A5: Trends in job control (control over task order/choice)

| Question (code) | Did control rise or fall 1 | Sig 1 | Comparability | Years | *Further detail* | |
| --- | --- | --- | --- | --- | --- | --- |
| Greatest change in response categories | Sources |
| *1980s to 2000* |  |  |  |  |  |  |
| Do you decide the specific tasks that you carry out or does someone else? (C38) | ↑↑ | \*\*\* | Moderate  Different sampling frames (electoral register vs. PAF) but both are multistage stratified random samples. See also general WiB note. | 1985-2000 | Respondent does:  53.4% to 61.8% | BSA85 (n=740)  WiB (n=2031) |
| Do you decide the specific tasks that you carry out or does someone else? (C38) | ↑↑↑ | \*\*\* | Low  Design weights are not available for SCMB, which makes establishing the representativeness of the sample difficult. See also general WiB note. | 1984-2000 | Respondent does:  47.8% to 61.8% | SCMB (n=923)  WiB (n=2031) |
| Can you decide on your own to introduce a new task that you will do? (C39) | ↑↑↑ | \*\*\* | Low  Design weights are not available for SCMB, which makes establishing the representativeness of the sample difficult. See also general WiB note. | 1984-2000 | Yes:  47.3% to 63.2% | SCMB (n=933)  WiB (n=2086) |
| *Early to late 1990s* |  |  |  |  |  |  |
| How much influence do you have on deciding what tasks to do? (C33\_B) | ↓↓ | \*\*\* | Moderate  Question wording is identical but different preamble ("How much influence does (i) your work group; (ii) you personally; (iii) your supervisor or superior have on..." , RATHER THAN "How much influence do you personally have on...") | 1992-1997 | A great deal:  42.4% to 33.1% | EiB (n=3391)  SS97 (n=2194) |
| *Mid 1990s to early/mid 2000s* |  |  |  |  |  |  |
| How much influence do you have on deciding what tasks you to do? (C33) | ↓ | \* | High | 1997-2001 | A great deal:  37.3% to 34.1% | SS97 (n=2466)  SS01 (n=4468) |
| Are you able, or not, to choose or change...your order of tasks? (C42) | ↓ | ns | Moderate  See general EWCS note. No information on response rate provided (only the 'cooperation rate' among those confirmed as eligible, which stays constant). Unclear whether design weight used in 1995. | 1995-2000 | Yes:  73.5% to 70.9% | EWCS95 (n=1008)  EWCS00 (n=1432) |
| *Early 2000s to mid 2000s* |  |  |  |  |  |  |
| How much influence do you have on… deciding what tasks to do? (C33) | ≈≈ | ns | High | 2001-2006 | A great deal:  34.1% to 34.3% | SS01 (n=4468)  SS06 (n=6930) |
| Are you able, or not, to choose or change...your order of tasks? (C42) | ↓ | \* | Moderate  See general EWCS note. No information on response rate provided (only the 'cooperation rate' among those confirmed as eligible, which falls by 10% 2000-2005). | 2000-2005 | Yes:  70.9% to 66.4% | EWCS00 (n=1432)  EWCS05 (n=927) |
| How much mngment allow you to decide how daily wrk is organised? (C46) | ↑ | †  (\* 02-  08) | High  2002-2008 trend is significant and larger (6+ out of 10 (10=Complete Control) 65.5% to 76.8%) | 2002-2004 | Complete control:  17.5% to 19.8% | ESS02 (n=860)  ESS04 (n=740) |
| *Mid 2000s to late 2000s* |  |  |  |  |  |  |
| I have a choice in deciding what I do at work (C36) | ↑ | ns | High | 2004-2008 | Never or seldom:  35.0% to 40.3%  (Scale change is smaller) | PWCS04-4 (n=795)  PWCS08-4 (n=547) |
| How much mngment allow you to decide how daily wrk is organised? (C46) | ↑ | ns | High  2002-2008 trend is significant and larger (6+ out of 10 (10=Complete Control) 65.5% to 76.8%) | 2004-2008 | 8+ out of 10 (10=Complete Control):  51.6% to 55.8% | ESS04 (n=740)  ESS08 (n=966) |
| *MULTIPLE PERIODS* |  |  |  |  |  |  |
| How much influence do you have on deciding what tasks to do? (C33\_B) | ↓↓↓ | \*\*\* | Moderate  Question wording is identical but different preamble ("How much influence does (i) your work group; (ii) you personally; (iii) your supervisor or superior have on..." , RATHER THAN "How much influence do you personally have on...") | 1992-2006 | A great deal:  42.4% to 28.1% | EiB (n=3391)  SS06 (n=6115) |
| Are you able, or not, to choose or change...your order of tasks? (C42) | ↓↓ | \*\* | Moderate  See general EWCS note. No information on response rate provided (only the 'cooperation rate' among those confirmed as eligible, which stays constant). Unclear whether design weight used in 1995. | 1995-2005 | Yes:  73.5% to 66.4% | EWCS95 (n=1008)  EWCS05 (n=927) |

**Key:**

**↑↑↑/↓↓↓** Very large rise/fall in demands (>10 percentage points in at least one category); **↑↑** Large rise/fall in demands (5-10%); **↑** Moderate rise/fall in demands (2-5%); **≈**↑ Small rise/fall in demands (1-2%); **≈≈** no change (<1%).

† p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001. For binary outcomes this is from a logistic regression model, for ordinal outcomes this is from an ordered logistic regression model.

Table A6: Trends in job control (say in decisions at work)

| Question (code) | Did control rise or fall 1 | Sig 1 | Comparability | Years | *Further detail* | |
| --- | --- | --- | --- | --- | --- | --- |
| Greatest change in response categories | Sources |
| *Mid to late 1980s* |  |  |  |  |  |  |
| If a change in the way you do your job, would you have any say? (C16\_1) | ↓↓↓ | \*\*\* | High  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. | 1985-1987 | Yes:  62.7% to 50.9% | BSA85 (n=770)  BSA87 (n=1213) |
| Do you have a good deal of say in decisions about work? (C13) | **↑** | ns | Moderate  Poor documentation makes it difficult to check details (e.g. response rates are not provided). But question wording and sampling appear to be identical. | 1981-1990 | 7+ out of 10  (10=’A great deal’):  63.8% to 66.6% | WVS81 (n=535)  WVS90 (n=784) |
| *Late 1980s to early 1990s* |  |  |  |  |  |  |
| If a change in the way you do your job, would you have any say? (C16\_1) | ≈≈ | ns | Moderate  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. See also general BSA note. | 1987-1993 | Yes:  50.9% to 51.9%  (Scale change is negligible) | BSA87 (n=1213)  BSA93 (n=1060) |
| *Early to mid 1990s* |  |  |  |  |  |  |
| Do you have a good deal of say in decisions about work? (C14) | ≈↓ | ns | Moderate  Change in screening question to more explicitly include the self-employed in 1994 | 1993-1994 | Often:  46.1% to 44.6% | HSE93 (n=4358)  HSE94 (n=4813) |
| If a change in the way you do your job, would you have any say? (C16\_1) | ↑ | \* | High  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. | 1993-1996 | Yes:  51.9% to 56.3% | BSA93 (n=1060)  BSA96 (n=1414) |
| *Early/mid 1990s to early/mid 2000s* |  |  |  |  |  |  |
| I have a lot of say over what happens in my job (C12) | ↑↑ | \*\*\* | High | 1992-2001 | Not at all true:  25.3% to 17.9% | EiB (n=3431)  SS01 (n=4004) |
| I have a lot of say over what happens in my job (C12\_B) | ↓↓ | \*\* | Low  EB56-1 survey was conducted immediately following 9/11, and therefore has a very low response rate (21% vs. 53% in 1996) | 1996-2001 | Quite or very true:  51.4% to 43.2% | EB44-3 (n=632)  EB56-1 (n=574) |
| My job allows me to take part in making decisions that affect my work (C15) | ≈↑ | ns | High | 1992-2001 | Not at all true:  17.5% to 13.6%  (Scale change is much smaller) | EiB (n=3438)  SS01 (n=4004) |
| My job allows me to take part in making decisions that affect my work (C15\_B) | ↓↓ | ns | Low  EB56-1 survey was conducted immediately following 9/11, and therefore has a very low response rate (21% vs. 53% in 1996) | 1996-2001 | Very true:  34.5% to 29.5% | EB44-3 (n=634)  EB56-1 (n=573) |
| If a change in the way you do your job, would you have any say? (C16\_1) | ↑↑ | \*\*\* | Moderate  See general WiB note. ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’, and is roughly constant (8.1% and 7.0% in EiB and WiB respectively). | 1992-2000 | Yes:  51.9% to 60.0% | EiB (n=3430)  WiB (n=2040) |
| If a change in the way you do your job, would you have any say? (C16\_1) | ↓↓ | \*\*\* | Moderate  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. However, it may have been presented differently in SS01 as it shows a sharp rise (7.0% to 14.9%) that is not visible in the BSA series (in which ‘it depends’ slightly declines); the proportion of ‘no’ answers in WiB & SS01 is almost identical. | 2000-2001 | Yes:  60.0% to 53.0% | WiB (n=2040)  SS01 (n=3905) |
| If a change in the way you do your job, would you have any say? (C16\_1) | ≈≈ | ns | High  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. | 1996-2005 | Yes:  56.3% to 57.1% | BSA96 (n=1414)  BSA05 (n=1336) |
| If a change in the way you do job, would you have any say in decision? (C16\_1) | ↑↑↑ | \*\*\* | Low  EB56-1 survey was conducted immediately following 9/11, and therefore has a very low response rate (21% vs. 53% in 1996). ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. | 1996-2001 | Yes:  57.3% to 74.7% | EB44-3 (n=395)  EB56-1 (n=347) |
| *Early 2000s to mid 2000s* |  |  |  |  |  |  |
| My job allows me to take part in making decisions that affect my work (C15) | ↑↑ | \*\*\* | High | 2001-2006 | Very true:  24.7% to 32.9% | SS01 (n=4004)  SS06 (n=6113) |
| If a change in the way you do your job, would you have any say? (C16\_1) | ↓↓ | \*\*\* | Moderate  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. However, it may have been presented differently in SS06 as it shows a slight rise (14.9% to 18.3%) that is not visible in the BSA series (in which ‘it depends’ slightly declines). | 2001-2006 | Yes:  53.0% to 44.0% | SS01 (n=3905)  SS06 (n=5594) |
| *WHOLE PERIOD* |  |  |  |  |  |  |
| If a change in the way you do your job, would you have any say? (C16\_1) | ↓↓ | \*\* | Moderate  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. See also general BSA note. | 1985-2005 | Yes:  62.7% to 57.1% | BSA85 (n=770)  BSA05 (n=1336) |
| If a change in the way you do your job, would you have any say? (C16\_1) | ↑↑ | \*\* | High  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. | 1985-2005 | Yes:  51.9% to 57.1% | BSA93 (n=1060)  BSA05 (n=1336) |
| My job allows me to take part in making decisions that affect my work (C15) | ↑↑ | \*\*\* | High | 1992-2006 | Very true:  26.5% to 32.9% | EiB (n=3438)  SS06 (n=6113) |
| If a change in the way you do your job, would you have any say? (C16\_1) | ≈↓ | ns | Moderate  ‘It depends’ treated as intermediate ordinal category between ‘yes’ and ‘no’. However, it may have been presented differently in SS06 as it shows a sharp rise (7.0% to 18.3%) that is not visible in the BSA series (in which ‘it depends’ slightly declines), with reduced answers for both ‘yes and ‘no’. | 1992-2006 | Yes:  50.6% to 44.0%  (Scale change is much smaller, see left on ‘it depends’) | EiB (n=3430)  SS06 (n=5594) |

**Key:**

**↑↑↑/↓↓↓** Very large rise/fall in demands (>10 percentage points in at least one category); **↑↑** Large rise/fall in demands (5-10%); **↑** Moderate rise/fall in demands (2-5%); **≈**↑ Small rise/fall in demands (1-2%); **≈≈** no change (<1%).

† p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001. For binary outcomes this is from a logistic regression model, for ordinal outcomes this is from an ordered logistic regression model.

### Appendix 2B: Construction of the demands/control scales

Three questions on demands and five questions on control are available in all of the Employment in Britain survey 1992 and the Skills Surveys 2001/2006, from which we can construct scales of demands and control. These are:

**Demands**

D1: My job requires that I work very hard

1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree

D4: How often does your work involve working at very high speed?

1. All the time 2. Almost all the time 3. Around three quarters of the time 4. Around half the time 5. Around quarter of the time 6. Almost never 7. Never

D12: I work under a great deal of tension

1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree

**Control**

C1: How much influence do you personally have on how hard you work?

1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree

C2: Which, if any, of the things on this card are important in determining how hard you work in your job?

...Your own discretion

C25: (And how much influence do you personally have on … deciding how you are to do the task?

1. A great deal 2. A fair amount 3. Not much 4. None at all

C33: And how much influence do you personally have on… deciding what tasks you are to do?

1. A great deal 2. A fair amount 3. Not much 4. None at all

C49: And how much influence do you personally have on … deciding the quality standards to which you work?

1. A great deal 2. A fair amount 3. Not much 4. None at all

Initial analyses used all of these questions. However, Principal Components Analysis suggested that one of the items for control (C2) was only weakly related to the other questions; this question was therefore dropped. (The question itself was designed to measure effort pressures rather than job control, so it seems reasonable to consider that the responses reflect wider factors than the other questions).

There are various ways of constructing scales from measures such as these based on the degree to which they appear to measuring the same underlying construct (rather than arbitrary summary scores). Much of the previous work using the Skills Surveys has used Principal Components Analysis (PCA), but while this is a widely-used approximation, PCA makes assumptions that are not technically valid for the ordinal data used here.

Instead, I here used Item Response Theory (IRT) models. IRT models are conceptually distinct from PCA/factor analysis in certain ways ([Bartholomew et al 2008](#_ENREF_3)), but they are used here for the practical reason that they are more suited to ordinal data. The IRT model for ordinal data is an extension of the model for binary data, where is the probability of observing a specific binary response given the latent trait . For the ordinal model, we model the cumulative response probability of a response falling into group ‘*s*’ or lower ([Zheng and Rabe-Hesketh 2007](#_ENREF_37); [Bartholomew et al 2008](#_ENREF_3)):

and therefore

For example, would be the probability of a person giving the response ‘strongly agree’ (s=1) or ‘agree’ (s=2) to item D1 (‘My job requires that I work very hard’). The univariate two-parameter model is usually parameterised based on the logistic distribution, showing the odds of being in any category up to and including category s, based on the difficulty parameter and the discrimination parameters for each question in the overall scale:

While Stata is not the ideal package in which to do latent variable analyses, the capability to estimate ordinal IRT models has been provided using the GLLAMM command of Skrondal, Pickles and Rabe-Hesketh ([Skrondal and Rabe-Hesketh 2004](#_ENREF_34); [Zheng and Rabe-Hesketh 2007](#_ENREF_37)). While some authors present relatively little of the GLLAMM output ([Li et al 2005](#_ENREF_21); [Ng et al 2010](#_ENREF_23)), I here follow other authors ([Bertelli and Richardson Jr 2008](#_ENREF_4); [Raileanu Szeles and Fusco 2009](#_ENREF_25); [Faye et al 2011:13](#_ENREF_12)) in presenting the difficulty and discrimination parameters for transparency at the end of this Appendix. The trait score for each possible combination of answers was then estimated and assigned to each person in the Skills Surveys.

Finally, for the main demands and control scales, I checked the two-parameter IRT models against other possible ways of constructing latent scales, including less flexible (one-parameter) IRT models, PCA adapted for ordinal variables (polychoric PCA, [Kolenikov and Angeles 2004](#_ENREF_20); [Holgado-Tello et al 2008](#_ENREF_17)) and conventional PCA analysis. These other methods all produced a control scale that was more highly non-normal than the two-parameter IRT models, but otherwise the scales all correlated with one another very highly.

Unless otherwise specified, the results below account for differential non-response by gender.[[2]](#footnote-2) The combined sample size for these results is 15,654 for control and 13,397 for demands (the difference between them being the absence of the 1997 Skills Survey for the demands results). Some caution should be used with linear forms of the control scale as it leads to a highly non-normal distribution of the latent scale, as shown in Figure 1 (although this is not a problem for the categorical job strain measures).

Figure 1: Normality of the demands and control scales

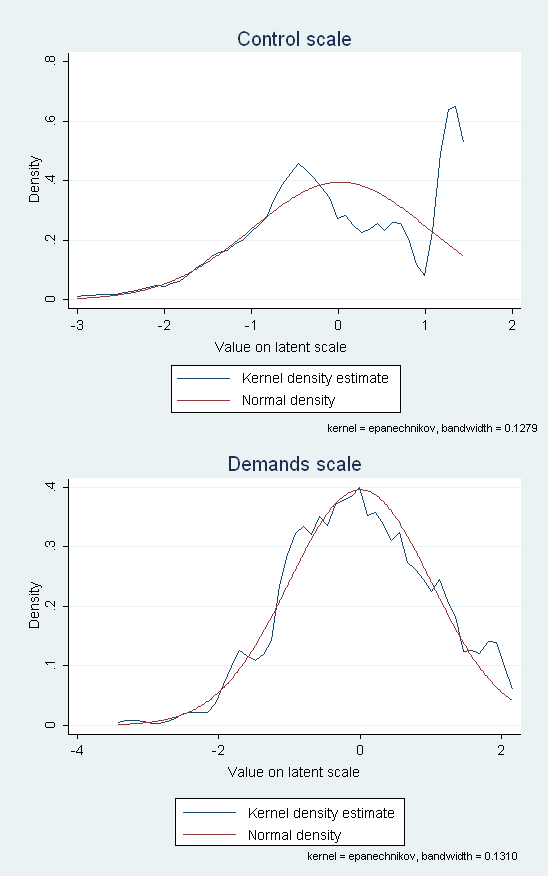


Table A7: Item Response Theory model for Demands

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | One-parameter model | | Two-parameter model | |
|  | Estimate | SE | Estimate | SE |
| *Difficulty parameter* |  |  |  |  |
| D1 work very hard |  |  |  |  |
| Strongly agree->Agree | 0.37 | 0.02 | 0.51 | 0.03 |
| Agree->Disagree | -2.11 | 0.03 | -2.69 | 0.06 |
| Disagree->Strongly disagree | -3.58 | 0.11 | -4.85 | 0.15 |
| D4 very high speed |  |  |  |  |
| All the time->Almost all the time | 1.31 | 0.04 | 0.77 | 0.04 |
| Almost all the time->Around ¾ of the time | 0.16 | 0.03 | -0.18 | 0.03 |
| Around ¾ of the time->Around ½ the time | 0.79 | 0.03 | 0.65 | 0.03 |
| Around ½ the time->Around ¼ of the time | -0.23 | 0.03 | -0.19 | 0.03 |
| Around ¼ of the time->Almost never | -0.21 | 0.03 | 0.04 | 0.03 |
| Almost never->Never | -1.15 | 0.03 | -0.68 | 0.03 |
| D12 great deal of tension |  |  |  |  |
| Strongly agree->Agree | 1.02 | 0.03 | 1.28 | 0.04 |
| Agree->Disagree | -0.08 | 0.02 | -0.14 | 0.02 |
| Disagree->Strongly disagree | -2.31 | 0.04 | -2.70 | 0.06 |
| *Discrimination parameter* |  |  |  |  |
| D1 work very hard | - | - | 1 | Fixed |
| D4 very high speed | - | - | 0.26 | 0.01 |
| D12 great deal of tension | - | - | 0.81 | 0.05 |
| *Log likelihood* | *-63,644* |  | *-63,353* |  |
| *n* | *15,972* |  | *15,972* |  |

The ‘difficulty parameters’ refer to the point at which the item is most sensitive to changes in latent demands – higher values of difficulty parameters here mean that the items respond most strongly to changes at higher levels of demands on the latent scale. For example, the cut-point between ‘disagree’ and ‘strongly disagree’ for D1 (‘work very hard’) is -3.58, whereas the cut-point between ‘strongly agree’ and ‘agree’ is higher at +0.37, showing that the former cut-point is primarily about low levels of demands, while the higher cut-point is about higher levels of demands (analogously to a more difficult exam question).

The ‘discrimination parameter’ is introduced in the two-parameter model. A high value on the discrimination parameter means that the variable is more sensitive to changes in the latent scale (i.e. it discriminates better between different levels of WLD). In other words, D1 (‘work very hard’) has a discrimination parameter of 1, and is therefore better at predicting latent demands than D4 (‘very high speed’) with a discrimination parameter of 0.26 ([see Bartholomew et al 2008](#_ENREF_3)).

Table A8: Item Response Theory model for Control

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | One-parameter model | | Two-parameter model | |
|  | Estimate | SE | Estimate | SE |
| *Difficulty parameter* |  |  |  |  |
| C1 Influence on how hard work |  |  |  |  |
| Strongly agree->Agree | -0.74 | 0.02 | -0.67 | 0.02 |
| Agree->Disagree | -3.09 | 0.04 | -2.58 | 0.05 |
| Disagree->Strongly disagree | -4.02 | 0.08 | -3.17 | 0.08 |
| C25 Influence on how to do task |  |  |  |  |
| A great deal->A fair amount | -0.21 | 0.02 | -0.19 | 0.03 |
| A fair amount->Not much | -2.40 | 0.03 | -3.57 | 0.08 |
| Not much->None at all | -2.89 | 0.05 | -4.96 | 0.14 |
| C33 Influence on what tasks to do |  |  |  |  |
| A great deal->A fair amount | 0.52 | 0.02 | 0.62 | 0.03 |
| A fair amount->Not much | -1.26 | 0.03 | -1.39 | 0.04 |
| Not much->None at all | -2.40 | 0.04 | -2.72 | 0.06 |
| C49 Influence on quality standards |  |  |  |  |
| A great deal->A fair amount | -0.73 | 0.02 | -0.70 | 0.02 |
| A fair amount->Not much | -2.34 | 0.04 | -1.92 | 0.04 |
| Not much->None at all | -2.75 | 0.05 | -2.07 | 0.06 |
| *Discrimination parameter* |  |  |  |  |
| C1 Influence on how hard work | - | - | 1 | Fixed |
| C25 Influence on how to do task | - | - | 2.38 | 0.09 |
| C33 Influence on what tasks to do | - | - | 1.55 | 0.05 |
| C49 Influence on quality standards | - | - | 1.00 | 0.03 |
| *Log likelihood* | *-69,150* |  | *-68,700* |  |
| *n* | *18,059* |  | *18,059* |  |

For explanation of ‘difficulty’ and ‘discrimination’ parameters, see Table A7.

## Chapter 3 Appendices

### Appendix 3A: Construction of WLD scale

As described in Chapter 3, I used item response theory to estimate the level of latent WLD underlying the eight individual WLD questions ([Bartholomew et al 2008](#_ENREF_3)). This was conducted using the GLLAMM commands in Stata; further detail on IRT modelling is provided in Appendix 2B. Data from all four waves containing WLD data (waves 3-6) were used,[[3]](#footnote-3) although no adjustment was made for the clustering of observations in particular people. The parameters of the IRT model are presented below.

Table A9: Item Response Theory model for WLD

|  |  |  |
| --- | --- | --- |
|  | Two-parameter model | |
|  | Estimate | SE |
| *Difficulty parameter* |  |  |
| WLD: mh-cut time | -3.22 | 0.04 |
| WLD: mh-do less | -2.07 | 0.03 |
| WLD: mh-careless | -2.81 | 0.03 |
| WLD: ph-cut time | -6.48 | 0.15 |
| WLD: ph-do less | -5.23 | 0.13 |
| WLD: ph-limit type | -6.01 | 0.14 |
| WLD: ph-difficulty | -5.67 | 0.13 |
| WLD pain (mild or above) | -1.28 | 0.02 |
| *Discrimination parameter* |  |  |
| WLD: mh-cut time | 1 | Fixed |
| WLD: mh-do less | 0.83 | 0.02 |
| WLD: mh-careless | 0.86 | 0.02 |
| WLD: ph-cut time | 2.73 | 0.09 |
| WLD: ph-do less | 3.03 | 0.10 |
| WLD: ph-limit type | 2.80 | 0.09 |
| WLD: ph-difficulty | 2.87 | 0.09 |
| WLD pain (mild or above) | 0.96 | 0.03 |
| *Log likelihood* | *-75260* |  |
| *n* | *29,502* |  |

Higher values on the ‘difficulty parameters’ refer to the point at which the item is most sensitive to changes in latent WLD – higher values of difficulty parameters here mean that the items respond most strongly to changes at higher levels of WLD on the latent scale. For example, WLD: mh-cut time has a cut point of -3.22, while WLD: ph-limit type has a cut point of -6.01, showing that the former cut-point is differentiating between higher levels of WLD than the former (analogously to a more difficult exam question).

A high value on the ‘discrimination parameter’ means that the variable is more sensitive to changes in the latent scale (i.e. it discriminates better between different levels of WLD). In other words, the mental health variables (discrimination parameters 0.83-1) are less strongly associated with the latent scale than the physical health vairables (2.73-3.03) ([see also Bartholomew et al 2008](#_ENREF_3)).

### Appendix 3B: Description of control variables

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** | **Waves  available** |
| ***Demographics and administrative*** | | |
| Lag to next wave | Length of follow-up to following wave, in years *Where this was not supplied in the data, it was estimated as: (i) the difference in ages between successive waves; or (ii) if age was not available, assuming that missing questionnaires were completed at the mid-point for that wave. (This procedure is used by the WII team).* | 3-5 |
| *Wave dummies*  Wave 3  Wave 4  Wave 5  Wave 6 | WII study wave, numbered according to the WII sweeps. *The central year for each wave was 1991 (wave 3), 1995 (wave 4), 1997 (wave 5) and 2001 (wave 6).* | 3-6 |
| *Age dummies*  Age 39-45  Age 45-50  Age 50-55  Age 55-60  Age 60-65 | Age group  *Where age was not supplied in the data, it was estimated as the sum of initial age and lag to following wave, as described above.* | 3-6 |
| Male | Gender (dummy where Male=1, Female=0) | 3-6 |
| Married | Marital status (dummy where married or cohabiting=1, Other=0) | 3-6 |

|  |  |  |
| --- | --- | --- |
| ***SES*** | | |
| **Label** | **Description** | **Waves** |
| *Last Civil Service Grade* Last grade: Clerical/Support Last grade: Prof/Exec Last grade: Administrative | Civil Service grade *The initial 12 non-industrial grades have been recoded by the Whitehall II team into administrative (unified grades 1-7), professional & executive (SEO, HEO and EO), and clerical & support (clerical officer and clerical assistant) grades. For this study, individuals working outside the Civil Service have been assigned their last Civil Service grade recorded in the data.* | 3-6 |
| *Class*Non-CS job: RG1/2 Non-CS job: RG3/6 [Reference category: CS job] | Registrar General's social class for those that have left the Civil Service *Note that this is not applicable to models that predict retirement, as by definition all of those who report a social class have already left the Civil Service.* | 3,5,6 |
| *Financial problems* Finance problems: none Finance problems: slight Finance problems: moderate/high | Financial problems, derived from two questions (on not having enough money for food/clothing that 'you should have', and difficulty in paying bills) - *Respondents were classified as having 'moderate/high' financial problems if they said they always/often/sometimes (rather than seldom/never) did not have enough money for suitable food/clothing, AND they had very great/great/some (rather very little/no) difficulty in paying bills.  - Respondents were classified as 'none' if they said that it never happened that they did not have enough money for suitable food/clothing AND they had very little or no diffiuclties in paying their bills. - Remaining respondents were classified as having 'slight' problems.* | 3-6 |
| Housing problems | Reported housing problems (e.g. too small, repairs, damp) *Respondents were classified as having housing problems if they had very great/great/some problems with housing (rather than slight/very little/none).* | 3-6 |

|  |  |  |
| --- | --- | --- |
| ***Job characteristics (excluding demands and control)*** | | |
| **Label** | **Description** | **Waves** |
| *Job satisfaction* JobSat: Dissatisfied JobSat: Satisfied JobSat: V Satisfied | Job satisfaction. *Note that this variable is constructed from two differently worded questions: - Wave 3: "All things considered, how satisfied or dissatisfied are you with the following areas of your life? ...(d) your job [responses on a 7-point scale]; - Wave 5: "About your job in general. How satisfied have you been with the following? ...(h) your job as a whole taking everything into consideration" [responses on a 4-point scale]  The different response categories were then converted to similar distributions, on the assumption that (i) the question meaning was effectively identical at each wave; and (ii) that the distribution of responses should be similar at waves 3 and 5. On these assumptions, this derived variable was created such that: - Dissatisfied = 'very/moderately/slightly dissatisfied 'at wave 3 / dissatisfied/very dissatisfied at wave 5 - Satisfied = 'no feelings' or 'slightly/moderately satisfied' at wave 3 / 'satisfied' at wave 5 - Very satisfied = 'very satisfied' at wave 3 / 'very satisfied' at wave 5* | 3, 5 |
| *Social support* Support: Low Support: Moderate Support: High | Social support at work *This is based on a standard WII scale consisting of six individual questions: How often do you get help and support from your colleagues? How often are your colleagues willing to listen to your work related problems? How often do you get help and support from your immediate superior? How often is your immediate superior willing to listen to your problems? Do you get sufficient information from line management (your superiors)? Do you get consistent information from line management (your superiors)?  This variable was supplied by the WII team to the author in the form of a 0-100 scale score, which was simply the sum of the individual item scores re-scaled to 0-100. This was then converted into three equal tertiles.* | 3,5 |
| *Variety and skill use* Variety: Low Variety: Moderate Variety: High | Variety and skills use at work. *This is the 'skill discretion' sub-scale of the Karasek 'decision authority' scale. It consists of six questions: Do you have to do the same thing over and over again? Does your job provide you with a variety of interesting things? Is your job boring? Do you have the possibility of learning new things through your work? Does your work demand a high level of skill or expertise? Does your job require you to take the initiative?   This variable was supplied by the WII team to the author in the form of a 0-100 scale score, which was simply the sum of the individual item scores re-scaled to 0-100. This was then converted into three equal tertiles.* | 3,5 |

|  |  |  |
| --- | --- | --- |
| ***Longstanding illness*** | | |
| **Label** | **Description** | **Waves** |
| LSI | Longstanding illness *At each wave, respondents were asked if they had a longstanding illness (LSI), although the wording varied ('longstanding illness, disability or infirmity' at waves 3 and 4, 'longstanding illnesses, diseases or medical conditions for which you have sought treatment in the last 12 months' at waves 5 and 6).  Respondents were then asked to describe these LSIs in their own words. These have been coded at a fine-grained level by the WII team, but using two different coding schemes (a specially constructed WII coding scheme for waves 3-5, and ICD-10 for wave 6). LSI does not appear to have been coded further, nor used in any WII publications.  For the purposes of this study, I have developed a broader coding scheme that enabled me to consistently code LSIs across all four waves, and which also provided large enough groups for analysis. The individual dummy variables for each LSI are described further below. An Excel table detailing the specific matching between the three coding schemes is available from the author.* | 3-6 |
| LSI Brain (oth+stroke) | Longstanding illlness: brain problems, including stroke but excluding migraines | 3-6 |
| LSI Cancer | Longstanding illness: cancer | 3-6 |
| LSI Depression | Longstanding illness: depression | 3-6 |
| LSI Ear | Longstanding illness: ear problems | 3-6 |
| LSI Eye | Longstanding illness: eye problems | 3-6 |
| LSI Gastro | Longstanding illness: digestive or excretory system problems | 3-6 |
| LSI Genitourinary | Longstanding illlness: genitourinary (reproductive system/prostate) problems | 3-6 |
| LSI Heart (maj+angina) | Longstanding illness: major heart problems, including angina *Respondents were classified as having a major heart problem if they EITHER reported angina in a separate, specific question, OR if they described their longstanding illness as a major heart problem: - Separate question: this is a derived variable supplied by the WII team for angina pectoris (including pains only in the left chest) - Longstanding illness: major heart conditions included valve disease, heart attack, ischaemic heart disease, heart failure, coronary artery bypass graft, atrial filibration, and heart problems not otherwise specified.* | 3-6 |
| LSI Heart (minor) | Longstanding illness: minor heart problems *This included hypertension, circulation diseases, haemorrhoids, high cholesterol, and varicose veins.* | 3-6 |
| LSI Infect disease | Longstanding illness: infectious diseases | 3-6 |
| LSI MH (oth+addiction) | Longstanding illness: mental health problems, excluding depression | 3-6 |
| LSI Migraines | Longstanding illness: migraines | 3-6 |
| LSI Musculo | Longstanding illness: musculoskeletal problems | 3-6 |
| LSI Resp (all) | Longstanding illness: respiratory problems | 3-6 |
| LSI Skin/allergy | Longstanding illness: skin or allergy problems | 3-6 |
| LSI Other+Blood | Longstanding illness: longstanding illnesses not otherwise classified, including blood disorders | 3-6 |
| ***Health (other)*** | | |
| **Label** | **Description** | **Waves** |
| Diabetes | Diabetes *Binary derived variable supplied by the WII team, based on self-reported diabetes or self-reported diabetes medication.* | 3-6 |
| Resp illness | Respiratory illness *Binary derived variable supplied by the WII team, based on three questions that make up part of the MRC chronic bronchitis questionnaire (bringing up phlegm from your chest first thing in the morning in winter, bringing up phlegm in the morning on most days for three months in winter, increased cough/phlegm for 3+ weeks in past 3 years).* | 3, 5 |
| *Bodily pain* Pain: None Pain: Very mild Pain: Mild Pain: Moderate to severe | Bodily pain within the past 4 weeks *Recoded from the original five response categories, with 'moderate', 'severe' and 'very severe' being combined.* | 3-6 |
| Phys functioning  (scale SF36) | Physical functioning scale from SF-36 (0-100 range) *Derived variable supplied by the WII team, using the established scale from the Short-Form 36 (SF-36) questionnaire. This is based on 10 individual responses as to how far people's health limits them in the following activities: vigorous activities (e.g. running), moderate activities (e.g. moving a table), lifting/carrying groceries, climbing several flights of stairs, climbing one flight of stairs, bending/kneeling/stooping, walking >1 mile, walking 1/2 mile, walking 100 yards, bathing/dressing yourself.* | 3-6 |
| GHQ Caseness | Psychiatric morbidity, as measured through GHQ caseness *The General Health Questionnaire (GHQ-30) is a common 30-question scale that screens for psychiatric morbidity. For WII, a cut-off score of 5 is conventionally used for a 'case' of psychiatric morbidity, following the testing and validation in Stansfeld and Marmot 1992.* | 3, 5, 6 |
| MH (scale SF36) | Mental health scale from SF-36 (0-100 range) *Derived variable supplied by the WII team, using the established scale from the Short-Form 36 (SF-36) questionnaire. This is based on 5 individual responses as to how much of the time (i) 'have you been a very nervous person; (ii) have you felt so down in the dumps that nothing could cheer you up' (iii) have you felt calm and peaceful; (iv) have you felt downhearted and low; (v) have you been a happy person?' The 6 answer categories for each of the 5 questions are converted to a 0-100 scale, which was supplied by the WII team.* | 3-6 |

|  |  |  |
| --- | --- | --- |
| ***Variables used in sensitivity analyses only*** | | |
| ***Demographics*** | | |
| **Label** | **Description** | **Waves** |
| Age 39-50 | Age group (replaces Age 39-45 and Age 45-50) *Necessary for analyses on wave 5, where very few people are aged 35-45.* | 5 |
| *Children dummies*  Num children: 0  Num children: 1-2  Num children: 3+ | Number of own children, reported at wave 3 | 3 |
| ***SES*** | | |
| **Label** | **Description** | **Waves** |
| *Education* Education: NoQual/O-lev Education: Alev/HND Education: BA+ | Education, reported at wave 5:  1=No qualifications / O-levels / GCSEs 2=A-levels / HNDs 3=Degree or above | 5 |
| Council house | Respondent lives in accommodation that is rented from local authority, reported at wave 3 | 3 |
| Car available | Car or van normally avilable for own (or household) use, reported at wave 3 | 3 |
| *Assets* Assets: £100k+ Assets: 40-100k Assets: £<40k | Household assets (inc. house) | 5 |
| *Income* Income: <£20k [Low income] Income: £20-40k Income: £40-60k Income: £60+k | Household income *For a few sensitivity models that had separation problems, only the lowest income dummy has been used, and this has been labelled 'low income'.* | 5 |
| Partner working | Employment status of spouse/partner  *1 = Employed, either full-time or part-time;  0 = Not-employed or no partner.* | 5 |

|  |  |  |
| --- | --- | --- |
| ***Health/negative affectivity*** | | |
| **Label** | **Description** | **Waves** |
| *Negative affectivity* Negative affect: High(4-15) Negative affect: Mod(2-3) Negative affect: Low(0-1) | Negative affectvity *Negative affectivity (NA) is a derived variable supplied by the WII team, and is measured using a sub-scale of the Bradburn affect balance scale. The five questions ask: "During the past few weeks, how often have you felt: (i) very lonely or remote from other people; (ii) depressed or very unhappy; (iii) bored; (iv) so restless you couldn't sit long in a chair; (v) upset because someone criticized you.   Each question is scored 0-3, and the 0-15 score was supplied by the WII team. This was converted into three groups of roughly equal size for the purposes of analysis: low (a score of 0-1), moderate (2-3) and high (4-15).  Only 75% of respondents were given negative affectivity questions at wave 1, while all were asked at wave 2.* | 1+2 *(see left)* |
| Vitality (scale SF36) | Vitality scale from SF-36 (0-100 range) *Derived variable supplied by the WII team, using the established scale from the Short-Form 36 (SF-36) questionnaire. This is based on 4 individual responses as to how much of the time (i) 'did you feel full of life?; (ii) did you have a lot of energy; (iii) did you feel worn out; (iv) did you feel tired?' The 6 answer categories for each of the 4 questions are converted to a 0-100 scale, which was supplied by the WII team.* | 3-6 |
| *General health* Health: Excellent Health: V Good Health: Good Health: Fair/Poor | Self-reported general health *This is recoded from the original question such that 'fair' and 'poor' health are combined. This general health question is preferred to the SF-36 general health scale as some parts of the scale seem to relate more to optimism/pessimism than health states (e.g. ‘I expect my health to get worse’).* | 3-6 |
| Asthma | Ever been told by doctor that have asthma *Question was asked at wave 4, together with year of first diagnosis. Asthma at wave 3 is derived from this.* | 3-4 |
| Back-ache | Backache or pains in the back in last 14 days | 3 |
| Cancer | Ever been diagnosed as having cancer *Question was asked at wave 5, together with year of first diagnosis. Cancer at waves 3 and 4 is derived from this. In sensitivity analyses, this is used in place of LSI cancer.* | 3-5 |
| LSI Diabetes | Longstanding illness: diabetes *See LSI section above for details on how longstanding illness vaariables were derived. In sensitivity analysis this is used in place of the standalone diabetes question.* | 3-6 |
| Hearing problems | Hearing problems (binary) *Derived variable using thre individual questions on hearing problems (wearing a hearing aid, difficulty hearing someone in quiet room (with hearing aid), difficulty following conversation with background noise (with hearing aid)). Individuals reporting any hearing problem are scored 1, others are scored 0.* | 5 |
| Vision problems | Difficulty recognising a friend across the road | 5 |
| Osteoporosis | Ever been told by doctor that have osteoporosis, brittle bones or bone loss *Question was asked at wave 6, together with year of first diagnosis. Osteoporosis at waves 3-5 is derived from this.* | 3,5 |
| Osteo-arthritis | Ever been told by doctor that have osteoarthritis ('wear and tear' arthritis) *Question was asked at wave 4, together with year of first diagnosis. Osteoarthritis at wave 3 is derived from this.* | 3-4 |
| Rheumatoid arthritis | Ever been told by doctor that have rheumatoid arthritis *Question was asked at wave 4, together with year of first diagnosis. Rheumatoid arthritis at wave 3 is derived from this.* | 3-4 |
| LSI Heart (major) | Longstanding illness: major heart problems *For details see 'LSI Heart (maj+angina)' above. This version - excluding the separate question on angina - was used in sensitivity analyses alongside the angina variable.* | 3-6 |
| Angina | Angina pectoris *Derived variable supplied by the WII team for angina pectoris (including pains only in the left chest)* | 3-6 |
| High blood pressure | Ever been told by doctor that have high blood pressure (hypertension) *Question was asked at wave 5, together with year of first diagnosis. High blood pressure at waves 3-4 is derived from this.* | 3-5 |
| LSI Epilepsy | Longstanding illness: epilepsy *See LSI section above for details on how longstanding illness vaariables were derived. LSI Epilepsy was not used in the main analyses due to very low numbers and even lower rates of change.* | 3-6 |
| Epilepsy | Ever been told by doctor that have epilepsy (fits or convulsions) *Question was asked at wave 4, together with year of first diagnosis. High blood pressure at wave 3 is derived from this.* | 3-4 |
| LSI Brain (other) | Longstanding illness: brain problems, excluding stroke and migraines *See LSI section above for details on how longstanding illness vaariables were derived.* |  |
| Stroke/TIA | Ever been told by doctor that have stroke or Transient Ischaemic Attack (TIA) *Stroke/TIA symptoms were asked in a series of questions (ever had sudden attack of weakness/numbness on one side of body, ever had sudden attack of slurred speech, ever had sudden of vision loss or blurred vision), and if responded positively were asked if doctor had diagnosed this as stroke, TIA (or 'mini stroke'), or other. Respondents reporting both symptoms and doctor diagnosis of stroke or TIA were treated as having had stroke/TIA.   Question was asked at wave 5, together with year of first diagnosis. Stroke/TIA at waves 3-4 is derived from this.* | 3,5 |
| LSI Blood | Longstanding illness: blood disorders *See LSI section above for details on how longstanding illness vaariables were derived.* | 3-6 |
| LSI Other | Longstanding illness: longstanding illnesses not otherwise classified, excluding blood disorders *See LSI section above for details on how longstanding illness vaariables were derived.* | 3-6 |

### Appendix 3C: Multiple Imputation

The sample for the multiple imputation analyses was anyone who reported their job strain at wave 3 or wave 5. For all analyses, those out of the age range, not working at baseline, known to die between waves or not reporting their job demands/control at one wave were excluded.

For the imputation I use multiple imputation by chained equations, which has been shown to perform well in simulations ([Kenward and Carpenter 2007](#_ENREF_19); [Carpenter and Kenward 2008](#_ENREF_8)). This was done with the ICE and MIM commands in Stata ([Royston 2009](#_ENREF_30); [Royston et al 2009](#_ENREF_32)) with 20 imputations and 20 cycles per imputation – in both cases above the recommendations from the help file ([Royston 2004](#_ENREF_27); [2005](#_ENREF_28)). Following the MIM estimates ([Royston et al 2009](#_ENREF_32)), the Monte Carlo error for this model was acceptable at around 4-8% of the size of the coefficients ([Royston 2004](#_ENREF_27)).

Ideally data would be imputed in ‘wide’ format to reflect the longitudinal structure of the data ([Rose and Fraser 2008](#_ENREF_26)); that is, job demands at wave 3 and job demands at wave 5 are treated as two completely different variables within the same person. However, practical difficulties in applying this to health retirement in Chapter 4 meant that imputation took place in ‘long’ format (i.e. the different baseline waves were treated as independent from one another). The imputation model used as many variables that were possible ([Royston 2007:461](#_ENREF_29); [Rose and Fraser 2008](#_ENREF_26)) without causing computational problems – including job satisfaction, social support and skill discretion.[[4]](#footnote-4) All covariates were also turned into categorical variables as ICE is sometimes claimed to perform poorly with non-normally distributed variables.

### Appendix 3D: Results of sensitivity analyses

Firstly, I used*different forms of demands, control and strain*[[5]](#footnote-5) to see if the results were sensitive to the particular way that I operationalised these variables (see Web Table 3.ConDem). In most cases the size of the effect was slightly reduced, and specifications with linear demands/control/strain suggested a role only for job demands; the term for the demands\*control interaction was non-significant. However, in all cases the contrast between high- and low-strain jobs was still significant at the 0.01% level, suggesting that the main conclusions above are robust to alternate ways of including job strain in the model.

Second, in the models above I use categorical forms of most continuous controls, and linear forms of the continuous SF-36 physical functioning and mental health scales. However, both categorising variables (through a loss of information) and linear forms (through misspecification) can lead to residual confounding ([Davey Smith and Phillips 1992](#_ENREF_9); [Fewell et al 2007](#_ENREF_13)). To test for this, I used a technique known as ‘multivariable fractional polynomial models’,[[6]](#footnote-6) which tests whether a parsimonious set of transformations of continuous variables offer any significant improvements in model fit. The differences between this and the main model were negligible, suggesting that there is no residual confounding due to functional form of continuous predictors.

Third, for the pathways models I checked the results using a related technique recently developed by Karlson, Holm and Breen ([Karlson and Holm 2011](#_ENREF_18)).[[7]](#footnote-7) This again shows that the job requirements pathway is much larger than the health pathway – but unlike the main results, the health pathway is not statistically significant.

Finally, I used multiple imputation to test for the importance of missing data as described in Appendix 3C, where the resulting models have a minimum of 10,010 observations.[[8]](#footnote-8) This shows that the main results are robust to missing data (assuming that data is missing at random conditional on the observed data).[[9]](#footnote-9)

## Chapter 4 Appendices

### Appendix 4A: Results of sensitivity analyses

Firstly, I tested whether different forms of demands, control and strain[[10]](#footnote-10) produced different effects. The results for linear forms of demands/control were the same as those in the main models (see Web Table 4.ConDem): the demands\*control interaction was marginally significant for health retirement but none of the other individual terms were significant, and nor was the contrast between high- and low-strain jobs. Furthermore, none of the coefficients were significant in the absence of an interaction term. The results are therefore robust to alternate ways of operationalising job strain.

Secondly, among the controls supplied as continuous variables, I have either used these as simple linear effects (SF-36 mental health and physical functioning scales) or categorised them (all other variables). In both cases this can lead to residual confounding, in the former case due to misspecification, and in the later due to measurement error from the loss of information. I therefore tested whether better-fitting covariates changed the results, using multivariable fractional polynomials (see previous chapter). For both forms of health-related job loss the linear forms of the variables were the best-fitting, hence the results are unchanged from those presented above.

Third, I tested whether my bespoke technique for testing mediation was replicated when using a recently-developed command designed for mediation analysis using binary variables ([Karlson and Holm 2011](#_ENREF_18)).[[11]](#footnote-11) For both health retirement and long-term sickness at the following wave, this technique found that the WLD pathway was negligible and non-significant – yet looking two waves into the future it was strongly significant in both cases, including when job satisfaction/support/variety were controlled for.

Given the divergent results, I also tested the mediation analyses with another recently developed binary mediation technique ([Erikson et al 2005](#_ENREF_11); [Buis 2010](#_ENREF_6)). This confirmed that the WLD pathway is significant for both health retirement and long-term sickness (the significance levels are different to the main analyses however: p<0.001 rather than p<0.10 for health retirement and p<0.10 rather than p<0.05 for long-term sickness).

Techniques for binary mediation analysis are an emerging area and it is unclear why the Karlson/Holm/Breen method gives different results to the other two. Given that there is some support for the WLD pathway using all three techniques – particularly when looking two waves into the future (Sequence B4) – I interpret this as providing some support for the existence of the WLD pathway (conditional on the assumptions set out in the Discussion).

Finally, to test whether the results are sensitive to missing data, I re-estimated the models following multiple imputation – although for computational reasons (relating to small numbers of cases), this was only possible for health retirement. Looking at the highest-powered analysis (health retirements *two* waves into the future, looking at baseline waves 1, 3 and 5), I find effectively the same results in the main and missing data analyses (high-strain vs. low-strain coefficient = 0.48 (p<0.05) for main analysis and 0.42 (p<0.05) for missing data analysis).[[12]](#footnote-12)

## Chapter 5 Appendices

### Appendix 5A: Construction of the demands/control scales

The scales of job demands and control used in Chapter 5 are similar to those in Chapter 2 (for details, see Appendix 2B). The main difference is that only the 2001 and 2006 Skills Surveys were used, which means that I could create scales based that include the following additional questions:

**Demands**

D15: How often does your work involve working to tight deadlines?

1. All the time 2. Almost all the time 3. Around three quarters of the time 4. Around half the time 5. Around quarter of the time 6. Almost never 7. Never

**Control**

C15: How true would you say each of the following statements is about your job? ‘My job allows me to take part in making decisions that affect my work'

1. Very True 2. True 3. Somewhat true 4. Not at all true

C23: How much choice do you have over the way in which you do your job…

1. a great deal of choice, 2. some choice, 3. hardly any choice, 4. or no choice at all?

This is in addition to the questions underlying the scales in Chapter 2:

**Demands**

D1: My job requires that I work very hard

1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree

D4: How often does your work involve working at very high speed?

1. All the time 2. Almost all the time 3. Around three quarters of the time 4. Around half the time 5. Around quarter of the time 6. Almost never 7. Never

D12: I work under a great deal of tension

1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree

**Control**

C1: How much influence do you personally have on how hard you work?

1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree

C2: Which, if any, of the things on this card are important in determining how hard you work in your job?

...Your own discretion

C25: (And how much influence do you personally have on … deciding how you are to do the task?

1. A great deal 2. A fair amount 3. Not much 4. None at all

C33: And how much influence do you personally have on… deciding what tasks you are to do?

1. A great deal 2. A fair amount 3. Not much 4. None at all

C49: And how much influence do you personally have on … deciding the quality standards to which you work?

1. A great deal 2. A fair amount 3. Not much 4. None at all

I again tested the extent to which these questions (together with the original 3 demands questions and 5 control questions) appeared to be measuring the same construct. As before, question C2 appeared to be measuring a different construct so was excluded. The conventional test of internal reliability is Cronbach’s alpha, which for the remaining questions produces estimates of 0.82 for control (for 6 questions) and 0.62 for demands (for 4 questions), compared to a conventional guideline of around 0.70-0.80 for a reliable scale. However, Cronbach’s alpha should not really be used for ordinal measures and tends to produce underestimates if it is used ([Zumbo et al 2007](#_ENREF_38)). If we instead use the more appropriate ordinal theta measure, then we find values of 0.88 for control and 0.71 for demands, which are acceptable.

The scales were then constructed using Item Response Theory (IRT) models in identical fashion as in Chapter 2. The IRT models produced scales that were very similar to those using conventional Principal Components Analysis; for example the correlation of Green’s method with the two-parameter IRT model was 0.99 for demands and 0.97 for control. The parameters of the IRT models are presented in the following tables.

Table A10: Item Response Theory model for Demands

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | One-parameter model | | Two-parameter model | |
|  | Estimate | SE | Estimate | SE |
| *Difficulty parameter* |  |  |  |  |
| D1 work very hard |  |  |  |  |
| Strongly agree->Agree | 0.24 | 0.02 | 0.29 | 0.02 |
| Agree->Disagree | -2.03 | 0.03 | -2.27 | 0.05 |
| Disagree->Strongly disagree | -3.57 | 0.13 | -4.09 | 0.14 |
| D4 very high speed |  |  |  |  |
| All the time->Almost all the time | 1.23 | 0.04 | 1.22 | 0.05 |
| Almost all the time->Around ¾ of the time | 0.10 | 0.04 | 0.09 | 0.04 |
| Around ¾ of the time->Around ½ the time | 0.68 | 0.03 | 0.68 | 0.03 |
| Around ½ the time->Around ¼ of the time | -0.37 | 0.03 | -0.37 | 0.03 |
| Around ¼ of the time->Almost never | -0.34 | 0.03 | -0.33 | 0.04 |
| Almost never->Never | -1.49 | 0.04 | -1.48 | 0.05 |
| D12 great deal of tension |  |  |  |  |
| Strongly agree->Agree | 0.91 | 0.03 | 1.00 | 0.03 |
| Agree->Disagree | -0.14 | 0.02 | -0.17 | 0.02 |
| Disagree->Strongly disagree | -2.28 | 0.04 | -2.42 | 0.05 |
| D15 tight deadlines |  |  |  |  |
| All the time->Almost all the time | 0.60 | 0.03 | 0.42 | 0.03 |
| Almost all the time->Around ¾ of the time | -0.57 | 0.04 | -0.64 | 0.04 |
| Around ¾ of the time->Around ½ the time | 0.38 | 0.04 | 0.40 | 0.04 |
| Around ½ the time->Around ¼ of the time | -0.51 | 0.04 | -0.41 | 0.04 |
| Around ¼ of the time->Almost never | -0.43 | 0.04 | -0.26 | 0.04 |
| Almost never->Never | -1.23 | 0.04 | -0.95 | 0.05 |
| *Discrimination parameter* |  |  |  |  |
| D1 work very hard | - | - | 1 | Fixed |
| D4 very high speed | - | - | 0.72 | 0.04 |
| D12 great deal of tension | - | - | 0.86 | 0.03 |
| D15 tight deadlines | - | - | 0.52 | 0.03 |
| *Log likelihood* | *-69,614* |  | *-69,529* |  |
| *n* | *12,172* |  | *12,172* |  |

Higher values on the ‘difficulty parameters’ mean that these categories are associated with higher levels of demands on the latent scale. For example, for D1 (work very hard), the cut point between strongly agree and agree is at -0.24 on the latent scale, while the cut point between disagree and strongly disagree is at -3.5, showing that the former variable is associated with higher demands.

A high value on the ‘discrimination parameter’ means that the variable is more strongly associated with the latent scale, i.e. it discriminates more between high and low demands. D1 (work hard, discrimination parameter=1) is therefore more strongly associated with latent demands than D15 (tight deadlines, discrimination parameter=0.52) ([see also Bartholomew et al 2008](#_ENREF_3)).

Table A11: Item Response Theory model for Control

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | One-parameter model | | Two-parameter model | |
|  | Estimate | SE | Estimate | SE |
| *Difficulty parameter* |  |  |  |  |
| C1 Influence on how hard work |  |  |  |  |
| Strongly agree->Agree | 0.45 | 0.03 | 0.44 | 0.02 |
| Agree->Disagree | 2.74 | 0.04 | 2.49 | 0.05 |
| Disagree->Strongly disagree | 3.62 | 0.09 | 3.18 | 0.10 |
| C15 Allowed to take part in decisions |  |  |  |  |
| A great deal->A fair amount | -0.59 | 0.03 | -0.43 | 0.03 |
| A fair amount->Not much | 1.19 | 0.03 | 1.01 | 0.03 |
| Not much->None at all | 1.94 | 0.04 | 1.50 | 0.05 |
| C23 Influence over way do job |  |  |  |  |
| A great deal->A fair amount | -0.33 | 0.03 | -0.31 | 0.03 |
| A fair amount->Not much | 2.17 | 0.04 | 2.10 | 0.04 |
| Not much->None at all | 2.28 | 0.05 | 2.16 | 0.07 |
| C25 Influence on how to do task |  |  |  |  |
| A great deal->A fair amount | 0.08 | 0.03 | -0.03 | 0.04 |
| A fair amount->Not much | 2.10 | 0.04 | 3.27 | 0.08 |
| Not much->None at all | 2.55 | 0.06 | 4.71 | 0.13 |
| C33 Influence on what tasks to do |  |  |  |  |
| A great deal->A fair amount | -0.54 | 0.03 | -0.74 | 0.04 |
| A fair amount->Not much | 1.02 | 0.03 | 1.23 | 0.04 |
| Not much->None at all | 2.11 | 0.04 | 2.67 | 0.07 |
| C49 Influence on quality standards |  |  |  |  |
| A great deal->A fair amount | 0.57 | 0.03 | 0.57 | 0.02 |
| A fair amount->Not much | 2.04 | 0.04 | 1.75 | 0.04 |
| Not much->None at all | 2.38 | 0.06 | 1.89 | 0.07 |
| *Discrimination parameter* |  |  |  |  |
| C1 Influence on how hard work | - | - | 1 | Fixed |
| C15 Allowed to take part in decisions | - | - | 0.89 | 0.03 |
| C23 Influence over way do job | - | - | 1.12 | 0.04 |
| C25 Influence on how to do task | - | - | 2.29 | 0.08 |
| C33 Influence on what tasks to do | - | - | 1.56 | 0.05 |
| C49 Influence on quality standards | - | - | 0.93 | 0.03 |
| *Log likelihood* | *-73203* |  | *-72653* |  |
| *n* | *12,198* |  | *12,198* |  |

For details on ‘difficulty’ vs. ‘discrimination’ parameters, see Table A10.

### Appendix 5B: Empirical Bayes estimates

The general rationale for Empirical Bayes estimation is given in the main body of Chapter 5. More formally, I am using a random intercept model ([Rabe-Hesketh and Skrondal 2008](#_ENREF_24)) for each of job demands and job control separately:

...where is the level of demands/control for person i in occupation j, arethe coefficients on the dummy variables for each SOC major group k (of 9 total groups), is the random effect for the detailed occupation j, and is the residual individual error term.

The random effects are therefore the deviations of the occupational averages from the fixed part of the model. To obtain Empirical Bayes estimates, we then specify our prior assumptions about (that the random effect is normally distributed with mean 0, and with the observed variance of the random effect – with this ‘empirical’ estimate of the prior being the reason that this is ‘empirical Bayes’ (rather than full Bayesian) estimation). We then combine this prior distribution with the likelihood to get the posterior distribution of given the observed values of y ([Rabe-Hesketh and Skrondal 2008:80](#_ENREF_24)). The EB estimate for each occupation is simply the mean of the posterior distribution with the parameter estimates plugged in.

In practice this was done by using the GLLAMM command to estimate the random effects model, and using GLLAPRED to obtain the Empirical Bayes estimates. Unweighted data were used to ease the computational burden; tests suggested weights would have little impact on the results.

#### An example

To show how the EB estimates work in practice, we can compare the conventional and EB estimates using the dataset in Chapter 5 for a single occupation:[[13]](#footnote-13) Restaurant and Catering Managers:

* The pooled 2001-6 dataset contains only 47 restaurant and catering managers. This compares to 1,647 people in the occupational major group *‘managers and senior officials’*.
* Based on these 47 people, the conventional model estimates that job demands among restaurant and catering managers is +0.87 – a very high level, and far higher than the +0.25 among the 1,647 managers and senior officials as a whole. This might be because job demands is genuinely very high in this group, but given the small number of respondents it may instead reflect sampling error.
* The Empirical Bayes estimate is +0.72 – a compromise between the precise estimate of job demands among managers and senior officials (+0.25) and the unbiased but imprecise estimate among restaurant and catering managers (+0.87).
* If we can compare the confidence intervals for the two estimates, we can see that the EB confidence interval is smaller than the conventional one – the size of the interval is 0.42 (from 0.47 to 0.89) rather than 0.56 (from 0.59 to 1.15).

Looking across all the occupations, the confidence intervals for the EB estimates are 12% smaller than the conventional estimates for job control, and 18% smaller than conventional estimates for job demands. Moreover, for control the improved precision is particularly great for smaller occupations; confidence intervals were reduced by 10-40% in occupations where n<40, while EB estimates offered no net benefit for occupations where n>100. Another advantage is that particularly high/low estimates are more likely to be meaningful, rather than simply the result of random sampling error in smaller occupations. For these reasons, the occupational map in this chapter uses EB estimates.

## Chapter 6 Appendices

### Appendix 6A: Control variables used in BHPS

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Description** | | **Waves  available** |
| ***Demographics and administrative*** | | | |
| *Wave dummies* Each year of 1992-2006 | | BHPS study wave | All |
| *Age dummies* Five-year bands from 20-24 to 60-64 | | Age group at date of interview | All |
| Male | | Gender (dummy where Male=1, Female=0) | All |
| Minority ethnic group | | Black and minority ethnic group membership *Based on wRACE (1991-2002) and wRACEL (2003-2008). Binary indicator of white vs. others due to small sample sizes individual ethnic groups and incomparability across years (for wRACEL, this is 'white British', 'white Irish', 'white Welsh', 'white Scottish' and 'other white background' for comparability with the 'white' category in wRACE).  Previous research suggests that certain ethnic minorities show elevated job stress after controlling for demands-control (*[*Smith et al 2005*](#_ENREF_35)*). In a sensitivity analysis, I therefore used a five-banded version of ethnicity:* ***White*** *(wRACE=White; wRACEL=White British, White Irish, White Welsh, White Scottish, or other White background),* ***Black*** *(wRACE=Black Carribean, Black African,or Black Other; wRACEL=mixed White and Black Carribean, mixed White and Black African, Black/British Carribean, Black/British African),* ***Indian*** *(wRACE=Indian, wRACEL=Asian/British Indian),* ***Pakistani/Bangladeshi*** *(wRACE=Pakistani, Bangladeshi; wRACEL=Asian/British Pakistani, Asian/British Bangladeshi) and* ***Other****. The choice of these groups was primarily determined by sample size considerations rather than any a priori theory. In these sensitivity analyses I see no difference in the estimated effects.* | All |
| *Marital status* Married Separated/divorced /widowed Never married | | Marital status, grouped *From wMLSTAT (self-reported marital status). In later waves, this includes 'in a civil partnership' (under 'married') and 'have a dissolved civil partnership' (under 'separated/divorced/widowed')* | All |
| *Children* | | Number of own children in household, grouped *From wNCHILD, with 3-9 children grouped into 3+* | All |
| *Region* | | Region of residence *From wREGION2, Government Office Region: 1=North East | 2=North West | 3=Yorkshire & Humber | 4=East Midlands | 5=West Midlands | 6=East of England | 7=London | 8=South East | 9=South West | 10=Wales | 11=Scotland* | All |

|  |  |  |  |
| --- | --- | --- | --- |
| ***SES*** | | | |
| **Label** | **Description** | | **Waves** |
| *Education* No qualifications Level 1 Level 2 Level 3 Level 4 - Other Level 4 - Degree | Highest educational qualification, grouped *Recoded version of wQFEDHI; see BHPS User Guide:253 for details. This is then recoded as follows: - Level 4 - Degree: University or higher degree - Level 4 - Other: teaching qualifications; nursing qualifications; any other technical, professional or higher qualifications - Level 3: A levels or recognised trade apprenticeship - Level 2: GCSE grades A-C; clerical or commercial qualifications - Level 1: GCSE grades D-G; youth training certificate; any other qualifications* | | All |
| Log household income | Log of net household income *From wHHNETDE2, supplied by BHPS team in supplementary dataset based on HBAI definitions; see Jenkins 2010. Household net income is equivalised using the Modified OECD scale, and deflated to Jan 2008 prices* | | 1991- 2006 |
| *Perceived financial situation* Doing alright Getting by Finding it difficult | Financial situation, regrouped *From wFISIT, "How well would you say you yourself are managing financially these days?" This is then recoded as follows: - Doing alright = 'Living comfortably' or 'Doing alright' - Getting by = 'Just about getting by' - Finding it difficult = 'Finding it quite difficult' or 'Finding it very difficult'* | | All |
| *Tenure* Owned Social housing Other rented housing | Housing tenure, recoded *From wTENURE, recoded as follows: - Owned = owned outright or owned with mortgage - Social housing = Local Authority rent or Housing Association renter - Other rented housing = all other options* | | All |
| ***Other work characteristics*** | | | |
| **Label** | **Description** | | **Waves** |
| Occupation | Standard Occupational Classification 1990 *Based on wJBSOC* | | All |
| Industry | Standard Industrial Classification 1992 *wJBSIC92 is available in 1994, 1997, and 2001-2008. For other years, I recoded SIC80 (wJBSIC) using the BHPS dual-coded data in 1994, 1997 and 2001. SIC80 codes were assigned a SIC92 code based on the most frequent SIC92 code in the data. The person-weighted reliability of the recoding in the dual-coded BHPS data is 94.3%.* | | All |
| *Sector* Private sector Public sector Non-profit and other | Sector of work *From wJOBSECT, recoded as follows: - Private sector = 'private firm/company' - Public sector = 'Civil Service/Central Govt', 'Local govt/town hall', 'NHS or higher educ', 'nationalised industry' or 'armed forces' - Non-profit and other = 'non-profit orgs' or 'other'* | | All |
| Self-employed | Self-employment status *From wJBSEMP. Self-employed respondents were excluded from all analyses* | | All |
| Temporary job | Contractual status, permanent vs. temporary *From wJBTERM (1991-1998) and wJBTERM1 (1999-2008). wJBTERM was recoded such that a temporary job = 'seasonal/tmp job' or 'contract/fixed time'; for wJBTERM1 a temporary job = 'non-permanent job'* | | All |
| *Size of workplace* Small Medium Large | Size of workplace *From wJBSIZE, recoded into <25 employees (small), 25-199 employees (medium), and 200+ employees (large)* | | All |
| *Hours of work* <16hrs 16-29hrs 30-45hrs >45hrs | Total hours of work per week, including overtime *Sum of wJBHRS and wJBOT, recoded into bands* | | All |
| Occupational pension | Member of employer's occupational pension scheme *Coded as 1 if respondent reports being a member of employer's pension scheme (wJBPENM), and 0 if either not a member, employer doesn't run an pension scheme, or respondent is uncertain if employer runs a scheme (wJBPEN)* | | All |
| ***Health*** | | | |
| **Label** | **Description** | **Waves** | |
| GHQ caseness | Minor psychistric morbidity (GHQ caseness) *Based on GHQ-12 score (wHLGHQ2), with scores of 3+ being considered a 'case' (following the recommendations of Goldberg et al 2000,* Psychological Medicine *28:915-921). The original scale score is used in a sensitivity analysis.* | All | |
| Musculoskeletal problem | Problem with arms, legs, hands, feet or back *Based on wHLPRBA, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Vision problem | Difficulty in seeing *Based on wHLPRBB, part of the list of problem types. Respondents were asked to exclude temporary conditions. Excludes needing glasses to read normal size print.* | All | |
| Hearing problem | Difficulty in hearing *Based on wHLPRBC, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Allergy problem | Problem with skin/allergies *Based on wHLPRBD, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Breathing problem | Problem with chest/breathing, asthma or bronchitis *Based on wHLPRBE, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Heart problem | Problem with heart, circulation, or high blood pressure *Based on wHLPRBF, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Digestive problems | Problem with stomach, liver, kidneys, or digestion *Based on wHLPRBG, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Diabetes | Diabetes *Based on wHLPRBH, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Anxiety/depression | Anxiety, depression, or psychiatric problems *Based on wHLPRBI, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Addiction problems | Alcohol or drug-related problems *Based on wHLPRBJ, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Epilepsy | Epilepsy *Based on wHLPRBK, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Migraine | Migraine or frequent headaches *Based on wHLPRBL, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Other health problem | Other health problems *Based on wHLPRBM, part of the list of problem types. Respondents were asked to exclude temporary conditions.* | All | |
| Activity limitation: housework | Limitations in doing the housework *From wHLLTA, asked if respondents have an LLSI (those without LLSI are coded as 0).* | All bar 1999 and 2004 | |
| Activity limitation:  climbing stairs | Limitations in climbing stairs *From wHLLTB, asked if respondents have an LLSI (those without LLSI are coded as 0).* | All bar 1999 and 2004 | |
| Activity limitation: dressing | Limitations in dressing yourself *From wHLLTC, asked if respondents have an LLSI (those without LLSI are coded as 0).* | All bar 1999 and 2004 | |
| Activity limitation: walking | Limitations in walking for at least 10 minutes *From wHLLTD, asked if respondents have an LLSI (those without LLSI are coded as 0).* | All bar 1999 and 2004 | |
| Hospital in-patient | Hospital in-patient in past year *From wHOSP and wHOSPCH, including all hospital visits except those for childbirth* | All waves | |
| ***Partner's characteristics*** | | | |
| **Label** | **Description** | | **Waves** |
| Partnered | Whether has partner in household | | All |
| Partner's WLD | Whether partner reports a WLD *See under WLD in text for details of WLD measure. Respondents without partners coded to zero; respondents whose partners do not respond are coded missing* | | All |
| Partner's employment | Whether partner is working *Derived from wJBHAS (whether did any paid work in the past week) and wJBOFF (whether had a job that was away from, even though wasn't working in the past week). Respondents without partners coded to zero; respondents whose partners do not respond are coded missing* | | All |
| Partner's IB receipt | Whether partner claims an incapacity benefit *See under IB in text for details of IB measure. Respondents without partners coded to zero; respondents whose partners do not respond are coded missing* | | All |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Variables used in sensitivity analyses only*** | | | |
| ***Other work characteristics*** | | | |
| **Label** | | **Description** | **Waves** |
| Cambridge scale | | Occupation-based Cambridge score of social distance *Social distance score (wJBCSSF and wJBCSSM) using marriage/friendship patterns across occupations (see Prandy 1990,* Sociology *24:629-655). In some analyses this is used in gender-disaggregated form, in others this is pooled across both genders (sometimes as a continuous variable, and at other times grouped into four categories).* | All |
| CAMSIS scale | | Occupation-based CAMSIS score of social distance *The updated 1991 social distance score imported from CAMSIS (http://www.camsis.stir.ac.uk/) using marriage/friendship patterns across occupations (an update to the Cambridge scale score.* | All |
| Chan-Goldthorpe scale | | Occupation-based Chan-Goldthorpe stats score  *Occupational status score using friendship patterns across occupations (see Chan & Goldthorpe 2004, 2007). This is imported into BHPS using the occupation scores given in Chan & Goldthorpe 2004* | All |
| NS-SEC (7-band) | | Social class using National Statistics Socio-Economic Classification *Based on wJBSEC, recoded following Chan & Goldthorpe 2007 as follows: 1 Higher managers and professionals 2 Lower managers and professionals 3 Intermediate employees 4 Small employers and own account workers 5 Lower supervisors and technicians 6 Semi-routine workers 7 Routine workers* | All |
| NS-SEC (3-band) | | Grouped social class using National Statistics Socio-Economic Classification *Based on Rose & Pevalin (2005:37), I grouped the NS-SEC into managerial & professional occupations (1 & 2 from NS-SEC 7 band), intermediate occupations (3 & 4), and routine and manual occupations (5-7).* | All |
| *Job satisfaction* Dissatisfied Neither satisfied /dissatisfied Satisfied Completely satisfied | | Satisfaction with present job overall *Based on wJBSAT, recoded from the original 7-point scale as follows: - 1/3 = Dissatisfied - 4/5 = Neither satisfied nor dissatisfied - 6 = Satisfied - 7 = Completely satisfied Note that the labelling of the response options changed during the lifetime of the BHPS, with significant effects on the distribution of the results (see ISER Working paper 2008-39).* | All |
| *Satisfaction with security* Dissatisfied Neither satisfied /dissatisfied Satisfied Completely satisfied | | Satisfaction with job security *Based on wJBSAT4, with same coding as job satisfaction* | All |
| *Satisfaction with pay* Dissatisfied Neither satisfied /dissatisfied Satisfied | | Satisfaction with relations with supervisor/manager *Based on wJBSAT2, recoded from the original 7-point scale as follows: - 1/3 = Dissatisfied - 4/5 = Neither satisfied nor dissatisfied - 6/7 = Satisfied* | All |
| *Satisfaction with boss* Dissatisfied Neither satisfied /dissatisfied Satisfied Completely satisfied | | Satisfaction with relations with supervisor/manager *Based on wJBSAT3, with same coding as job satisfaction* | 1991- 1997 |
| *Satisfaction with initiative* Dissatisfied Neither satisfied /dissatisfied Satisfied Completely satisfied | | Satisfaction with being able to use own initiative *Based on wJBSAT5, with same coding as job satisfaction* | 1991- 1997 |
| *Managerial duties* Manager Foreman/supervisor Not manager /supervisor | | Managerial or supervisory duties *Based on wJBMNGR, "Do you have any managerial duties or do you supervise any other employees?"* | All |
| Opportunities for promotion | | Whether has current opportunities for promotion *Based on wJBOPPS. Note that wasn't asked in 1992-1994 if respondent had not changed jobs since previous wave; response therefore copied from previous response* | All |
| *Self-reported class* Working class Middle class Other | | Self-reported social class *Based on wOPCLS2, "which social class would you say you belong to?" Responses recoded into working class (inc. lower/upper w/c), middle class (inc lower/upper m/c and upper class) and other (inc no particular class, refuses class category, other not class, and other)* | 1991,  1996,  2000,  2005 |
| *Value of work* Not important Quite unimportant Quite important Very important | | Importance of having a fulfilling job *Based on wLFIMPD, with the 1-10 scale recoded into not important (1-4), quite unimportant (5-6), quite important (7-8) and very important (9-10). Because this variable was not available at any baseline wave at which WLD was asked, respondents at all waves from 1998 onwards were given their last reported value of work score (for most respondents, this means the 1998 score in 1999-2002, and the 2003 score in 2004-6).* | 1998,  2003,  2008 |
| ***Health*** | | | |
| **Label** | **Description** | | **Waves** |
| LLSI | Limiting Longstanding Illness *From wHLLT, "Does your health in any way limit your daily activities compared to most people of your age?"* | | All bar 1999 and 2004 |
| GP visits | GP visits in past year *From wHL2GP* | | All waves |

### Appendix 6B: Physical demands

The physical demands scale was created and imputed in exactly the same way as demands-control.

Firstly, a composite physical demands scale was constructed from the following two questions in the 1997, 2001 and 2006 Skills Surveys:

P6: [In your job, how important is…] physical strength (for example, to carry, push or pull heavy objects?

1. Essential 2. Very important 3. Fairly important 4. Not very important 5. Not at all important/does not apply

P8: [In your job, how important is...] physical stamina (to work for long periods on physical activities)?

1. Essential 2. Very important 3. Fairly important 4. Not very important 5. Not at all important/does not apply

An Item Response Theory (IRT) model was created as described in Appendix 2B, with the results shown in Table A12.

Table A12: Item Response Theory model for Physical demands

|  |  |  |
| --- | --- | --- |
|  | Two-parameter model | |
|  | Estimate | SE |
| *Difficulty parameter* |  |  |
| P6 (stamina) |  |  |
| Essential -> Very important | 2.83 | 0.23 |
| Very important -> Fairly important | 1.49 | 0.10 |
| Fairly important -> Not very important | 0.09 | 0.04 |
| Not v important -> Not at all important | -1.55 | 0.14 |
| P8 (strength) |  |  |
| Essential -> Very important | 2.22 | 0.14 |
| Very important -> Fairly important | 0.73 | 0.06 |
| Fairly important -> Not very important | -0.64 | 0.04 |
| Not v important -> Not at all important | -1.73 | 0.13 |
| *Discrimination parameter* |  |  |
| P6 (strength) | 1 | Fixed |
| P8 (stamina) | 0.83 | 0.11 |
| *Log likelihood* | *-41,113* |  |
| *n* | *14,695* |  |

For explanation of ‘difficulty’ and ‘discrimination’ parameters, see Appendix 2B.

This was then imputed into BHPS on the basis of occupation (and in some sensitivity analyses, also year and other covariates), as described in Chapter 5. Unlike demands-control, I extrapolated the 1997-2001 trend in physical demands back to 1991 in those specifications that took year into account; given that there is little trend in physical demands, though, this assumption makes little impact.

### Appendix 6C: Sensitivity analyses

A number of other sensitivity analyses were conducted to test the robustness of the findings to various methodological decisions. While these are not of interest in themselves, they are reported here insofar as they provide evidence on the robustness of the findings in the main part of the chapter.

Firstly, I used different forms of demands and control to see if the results were sensitive to the particular way that I operationalised these variables (see Web Table 6.CDlin). Using continuous linear forms of demands and control produces similar results to the main results, except that demands/control were only significant at the 10% level. Given the problems described in the main chapter when looking at multiple occupation-level variables simultaneously, I additionally looked at the JEM-imputed demands and control separately, without controlling for one another (see Web Table 6.CDlin) – although still controlling for physical job demands. These results were very similar to the main results, except that the effects were slightly larger for each alone (AMEcontrol=1.2% vs. 0.9%, AMEdemands=1.1% vs. 0.9).

Second, given the likelihood that WLD reports are influenced by respondents’ work status, I re-ran the main model on the subsample of people who were still working at the follow-up wave. The effect of control was here slightly larger and more highly significant (p=0.03), while job demands had a slightly smaller and less significant effect (p=0.06) – but generally it seemed that the main results were not being driven by the endogeneity of work status and WLD (see Web Table 6.CDspec).

Third, if I exclude people who say they were away from work due to ‘temporary sickness’ in the week prior to the interview, the statistical significance of job control declines below conventional levels of significance (p=0.06) when looking one wave into the future, but the effects are similar in size. I expected this to be because there was insufficient time for the effects of demands-control to become apparent (Chapter 3), and therefore tested the effects of demands/control were similar on WLD two waves later (rather than one) after excluding the baseline temporary sick – which also provides greater comparability with the WII results (where waves were 2-3 years apart). In this model, job control becomes strongly significant (p<0.001), while there is no apparent effect of job demands – a finding that leans against a genuine role of job demands, as I discuss below.

Fourth, I used slightly different methods of imputing demands, control and physicality, as described in Chapter 5 (see Web Table 6.CDJEM). In the main analyses above I impute based on occupation and the population-wide trend 1992-2006, but these sensitivity analyses consider imputations using (i) occupation only; (ii) occupation and the trend within each major occupational group; (iii) occupation, the population-wide trend and covariates that are likely to be associated with working conditions (industry and education; (iv) occupation, the population-wide trend, and other covariates (age, gender, and hours of work as well as industry and education); (v) industry-occupation subgroups and the population-wide trend; and (vi) Empirical Bayes estimates using occupation only.

For job control the results were similar across most of the JEMs, with the effect varying between 0.7 and 1.2 percentage points, and the significance level varying from p<0.01 to p=0.08, with only one exception.[[14]](#footnote-14) The effect of job demands on WLD varied much more – in four JEMs it was strong and highly significant (AME=0.9-1.3%, p<=0.03 to p<0.001), while in three it was much smaller and non-significant (AME=0.3-0.6%, p=0.15 to p=0.51). This does not appear to be explain by the different characteristics of the JEMs.[[15]](#footnote-15) This suggests that the effect of job control is relatively robust but that of job demands is sensitive to the way that demands are imputed – and more generally, that the form of imputation technique can impact on the results.

Fourth, I checked the results using different estimation techniques in several ways. I re-ran the pathways models using other recently developed techniques,[[16]](#footnote-16) and the main models using random effects models and other GEE working correlation matrices (unstructured and autoregressive(1)) rather than exchangeable GEE matrix. I also checked whether the loss of information in categorising the continuous covariates (age, income, GHQ and hours of work) led to residual confounding ([Altman and Royston 2006](#_ENREF_1); [Royston et al 2006](#_ENREF_31)), by using fractional polynomial transformations of these variables ([Royston and Sauerbrei 2008](#_ENREF_33)). None of these variations led to even slight changes to the conclusions above.[[17]](#footnote-17) After finding the goodness-of-fit of the model was poor (using the technique described in the main chapter), I also used a slightly different parameterisation that produced a better-fitting model, and again this made little difference to the results.[[18]](#footnote-18)

Fifth, I checked if accounting for the clustering of demands-control within occupations led to more tentative conclusions. As described in the main chapter, it was not possible to include these in a full crossed-level logistic model (the analyses were sufficiently burdensome that even in LSE’s ABACUS system, the first iteration – let alone convergence – had not been reached within several days). Instead, I tested whether the results were affected by (i) accounting for the clustering of occupations at a single baseline wave (1991); and (ii) by conducted a crossed-level model that treats the binary outcomes as normally-distributed continuous variables (which is easier computationally, but is particularly problematic for rare outcomes like incapacity benefit receipt). In both cases, the coefficients or standard errors were effectively unchanged when accounting for clustering within occupations.[[19]](#footnote-19)

Finally, I checked for the importance of missing data. I used the supplied BHPS weights to make the sample nationally representative and to compensate for differential attrition (see Web Table 6.CDmethods). The weights themselves had little effect, but restricting the analysis to those who responded at every wave 1991-2006 changed the results: job control had a smaller, non-significant effect, while job demands had a larger and marginally significant effect.[[20]](#footnote-20)

The multiple imputation models allow us to see if this is due to selection effects (the types of people responding at every wave are atypical) or whether non-random attrition between the baseline and follow-up waves lead to biases – although the resulting estimates will slightly underestimate the uncertainty due to missing data.[[21]](#footnote-21) For any given wave, those in low-control jobs were more likely to non-respond at the following wave (primarily due to the sorts of people in low-control jobs). Taking this into account, baseline job control had a smaller (and non-significant[[22]](#footnote-22)) effect on later WLD, while the effect of job demands was unchanged.

## Chapter 7 Appendices

### Appendix 7A: Sensitivity analyses

As in the previous chapters, several other sensitivity analyses focused on different methodological decisions.

Firstly, I used different forms of demands and control to see if the results were sensitive to the particular way that I operationalised these variables (see Web Table 7.CDlin). The results for linear demands and control were similar to the main analyses.[[23]](#footnote-23) When job demands and control were not mutually adjusted for one another, control was significant while demands was not, suggesting that an effect of job demands is not being concealed by any collinearity with control.

Second, I checked the importance of different time lags between control and incapacity benefits receipt. If we exclude those who said they were away from work due to ‘temporary sickness’ in the week prior to the baseline interview, then the effect of job control becomes non-significant (although still large in size). This removes from consideration some people who had already started down the pathway from job control to benefits receipt – and a one-year gap may be too short for other people to complete this pathway. If we therefore repeat the analysis looking at effects *two* waves into the future (rather than one), and still excluding the temporarily sick, then we again see a large and statistically significant effect of job control.[[24]](#footnote-24)

Third, I used slightly different methods of imputing demands, control and physicality, as described in Chapter 5 (see Web Table 7.CDJEM). As the imputation used a greater amount of information, the strength of the effect of job control slightly increased, with the estimated effect being largest for the imputations based on the covariates that we would expect to be genuinely associated with control (rather than reporting effects), and for the more fine-grained industry-occupation groups, where the effects were 0.46% and 0.42% respectively (compared to 0.34% in the main analyses). Where the imputation used less information – it was based on occupation averages across 1992-2006, without adjusting for declining control over time – the effect was smaller (0.17%) and non-significant (p=0.29). No effect was found for job demands in any model. These results are as we would expect: the less measurement error in the imputation technique, then the less attenuation we would expect to see.

Fourth, I checked the results using different estimation techniques in several ways. I re-ran the pathways models using other recently developed techniques,[[25]](#footnote-25) and the main models using random effects models and other GEE working correlation matrices (unstructured and autoregressive(1) rather than exchangeable). I also checked whether the loss of information in categorising the continuous covariates (age, income, GHQ and hours of work) led to residual confounding ([Altman and Royston 2006](#_ENREF_1); [Royston et al 2006](#_ENREF_31)), by using fractional polynomial transformations of these variables ([Royston and Sauerbrei 2008](#_ENREF_33)). None of these variations led to even slight changes to the conclusions above.[[26]](#footnote-26) After finding the goodness-of-fit of the model was poor (using the technique described in Chapter 7), I also used a slightly different parameterisation that produced a better-fitting model, and again this made little difference to the results.[[27]](#footnote-27)

Fifth, I checked if accounting for the clustering of demands-control within occupations led to more tentative conclusions. As described in the previous chapter, it was not possible to include these in a full crossed-level logistic model. Instead, I tested whether the results were affected by (i) accounting for the clustering of occupations at a single baseline wave (1991); and (ii) by conducted a crossed-level model that treats the binary outcomes as normally-distributed continuous variables. In both cases, the standard errors were often (but not consistently) raised when accounting for clustering within occupations, but only by small amounts – and the impact on the conclusions in this chapter will therefore be minimal.[[28]](#footnote-28)

Finally, I checked for the importance of missing datausing weighting. Using the supplied BHPS weights had little effect on the results (see Web Table 7.Weights), and (after weighting) job control had strong and marginally significant effects in both of the subsamples that allowed weighting. Similarly after multiple imputation to account for missing data and drop-out between the baseline and follow-up waves, the difference between high and low control was similar in size and significant at the 1% level.

## Appendix Bibliography

Altman, D. G. and P. Royston (2006). "The cost of dichotomising continuous variables." *BMJ,* 332: 1080.

Bajekal, M., T. Harries, et al. (2004). Review of Disability Estimates and Definitions: A study carried out on behalf of the Department for Work and Pensions. *In-House Report 128*. <http://www.dwp.gov.uk/asd/asd5/ih2003-2004/IH128.pdf>

Bartholomew, D. J., F. Steele, et al. (2008). *Analysis of Multivariate Social Science Data [2nd edition]*. Boca Raton, FL, Chapman & Hall/CRC.

Bertelli, A. M. and L. E. Richardson Jr (2008). "The Behavioral Impact of Drinking and Driving Laws." *Policy Studies Journal,* 36(4): 545-569.

Breen, R., K. B. Karlson, et al. (2010). Total, direct, and indirect effects in logit models. *Working paper*. Available at SSRN: <http://ssrn.com/abstract=1730065>

Buis, M. (2010). "Direct and indirect effects in a logit model." *The Stata Journal,* 10(1): 11-29.

Carpenter, J. and J. Bithell (2000). "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians." *Statistics in Medicine,* 19: 1141-1164.

Carpenter, J. and M. Kenward (2008). *Missing data in clinical trials – a practical guide*. Birmingham, NHS Co-ordinating Centre for Research Methodology.

Davey Smith, G. and A. Phillips (1992). "Confounding in epidemiological studies: why 'independent' effects may not be all they seem." *BMJ,* 305: 757-759.

Efron, B. and R. J. Tibshirani (1993). *An introduction to the bootstrap*. London, Chapman and Hall.

Erikson, R., J. Goldthorpe, et al. (2005). "On class differentials in educational attainment." *PNAS (Proceedings of the National Academy of the United States of America),* 102(27): 9730-9733.

Faye, O., A. Baschieri, et al. (2011). "The New York Academy of Medicine Hunger and Food Insecurity in Nairobi’s Slums: An Assessment Using IRT Models." *Journal of Urban Health: Bulletin of the New York Academy of Medicine*.

Fewell, Z., G. Davey Smith, et al. (2007). "The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study." *American Journal of Epidemiology,* 166: 646-655.

Forth, J., H. Bewley, et al. (2010). "Survey errors and survey costs: a response to Timming’s critique of the Survey of Employees Questionnaire in WERS 2004." *Work, Employment and Society,* 24(3): 578-590.

Haukoos, J. S. and R. J. Lewis (2005). "Advanced Statistics: Bootstrapping Confidence Intervals for Statistics with ‘‘Difficult’’ Distributions." *Academic Emergency Medicine,* 12(4): 360-365.

Hoffmeyer-Zlotnik, J. (2003). New Sampling Designs and the Quality of Data. *Developments in Applied Statistics*. A. Ferligoj and A. Mrvar. Ljubljana, FDV.

Holgado-Tello, F. P., S. Chacón-Moscoso, et al. (2008). "Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables." *Quality and Quantity,* 44(1): 153-166.

Karlson, K. B. and A. Holm (2011). "Decomposing primary and secondary effects: A new decomposition method." *Research in Social Stratification and Mobility,* 29(2): 221-237.

Kenward, M. and J. Carpenter (2007). "Multiple imputation: current perspectives." *Statistical Methods in Medical Research,* 16(3): 199-218.

Kolenikov, S. and G. Angeles (2004). The Use of Discrete Data in Principal Component Analysis: Theory, Simulations, and Applications to Socioeconomic Indices. *Working Paper of MEASURE/Evaluation project, No. WP-04-85*, Carolina Population Center, UNC. https://[www.cpc.unc.edu/measure/publications/pdf/wp-04-85.pdf](http://www.cpc.unc.edu/measure/publications/pdf/wp-04-85.pdf)

Li, Y., A. Pickles, et al. (2005). "Social capital and social trust in Britain." *European Sociological Review,* 21(2): 109-123.

Lynn, P., S. Häder, et al. (2004). Methods for achieving equivalence of samples in cross-national surveys: the European Social Survey experience. *ISER Working Papers Number 2004-09*. <http://www.iser.essex.ac.uk/files/iser_working_papers/2004-09.pdf> [accessed 11/10/2010]

Ng, N., P. Kowal, et al. (2010). "Health inequalities among older men and women in Africa and Asia: evidence from eight Health and Demographic Surveillance System sites in the INDEPTH WHO-SAGE Study." *Global Health Action,* 3.

Rabe-Hesketh, S. and A. Skrondal (2008). *Multilevel and Longitudinal Modeling Using Stata (Second Edition)*. College Station, TX, Stata Press.

Raileanu Szeles, M. and A. Fusco (2009). Item response theory and the measurement of deprivation: Evidence from PSELL-3. *IRISS-C/I 2009-05*. Luexmbourg, Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques / International Network for Studies in Technology, Environment, Alternatives, Development (CEPS/INSTEAD).

Rose, R. A. and M. W. Fraser (2008). "A simplified framework fo rusing multiple imputation in social work research." *Social Work Research,* 32(3): 171-178.

Royston, P. (2004). "Multiple imputation of missing values." *The Stata Journal,* 4(3): 227-241.

Royston, P. (2005). "Multiple imputation of missing values: Update of ice." *The Stata Journal,* 5(4): 527-536.

Royston, P. (2007). "Multiple imputation of missing values: further update of ice, with an emphasis on interval censoring." *The Stata Journal,* 7(4): 445-464.

Royston, P. (2009). "Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables." *The Stata Journal,* 9(3): 466-477.

Royston, P., D. G. Altman, et al. (2006). "Dichotomizing continuous predictors in multiple regression: a bad idea." *Statistics in Medicine,* 25(1): 127-141.

Royston, P., J. Carlin, et al. (2009). "Multiple imputation of missing values: New features for mim." *The Stata Journal,* 9(2): 252-264.

Royston, P. and W. Sauerbrei (2008). *Multivariable Modelling: A pragmatic approach based on fractional polynomials for continuous variables*, Wiley.

Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL, Chapman & Hall/CRC.

Smith, A., E. Wadsworth, et al. (2005). Ethnicity, work characteristics, stress and health. Cardiff, Cardiff University and Queen Mary, University of London for the Health and Safety Executive.

Welch, C., I. Petersen, et al. (2010). Simulation of “forwards-backwards” multiple imputation technique in a longitudinal, clinical dataset. *2010 UK Stata Users Group meeting*. London. <http://www.statacorp.com/meeting/uk10/UKSUG10.Welch.ppt>

Zheng, X. and S. Rabe-Hesketh (2007). "Estimating parameters of dichotomous and ordinal item response models with gllamm." *The Stata Journal,* 7(3): 313-333.

Zumbo, B., A. Gadermann, et al. (2007). "Ordinal versions of coefficients alpha and theta for Likert rating scales." *Journal of Modern Applied Statistical Methods,* 6: 21-29.

1. LFS Performance and Quality Monitoring Report Jan-Mar 2008,

   <http://www.statistics.gov.uk/downloads/theme_labour/JM08PQM.pdf>, accessed 14/5/08; LFS Data Archive document ‘LFS variables 1975-1991’, p7. [↑](#footnote-ref-1)
2. All of these surveys include both design weights and non-response weights. However, EiB includes two different sets of non-response weights. I have used the weights that are most likely to be comparable to the Skills Surveys (adjusting for non-response by gender), rather than the weights that account for further forms of non-response (adjusted for full-time vs. part-time, sex, age, and socio-economic group, compared to the Labour Force Survey from the same period, cf. the weights for WiB). [↑](#footnote-ref-2)
3. For the purposes of deriving the summary variable, WLD-pain was turned into a binary variable (no interference vs. any interference with normal work), to overcome computational problems. [↑](#footnote-ref-3)
4. The variables were: job demands/control (job demands, job control, job strain), demographics (age, gender, marital status, number of children), SES (Civil Service Grade, education measured at wave 5, material problems, housing problems, council housing, car ownership), physical health (bodily pain, SF-36 physical functioning, three broad categories of longstanding illness (musculoskeletal, heart, and other)), mental health (SF-36 general mental health, GHQ, negative affectivity), all measured at the baseline wave only, WLD (any WLD and latent WLD at both baseline and follow-up waves) and health retirement (at follow-up). This includes several variables excluded in the main models due to substantial missing data as they were copied across from answers at specific waves (negative affectivity, education, number of children, council housing and car ownership). [↑](#footnote-ref-4)
5. I used the following alternate paramaterizations: (i) categorical demands/control without an interaction term for strain; (ii) linear demands/control, both with and without a linear interaction between them. [↑](#footnote-ref-5)
6. Fractional polynomials are ways of balancing parsimony and flexibility in modelling continuous covariates; they use a limited set of power transformations (in this case, the powers -2 -1 -0.5 0 0.5 1 2), and only use more complex forms if these lead to significantly better model fit ([Royston and Sauerbrei 2008](#_ENREF_33)). The Stata command MFP\_PLUS was used, setting the p-values for choosing between FP models to 0.05, but forcing all of these variables to be included in the model. [↑](#footnote-ref-6)
7. This was done using the KHB command in Stata. KHB has the advantage of being an established (and peer-reviewed) rather than bespoke method, but the disadvantage that it produces estimates at the mean value of the covariates rather than averaging across them. (KHB can supply average marginal effects, but without confidence intervals). [↑](#footnote-ref-7)
8. The sample size varies according to whether people are imputed as being non-employed (and therefore out-of-sample) at baseline. [↑](#footnote-ref-8)
9. The coefficient for high vs. low job-strain is 0.48 (p<0.001) in both the main and missing data analysis. [↑](#footnote-ref-9)
10. I used the following four alternate paramaterizations: (i) categorical demands/control without an interaction term for strain; (ii) linear demands/control, and a linear interaction between them; (iii) linear demands/control without an interaction term; (iv) a single term for high-strain jobs without looking at demands/control. [↑](#footnote-ref-10)
11. For both the Karlson/Holm/Breen (KHB) and Buis (LDECOMP) commands, it was necessary to use simple dummy variables for high-strain jobs vs. all other jobs, rather than the dummy variables for demands, control and demands\*control in the main models. If my bespoke mediation technique is repeated using the high-strain vs. other jobs comparison, however, the results are similar to the results in the main part of the chapter; the different paramaterizations do not explain the divergent results. [↑](#footnote-ref-11)
12. n for main analysis = 14, 289; minimum n for missing data analysis = 21,671. [↑](#footnote-ref-12)
13. To get confidence intervals for the EB estimates, I used a non-parametric bootstrapping technique ([Efron and Tibshirani 1993](#_ENREF_10); [Carpenter and Bithell 2000](#_ENREF_7); [Haukoos and Lewis 2005](#_ENREF_15)). Due to the considerable computational time required I restricted this to 100 bootstrap replications; the confidence intervals shown in this section are therefore approximate only. For the four-category strain results, the bootstrap estimates were obtained for a series of four binary logistic models rather than a multinomial logistic model (again because of the computation time required to bootstrap the multinomial model that provides the main estimates); the confidence intervals should be treated as illustrative only here. [↑](#footnote-ref-13)
14. Using the Empirical Bayes method of imputation, the effect was slightly smaller (AME=0.6%) and non-significant (p=0.16). [↑](#footnote-ref-14)
15. Demands was non-significant for the occupation-only, occupation + occupation-specific trend, and full covariate JEMs. [↑](#footnote-ref-15)
16. Namely the LDECOMP ([Erikson et al 2005](#_ENREF_11); [Buis 2010](#_ENREF_6)) and KHB ([Breen et al 2010](#_ENREF_5)) commands. [↑](#footnote-ref-16)
17. One of the alternative pathways models (LDECOMP) showed that the health pathway was larger than the job requirements pathway for job control. It seems likely that this is due to some problem with this recently-introduced command; for example, the overall effect of job demands seems to be implausibly low compared to other techniques. It is therefore reassuring that the KHB technique produces the same results as my pathways models in the main chapter. [↑](#footnote-ref-17)
18. The main model (containing controls for individual-level SES and physicality) was poorly-fitting (adapted Hosmer-Lemeshow chi2(9)=83.1, p<0.0001), primarily because it overestimated the probability of reporting a WLD among those with low predicted probabilities. I therefore created a latent health scale by regressing the specific health measures on self-reported general health, and using a multivariate fractional polynomial model to find the best-fitting form of this latent health scale (alongside the non-health controls and a dummy for zero health problems). This revised model fitted the data acceptably (chi2(9)=14.6, p=0.10), and produced similar results (AMEcontrol=0.78% vs. 0.82% in the main model, AMEdemands=0.91% vs. 0.85%). [↑](#footnote-ref-18)
19. For the 1991 wave, the coefficients and standard errors for high control, high demands and high physicality were 0.06 (se=0.27), 0.56 (0.26) and -0.02 (0.25) respectively. When accounting for clustering within occupations. these were effectively identical at 0.06 (se=0.27), 0.56 (0.25) and -0.02 (0.25).

    For the linear panel model on the full sample, the coefficients accounting for the clustering of observations within people were for high control, high demands and high physicality were -0.0075 (se=0.0043), -0.0049 (0.0040), and 0.0040 (0.0041) respectively. When the clustering within occupations was additionally incorporated in a crossed-effects model, the coefficients were -0.0069 (0.0044), -0.0049 (0.0042) and 0.0052 (0.0044).

    For these analyses, I used the version of the JEM that only imputed based on the 133 occupations (rather than occupations + years + covariates). Random effects rather than GEE models were used in both cases (the XTLOGIT command for the 1991 wave, and the XTMIXED command for the linear crossed model). [↑](#footnote-ref-19)
20. It was necessary to restrict the sample to those who had responded at every wave as these were the only cases where weights were supplied (as described in the main chapter). I also checked this by looking at the Scottish and Welsh extension samples 1999-2006 (starting at 1999 enables me to use the longitudinal weights that include this larger sample). Here there was no effect whatsoever of job control, but job demands had a greater effect after weighting (AMEdemands=2.1%, p<0.01). [↑](#footnote-ref-20)
21. The usual suggestion for multiply imputing longitudinal data is to convert it to ‘wide’ format – that is, treat the variables in each year as different variables. However, with repeated waves and large numbers of covariates such estimation becomes unfeasible; it would here lead to 16-times the original number of covariates, leading to substantial problems of multicollinearity and separation. A way of overcoming such problems has been suggested, but as yet the ‘forwards-backwards’ technique has not been implemented in any software package ([Welch et al 2010](#_ENREF_36)). As a result, I here conduct multiple imputation in ‘long’ format as described in Chapter 4. This does not take into account the clustering of responses within individuals, and only imputes for item-missing data and attrition between baseline and follow-up (ignoring the substantial attrition between baseline waves). This should produce unbiased estimates of the effects of baseline characteristics, but the sample will become progressively more selected over time. [↑](#footnote-ref-21)
22. The difference of high (vs. low) control jobs was not significant (p=0.26), but the difference of moderate (vs. low) control jobs was still marginally significant (p=0.06), and the coefficients were similar in both. It is also hard to evaluate the impact of the missing data analysis as the estimated effect of high vs. low job control was smaller with the reduced covariate list used for the imputation. [↑](#footnote-ref-22)
23. Job demands was non-significant, while job control was significant (p<0.05), and the estimated AME for the change from the average control in the lowest-control tertile to the average control in the highest-control tertile was also virtually identical to the AMEs reported above. [↑](#footnote-ref-23)
24. AMEcontrol=0.49% against an incidence of transitions to incapacity benefits of 1.4%. [↑](#footnote-ref-24)
25. Namely the LDECOMP ([Erikson et al 2005](#_ENREF_11); [Buis 2010](#_ENREF_6)) and KHB ([Breen et al 2010](#_ENREF_5)) commands. [↑](#footnote-ref-25)
26. In the subsample for the autoregressive(1) model that excludes the 1991 wave, the effect of control was slightly weaker and non-significant, but still strong; however, the AR(1) model produced identical results to the exchangeable correlation matrix. The fractional polynomial (MFP) results are available in Web Table 7.CD\_spec, while the random effects etc. models are available in Web Table 7.CD\_methods. [↑](#footnote-ref-26)
27. The main model (containing controls for individual-level SES and physicality) was poorly-fitting (adapted Hosmer-Lemeshow chi2(9)=17.8, p=0.04), primarily because it overestimated the probability of claiming incapacity benefits among those with the very lowest predicted probabilities. I therefore created a latent health scale by regressing the specific health measures on self-reported general health, and using a multivariate fractional polynomial model to find the best-fitting form of this latent health scale (alongside the non-health controls and a dummy for zero health problems). This revised model fitted the data acceptably (chi2(9)=11.9, p=0.22), and produced similar results (AMEcontrol=0.32% vs. 0.30% in the main model). [↑](#footnote-ref-27)
28. For the 1991 wave, the coefficients and standard errors for high control, high demands and high physicality were -0.90 (se=0.60), 1.21 (0.48) and 1.71 (0.61) respectively. When the clustering within occupations was incorporated, this changed to -0.90 (se=0.62), 1.21 (0.55) and 1.71 (0.57) respectively.

    For the linear panel model on the full sample, the coefficients accounting for the clustering of observations within people were for high control, high demands and high physicality were -0.0019 (se=0.0016), -0.0007 (0.0015) and 0.0054 (0.0015) respectively. When the clustering within occupations was additionally incorporated in a crossed-effects model, the coefficients were -0.0017 (se=0.0017), -0.0013 (0.0016) and 0.0054 (0.0017).

    For these analyses, I used the version of the JEM that only imputed based on the 133 occupations (rather than occupations + years + covariates). Random effects rather than GEE models were used in both cases (the XTLOGIT command for the 1991 wave, and the XTMIXED command for the linear crossed model). [↑](#footnote-ref-28)