Working Paper

Predicting likelihood of long-term unemployment: the development of a UK jobseekers' classification instrument

by Simon Matty

DWP Department for Work and Pensions

Department for Work and Pensions

Working paper No 116

Predicting likelihood of long-term unemployment: the development of a UK jobseekers' classification instrument

Simon Matty

A report of research carried out by analysts in DWP Insight on behalf of the Department for Work and Pensions

This document/publication is also available on our website at:
http://research.dwp.gov.uk/asd/asd5/rrs-index.asp

Any enquiries regarding this document/publication should be sent to us at:
Central Analysis Division, Department for Work and Pensions, Upper Ground Floor, Steel City House, West Street, Sheffield, S1 2GQ

First published 2013.

ISBN 978 1 909532 22 9

Contents

List of tables

List of figures

Acknowledgements

I would like to thank colleagues in the Department for Work and Pensions (DWP) who provided comments on various drafts of this report, notably Simon Palmer, Julian Williamson, Mike Daly and members of the Labour Market Division. Particular thanks to Dominic Pinto and Alex Hart for comments on the technical aspects of the modelling, including undertaking earlier analysis.

The Author

Simon Matty is a Senior Research Officer at the Department for Work and Pensions.

Abbreviations and glossary of terms

AUC	Area Under the Curve
CATI	Computer-Aided Telephone Interviewing
DWP	Department for Work and Pensions
ESA	Employment and Support Allowance
GMS	General Matching Service
IB	Incapacity Benefit
IS	Income Support
JSA	Jobseeker's Allowance
JSCI	Jobseekers' Classification Instrument
LTU	Long-term unemployment
NBD	National Benefits Database
OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics
ROC	Receiver Operating Characteristic

Summary

Aim

This report describes work undertaken by the Department of Work and Pensions (DWP) to explore the feasibility of developing a profiling tool to predict, at the moment of first claim, the likelihood of a new claimant reaching long-term unemployment (LTU)[1].

Methodology

A telephone survey of new claimants was carried out in 2010 by an external contractor, TNS-BMRB, that asked a range of questions covering demographics, work experience, education/skills and health and barriers to work. These questions were developed from earlier attempts to build such predictive models, along with questions adapted from the claimant profiling approach currently used in Australia.

The survey respondents were then tracked for 12 months to capture their benefit claim journeys and to confirm which individuals reached 12 months with a continuous Jobseeker's Allowance (JSA) claim.

The variables collected as part of the survey were combined with administrative data held by the Department and these were then used to build and test a number of predictive models using a binary logistic regression approach.

The results from the model building and testing were then considered in terms of model accuracy and how the model might be applied operationally.

Results

The best performing model included a mix of attribute data, administrative data and attitudinal data[2], with 59 per cent of the variation in the data explained by the model – that is, the model is 59 per cent of the way between a random model and a model with perfect prediction[3].

---

1 LTU is defined here as 12 months or more continuous Jobseeker's Allowance (JSA) claim.

2 **Attribute data** are data that describe the characteristics or circumstances of individuals – in this context drawn from the survey, as distinct from **Administrative data** that are data drawn from DWP databases relating to benefit history along with other data drawn from national sources; **Attitude data** are data that cover the attitudes of respondents covering a range of topics, also drawn from the survey.

3 There are a number of different ways of measuring the performance of a model, described later in the report, as well as a precise definition of this measure.

The variables included in this model, and which are, therefore, significant predictors of LTU, are as follows:

Table 1 Significant predictors of LTU from the Jobseekers' Classification Instrument (JSCI) model

Attribute variable	**Attitudinal variable (5 point scale)**
Age (banded)	How much do you agree with the following statement: Getting a job is more down to luck than the effort you put in
Age (continuous)	
Gender	
Whether English is first language	
Household composition: Live alone	
Household composition: Live with partner/spouse	How much do you agree with the following statement: I am confident I can get a job within three months
Household composition: Live with a dependent under 18-years-old	
If one parent/guardian was in paid work until respondent was 16-years-old	How much do you agree with the following statement: Having almost any job is better than being unemployed
Has or does a lack of references from a previous employer make it difficult to find or keep jobs	
Any paid work in the last two years	
How much earn per week in last job	**Administrative variable**
Have driving licence for car/motorcycle	Type of JSA claim
Public transport available for commuting to work	Number of days on all out-of-work benefits in the last two years
Have or do problems with public transport make it difficult to find or keep a job	Proportion of working-age population employed in the public sector in local area
Has or is a lack of experience making it difficult to find or keep a job	Average house price for local area

The output from the logistic regression model was used to calculate the predicted probability of reaching 12 months for each individual case.

Applying the model

This approach means that the JSCI can be used as a measure of relative risk, with individuals being placed into groupings defined by their likelihood of long-term unemployment. Most straightforwardly they can be split into two groups: a 'high risk' group and a 'lower risk' group, with those in the high risk group having a JSCI score over a specified level and those in the lower risk group having a score below this threshold. This approach to claimant segmentation could assist the more efficient targeting of support to those with the greatest need, with those included in the high risk category receiving appropriate and possibly early intervention. The effectiveness of such segmentation, as discussed later, depends on a number of factors in addition to the accuracy of the model – including what we mean by 'effectiveness'.

Table 2 presents this approach for a range of different cut off points, along with the accuracy of the model at each cut off point. Therefore, if we choose to segment with the top ten per cent JSCI scores and classify those individuals as at high risk of LTU, the model accuracy for this high risk group would be 31 per cent and 94 per cent for the lower risk group. That is, the model correctly predicts 31 per cent of those who reach LTU (compared with 8 per cent for a random model).

Table 2 Model accuracy rates at different cut off points

Percentage segmented	JSCI score (%)	High risk long term unemployed accuracy (%)	Lower risk long term unemployed accuracy (%)	Overall accuracy (%)
-	70	-	92	92
10	24	31	94	88
20	14	23	95	81
30	8	20	96	73
40	5	17	98	65
50	3	15	98	56
60	2	13	98	47
70	2	11	99	37
80	1	10	99	28
90	1	9	100	18
100	-	8	-	8

Operational implications of model accuracy

When thinking about the operational implications of the accuracy of the model, it is less useful to consider accuracy only as a single value for the proportion of correct predictions, as described above. How we define the operational accuracy (and what is sufficient accuracy in this instance) depends upon what decisions are based on the segmentation – for example, screening individuals into or out of an intervention – and the cost of any such intervention in relation to the cost of remaining on JSA. The effectiveness of interventions also need to be considered, as screening claimants for LTU as an isolated process becomes a nugatory activity without an effective intervention to reduce the likelihood of a claimant becoming LTU.

To illustrate the 'operational accuracy' point, consider if we ranked all claimants by their risk scores and then targeted the top eight per cent to whom we would then offer additional support. Doing so would reach 32 per cent of all those who end up long-term unemployed. If we targeted the top 30 per cent to whom we then offered additional support we would reach 70 per cent of all those who would end up long-term unemployed (see Table 3). It is worth noting, however, that using the eight per cent segmentation means that for every individual we correctly identify and treat we would be treating two individuals who would not become long-term unemployed. If we used the thirty per cent segmentation then for every correctly identified individual we would be treating four individuals who would not become long-term employed. If claimants were selected at random for support then choosing eight per cent would only reach eight per cent of the LTU and choosing 30 per cent would reach 30 per cent of the LTU.

Table 3 Operational effectiveness of ranking

	Size of targeted segment	
	Top 8%	**Top 30%**
Number of target segment that reach LTU	29	64
Number of target group that do not reach LTU	58	262
Total number of LTU	91	91
Total number of non-LTU	994	994
Proportion of all LTU captured by segmentation	32%	70%
Proportion of all non-LTU receiving 'unnecessary' support	6%	26%

Note: Model applied to test dataset (n = 1,085).

Conclusions

We have described the initial results from the development of a profiling model that aims to predict at first claim the likelihood of a new JSA claimant reaching LTU.

The model that has been developed is based on statistically significant predictors and is a combination of administrative data, attribute data and attitudinal data.

When considering the application of the model we have discussed using an approach that ranks claimants based on their JSCI score prior to segmentation as an efficient way of targeting support.

There are still a number of significant steps that need to be taken to develop our understanding of such an approach and its application in an operational setting. These include:

- cost benefit analyses/intervention testing to determine the optimal approach(es);
- pilot testing in a live, operational setting to identify real world challenges; and,
- model refinement and revision to consider application of the model at different points in a claimant journey as well as possible alternative outcome variables.

The Department is undertaking further work in this area to determine a high-level approach to claimant segmentation and to generate the necessary material to inform strategic decision-making about the use of segmentation approaches within this context. The lessons learnt from the JSCI modelling will feed into this broader thinking.

1 Introduction – background and previous research

The Department for Work and Pensions (DWP) traditionally has operated a segmented service provision for Jobseeker's Allowance (JSA) claimants based on their claim duration[4]. Access to different, and more intensive, employment support is based, therefore, on the length of time an individual has received the benefit. This approach recognises that 'most people who become unemployed find work within a few months with minimal assistance from public sources' (Payne and Payne, 2000, p7). Indeed, current figures from the Department confirm this, showing that around half of new JSA claimants flow off the benefit by 13 weeks and nearly three-quarters by 26 weeks[5].

Targeting intensive support on JSA claimants based on their duration of unemployment has a number of advantages. It is relatively simple to implement and operate, with resources being 'directed only to those who have actually experienced difficulty in obtaining employment' (Bimrose *et al.*, 2007, p4). The system should be equitable too, with all claimants treated according to the same rules.

However, it is recognised that this approach can be seen as a rather blunt instrument and that improvements could be made in terms of achieving a better match between individual claimant needs and provision of support, if it were possible to identify in advance those who were going to go on to become long-term claimants, and that this would be a highly desirable goal. Partly this is because evidence suggests that the longer an individual is out of work the less likely they are to return to work, due in part to changes in their attitude, loss of current skills and an increasing view from employers that they represent a poor employment prospect (see Chapman and Smith, 1993; Hasluck, 2004; Payne and Payne, 2000 and others). More straightforwardly, though, the sooner an individual claimant leaves benefit to enter work, the lower the cost to them of lost earnings and the lower the cost to the economy of the benefits paid to them.

Operating a system that defines eligibility for support on a specified length of unemployment can run the risk, therefore, of 'allowing' unemployed people to begin to suffer the extra challenges longer term unemployment brings, with the result that by the time they receive the additional help, the task of getting them back into work is all the more challenging (Payne and Payne, 2000). Additionally, the Organisation for Economic Co-operation and Development (OECD) notes that '*the long-term unemployed face substantial declines in well-being as a result of a greater risk of poverty, health problems and school failure of their children*' (OECD, 2012) which highlights that there are other linked social issues that could benefit from refining the match between claimant need and support to help minimise the numbers of unemployed who drift into long periods of unemployment. This serves to underline the potential significant gains to be made simply from reducing duration of claim.

[4] Although the Department has offered early intervention in a number of situations, for example in defining categories for early entry to the New Deal for Young People programme as well as supporting adviser discretion which allows Personal Advisers in Jobcentre Plus to recommend early interventions.

[5] Source: DWP. Actual figures are 52 per cent off flow by 13 weeks and 72 per cent off flow by 26 weeks (July, 2012).

Table 1.1 presents the figures for UK long-term unemployment (LTU) [6] since 1997 and highlights that in July 2012 around 419,000 individuals were claiming JSA, with just over one-quarter (26 per cent) of current JSA claimants being long-term unemployed.

Table 1.1 Seasonally adjusted JSA claimant count by duration, all aged 18 and over, UK

Year*	All claimants ('000s)	All over 12 months ('000s)	Percentage claiming over 12 months
1997	1,554.1	522.0	33.8
1998	1,344.0	362.8	27.1
1999	1,244.7	305.4	24.8
2000	1,074.7	238.2	22.3
2001	951.8	189.0	20.0
2002	950.0	151.9	16.2
2003	940.2	140.3	15.1
2004	836.4	134.7	16.2
2005	866.7	121.1	14.1
2006	956.7	153.9	16.2
2007	856.3	143.1	16.8
2008	882.7	100.7	11.5
2009	1,585.8	145.7	9.2
2010	1,457.6	265.6	18.3
2011	1,557.6	229.4	14.8
2012	1,593.2	418.6	26.3

* Figures are for the month of July in each year.
Source: Office for National Statistics (ONS) labour market statistics, August 2012 release.

Considerable research evidence is already available about the factors that are related to LTU and the Department has in the past examined the possible effectiveness of screening or profiling claimants to identify those at greatest risk. However, as Hasluck (2004) noted:

> *'Statistical modelling of the risk of long-term unemployment has been attempted in the UK but such studies have generally concluded it was not possible to accurately predict which clients were at risk.'*
>
> (Hasluck, 2004, pii)

and that:

> *'To date, the evidence relating to statistical profiling in the UK has not provided support for the introduction of such systems of decision-making in practice.'*
>
> (Hasluck, 2004, pii)

[6] LTU is defined as reaching 12 months on a continuous JSA claim.

In his report, Hasluck did note, however, that the DWP-commissioned work by Bryson and Kasparova (2003) perhaps reopened the debate about the extent to which it was possible to predict with acceptable accuracy those new claimants that were at risk of LTU[7].

Also emerging at that time was information relating to the apparently successful implementation of profiling systems in other countries, notably Australia. In their review of the literature, Bimrose *et al.* identify profiling systems that have been used in other countries and remark that '*the one emerging with the most consistently robust track record is the JSCI.*' (Bimrose *et al.* 2007, p6). The JSCI – which stands for Job Seeker Classification Instrument – is the system that was introduced in Australia[8, 9]. (Despite this 'robust track record', the Productivity Commission of Australia, in its review of the Australian job network, whilst recommending the continued use of the JSCI highlighted issues relating to its implementation that needed to be addressed (Productivity Commission, 2002)).

The DWP has always recognised that screening and profiling methods, though technically challenging, are potential approaches to refine service allocation through a more precise definition and identification of eligibility groups, and this evidence of potentially successful implementation in Australia, as well as positive information on profiling in other countries – for example, Ireland (see O'Connell *et al.* 2009 and O'Connell *et al.* 2012) – prompted the Department to once again examine the issue.

There was a further, policy dimension to the work, which provided additional focus and impetus to the issue. In his 2007 review of the Government's welfare-to-work strategy, David Freud recommended that the Department for Work and Pensions investigated the introduction of a system similar to the Australian JSCI (Freud, 2007). The 2008 review by Professor Paul Gregg on conditionality and support for those claiming benefits set out recommendations that included an emphasis on personalised support (for benefit claimants) on the basis of need. The Review also noted that '*developing an accurate early identification tool for all jobseekers is a major challenge*' but recommended that '*the Government should look closely at emerging findings and use these to assess whether it is possible to develop a more accurate and individualised screening tool...*' even though the '*conclusion to be drawn from these profiling models is that the tools that are currently available are unlikely to be accurate enough to be cost-effective*'. (Gregg, 2008).

Indeed, current policy and approaches to service delivery, such as the Get Britain Working initiative (including the Work Programme) and the introduction of Freedom and Flexibilities pilots in Jobcentre Plus reinforce the potential role and value of such early identification models[10].

It was within this context, then, that the Department looked at the feasibility of developing a profiling tool that aimed to predict at the moment of first claim, the likelihood of a new claimant reaching LTU.

---

7 It should be noted that 'acceptable accuracy' in this context relates specifically to the efficacy of the profiling – the degree to which individuals are correctly assigned to a particular group. Later on in the report the issue of 'operational accuracy' will be discussed which highlights the need for effective interventions to be available to make use of profiling activity.

8 Further information on the Australian JSCI can be found at: http://www.deewr.gov.au/Employment/JSCI/Pages/JSCI.aspx

9 A good, short and up-to-date review of international experiences can be found in the paper by O'Connell *et al.* (2012).

10 See DWP website for more information: http://www.dwp.gov.uk/policy/welfare-reform/get-britain-working/

2 Methodological approach

The development of the profiling tool – called the Jobseekers' Classification Instrument (JSCI) – consisted of three broad stages, as follows: a) choice of method; b) data collection; and, c) model development and testing.

2.1 Choice of method

Earlier DWP internal analysis in this area had explored building predictive models using only the administrative data held by the Department. This analysis aimed to predict, at the beginning of a Jobseeker's Allowance (JSA) claim, unemployment of different durations. The work revealed that the best results were achieved when predicting a continuous JSA claim of over 12 months, which fitted well with the current aspiration to identify those at risk of long-term unemployment (LTU).

Some of the more successful attempts at profiling to predict LTU have used logistic regression as the method of analysis (see for example, Payne and Payne, 2000; Bryson and Kasparova, 2003; O'Connell *et al.* 2009). Using a prototype dataset, four possible approaches were tested and this confirmed that logistic regression performed no worse than other methods (see Appendix A for further description of the tests). Given this result, along with its successful application in the previous research noted above, meant that logistic regression was chosen as the analytical approach to build the JSCI model. Additionally, the parameters produced using logistic regression lend themselves more naturally to generating a score for each attribute, which is the approach followed in the Australian model.

2.2 Data collection

There are a wide range of observed and unobserved factors that potentially can affect unemployment duration, some of which are readily observed and some which require additional data collection. These can include external factors (for example, related to the local job market) and factors relating to an individual's changes in circumstances (many of which cannot be predicted in advance). The predictive value of profiling depends in part on the type and quality of data available to predict outcomes. As noted above, previous DWP attempts at modelling were based on administrative data with limited availability of characteristics such as attitudinal variables.

One aim of the JSCI profiling project was to consider the value of attitudinal and attribute data in addition to administrative data in predicting outcomes. This recognises that the goal of the modelling here is not to identify (or confirm) the **most important factors** behind LTU but instead to identify the **most efficient predictors** of future LTU (Payne and Payne, 2000).

2.2.1 Questionnaire development

To capture these data a questionnaire was developed, initially based on the questions used in the Australian profiling tool (see Australia. Department of Education, Employment and Workplace Relations, 2012) but amended to reflect the UK context, as well as drawing on a review of evidence and contributions from DWP psychologists. The questionnaire was refined through input from policy experts and quality assured by a Steering Group. TNS-BMRB were commissioned to help develop the questionnaire and to administer the final version via CATI (Computer-Aided Telephone Interviewing) to a suitable sample of new JSA claimants. As part of the questionnaire development TNS-BMRB carried out a pilot study of 50 interviews using the same methodology as would be used for the main survey. Following the pilots a number of minor changes were made to the questionnaire. (See Appendix E for the survey questionnaire).

The questionnaire included two screening questions to establish eligibility; the first question was used to check whether the respondent had made a successful claim for JSA and the second whether the respondent was happy for their answers to the survey data to be linked with administrative data held by the Department. Any respondent who had not made a successful claim or who did not agree to their data being linked was then screened out of the survey (14 per cent). The rest of the questionnaire was structured around the following themes: demographics, work experience, education/skills and health and barriers.

2.2.2 Sampling strategy

Sample size

To determine how big the sample of new claimants should be it was necessary to consider a number of issues. For example, an estimate of the proportion of new JSA claimants expected still to be unemployed after 12 months was needed. We settled on obtaining an achieved sample of 5,000; the rationale for doing so is presented in Appendix A.

Sample strategy

Discussion and calculations with TNS-BMRB suggested that a total number of 15,000 new claimants should be selected for the research to deliver the required achieved sample of 5,000. A random sample of new JSA claimants was drawn, therefore, from DWP administrative records – specifically, the JSA General Matching Service (GMS) scans.

The intention was to interview customers as soon as possible after submitting a claim, to capture their attitudes at the start of their claim period and, therefore, to minimise any duration effects on their responses. The complexity of the recruitment process meant that the length of time from sampling to date of interview for the respondents was between 52 and 100 days.

A comprehensive check for duplicate records and incomplete records was carried out by TNS-BMRB which resulted in a final set of records from which the sample was drawn. The fieldwork for the survey took place between 24 May 2010 and 11 July 2010 and covered new JSA claimants who made a claim between 2 April 2010 and 23 April 2010. Further details on the sample strategy are presented in Appendix A.

Response rate

A total of 5,629 interviews with new claimants were completed between May and July 2010 (a response rate of 60 per cent) and these claimants were then tracked over a 12-month period to allow the DWP to use the survey data, linked with administrative data to retrospectively check who did, in fact, reach 12 months on JSA (see Appendix A for further information).

2.3 Model development and testing

After cleaning the data, 5,417 claimants remained eligible for the modelling. Combining the survey responses with DWP administrative data allowed us to build a number of different models that used different combinations of data to explore significant predictors of LTU and test for the optimal combination of predictors.

There were three broad types of data available, as follows:

- **Attribute data** – these are data derived from the survey that describe the characteristics or circumstances of individuals, for example, age, gender, household composition and availability of public transport.
- **Admin data** – these are data drawn from DWP databases that relate to benefit history along with other data drawn from national sources such as average house price for local area.
- **Attitude data** – these are data that cover the attitudes of respondents covering a range of topics and are generally captured via five point agree/disagree Likert scales.

Six different models were built using these data, as follows:

1 Admin – based on administrative data alone (drawn from DWP administration and other national sources[11]).

2 Attribute – based on attribute data alone (drawn from the survey).

3 Attitude – based on attitudinal data alone (drawn from the survey).

4 Attribute and Admin – using both attribute and administrative data.

5 Attribute and Attitude – using both attribute and attitudinal data.

6 All data – using all available data.

As we are interested in the prospective fit of the models (i.e. how well the predictions will fit future data sets) we split the dataset into two random, uneven sets. The first consisted of 80 per cent of the responses from the survey data and was used to build the models (called here the 'training' dataset). The second consisted of the remaining 20 per cent of survey respondents (called here the 'test' dataset) and this was used to test the predictions from the models built with the training dataset. Adopting this approach, called cross validation, ensures that the effects of sampling and measurement errors on predictive performance are minimised (Payne and Payne, 2000).

2.3.1 Results

Although the primary purpose of the research was to build a model to predict the likelihood of LTU (12 months on JSA), an interest in the ability to predict different durations on JSA as well as a desire to test the approach meant that, after three months and six months had elapsed from the date of last claim, models were built to predict likelihood of three months on JSA and six months on JSA.

Tables 2.1, 2.2 and 2.3 show, therefore, the accuracy of each model on the training and test datasets, quoted in terms of percentage variation explained by the data, for predicting the likelihood of a three months continuous JSA claim, a six months continuous JSA claim and likelihood of 12 months on JSA respectively[12]. These figures are best understood in terms of how much of an improvement the model is from a random model, therefore, 40 per cent variation explained means that the model accuracy is 40 per cent of the way between that of a random model and a perfect model. It should be noted here that the final models built for each data type or combination only

11 The DWP administrative data for building the six-month models were drawn from the GMS database and were drawn from the DWP's National Benefits Database (NBD) for the 12-month models. The NBD includes data from the GMS, but these have been cleaned and validated. This is particularly true of claim start and finish dates.

12 The percentage variation figure is calculated using the c statistic, with percentage variation = 2*c-1

include variables that are statistically significant predictors. For example, the 'attribute' model, whilst having access to all attribute variables at the model building stage, only includes the attribute variables that were found to be statistically significant.

Table 2.1 Percentage variation explained by each model – predicting three months continuous JSA

	Training dataset (Retrospective fit)	Test dataset (Prospective fit)
All data	46	42
Admin	28	23
Attribute	42	39
Attitude	24	27
Attribute and Admin	43	38
Attribute and Attitude	45	41

Note: Sixty-two per cent of respondents reached three months on JSA.

Table 2.2 Percentage variation explained by each model – predicting six months continuous JSA

	Training dataset (Retrospective fit)	Test dataset (Prospective fit)
All data	47	41
Admin	33	33
Attribute	42	34
Attitude	27	26
Attribute and Admin	45	39
Attribute and Attitude	45	37

Note: Thirty-three per cent of respondents reached six months on JSA.

Table 2.3 Percentage variation explained by each model – predicting 12 months continuous JSA

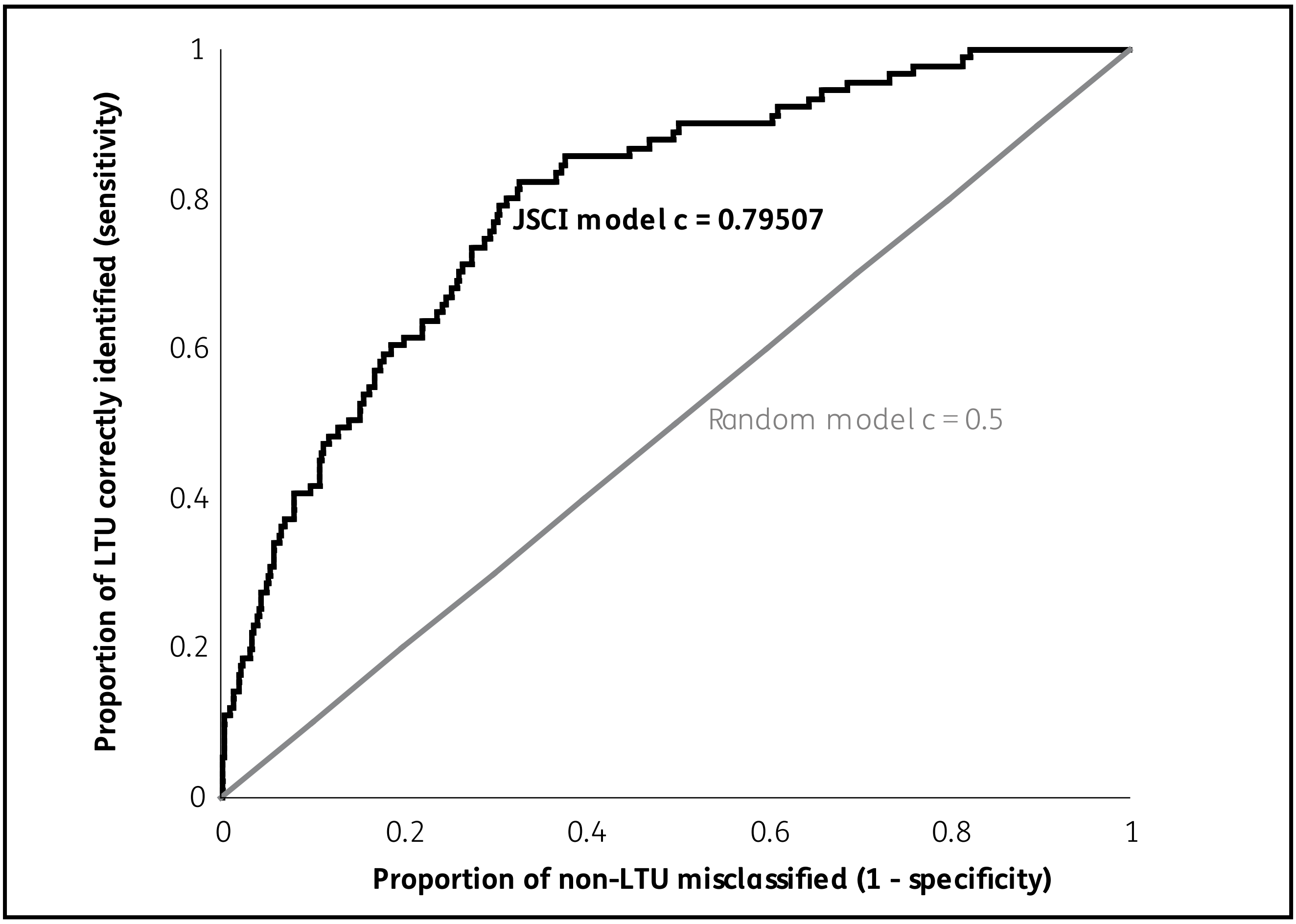
	Training dataset (Retrospective fit)	Test dataset (Prospective fit)
All data	69	59
Admin	55	51
Attribute	64	56
Attitude	36	27
Attribute and Admin	67	57
Attribute and Attitude	68	55

Note: Eight per cent of respondents reached 12 months on JSA

The all data model is shown to perform the best for each duration being predicted, explaining 59 per cent of the variation in the test dataset for the 12-month model, 41 per cent for the six-month model and 42 per cent for the three-month model. The results show that attribute data are key to obtaining the best 12-month predictions. For example, when looking at the 12-month models, Attitude and Admin data models explain 27 per cent and 51 per cent of variation in the test dataset but when attribute data are included these figures rise to 55 per cent and 57 per cent respectively. It should be noted, however, that Admin data also appear to be significant in obtaining the best possible model performance.

Receiver Operating Characteristic (ROC) curves can be used to measure the discrimination, or goodness-of-fit, of a predictive model and Figure 1.1 presents the ROC curve for the All Data model for predicting likelihood of reaching 12 months JSA, henceforth called the JSCI model. The c statistic for the model is 0.8 which is at the high end of what Hosmer and Lemeshow (2000) suggest is acceptable discrimination (a value of 0.8 – 0.9 would be considered excellent discrimination). The ROC curve for a random model has been included for comparison (see Appendix B for the model technical data).

Figure 1.1 ROC curve for JSCI model



All models were built using the logistic procedure in the SAS software package. A stepwise selection was used, initially using all of the variables. A backward selection was also tested as a check. The stepwise model had greater accuracy than the backward selection model, so the stepwise approach was used to identify which variables should be used in the final model. Table 2.4 presents the variables that are included in the final 12-month model as these are the ones that were found to be significant predictors at the 95 per cent level.

Table 2.4 Significant predictors of LTU from the JSCI model

Attribute variable	Attitudinal variable (5 point scale)
Age (banded)	How much do you agree with the following statement: Getting a job is more down to luck than the effort you put in
Age (continuous)	
Gender	
Whether English is first language	
Household composition: Live alone	
Household composition: Live with partner/spouse	How much do you agree with the following statement: I am confident I can get a job within three months
Household composition: Live with a dependent under 18-years-old	
If one parent/guardian was in paid work until respondent was 16-years-old	How much do you agree with the following statement: Having almost any job is better than being unemployed
Has or does a lack of references from a previous employer make it difficult to find or keep jobs	
Any paid work in the last two years	
How much earn per week in last job	**Administrative variable**
Have driving licence for car/motorcycle	Type of JSA claim
Public transport available for commuting to work	Number of days on all out-of-work benefits in the last two years
Have or do problems with public transport make it difficult to find or keep a job	Proportion of working-age population employed in the public sector in local area
Has or is a lack of experience making it difficult to find or keep a job	Average house price for local area

2.3.2 Development of JSCI scoring

The project used the output from the logistic regression to develop a straightforward scoring system that converts responses to the significant predictors detailed above into a probability of an individual in reaching LTU. This was done by using the coefficients for each response as ‘scores’ for each variable and then summing these scores to generate an overall score. The overall score was then transformed to generate a probability that the claimant will reach LTU. (See Appendix C for an explanation of the logistic regression methodology). To make the scoring more immediately understandable, any negative coefficients for categorical variables were adjusted so that they were positive, which means that the larger the value of the score the higher the likelihood of reaching LTU.

The individual scores for the variables found to be significant predictors of LTU are presented in Table 2.5 along with the original estimates from the model and the adjusted values:

Table 2.5 JSCI model scores for variables that are significant predictors of reaching a 12-month JSA claim

Variable	Response	Estimate[a]	Adjusted	Score[b]
Intercept	-4.6173	-10.93456	-10.93456	
Number of days on all out-of-work benefits in the last two years (ESA, IB, IS, JSA)	0.00118	0.00118	0.0354 per 30 days	
Age of respondent	16–24	Reference	0	0.0
	25–34	1.7141	1.7141	1.7
	35–44	1.6769	1.6769	1.7
	45–54	1.2848	1.2848	1.3
	55+	0.7534	0.7534	0.8
	No response	3.3956	3.3956	3.4
Type of JSA claim	Contributory only	-1.1853	0	0.0
	Includes income component	Reference	1.1853	1.2
Agreement with statement: I am/was confident I can/could get a suitable job within three months	Strongly agree	Reference	0	0.0
	Slightly agree	0.6615	0.6615	0.7
	Neither agree nor disagree	0.7722	0.7722	0.8
	Slightly disagree	0.8727	0.8727	0.9
	Strongly disagree	0.7042	0.7042	0.7
	Don't know	1.0978	1.0978	1.1
Work status for the last 2 years	Have had paid work	Reference	0.7244	0.7
	Have not had paid work	0.9212	1.6456	1.6
	Don't know/refused	-0.7244	0	0.0
Whether have problems that make it difficult to keep or find jobs – problems with public transport	No	Reference	0	0.0
	Yes	2.8771	2.8771	2.9
Proportion of working-age population employed in the public sector		-8.1437	-8.1437	-0.8 per 10% increase
Average house price for local area		-1.46E-06	-1.46E-06	-0.1 per £40,000 value
Whether have problems that make it difficult to keep or find jobs – lack of experience	No	Reference	0	0.0
	Yes	1.3891	1.3891	1.4
Whether could use public transport to travel to and from work	Yes	Reference	0.06866	0.1
	No	-0.06866	0	0.0
	Don't know	0.2745	0.34316	0.3

Continued

Table 2.5 Continued

Variable	Response	Estimate[a]	Adjusted	Score[b]
Whether parent/guardian was regularly in paid work	Yes	Reference	1.2858	1.3
	No	0.4583	1.7441	1.7
	Not applicable (e.g. respondent was raised in care)	-1.2858	0	0.0
	Not stated/Refused/Don't know	0.3376	1.6234	1.6
Whether English is first language	Yes	Reference	0.5773	0.6
	No	-0.5773	0	0.0
Weekly income (Employed) – banded or How much did you earn per week in your last job?	£1 – £100	0.6486	1.0942	1.1
	£101 – £200	0.6737	1.1193	1.1
	£201 – £300	0.2094	0.655	0.7
	£301 – £400	-0.4456	0	0.0
	£401 – £500	0.6312	1.0768	1.1
	£501 +	0.2246	0.6702	0.7
	Other	-0.1236	0.322	0.3
	Don't know	1.4705	1.9161	1.9
Gender of respondent	Male	Reference	0.5578	0.6
	Female	-0.5578	0	0.0
Household composition: Live with partner/spouse	No	Reference	0.6294	0.6
	Yes	-0.6294	0	0.0
Household composition: Live alone	No	Reference	0	0.0
	Yes	0.3792	0.3792	0.4
Household composition: Live with dependent child under 18	No	Reference	0	0.0
	Yes	0.654	0.654	0.7
Agreement with statement – getting a job is more down to luck than the effort you put in	Strongly agree	0.5279	0.8508	0.9
	Slightly agree	0.0841	0.407	0.4
	Neither agree nor disagree	-0.3229	0	0.0
	Slightly disagree	0.1557	0.4786	0.5
	Strongly disagree	Reference	0.3229	0.3
	Don't know	-0.0689	0.254	0.3
Whether hold a current full driving licence for a car or motorcycle	Yes	Reference	0	0.0
	No	0.3396	0.3396	0.3
	Don't know	2.4101	2.4101	2.4
Agreement with statement – having almost any job is better than being unemployed	Strongly agree	Reference	0.1884	0.2
	Slightly agree	-0.029	0.1594	0.2
	Neither agree nor disagree	0.5415	0.7299	0.7
	Slightly disagree	0.2218	0.4102	0.4
	Strongly disagree	0.6063	0.7947	0.8
	Don't know	-0.1884	0	0.0

Continued

Table 2.5 Continued

Variable	Response	Estimate[a]	Adjusted	Score[b]
Whether have problems that make it difficult to find or keep jobs – Lack of references	No	Reference	0.3317	0.3
	Yes	-0.3317	0	0.0
Age of respondent (continuous)		0.0439	0.0439	0.04 per year

a: Each categorical variable is assigned a reference category to which the other responses are compared. In all cases the reference category has been assigned to the response category that had the highest number of responses.
b: Categorical variables are rounded to the nearest 0.1. Numerical variables to one significant figure.

The scores for each category of the risk factors identified above highlight aspects of an individual's circumstances that are likely to suggest their increased likelihood of becoming long-term unemployed. As mentioned earlier, a higher score indicates an increased chance of LTU which means the model suggests, all things being equal, the following:

- Living alone or with a dependent under 18-years-of-age increases the chances of reaching 12 months on JSA, whilst those living with a partner/spouse are less likely to reach 12 months.
- If one or both of the individual's parents were regularly in paid work during the claimant's childhood then there is a decreased risk of the longer JSA claim.
- Not having worked at all in the last two years is a strong sign of spending at least 12 months on JSA.
- Men are more likely to have a 12-month claim than women.

The above points are intuitive and, therefore, not unexpected. Perhaps less intuitive are the following:

- Having English as one's first language increases the chances of reaching 12 months on JSA.
- Not having access to public transport for travelling to work increases the chances of leaving JSA.

It is possible that the values of the predictors relating to the English language could relate to the need or desire to learn a second language which might make people more employable and this demonstrates a level of intelligence and willingness to apply oneself to a task. The scoring related to access public transport might be the result of personal car ownership. It also needs to be remembered that the model is predicting exit from JSA not movement into employment. This means that some of the factors may be predicting exits to other destinations, such as the Employment and Support Allowance (ESA) or Training Allowance.

However, one should bear in mind that in predicting the likelihood of LTU it is the score generated by the full range of variables that should be used. Table 2.6 presents two hypothetical examples that serve to demonstrate this:

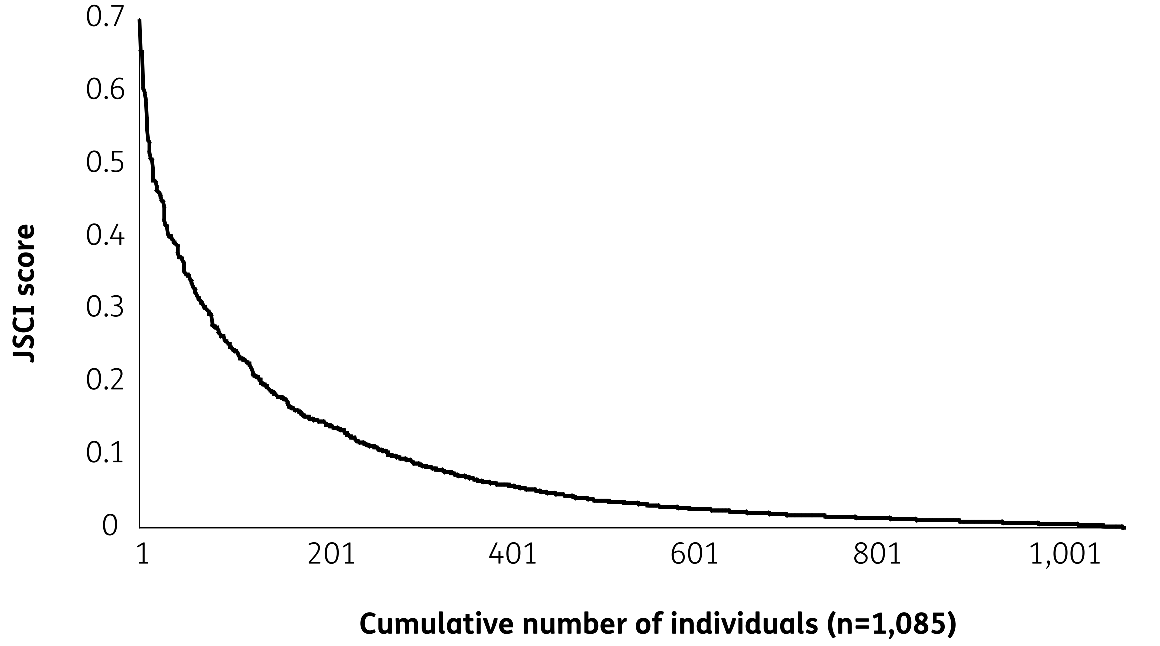
Table 2.6 Examples of how the JSCI score is generated for individuals

	Individual A		Individual B	
DWP JSCI variable	**Response**	**Score**	**Response**	**Score**
Age	28	1.2	50	2.2
Age band	25-34	1.7	45-54	1.3
Gender	Male	0.6	Male	0.6
English first language	Yes	0.6	Yes	0.6
Live alone	Yes	0.4	No	0.0
Live with partner/spouse	No	0.6	Yes	0.0
Live with a dependent under 18 yrs old	No	0.0	No	0.0
One parent/guardian in paid work until 16	Yes	1.3	No	1.7
Any paid work in the last two years	Yes	0.7	Yes	0.7
How much earn per week in last job	£201-£300	0.7	£301-400	0.0
Have driving licence for car/motorcycle	Yes	0.0	Yes	0.0
Public transport available for commuting to work	Yes	0.1	Yes	0.1
Have or do problems with public transport make it difficult to find or keep a job	No	0.0	No	0.0
Has or is a lack of experience making it difficult to find or keep a job	Yes	1.4	Yes	1.4
Has or is a lack of references from a previous employer making it difficult to find or keep a job	Yes	0.0	No	0.3
How much do you agree with the following statement: getting a job is more down to luck than the effort you put in	Strongly agree	0.9	Slightly agree	0.4
How much do you agree with the following statement: I am confident I can get a job within 3 months	Strongly disagree	0.7	Neither agree nor disagree	0.8
How much do you agree with the following statement: Having almost any job is better than being unemployed	Strongly disagree	0.8	Slightly agree	0.2
Type of JSA claim	Contributory only	0.0	Contributory only	0.0
Number of days on all out of work benefits in the last two years (ESA, IB, IS, JSA)	50	0.1	45	0.1
Proportion of working aged population employed in the public sector in local area	Live in Colchester	-1.4	Live in Oldham	-1.7
Average house price for local area	Live in Colchester	-0.3	Live in Oldham	-0.2
Total Score*		**9.9**		**8.4**
Probability of reaching LTU*		**26%**		**7%**

* Total score and probability calculation using full scores, not the rounded numbers

The examples above demonstrate how changes in the responses to questions can result in different scores. When the model was applied to the individuals in the test dataset the scores generated ranged from 70 per cent to less than one per cent. The distribution of the scores is presented in Figure 2.1 below.

Figure 2.1 JSCI score distribution



3 Applying the model and model accuracy

3.1 Segmentation using the JSCI scores

One of the primary purposes of early identification approaches is to identify individual claimants with particular types of outcome so that services can be better targeted. In this instance we are interesting in identifying, at the start of the new claim, those individuals who are likely to reach long-term unemployement (LTU). Binary logistic regression places an individual (or case) into one of two outcome categories based on a selected cut point – in this instance the two outcome categories are 'long-term unemployed' and 'not long-term unemployed'. The logistic regression approach also allows one to provide a score and probability of LTU which means that it can operate as a relative scale – one individual's risk of LTU can be compared with another's – and allows for a straightforward mechanism for applying different cut points to suit the needs of the Department.

If we want then to deliver a differentiated service, we could segment claimants into different groups by ranking individuals on their JSCI model score and then choosing a cut off point (for two groups) or cut off points (for more than two groups).

If we consider segmenting claimants into two groups – those 'high risk' and those 'lower risk' – we would rank the individuals as described above and then choose our cut off point. Anyone with a JSCI score higher than the cut off point would be placed in the 'high risk' category and anyone with a lower score would be placed in the 'lower risk' category. The significant point to note here is that the level at which the cut off point is set influences the accuracy of the actual outcomes.

In the sample of claimants that completed the original survey, eight per cent of individuals reached 12 months on Jobseeker's Allowance (JSA). If we segment individuals based on this figure, so that the eight per cent with the highest JSCI score are classified as high risk and the remainder classified as a lower risk, the model produces the predicted outcomes compared with actual outcomes as presented in Table 3.1.

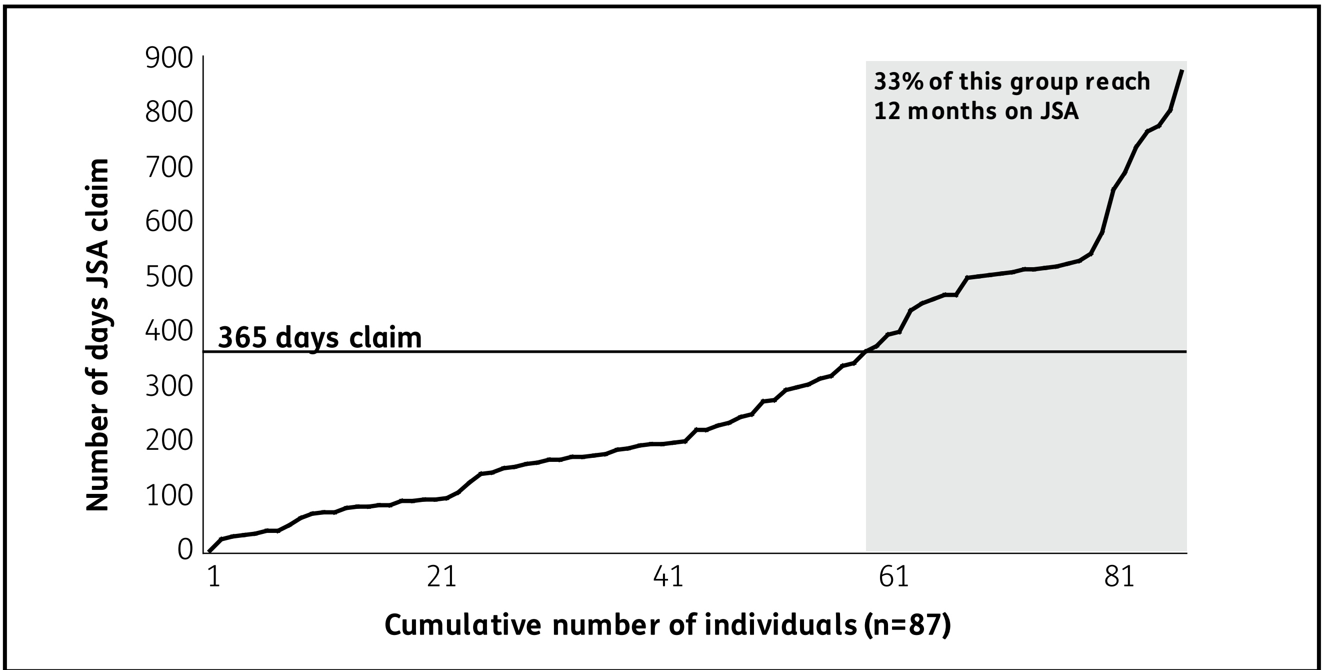
Table 3.1 Predicted outcome against actual outcome – eight per cent highest risk individuals segmentation

		Predicted outcome	
		High risk	**Lower risk**
Actual outcome	High risk	29	62
	Lower risk	58	936
Accuracy		33%	94%

Note: Model applied to test dataset (n = 1,085).

Of those 87 claimants predicted to be at high risk of LTU, 29 actually reach LTU and 58 flow off JSA before 12 months. This is a model accuracy rate of 33 per cent for identifying the high risk category. The full spread of JSA duration for this group is presented in Figure 3.1. For comparison, a random assignment to the two segments would achieve an accuracy figure of eight per cent for identifying the high risk category.

Figure 3.1 Length of JSA claim of the eight per cent highest scoring JSCI segment



In the lower risk category, 936 of those identified as lower risk of reaching LTU do indeed flow off JSA before the 12-month period, while 62 are still on JSA at 12 months. This is a model accuracy rate of 94 per cent in the lower risk category.

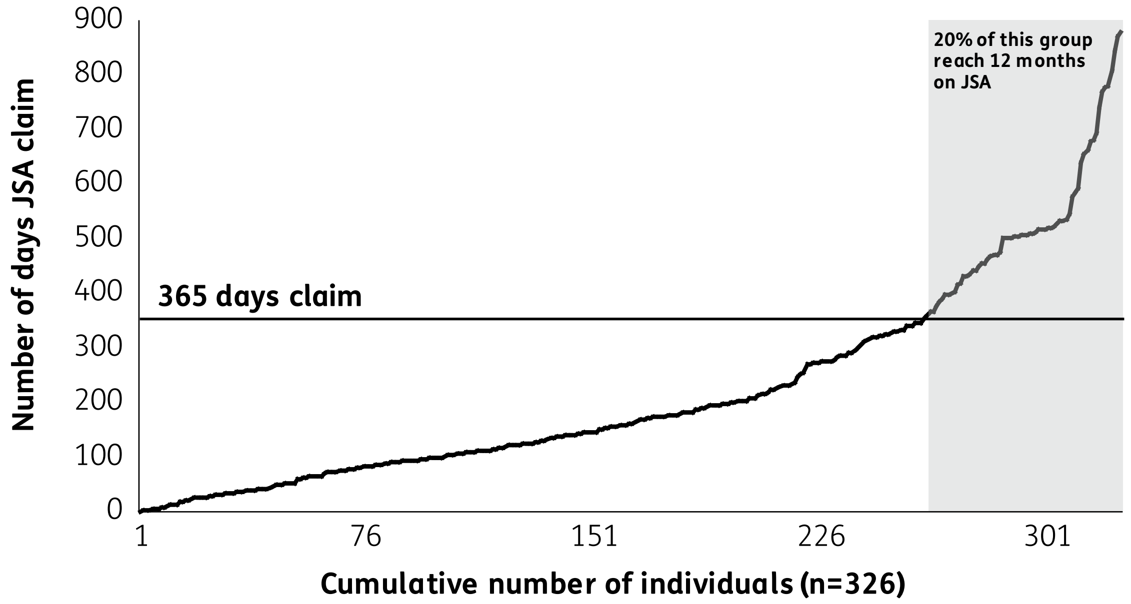
Different levels of accuracy for identifying a high risk category can be achieved by applying the segmentation cut off at different points. The implications of increasing the size of the high risk proportion are shown in Table 3.2 and Figure 3.2. Table 3.2 presents the accuracy of a 30 per cent segmentation, that is, the 30 per cent with the highest JSCI score are classed as high risk of LTU with the remaining 70 per cent classed as lower risk.

Table 3.2 Predicted outcome against actual outcome – 30 per cent highest risk individuals segmentation

		Predicted outcome	
		High risk	**Lower risk**
Actual outcome	High risk	64	27
	Lower risk	262	732
Accuracy		20%	96%

Note: Model applied to test dataset (n = 1,085).

Figure 3.2 Length of JSA claim of the 30 per cent highest scoring JSCI segment



This demonstrates that as we expand the cut off point for high risk (i.e. moving from eight per cent to 30 per cent) we slightly increase the accuracy in the lower risk group (from 94 per cent to 96 per cent), but this is outweighed by a larger decrease in accuracy in the high risk group (from 33 per cent to 20 per cent). As we strive to increase the total number of high risk individuals captured by the model we sacrifice some accuracy when predicting them (the number of false positives has increased significantly). Table 3.3 provides full information on accuracy rates for the model segmenting at each decile.

Table 3.3 Model accuracy rates at different cut off points

Percentage segmented	JSCI score (%)	High risk long term unemployed accuracy (%)	Lower risk long term unemployed accuracy (%)	Overall accuracy (%)
-	70	-	92	92
10	24	31	94	88
20	14	23	95	81
30	8	20	96	73
40	5	17	98	65
50	3	15	98	56
60	2	13	98	47
70	2	11	99	37
80	1	10	99	28
90	1	9	100	18
100	-	8	-	8

3.2 Operational implications for segmentation accuracy

When thinking about the operational implications of the accuracy of the model, it is less useful to consider accuracy only as a single value for the proportion of correct predictions, as described above. How we define the operational accuracy (and what is sufficient accuracy in this instance) depends upon what decisions are based on the segmentation – for example, screening individuals into or out of an intervention – and the cost of any such intervention in relation to the cost of remaining on JSA. The effectiveness of interventions also needs to be considered. Screening claimants for LTU as an isolated process, and without the availability of an effective intervention to reduce the likelihood of a claimant becoming LTU, becomes a nugatory activity. Additionally, there may be potential constraints on the availability of interventions which could influence the way in which the model might be applied.

We can illustrate this 'operational accuracy' point using the figures from the two segmentations presented above. If we ranked all claimants by their risk scores and then targeted the top eight per cent to whom we would then offer additional support we would reach 32 per cent of all those who end up long-term unemployed. If we targeted the top 30 per cent to whom we then offered additional support we would reach 70 per cent of all those who would end up long-term unemployed (see Table 3.4). It is worth noting, however, that using the eight per cent segmentation means that for every individual we correctly identify and treat we would be treating two individuals who would not become long-term unemployed. If we used the thirty per cent segmentation then for every correctly identified individual we would be treating four individuals who would not become long-term employed. If claimants were selected at random for support then choosing eight per cent would only reach eight per cent of the LTU and choosing 30 per cent would reach reach 30 per cent of the LTU

Table 3.4 Operational effectiveness of ranking

	Size of targeted segment	
	Top 8%	**Top 30%**
Number of target segment that reach LTU	29	64
Number of target group that do not reach LTU	58	262
Total number of LTU	91	91
Total number of non-LTU	994	994
Proportion of all LTU captured by segmentation	32%	70%
Proportion of all non-LTU receiving 'unnecessary' support	6%	26%

Note: Model applied to test dataset (n = 1,085).

Alternatively, we could use the figures to screen out individuals for support – if we screened out the bottom 70 per cent of individuals we would lose 30 per cent of the target group.

If we were concerned by the size of error in the high risk category and the associated cost (for example, if we were considering a high cost intervention for all those identified as high risk) we could increase the JSCI score cut off point required for claimants identified as high risk of LTU. This would result in fewer claimants being placed in the high risk category and also less error in those identified as high risk (as a percentage as well as in actual number). However, this means that we would reach fewer people who actually need the intervention and increases the error rate in the lower risk category.

In contrast, if we were trying to identify as many high risk individuals as possible (for example, to ensure a low cost intervention is offered widely), then lowering the cut off point results in less error in the lower risk group, but more error in the high risk group.

Alternatively, if there is a policy need to target a low cost intervention at a certain percentage of JSA claimants, by ranking claimants in order of their JSCI score we can ensure that the resources are targeted using a risk-based metric. By targeting the top 20 per cent of claimants according to their JSCI score, we would reach nearly 55 per cent of those who go on to become long-term unemployed (random targeting would reach 20 per cent). Targeting the top 40 per cent of claimants based on JSCI score would reach 82 per cent of those who become long-term unemployed (random targeting or allocation would reach 40 per cent).

There are other considerations one needs to be aware of with each approach. For example, when considering a 'high cost' intervention there could be unintended consequences for the non-LTU. Most obviously, a training course has lock-in effects, so that inappropriately sending non-LTU on the high cost intervention not only incurs additional expenditure, it could reduce incomes. When considering the low cost approach, in reality it would be possible to segment on more than one occasion. Therefore, we could re-segment those who were assessed as low-risk, but who nevertheless reach six months on JSA.

4 Conclusions

This paper outlines the initial results from the development of a profiling model that aims to predict, at first claim, the likelihood of a new Jobseeker's Allowance (JSA) claimant reaching long term unemployment (LTU). The data used were drawn from a telephone survey of new JSA claimants conducted between May and July 2010 and combined with DWP administrative data.

The model that has been developed is based on statistically significant predictors and is a combination of administrative data, attribute data and attitudinal data. The model testing suggests that, simply in terms of accurately predicting the likelihood of LTU, attitudinal data are less important than might have been assumed.

When considering the application of the model, the paper discusses using an approach that ranks claimants based on their JSCI score prior to segmentation as an efficient way of targeting support.

There are still a number of significant steps that need to be taken to develop our understanding of such an approach and its application in an operational setting. These include:

- **Cost benefit analysis/interventions**
  Judgements about the potential value of segmentation model are relative to the effectiveness and cost of 'treatment'. For instance, it is unlikely to be worth investing in the segmentation of low cost and near universally effective interventions. There is a risk of false positives within any predictive tool. In addition, there are potentially very large costs involved in sending people to expensive interventions that they may not actually need, and which may actually have a negative impact on their likelihood of finding work. This can only be fully explored by an operational test of a segmentation approach.

- **Operational setting**
  The model has been developed and tested using data drawn from a specific, one-off survey. This presents two challenges: a) the results from the questionnaire approach might not be replicated in a setting where the information is collected as a condition of claiming by Jobcentre Plus staff, and b) there are potentially significant hurdles to overcome if such a tool was to be introduced in a live, operational setting. For example, how the score for an individual was calculated would need to be determined, whether through a 'real time' IT solution or through a process solution that required new claimants to provide relevant information at the point of first contact, before any initial interview with an Adviser (at which time the JSCI score will have been generated). It would also be necessary to test whether the change in setting and mode of data collection affected model accuracy.

- **Model refinement and revision**
  a The model was built using data collected in 2010. It is possible that the model would change over time, and thought would be needed about how to keep an operational model under continuous review.

  b The development of a more dynamic model should be considered. This would allow for prediction and segmentation to take place, not just at the start of the claim, but at any point in the claim. Doing so would also raise the possibility of using additional data collected during the claim period that might relate to the claimants job-seeking behaviour as well as adviser assessments.

  c Modelling different outcome variables should be considered. These outcome variables (that might relate, for example, to claimants with complex needs or different claimant journeys) might be more useful operationally.

  d Modelling young people and older people separately should also be considered. This would recognise explicitly the differences between young and old and their likely benefit histories as well as acknowledge that young people are exposed to a different policy environment.

The DWP is undertaking further work in this area to determine a high-level approach to claimant segmentation and to generate the necessary material to inform strategic decision-making about the use of segmentation approaches within this context. The lessons learnt from the JSCI modelling will be fed into this broader thinking.

Appendix A Methodological approach

A.1 Choice of method

In order to identify the most appropriate method to use, four different technical approaches to predict the probability of a claimant remaining on Jobseeker's Allowance (JSA) for 12 months were tested, as follows:

- binary logistic regression;
- artificial neural networks;
- survival analysis (Cox proportional hazard rate regression); and
- decision trees.

The models generated by each of these methods were examined using receiver operating characteristic (ROC) curves. A ROC curve is generated by plotting for a predictive model the rate of true positives against the rate of false positives (in this case, the proportion of long-term unemployment (LTU) correctly identified against the proportion of non-LTU misclassified). For models with a binary outcome the area under the curve (AUC) is identical to the ROC summary statistic known as the c statistic (concordance statistic) and this is a measure of the discrimination of the model. Standard approaches to classifying the accuracy of a model in this way rate a c statistic value of between 0.7 and 0.8 as acceptable (between 0.8 and 0.9 is excellent, between 0.9 and 1.0 is outstanding, between 0.6 and 0.7 is poor and between 0.5 and 0.6 is a fail – no discrimination – note the lowest possible value is 0.5 and is the value that would be generated by random assignment).

These tests revealed that, as presented in Table A.1, logistic regression performed no worse than the other methods. Given this result and the reported use of logistic regression in previous research, logistic regression was chosen as the method to build the JSCI. Additionally, the parameters produced using logistic regression lend themselves more naturally to generating a score for each attribute.

Table A.1 ROC AUC for different models

Model	AUC
Logistic regression	0.772
Neural networks	0.773
Survival analysis	0.732
Decision tree	0.665

A.2 Data collection

A.2.1 Sample size

To determine how big the sample of new claimants should be it was necessary to consider a number of issues. Firstly, an estimate of the proportion of new JSA claimants expected still to be unemployed after 12 months was needed. This was derived from DWP administrative data and in the last year we had data this proportion was seven per cent.

To accurately ascertain the required sample size for a logistic regression model, we also needed to know the following)[13]:

- The number of variables to be fed into the model.
- The distribution of each variable (mean and standard deviation).
- How the variables interact with each other (variance inflation factor).

However, this information about the variables was not available and to obtain it would have required the collection of some data from the questions through a pre-pilot data collection exercise.

There are, however, some 'rules of thumb' covering sampling for logistic regression which were considered when determining the sample size. Peduzzi *et al.* (1996) suggest having at least ten data points in each outcome category for each parameter to be estimated in the model, whilst Pedhazur (1997) suggests 30 data points per parameter to be estimated. The Peduzzi *et al.* suggestion appears more robust when the frequency of the outcome variable is small, as is the case in this modelling exercise.

Based on the Peduzzi *et al.* suggestion, a necessary achieved sample of around 5,000 new claimants was calculated. This calculation took into account the number of factors in the original JSCI, as well as the need for a dataset that could be split into two parts to allow for model building (using one portion of the dataset) and model testing (using the other portion of the dataset)[14].

A.2.2 Sample strategy

Discussion and calculations with TNS-BMRB suggested that a total number of 15,000 new claimants should be selected for the research to deliver the required achieved sample of 5,000. A random sample of new JSA claimants was, therefore, drawn from DWP administrative records – specifically, the JSA General Matching Service (GMS) scans. These scans provide a snapshot of DWP customers who are claiming JSA on the date the scan was run. The scans are run each week, but two scans (two weeks of data) are required to provide complete data on the most recent claims as the data are refreshed for 53 per cent of Jobcentre Plus districts in one week and the remaining 47 per cent of Jobcentre Plus districts the following week. To ensure new claimant data covered all regions equally, the sample was drawn from two scans, the first on 16 April 2010 and the second on 23 April 2010.

The intention was to interview customers as soon as possible after submitting a claim, to capture their attitudes at the start of their claim period and, therefore, to minimise any duration effects on their responses. The complexity of the recruitment process meant that the length of time from sampling to date of interview for the respondents was between 52 and 100 days.

---

13 Having this information ensures that there are sufficient data points to carry out a hypothesis test to see whether any particular parameter is non-zero, subject to a specified maximum Type II error.

14 The 15 factors from the original JSCI would generate around 50 variables, which means a need for at least 500 LTU and 500 non-LTU. Assuming 15 per cent of new JSA claimants are still claiming after 12 months means 3,330 data points would be needed to build the regression model. If the total dataset was split using two-thirds to build the model and one-third to test the model then this results in a total sample requirement of around 5,000 individuals.

A comprehensive check for duplicate records and incomplete records was carried out by TNS-BMRB which resulted in a final set of records from which the sample was drawn. Advance letters were sent out to all in the sample along with a separate proforma and reply paid envelope for its return, to provide customers with a free and easy method for:

- opting out of the survey altogether;
- requesting a telephone interview in a non-English language; or
- requesting a face-to-face interview.

The fieldwork for the survey took place between 24 May 2010 and 11 July 2010 and covered new JSA claimants who made a claim between 2 April 2010 and 23 April 2010.

A.2.3 Response rate

Table A.2 presents the response figures for the survey at each stage of the process. A total of 5,629 interviews with new claimants were completed between May and July 2010 and these claimants were then tracked over a 12-month period to allow the DWP to use the survey data, linked with administrative data, to identify significant predictors of LTU and retrospectively check who did, in fact, reach 12 months on JSA.

Table A.2 JSCI survey response

Opt out letters sent	**15,000**	**100.0%**				
Opt outs received by post and telephone prior to fieldwork	523	3.5%				
Requests for a face-to-face interview prior to fieldwork	15	0.1%				
Surplus sample (not loaded)*	1,766	11.8%				
Sample issued for telephone interview	**12,696**	**100.0%**				
Requested face-to-face interview during fieldwork	22	0.2%				
Contact number invalid or unobtainable	1,916	15.1%				
	Telephone	**%**	**Face**	**%**	**Total**	**%**
Available for survey	10,758	100.0	37	100.0	10,795	100.0
Respondent ineligible** or moved or wrong number	1,467	13.6	1	2.7	1,468	13.6
Valid and eligible contacted respondents	9,291	100.0	36	100.0	9,327	100.0
Refused/unavailable/incapable of interview	3,689	39.7	9	25.0	3,698	39.6
Complete interviews***	5,602	60.3	27	75.0	5,629	60.4

* Only the amount of sample required to achieve the target number of interviews was loaded into the survey to ensure a high response.

** Eligibility was determined by screener questions. Respondents were deemed ineligible if their JSA claim was unsuccessful or if they did not consent to have their answers linked to administrative data held by DWP.

*** Average interview length for telephone interviews in English was 16 minutes. The average interview length for telephone interviews in a non-English language was 50 minutes. The average length for face-to-face interviews was 26 minutes.

Appendix B
JSCI model technical data

B.1 Summary measures of goodness-of-fit

B.1.1 Hosmer and Lemeshow Test

Hosmer and Lemeshow (2000) proposed a statistical test for testing the overall fit of a binary logistic regression model. The test is based on grouping observations into 'deciles of risk' of predicted probabilities. The ordinary chi-square test statistic is calculated using the predicted probability against the observed events and tests how well the model fits the data. If the chi-square goodness-of-fit is not significant (i.e. probability value (p) > .05) then the model has an adequate fit. In other words, the desirable outcome of non-significance indicates that the model prediction does not differ significantly from the observed values.

Table B.1 Decile partition for the Hosmer and Lemeshow Test on JSCI Model

		New_12MON = 1		NEW_12MON = 0	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	434	2	1.78	432	432.22
2	433	1	3.65	432	429.35
3	433	8	5.65	425	427.35
4	433	7	8.24	426	424.76
5	433	10	11.99	423	421.01
6	433	16	17.84	417	415.16
7	433	25	28.35	408	404.65
8	433	53	45.20	380	387.80
9	433	82	74.99	351	358.01
10	436	158	164.30	278	271.70

Table B.2 Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	p
6.7882	8	0.5597

B.1.2 Area Under the ROC Curve

A Receiver Operation Characteristic (ROC) curve is a method for describing classification accuracy for a predictive model and is generated by plotting for the model the rate of true positives (sensitivity) against the rate of false positives (1-specificity) – in this case, the proportion of long-term unemployment (LTU) correctly identified against the proportion of non-LTU misclassified). The area under the curve (AUC) generated by these points provides a measure of the model's discrimination.

The value for the AUC ranges from 0.5 to 1.0. Hosmer and Lemeshow (2000) provide the following general rule for interpreting the values:

Table B.3 Interpreting AUC values

Value	Interpretation
If AUC = 0.5	Suggests no discrimination. What a random allocation would deliver
If 0.7 ≤ AUC < 0.8	Acceptable discrimination
If 0.8 ≤ AUC < 0.9	Excellent discrimination
If AUC ≥ 0.9	Outstanding discrimination

For models with a binary outcome, the area under the curve (AUC) is identical to the ROC summary statistic known as the c statistic (concordance statistic).

Table B.4 C statistics for the JSCI modelling

Model	C statistic
Training dataset (used to build the model)	0.845
Test dataset (used to test the model)	0.79507

B.1.3 Other goodness-of-fit tests

A number of other measures of model fit for the overall model can be generated through the SAS logistic regression procedure.

The Likelihood Ratio Test and the Efficient Score Test both test the joint significance of the explanatory variables of the overall model, with the Likelihood Ratio Test testing the full model against the constant-only model. Significant p-values for these tests indicate a good model fit. (Tabachnick and Fidell, 2007). The **Akaike Information Criterion (AIC), the Schwarz Criterion (SC) and the negative of twice the log likelihood (-2 Log L)** are three model fit estimates that in this instance compare the intercept only model and for the full, fitted model. When comparing, smaller values of the AIC and SC are preferred and indicate superior models.

Table B.5 Likelihood ratio and efficient score tests for the JSCI model

Test	Chi-Square	DF	p value
Likelihood Ratio	583.2423	50	<.0001
Score	615.6701	50	<.0001

Table B.6 AIC, SC and -2 Log L fit estimates for JSCI model

Criterion	Intercept only	Intercept and covariates
AIC	2492.289	2009.046
SC	2498.663	2334.133
-2 Log L	2490.289	1907.046

B.2 Analysis of maximum likelihood estimates

The analysis of maximum likelihood estimates lists the parameter estimates, their standard errors, and the results of the Wald test for individual parameters. The estimates, also known as B coefficients, are the values that are used in the model as the basis of the score calculation.

Table B.7 Analysis of maximum likelihood estimates for the JSCI Model

Parameter	Response	Estimate	Standard Error	Wald	p-value
Intercept		-4.6173	0.6946	44.1834	<.0001
DAYS_CLAIM		0.00118	0.00032	13.5393	0.0002
BREAK2	2	1.7141	0.3105	30.4675	<.0001
BREAK2	3	1.6769	0.4855	11.9303	0.0006
BREAK2	4	1.2848	0.6694	3.6844	0.0549
BREAK2	5	0.7534	0.8726	0.7456	0.3879
BREAK2	99	3.3956	0.8667	15.3503	<.0001
CLAIM_TYPE	1	1.1853	0.226	27.5019	<.0001
QATT_2	2	0.6615	0.1766	14.0237	0.0002
QATT_2	3	0.7722	0.2618	8.6994	0.0032
QATT_2	4	0.8727	0.1951	20.0156	<.0001
QATT_2	5	0.7042	0.206	11.6872	0.0006
QATT_2	6	1.0978	0.3214	11.6672	0.0006
BREAK6	2	0.9212	0.2167	18.0764	<.0001
BREAK6	99	-0.7244	1.5749	0.2116	0.6455
_QPROB8	1	2.8771	0.9679	8.8358	0.003
PUBLIC		-8.1437	2.0669	15.5237	<.0001
AV_HOUSE		-1.46E-06	6.68E-07	4.7972	0.0285
_QPROB10	1	1.3891	0.6326	4.8225	0.0281
QTRANS	2	0.6866	0.2643	6.75	0.0094
QTRANS	3	0.2745	0.3968	0.4784	0.4891
QPARENT	2	0.4583	0.1668	7.5479	0.006
QPARENT	3	-1.2858	0.8278	2.4126	0.1204
QPARENT	5	0.3376	0.3587	0.8858	0.3466
QENGFIR	2	-0.5773	0.2181	7.0086	0.0081
WNETINC	1	0.6486	0.2704	5.7531	0.0165
WNETINC	2	0.6737	0.2283	8.7131	0.0032
WNETINC	3	0.2094	0.2366	0.7836	0.376
WNETINC	4	-0.4456	0.3733	1.4251	0.2326
WNETINC	5	0.6312	0.336	3.5292	0.0603
WNETINC	6	0.2246	0.3419	0.4315	0.5112
WNETINC	7	-0.1236	0.7696	0.0258	0.8724
WNETINC	8	1.4705	1.1152	1.7385	0.1873
QGENDER	2	-0.5578	0.1535	13.2059	0.0003
_QLIVE2	1	-0.6294	0.1967	10.2352	0.0014
_QLIVE3	1	0.654	0.1766	13.7088	0.0002
QATT1D	1	0.5279	0.167	9.9975	0.0016
QATT1D	2	0.0841	0.192	0.1918	0.6614

Continued

Table B.7 Continued

Parameter	Response	Estimate	Standard Error	Wald	p-value
QATT1D	3	-0.3299	0.2876	1.3154	0.2514
QATT1D	4	0.1557	0.1785	0.7605	0.3832
QATT1D	6	-0.0689	0.3913	0.031	0.8603
_QLIVE1	1	0.3792	0.1649	5.2883	0.0215
QDRIVE	2	0.3396	0.1373	6.1219	0.0134
QDRIVE	3	2.4101	1.3639	3.1225	0.0772
QATT1C	2	-0.029	0.1608	0.0326	0.8568
QATT1C	3	0.5415	0.3051	3.1504	0.0759
QATT1C	4	0.2218	0.1987	1.246	0.2643
QATT1C	5	0.6063	0.2115	8.2145	0.0042
QATT1C	6	-0.1884	0.7087	0.0707	0.7903
_QPROB1	1	-0.3317	0.1622	4.1847	0.0408
QAGE_1		0.0439	0.0221	3.9491	0.0469

B.3 Analysis of type 3 effects

The analysis of type 3 effects presents tests of significance for each of the predictors in the model.

Table B.8 Type 3 effects for the JSCI model

Effect	DF	Wald	p value
DAYS_CLAIM	1	13.5393	0.0002
BREAK2	5	71.1111	<.0001
CLAIM_TYPE	1	27.5019	<.0001
QATT_2	5	27.4914	<.0001
BREAK6	2	18.7479	<.0001
_QPROB8	1	8.8358	0.0030
PUBLIC	1	15.5237	<.0001
AV_HOUSE	1	4.7972	0.0285
_QPROB10	1	4.8225	0.0281
QTRANS	2	7.4535	0.0241
QPARENT	3	10.6143	0.0140
QENGFIR	1	7.0086	0.0081
WNETINC	8	19.0838	0.0144
QGENDER	1	13.2059	0.0003
_QLIVE2	1	10.2352	0.0014
_QLIVE3	1	13.7088	0.0002
QATT1D	5	14.3526	0.0135
_QLIVE1	1	5.2883	0.0215
QDRIVE	2	8.8369	0.0121
QATT1C	5	11.8118	0.0375
_QPROB1	1	4.1847	0.0408
QAGE_1	1	3.9491	0.0469

Appendix C
Logistic regression

In this instance, logistic regression fits the model:

$$\mathrm{Log}(p/(1-p)) = b_0 + b_1X_1 + b_2X_2 + \dots b_nX_n$$

Where p is the probability of having a continuous 12 month Jobseeker's Allowance (JSA) claim and $X_1, X_2 \dots X_n$ are the variables (e.g. age, gender, type of JSA claim). The logistic procedure algorithm then generates the coefficients for each variable ($b_1, b_2 \dots b_n$) that will give the best fit for the data.

The coefficients $b_1, b_2, b_3 \dots b_n$ can be regarded, therefore, as a ‘score’ for each variable (with a higher score indicating an increased chance of reaching long-term unemployment (LTU)) and with a total score, S, being the sum of the individual coefficients, that is:

$$S = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \dots b_nX_n$$

If one applies the transformation:

$$p = \exp(S)/(1+\exp(S))$$

to the total score, S, then p gives the probability that the claimant will become long-term unemployed.

Appendix D
Model building for different JSA durations

The report explains that the main purpose of the work was to build a model to predict the likelihood of a new claimant reaching long-term unemployment (12 months or more continuous Jobseeker's Allowance (JSA) claim). However, it was recognised that it could be helpful to gather information relating to predicting different JSA duration periods and, therefore, models were built to predict both three months and six months continuous JSA claims. The interesting point to note here is the difference in the variables that are significant predictors for three months, six months and twelve months.

D.1 Predicting three months continuous JSA claim

D.1.1 Administrative data

Rapid reclaim.
Type of JSA claim.
Time on JSA in last two years.
Time on ESA in last two years.
Number of other working-age benefit claims in the last two years.

D.1.2 Attribute data

Main activity in the last 12 months.
Any paid work at all in the last two years.
If worked in last two years, average number of hours worked per week.
If worked in last two years are you earning £1–£100 per week and not self-employed or earning over £100 per week or self-employed.
If worked in last two years and not self-employed whether manager/supervisor.
Lives alone.
Lives with dependant under 18.
Lives with family other than parent, partner, child, sibling, grandparent.
Do you have qualifications that will help you get a job.
Is English your first language.
Numeracy (self-assessed).
Public transport possible for commute.
Car/motorbike possible for commute.
Health issues that might affect the amount of work that can be done.
Refused to answer whether they have problems that make it difficult to find or keep jobs.

D.1.3 Attitude data

Agreement with ‘I am confident I can get a suitable job within three months.
Agreement with ‘Even if I don’t succeed first time, I can continue job seeking, setbacks don’t put me off.

D.2 Predicting six months continuous JSA claim

D.2.1 Administrative data

Rapid reclaim.
Type of JSA claim.
Time on JSA in last two years.
Number of JSA claims in last two years.
Number of other working age benefit claims in the last two years.
Government Office Region.
Method of payment.

D.2.2 Attribute data

Type of accommodation.
Main activity in last 12 months.
Any paid work in last two years.
If paid work in last two years, number of hours worked per week.
Self-employed in last job?
If not self-employed what was weekly income in last job.
If working but not self-employed did you have managerial responsibilities.
In childhood were your parents regularly in paid work.
Do you hold a full driving licence for car/motorbike.
Would commuting by public transport be possible.
Age of youngest child.
Lives with relative other than partner, child, parent, grandchild, sibling or dependent adult.
Any personal problems that make it difficult to find or keep jobs.
Gender.
Numeracy level (self-assessed).
Age.

D.2.3 Attitude data

Confidence in ability to get a job in the next three months.
Willingness to accept a job for less money than the previous one.

Appendix E
Survey questionnaire

Main questionnaire for JSCI (45109178)

Job seeker screening

Introduction/screening questions

Please may I speak to [CONTACT NAME?]

Good morning/afternoon/evening. My name is .... and I am calling on behalf of the British Market Research Bureau, an independent market research company.

We are conducting an important study for the Department for Work and Pensions, the government department that deals with benefit issues. We would like to learn more about people who recently used the Jobcentre Plus services so these services can be improved in the future.

You should have received a letter telling you what the survey is about and asking if you would be willing to participate, do you recall seeing it? (If no, briefly outline the purpose and content of the survey.)

The interview will take no longer than 20 minutes. Your help would be much appreciated, can we go ahead now?

IF NECESSARY: The findings from the research will be used for research purposes only.

Continue with interview
Refused – Terminates
Make appointment

IF INTRO = Continue with interview

I just want to reassure you that this is confidential, voluntary market research. Thank you for agreeing to participate.

I just need to start by asking you some questions about your recent claim for Jobseeker's Allowance.

QS1 [ASK ALL]

Can I just check, was the last claim you made successful? By successful I mean that you have received at least one Jobseeker's Allowance payment since then.

IF CLAIM IS CURRENTLY UNDER APPEAL CODE AS 4

1. Yes
2. No, claim not successful
3. No, never made claim
4. Don't know yet/haven't heard

IF S1 = NO/NOT SUCCESSFUL OR NO/NEVER MADE A CLAIM, TERMINATE INTERVIEW

QRAPID2 [ASK IF S1 = 1 OR 4]

When you made your recent claim for Jobseeker's Allowance, do you know if it was part of the rapid reclaim process? A claim is likely to be part of the rapid reclaim process if you have made a previous claim for Jobseeker's Allowance in the last six months AND your circumstances have NOT changed during this period.

IF NECESSARY: This is where you just need to answer a few questions, instead of giving full details about yourself and your circumstances.

1. Yes, rapid reclaim
2. No/gave full details
3. Don't know

QLINK [ASK IF S1 = YES]

We will be able to learn more about benefit claimants and how to improve services to meet their needs by linking your answers from these questions to administrative records held by the Department for Work and Pensions. This linked data will only be used by the research team at the Department for Work and Pensions.

Please remember that any claim for benefit either now or in the future will not be affected. Could we have your permission to link your answers to administrative data?

INTERVIEWER NOTE: If asked, the administrative data comes from the information provided when they made their claim.

1. Yes
2. No
3. Don't know

IF LINK = NO OR DON'T KNOW TERMINATE; IF YES CONTINUE WITH INTERVIEW

Household and Demographics

I'd like to ask you some questions about yourself and your household.

QLIVEWITH (JSC31) **[ASK ALL[15]]**

Who normally lives with you? PROMPT TO PRECOCES. MULTICODE

1. Live alone (Single code)
2. Partner/spouse (husband/wife/boyfriend/girlfriend), includes same-sex partner
3. Dependent child/children under 18
4. Dependent child/children over 18
5. Parent(s)
6. Dependent adult (someone respondent has caring responsibilities for)
7. Other
8. Don't know
9. Refused

15 In the remainder of the questionnaire, 'ask all' means 'ask all except those terminated at screening questions'.

QCHILD (JSC33) **[ASK IF DEPENDENT CHILD UNDER 18]**

Can you tell me the age of the youngest child?
TYPE IN AGE IN YEARS

QTENURE (not in JSCI) **[ASK ALL]**

What best describes the accommodation you currently live in? READ OUT

1. Accommodation owned outright
2. Being bought with a mortgage or loan
3. Rented from the council/housing association/new town
4. Rented from someone else (including parents/friend/private landlord)
5. Permanently living with relatives/friends rent free
6. Squatting/no permanent address
7. Other (specify)
8. Don't know
9. Refused

QMOVE1 **[ASK ALL]**

Have you moved house in the last 12 months?

1. Yes
2. No
3. Don't know
4. Refused

QMOVE2 (JSCI29) **[ASK IF MOVE1 = 1]**

How many times have you moved house in the last 12 months?

ENTER NUMBER
Don't know
Refused

QTENURE2 (JSCI28) **[ASK ALL EXCEPT TENURE = OWN OUTRIGHT OR WITH MORTGAGE/LOAN]**

Are you currently staying in emergency or temporary accommodation?

INTERVIEWER NOTE: This type of accommodation is **unplanned** and **excludes** people who have **chosen** to rent temporarily whilst trying to secure longer accommodation, such as between buying and selling a house.

1. Yes
2. No
3. Don't know
4. Refused

QETHNIC (JSCI16-20) **[ASK ALL]**

Would you describe yourself as....? READ OUT

1. White
2. Asian or Asian British
3. Black or Black British
4. Chinese

5. or from a mixed ethnic group
6. Other
7. Refused

QWHITE [IF WHITE]

And would you describe yourself as.? READ OUT

1. White British
2. White Irish
3. Or from another white background
4. Don't know
5. Refused

QASIAN [IF ASIAN]

And would you describe yourself as....? READ OUT

1. Pakistani
2. Indian
3. Bangladeshi
4. Or from another Asian background
5. Don't know
6. Refused

QBLACK [IF BLACK]

And would you describe yourself as....? READ OUT

1. Caribbean
2. African
3. Other Black background
4. Don't know
5. Refused

QMIXED [IF MIXED]

And would you describe your self as....? READ OUT

1. White and Black Caribbean
2. White and Black African
3. White and Asian
4. or from another mixed background
5. Don't know
6. Refused

Work experience

I would now like to ask you some questions about recent work experience you may have.

QWORKN (JSCI1) **[ASK ALL]**

Thinking about the last 12 months as a whole, which of these best describes what you have been doing?

READ OUT

1. In paid work
2. In unpaid work, including volunteering
3. Unemployed and looking for work or waiting to take up a job
4. On a government scheme (e.g. New Deal)
5. In training or education (school, college or vocational training)
6. Looking after the family or the home
7. In prison or other detention
8. Not working for other reason (specify)
9. Don't know
10. Refused

QWORKC (not in JSCI) **[ASK ALL]**

Are you currently in paid work?

ADD IF NECESSARY: That includes any casual jobs but NOT including holiday jobs? Please include any self-employed work.

By paid work I mean doing anything where you were actually paid a regular wage, no matter who this was paid by.

INTERVIEWER: TIME SPENT ON A GOVERNMENT TRAINING SCHEME SHOULD NOT BE COUNTED AS WORK. IF RESPONDENT IS CURRENTLY OFF SICK BUT HAS A JOB TO GO BACK TO PLEASE CODE THIS AS HAVING A JOB.

1. Yes
2. No
3. Don't know
4. Refused

QWORKL (JSCI3) **[ASK ALL EXCEPT WORKN = IN PAID WORK OR WORKC = YES]**

Have you done any paid work at all in the last two years, including any casual jobs but NOT including holiday jobs? Please include any time spent self-employed.

By paid work I mean doing anything where you were actually paid a regular wage, no matter who this was paid by.

INTERVIEWER: TIME SPENT ON A GOVERNMENT TRAINING SCHEME SHOULD NOT BE COUNTED AT WORK.

1. Yes
2. No
3. Don't know
4. Refused

QCONST1 (JSCI54) **[ASK IF LIVEWITH = 2]**

Is your partner currently in PAID work?

INTERVIEWER INSTRUCTION: IF PARTNER IS CURRENTLY OFF SICK BUT HAS A JOB TO GO BACK TO PLEASE CODE THIS AS HAVING A JOB.

1. Yes
2. No

QPARENT (JSCI46) **[ASK ALL]**

Up until you were 16, was at least one of your parents or guardians regularly in paid work?

IF NECESSARY: We are asking this question because values and beliefs on work can often be influenced from experiences as a child/young person.

1. Yes
2. No
3. Don't know
4. Refused
5. Not applicable (e.g. respondent was raised in care)

QHOURS (JSCI2) **[ASK ALL WHO HAVE WORKED IN LAST 2 YEARS AT WORKN, WORKC OR WORKL]**

Thinking about your current [IF CURRENTLY IN WORK]/most recent [IF NOT CURRENTLY IN WORK] paid job ...

About how many hours a week do/did you usually work in your job, excluding meal breaks but including any usual paid overtime?

READ OUT

1. 40 hours or more
2. 16-39 hours
3. Less than 16 hours
4. Irregular hours
5. Don't know

QNEMP (JSC50) **[ASK ALL WHO HAVE WORKED IN LAST 2 YEARS AT WORKN, WORKC OR WORKL]**

Still thinking about your current [IF CURRENTLY IN WORK]/most recent [IF NOT CURRENTLY IN WORK] paid job ...

Were you working as an employee or were you self-employed?

1. Employee
2. Self-employed

QFIRM [ASK ALL WHO HAVE WORKED IN LAST 2 YEARS AT WORKN, WORKC OR WORKL]

Still thinking about your current [IF CURRENTLY IN WORK]/most recent [IF NOT CURRENTLY IN WORK] paid job ...

What type of firm do/did you work for?

Open ended – may be DK or Refused

QINVOLVE [ASK ALL WHO HAVE WORKED IN LAST 2 YEARS AT WORKN, WORKC OR WORKL]

Still thinking about your current [IF CURRENTLY IN WORK]/most recent [IF NOT CURRENTLY IN WORK] paid job ...

What does/did the work involve?

Open ended – may be DK or Refused

QJOBTIT [ASK ALL WHO HAVE WORKED IN LAST 2 YEARS AT WORKN, WORKC OR WORKL]

Still thinking about your current [IF CURRENTLY IN WORK]/most recent [IF NOT CURRENTLY IN WORK] paid job ...

What is/was the job title of the person you report/reported to?

Open ended – may be DK or Refused

QMANA (JSCI50) **[ASK IF QNEMP=1]**

Do/did you have any managerial duties, or do/did you supervise any other employees?

1. Manager
2. Foreman/supervisor
3. Not manager supervisor
4. Don't know

QHOWMAN **[ASK IF MANA = 1 or 2]**

How many people are/were you responsible for?

0-999999
Answer may be DK or Refused

QEMPN (JSCI50) **[ASK IF QNEMP=1]**

How many employees work(ed) there at the place where you work(ed)? Is/was it under 25 or more?

INTERVIEWER: If the respondent is/was the only employee (e.g. a nanny) code as 1-24

1. 1-24
2. 25 or more
3. Don't know

QSOLO (JSCI50) **[ASK IF QNEMP=2]**

Do/did you work on your own or do/did you have employees?

1. On own/with partner(s) but no employees
2. With employees

QSEEM (JSCI50) **[ASK IF SOLO=2]**

How many people do/did you employ at the place where you work/ed? Is/was it under 25 or more?

1. 1-24
2. 25 or more
3. Don't know

QNETINC (JSCI52) **[ASK IF QNEMP = 1]**

When you were last paid in that job, how much take-home pay did you receive, that is after any deductions were made for tax, National Insurance and so on, but including overtime pay, bonus, commission, tips and so on?

IF DON'T KNOW OR CAN'T REMEMBER PROBE: Can you give me an approximate amount?

0.999999
Don't know
Refused

QNETINT (JSCI52) **[ASK IF NETINC>0]**

How long a period did that cover?

1. One week
2. Fortnight
3. Four weeks
4. One calendar month
5. Year
6. Other (specify)

QSEINC (JSCI52) **[ASK IF QNEMP = 2]**

On average, what was your WEEKLY or MONTHLY income from your self-employment over the last 12 months, that is after taking away all expenses and deductions such as taxes, National Insurance contributions, etc?

IF DON'T KNOW OR CAN'T REMEMBER PROBE : Can you give me an approximate amount?

0.999999
Don't know
Refused

QSEINCT (JSCI52) **[ASK IF SEINC>0]**

And was that weekly or monthly income?

1. Weekly income
2. Monthly income
3. Some other period (specify)

QJOBEND (JSC51) **[ASK ALL WHO HAVE WORKED IN LAST 2 YEARS BUT NOT WORKING CURRENTLY]**

What was the main reason that your last job came to an end? DO NOT PROMPT

1. Made Redundant
2. Dismissed/sacked
3. Left because I did not like it
4. Temporary job ended
5. I became ill and had to leave
6. Caring responsibilities
7. Moved area
8. Left for another job
9. Other reason (TYPE IN)
10. Don't know

QATT1 [ASK ALL]

I'm now going to read out some statements about work. Can you tell me how much you agree or disagree with each of these statements?

IF NECESSARY: Is that strongly or slightly agree/disagree

1. Strongly Agree
2. Slightly Agree
3. Neither Agree Nor Disagree
4. Slightly Disagree
5. Strongly Disagree
6. (Don't know/No Opinion)

RANDOMISE ORDER OF STATEMENTS

a. A person must have a job to feel a full member of society.
b. In general, the people I know think that if you are unemployed, it is important to keep looking for a job.
c. Having almost any job is better than being unemployed.
d. Getting a job is more down to luck than the effort you put in.
e. Benefits give a more stable income than trying to earn a wage.
f. Once you've got a job, it's important to hang on to it, even if you don't really like it.
g. You need to be in a job to get a better job.
h. I don't much care whether I work or not.
i. The right job is more important than any job.

Education/skills

I would now like to ask you some questions about education.

Qend (JSCI4) **[ASK ALL]**

At what age did you finish your continuous, full-time education at school or college?

ENTER ANSWER
Don't know
Refused

QQUAL (JSCI5) **[ASK ALL]**

Do you have any qualifications either from school, college or university, or connected with work or from government schemes?

1. Yes
2. No
3. Don't Know

QUALHIGH (JSCI6) **[ASK IF QUAL = 1]**

What is the highest qualification you have?

PROMPT AS NECESSARY. SINGLE CODE. PRIORITY CODE: IF TWO OR MORE ANSWERS GIVEN, CODE ANSWER WHICH IS HIGHER UP LIST.

1. Degree level qualification including foundation degrees, graduate membership of a professional institute, PGCE or higher
2. Diploma in higher education
3. HNC/HND
4. ONC/OND
5. BTEC/BEC/TEC/EdExcel
6. Teaching qualification (excluding PGCE)
7. Nursing or other medical qualification not yet mentioned
8. Other higher education qualification below degree level
9. A-level/Vocational A-level or equivalent
10. International Baccalaureate
11. NVQ/SVQ
12. GNVQ/GSVQ
13. AS-Level/Vocational AS-Level or equivalent
14. Access to HE
15. O-Level or equivalent
16. GCSE/Vocational GCSE
17. CSE
18. RSA/OCR
19. City and Guilds
20. YT Certificate
21. Key Skills
22. Basic Skills
23. Any other professional/vocational/foreign/other qualification
24. None
25. Don’t Know
26. Refused

QUALJOB (JSCI9) **[ASK IF QUAL = 1]**

[IF CURRENTLY IN PAID WORK: Have your qualifications helped you to get a job?][IF NOT CURRENTLY IN WORK: Do you think your qualifications will help you to get a job?]

1. Yes
2. No
3. Don’t Know

QDRIVE (JSCI34) **[ASK ALL]**

Do you hold a current full driving licence to drive a car or motorcycle?

1. Yes
2. No
3. Don’t know

QCAR (JSCI35) **[ASK ALL]**

[Although you don’t have a driving licence, do/Do] you have access to a car or motorcycle that you could use to travel to and from work?

1. Yes
2. No
3. Don't know

QTRANS (JSCI36) **[ASK ALL]**

Is there any public transport that you could use to travel to and from work?

1. Yes
2. No
3. Don't know

QENGFIRST (NOT IN JSCI) **[ASK ALL]**

Can I just check, is English your first language?

1. Yes, English is first language
2. No, English not first language

QSPEAK (JSCI12) **[ASK IF ENGFIRST = 2]**

How good are you at speaking English when you need to in daily life, for example to have a conversation on the telephone or talk to a professional, such as a teacher or a doctor? READ OUT

1. Very good
2. Fairly good
3. Below average
4. Poor
5. Don't know

QREAD (JSCI13) **[ASK ALL]**

How good are you at reading English when you need to in daily life? For example: reading newspapers and magazines or instructions for medicine or recipes? READ OUT

1. Very good
2. Fairly good
3. Below average
4. Poor
5. Cannot read English
6. Don't know

QWRITE (JSCI14) **[ASK ALL]**

And how good are you at writing in English when you need to in daily life? For example: writing letters or notes or filling in official forms? READ OUT

1. Very good
2. Fairly good
3. Below average
4. Poor
5. Cannot write English
6. Don't know

QNUMBER (JSCI53) **[ASK ALL]**

And how good are you at working with numbers when you need to in everyday life? For example working out your wages or benefits, or checking bills and statements? READ OUT

1. Very good
2. Fairly good
3. Below average
4. Poor
5. Don't know

Health and Barriers

I would now like to ask you a few questions about your health.

QHEALTH (JSCI21) **[ASK ALL]**

Do you have any physical or mental health conditions at present, that affect your ability to work?

1. Yes
2. No
3. Don't know
4. Refused

QHEALTH2 (not in JSCI) **[ASK IF HEALTH = 1]**

[Does this/do these] conditions affect....? READ OUT CODE ALL THAT APPLY

1. The KIND of paid work that you might do
2. The AMOUNT of paid work that you might do
3. Or neither

QDISTYPE (JSCI22) **[ASK IF HEALTH 2 =2]**

You said that the condition(s) affect(s) the amount of paid work that you might do. What is the most number of hours per week that you think you are able to work? PROMPT TO PRECODES

1. 40 hours or more
2. 16-39 hours
3. Less than 16 hours
4. (DO NOT PROMPT) Not able to work at all
5. Don't know

QHEALTH3 (JSCI25) **[ASK IF HEALTH = 1]**

How long do you think your condition will affect your ability to work? READ OUT

1. Less than three months, or
2. Three months or more
3. Don't know

QLEARNDI (JSCI4) **[ASK ALL]**

Do you think that you have a learning difficulty or has anyone ever mentioned that you have a learning difficulty?

IF NECESSARY; Some people have difficulty with learning. A learning difficulty may mean that you received special help at school, had a Statement of Special Educational Needs or attended a Special School. This also includes general difficulties with learning and specific difficulties such as dyslexia.

1. Yes
2. No
3. Don't Know
4. Refused

QATT [ASK ALL]

Can you tell me how much you agree or disagree with each of these statements? [IF IN WORK TO APPEAR AT THE TOP OF THE SCREEN THROUGHOUT: Please answer thinking about the last time you were out of work]

Is that strongly or slightly agree/disagree

1. Strongly Agree
2. Slightly Agree
3. Neither Agree Nor Disagree
4. Slightly Disagree
5. Strongly Disagree
6. (Don't know/No Opinion)

RANDOMISE ORDER OF STATEMENTS

1 I have/had [IF IN WORK] enough help and support from friends and family to keep looking for the right job.
2 I am/was [IF IN WORK] confident I can/could get a suitable job within three months.
3 I am/was [IF IN WORK] sure I can/could work and keep a job when I get one.
4 I have/had [IF IN WORK] a job to aim for that will suit me/suited.
5 The jobs that I have/had [IF IN WORK] in mind are available locally.
6 I have the knowledge, skills and experience to do my chosen job.
7 I can make applications and get an interview for my chosen job.
8 I don't have difficulty persuading an employer to take me on.
9 I can contact employers about jobs even when they are not advertising one.
10 Even if I don't succeed first time, I can keep going with my job seeking, setbacks don't put me off.

QAT [ASK ALL]

And can you tell me how much you agree or disagree with each of these statements?

[IF IN WORK: Please answer thinking about the last time you were out of work]

Is that strongly or slightly agree/disagree?

1. Strongly Agree
2. Slightly Agree
3. Neither Agree Nor Disagree
4. Slightly Disagree

5. Strongly Disagree
6. (Don't know/No Opinion)

RANDOMISE ORDER OF STATEMENTS

1 I have/had [IF IN WORK] difficulties that might make/have made it hard for me to find or keep a job.

2 I am/was [IF IN WORK] willing to accept a job for less money than my previous one.

QAT2 [ASK ALL]

And how much do you agree or disagree with the following statements?

Is that strongly or slightly agree/disagree?

1. Strongly Agree
2. Slightly Agree
3. Neither Agree Nor Disagree
4. Slightly Disagree
5. Strongly Disagree
6. (Don't know/No Opinion)

RANDOMISE ORDER OF STATEMENTS

1 If I can't get the type of job I want I will look for other types.

2 I would be prepared to move to a different area for the sake of a job, if it was alright in other ways.

Finally, I'd just like to ask you a few questions about yourself.

QINTER1 [ASK ALL]

Do you have access to the internet?

1. Yes
2. No

QINTER2 [ASK IF HAVE ACCESS TO THE INTERNET]

Have you used the internet for job searching or applying for jobs?

1. Yes
2. No

QPROB (JSCI48) **[ASK ALL]**

Have any of these problems made it difficult for you to find or keep a job in the past year? READ OUT

1. Lack of references from previous employer
2. Debt or money problems
3. No permanent place to live
4. Problems with the law, or a previous record

5. Problems with drugs or alcohol
6. Any other problems that have made it difficult for you to find or keep a job in the past year? (TYPE IN)
7. None
8. Don’t know

QCRIME (JSCI37) **[ASK ALL]**

And can I just check, have you ever had a criminal conviction, excluding minor motoring offences?

1. Yes
2. No
3. Don’t know
4. Refused

QCRIME2 (JSCI38) **[IF CRIME = 1]**

How many years ago was this (ADD IF NECESSARY: the last time that it happened)? PROBE FOR ESTIMATE IF NECESSARY. IF LESS THAN ONE YEAR, CODE ZERO

TYPE IN NUMBER OF YEARS

Don’t know

Refused

QCRIME3 (JSCI39) **[IF CRIME = 1]**

Have you ever spent time in prison at any point?

1. Yes
2. No
3. Refused

QCRIME4 (JSCI40) **[IF CRIME3 = 1]**

How long a prison sentence were you given?

IF NECESSARY: The last time you were sentenced.

TYPE IN NUMBER AND THEN CODE MONTHS OR YEARS ON NEXT SCREEN

Don’t know
Refused

QCRIM4A

CODE WHETHER MONTHS OR YEARS

1. Months
2. Years
3. Refused

QCRIME5 (JSCI41) **[IF CRIME = 1]**

Are you currently under probation supervision?

1. Yes
2. No
3. Don't know
4. Refused

QHOW [ASK ALL]

How do you think Jobcentre Plus could improve their services?

QGENDER [ASK ALL]

INTERVIEWER RECORD GENDER

1. Male
2. Female

QAGE [ASK ALL]

What was your age last birthday?
Refused

QRECON [ASK ALL]

BMRB will be conducting some further research in six to eight weeks time on behalf of DWP and may want to re-contact some people we have talked to in this survey. Would you be happy for us to re-contact you again?

IF NECESSARY: The interview is likely to take around ten minutes

1. Yes
2. No
3. Don't know

CONFIRM NUMBER

Did I correctly dial *********

CLOSE SCREENS

Thank you very much for your time and help.

I would like to confirm that my name is ..., calling from Kantar Operations. This interview was conducted within the Code of Conduct of the Market Research Society. All your replies will be treated in the strictest confidence.

If you would like to check any details about the interview, I can give you the relevant number to call.

IF YES:

1. To verify 'Kantar Operations' as a registered Market Research Organisation, with a professional Code of Conduct, please call the Market Research Society's verification service on Freephone 0500 39 69 99 – which will connect you free of charge.

2. For further information about my company or the nature of this particular survey, you may contact The Telephone Centre Manager ... (SELECT NAME FROM BELOW) during office hours on Freephone ... (SELECT FROM BELOW which will connect you free of charge.

TROJAN: Julie Avery 0500-126-542
EALING BROADWAY: Melanie Wymer 0800-015-1037
HULL: Lynn Stirling 0500-090-243
FARRINGDON: Danny Millar 0800-970-5450

THANK AND CLOSE

CONFIRM SEG

NOW WORK OUT DETAILED SOCIAL GRADE AND CODE AT NEXT SCREEN
TYPE OF FIRM ****
JOB ****
EMPLOYMENT STATUS ****
NO. OF PEOPLE WORK AT SAME PLACE ****
NO. OF PEOPLE RESPONSIBLE FOR ****
POSITION/RANK/GRADE ****
REPORT TO ****
CODE SOCIAL GRADE HERE

1 A

2 B

3 C1

4 C2

5 D

6 E

References

Australia. Department of Education, Employment and Workplace Relations (2012) *Job Seeker Classification Instrument – factors and points*. Canberra: Australian Government, Department of Education, Employment and Workplace Relations.

Bimrose, J., Barnes, S., Brown, A., Hasluck, C. and Behle, H. (2007) *Skills Diagnostics and Screening Tools: a Literature Review*. London: Department for Work and Pensions. Research Report no. 459.

Bryson, A. and Kasparova, D. (2003) *Profiling Benefit Claimants in Britain: a Feasibility Study*. London: Department for Work and Pensions. Research Report196.

Chapman, B. and Smith, P. (1993) *Predicting the Long-Term Unemployed: a Primer for the Commonwealth Employment Service*. Canberra: Centre for Economic Policy Research. Discussion Paper 285.

Freud, D. (2007) *Reducing Dependency, Increasing Opportunity: Options for the Future of Welfare to Work*. London: Department for Work and Pensions.

Gregg, P. (2008) *Realising Potential: a Vision for Personalised Conditionality and Support*. London: Department for Work and Pensions.

Hasluck, C. (2004) *Targeting Services in the Individual Customer Strategy: the Role of Profiling. A Review of Research Evidence*. Sheffield: Department for Work and Pensions. Research Report 192.

Hosmer, D. and Lemeshow, S. (2000) *Applied Logistic Regression* (2nd ed.). Chichester: Wiley.

O'Connell, P., McGuinness, S. and Kelly, E. (2012) *The Transition from Short- to Long-Term Unemployment: A Statistical Profiling Model for Ireland*. The Economic and Social Review, 43(1), 135–164.

O'Connell, P., McGuiness, S., Kelly, E., Walsh, J. (2009) *National Profiling of the Unemployed in Ireland.* Dublin: Economic and Social Research Institute. Research Series 10.

OECD (2012) OECD *Employment Outlook 2012*. [online] Available at: http://dx.doi.org/10.1787/empl_outlook-2012-en [Accessed 13 December 2012].

Office for National Statistics (2012) *Labour Market Statistics: CLA02: Claimant Count by Age and Duration*. [online] Available at: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-222507 [Accessed 13 December 2012].

Payne, C. and Payne, J. (2000) Early Identification of the Long-Term Unemployed. London: Policy Studies Institute. PSI Discussion Paper 4.

Pedhazur, E. (1997) *Multiple Regression in Behavioral Research: Explanation and Prediction*. London: Wadsworth.

Peduzzi, P., Concato, J., Kemper. E., Holford, T. And Feinstein, A. (1996) *A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis*. Journal of Clinical Epidemiology, 49(12), pp.1373-1379.

Productivity Commission. (2002) *Independent Review of the Job Network*. Canberra, AusInfo. Report no.21.

Tabachnick, B. and Fidell, L. (2007) *Using Multivariate Statistics*. (5th ed.). London: Pearson.

This Working Paper describes work undertaken by the Department for Work and Pensions to explore the feasibility of developing a profiling tool to predict, at the point of first claim, the likelihood of a new Jobseeker's Allowance (JSA) claimant reaching long-term unemployment, defined as 12 months or more with a continuous claim.

The work was undertaken in response to the Department's interest in the application of segmentation models to improve and refine service allocation, and the emergence of such tools in other countries. A predictive model was built using logistic regression and based on data collected from a 2010 telephone survey of 5,600 new claimants, combined with administrative data held by the Department. The paper presents the variables that the modelling suggests are the most efficient predictors of future long-term unemployment and discusses model accuracy along with the implications for implementation in an operational context.

If you would like to know more about DWP research, please email:
Socialresearch@dwp.gsi.gov.uk

DWP Department for Work and Pensions

Published by the
Department for Work and Pensions
March 2013
www.dwp.gov.uk
Working Paper no. 116
ISBN 978-1-909532-22-9