# Missing Spells and Fake Spells: Correcting Measurement Error in Duration Models using Validation Data*

Christopher R. Bollinger and Annette Jäckle

Fall 2013

Preliminary and Incomplete, Please do Not Cite or Circulate

**Abstract**

Using a unique administrative match to the British Household Panel Survey, we examine the impact of measurment error on estimates of duration models for three British Transfer programs: Income Supplement, Working Family Tax Credit and Job Seekers Allowance. Unlike previous research, our model addresses false positive spells, false negative spells as well as differences in reported duration for correctly reported spells. We demonstrate that in some programs, there exists substancial failure in reporting of the entire spell and that shorter spells are significantly more likely to go unreported. We demonstrate that this has the largest impact on the intercept and overdispersion parameters in a simple duration model.

* Christopher R. Bollinger, Professor of Economics, University of Kentucky, and Leverhulm Visiting Professor, ISER, University of Essex. Annette Jäckle, Senior Research Fellow, ISER, Essex University. Bollinger thanks the Leverhulm Foundation for its generous support of this research.

Missing Spells and Fake Spells: Correcting Measurement Error in Duration Models using Validation Data

# 1 Introduction

Estimation of duration models of transfer program participation are important for understanding what factors and policies determine how quickly individuals move off these programs. While in many cases, administrative records can be used for this purpose, restrictions on access to administrative records, and the paucity of detailed economic and demographic data in these records make the use of survey data for estimation of these models desirable. However, survey data are well known to suffer from measurement error, in particular in the detailed recall of dates of participation in transfer programs. Very few authors have documented recall errors in survey data (see Marquis et al., 1990; Jäckle & Lynn, 2008; Lynn et al, 2012; Jäckle, 2013). In contrast to the simple linear regression model, measurement error in the dependent variable, the duration measure, can have serious impacts on estimation of these models (see for example, Abrevaya and Hausman, 1999; Ham et al., 2012, Cheshire et al., 2002; Boudreau, 2003; Hill, 1994; Torelli and Trivallato, 1993; Holt, et al. 2004; Romeo, 1997; Augustin and Wolff, 2004; Paggiaro and Torelli, 2004; Dumangane, 2006). Most previous studies of measurement error in duration have focused on two models. The first (see for example Cheshire et al. 2002; Abrevaya and Hausman, 1999) is a simple additive white noise term, typically in the log-duration. The second approach, focuses on well documented seam bias issues, since it is has been observed that individuals disproportionately report the end of a participation period to coincide with an interview seam in the survey (see for example Ham et al. 2012, Romeo, 1997). However, very few researchers have examined the extent and structure of the response error in duration reports using validation data. Typically the modeling approaches taken by others has assumed that response errors are independent of demographic and other variables. Most other studies have ignored both failure to report spells which occurred and the reporting of spells which did not occur. False positive and negative reporting of program participation is well documented in cross section (see Bollinger and David, 1997; Ham and George, 2011 ), but has been largely ignored in estimation of duration models. Few studies have combined estimates of the structure and extent of measurement error in duration measurements with estimation of the actual duration model Using an administrative record match between the British Household Panel Survey and administrative records of participation in three income transfer programs (Job Seekers Allowance, Income Support and Working Family Tax Credit), we present detailed estimation of the measurement error structure in reporting duration of these programs. Further, we demonstrate how to use these results to correct survey estimates for the measurement error and present corrected estimates of simple duration models for these three programs.

We find that under-reporting of entire spells is extensive. We find that approximately 25% of actual Income Supplement spells are not reported, nearly 33% of Working Family Tax Credit spells go unreported, and over 50% of Job Seekers Spells are not reported. In all three programs we find that the probability a spell is reported is strongly positively associated with the length of the spell, thus the observed spells are a selected sample, with longer durations than the average spell. We also find, in particular for Income Supplement and Working Family Tax Credit programs that a substantial portion of the sample spells are false positives: there is no record match for these spells. Our estimates show that the distribution of durations for false positives differs from the distribution of true spells, further biasing estimates. We also find that false positives, false negatives and the mismeasurement in the duration of the spell is associated with various demographic characteristics. Failure to control for each of these can lead to substantial bias in estimation.

We find that our final estimation of the duration models, using a negative binomial model as a base, yield a number of changes in parameter estimates from a baseline model chosen for convenience. In particular, we find, as noted by many, that the overdispersion parameter than in the uncorrected model, suggesting in general that overdispersion is due to measurement error. We also find some differences in parameter estimates, although large standard errors due to small samples make it difficult to draw stark conclusions. Overall, our approach highlights the need for validation of survey measures of program participation and provides an approach which can be used in many settings. Section two of the paper discusses the data sources and in particular provides information on the validation data. Section

# 2 Data

The data in this study derive from the British Household Panel Survey (BHPS). The BHPS is one of the main data sources for measuring social and economic change in Britain (http://www.iser.essex.ac.uk/survey/bhps). The BHPS began in 1991 with a nationally representative sample of 5,538 households, containing 10,264 interviewed individuals. The sample is a stratified clustered design drawn from the Postcode Address File. All individuals resident at these addresses at the first interview are sample members and all aged 16+ are eligible for interviews. Sample members are re-interviewed annually and followed if they move out of their original household to form new households. Adult members of new households are also interviewed, as are new members joining sample households and children reaching the age of 16. Interviews are carried out face-to-face and last on average 45 minutes per individual, plus 10 minutes for a household-level questionnaire. The survey covers labour market activities, income, savings and wealth, household and family organisation, housing, consumption, health, social and political values, education and training. The primary sample, based on public use data of the BHPS, stem from four interviews, waves 9 (1999) to 12 (2002). This sample includes the individuals also used in the validation sample. We restricting the sample to individuals who were interviewed in each of those years and excluding proxy interviews, of which there are only 2-3% (Lynn et al. 2006 Table 19). We focus upon participation spells for three programs: Job Seekers Allowance, Income Supplement, and Working Family Tax Credit. We utilize the first reported participation spell starting after January 1st, 1999 and before September 1st, 2002, corresponding to the reference periods for waves nine through twelve. Although we allow for right censoring of spells, we do not allow for left censoring.

Table 1 presents descriptive statistics for these data by type of spell. There are 386 income supplement spells, 436 working family tax credit spells and 320 job seekers allowance spells. Spells for both Income Supplement and Working Family Tax Credits are typically longer, averaging slightly over 1 year, than spells on Job Seekers Allowance which average less than six months. The averages for completed spells are necessarily shorter, and both averages represent an understatement of the average spell length. Those on Job Seekers Allowance are typically male (61%). The Working Family Tax Credit is typically reported by a female in the household, but most of these families are headed by married or cohabiting couples. Individuals on Income Supplement are more typically women (nearly 70%) and are less typically married or cohabiting (41%). While this sample is certainly not representative of British individuals or households, it is consistent with participants in these programs.

The BHPS sample derived above is augmented by a validation sample drawn from the UK European Community Household Panel Survey (UK-ECHP). These data were collected since 1997 as a part of the BHPS, and have been used for methodological purposes. The sample was initially selected based on characteristics associated with low income, such as household reference person unemployed during the initial sampling frame, household reference person receiving lone parent benefit, rented housing , or receipt of means-tested welfare benefits. Hence the sample differs in some characteristics from the full sample above. Respondents in this sub-sample were asked for permission to obtain their records from the Department for Work and Pensions, the government department in charge of administering transfer program payments and tax credits such as the three studied here. The consent rate for the record linkage was 77.4%. Of these respondents, 74.1% were successfully linked to the records. Non-matches were most likely respondents without records, who had not received transfer payments or tax credits during the time frame of interest, although some failure to match due to problems with the identifying information used for the linkage cannot be excluded (see Jenkins et al. 2008). The low-income sample, attrition during prior panel waves and potentially selective consent for linkage (see Jenkins et al. 2006) mean that the validation sample is not representative of the general population. The over-representation of low-income groups, including those in receipt of transfer payments, is a considerable advantage for this study.

The matched validation sample contains both survey and administrative record measures of spells on the three programs of interest. Table 2 presents comparable descriptive statistics for these data. In constructing the matched validation sample, we attempted to match each spell in the survey to a spell in the administrative records. Spells were matched first on starting closest starting date for a specific program. The underlying assumption is that the administrative record is correct. The resulting sample has 237 individuals with either record spells, survey spells or matched spells across the three programs. Table 2 presents descriptive statistics for the validation sample. There are 77 individuals either reporting receiving or actually receiving (or both) Income Support, 76 individuals reporting or receiving the Working Family Tax Credit and 64 individuals reporting or receiving the Job Seekers Allowance. Income Support and Job Seekers Allowance have more censoring in the validation survey than in the full sample above. This may reflect the characteristics of the sample. In particular, since the sample was selected based on prior experience with both of these programs (although not continuing experience), one would expect longer spells conditional upon the characteristics. Characteristics such as age, gender, marital/cohabitation status and the number of children are quite comparable across the two samples. The validation sample has fewer individuals with high qualifications. Perhaps the most significant difference is the home ownership rate. In particular for those on Income Supplement it is quite low at 9.1% (compared to 28.5% in the full sample) and as well as those receiving the Working Family Tax Credit (24.7%% compared to 55.5%) and Job Seekers Allowance (27.0% compared to 57.5%). The differences in these samples is cause for some concern. However, the validation sample is known to be selected on these particular observable characteristics and thus conditioning upon them should help alleviate sample selection issues. Moreover, there is little evidence (as we shall investigate below) that this selection specifically effects response behavior conditional upon the true participation and duration state. Caution should be exercised in comparison of duration models estimated on the administrative data compared to the full sample, since previous participation in these programs was a selection criteria.

Table 3 presents more information on the matched spells in the validation sample data. Both Income Support and Working Family Tax credits are more likely to be reported when they are received (72.4% and 62.9%) than Job Seekers Allowance (44.4%). However, these programs also suffer from a higher false positive rate, only 63.6% of Income Supplement and 66.6% of Working Family Tax Credit reported spells are associated with a true spell, while 80% of reported Job Seekers Allowance spells are associated with a true spell. The paucity of false positive data, in particular for the JSA program, make it difficult to estimate the distribution of false positive spells. In all three programs, the agreement of censoring on matched spells is remarkably high. This suggests a high agreement on gross spell length (which we demonstrate more carefully below). Censoring appears to be more of a concern in the Income Supplement and Working Family Tax Credit spells than in the Job Seekers Allowance spells. This is consistent with the full sample data as well.

# 3 Modeling

Researchers are interested in recovering characteristics of the distribution of participation spell lengths conditional upon individual or program characteristics. However, in survey data one can only recover the distribution of reported spell lengths conditional upon these characteristics. Building off of work in Bollinger and David (2001), we construct the relationship between the observable distribution and the distribution of interest. Let $T$ be the spell length reported in the survey and let $T^*$ be the true spell length (when one exists). Let $R$ be the indicator for a true spell (according to the administrative records) and let $S$ be the indicator for a reported survey spell. Let $X$ be the conditioning characteristics of interest to the researcher. Then, formally, the researcher only observes

$$f\left(T|X,S=1\right),$$

the distribution of reported spells, but wishes to recover

$$f\left(T^*|X,R=1\right),$$

the distribution of true spells. We begin with the conditional distribution of the survey spells and using Bayes rule not that this can be written as

$$f\left(T|X,S=1\right)=f\left(T|X,S=1,R=1\right)\Pr\left[R=1|X,S=1\right]+f\left(T|X,S=1,R=0\right)*\Pr\left[R=0|X,S=1\right].$$

The observed distribution is a mixture of the observed distribution of reported spells when a true spell is actually associated and the distribution of false positive spells. Consider next, the reported spell distribution, conditional on a true spell being present, this can be further written as

$$f\left(T|X,S=1,R=1\right)=\int_0^\infty f\left(T|X,S=1,R=1,T^*\right)f\left(T^*|X,S=1,R=1\right).$$

The researcher, though, is not interested in simply the distribution of true spells for those which are reported. Note that repeated application of Bayes Rule allows us to write

$$f\left(T^*|X,S=1,R=1\right)=\frac{f\left(T^*,S=1|X,R=1\right)}{\Pr\left[S=1|X,R=1\right]}=\frac{\Pr\left[S=1|T^*,X,R=1\right]*f\left(T^*|X,R=1\right)}{\Pr\left[S=1|X,R=1\right]}.$$

Substitution of these expressions establishes the formal link between the observed distribution, $f\left(T|X,S=1\right)$ and the distribution of interest $f\left(T^*|X,R=1\right)$ :

$$\begin{aligned}f\left(T|X,S=1\right) &= \left(\int_0^\infty f\left(T|X,S=1,R=1,T^*\right)\frac{\Pr\left[S=1|T^*,X,R=1\right]}{\Pr\left[S=1|X,R=1\right]}*f\left(T^*|X,R=1\right)\right)\Pr\left[R=1|\right.\\ &\quad +f\left(T|X,S=1,R=0\right)*\Pr\left[R=0|X,S=1\right].\end{aligned}$$

This equation identifies five important terms which can be estimated only from the validation data. The first is the probability that a reported spell is associated with a true spell

$$\Pr\left[R=1|X,S=1\right]. \tag{2}$$

The second is the probability that a true spell is reported

$$\Pr[S = 1|X, R = 1] \tag{3}$$

while the third is the probability that a true spell is reported conditional upon the true length of the spell

$$\Pr[S = 1|T^{*}, X, R = 1]. \tag{4}$$

The final two are reported distributions: the distribution of reported spells conditional upon the true spell length

$$f(T|X, S = 1, R = 1, T^{*}) \tag{5}$$

and the distribution of reported spells conditional upon the spell being a false positive

$$f(T|X, S = 1, R = 0). \tag{6}$$

Our approach is straight forward. We utilize the validation data to estimate the five terms in the likelihood function which link observed data to the true spells. We then form the Quasi-likelihood function for the observed data in the full sample, replacing the five terms with their estimated counterparts. Then maximization of the likelihood will estimate the model for $f(T^{*}|X, R = 1)$. The estimation is, in a sense, conditional upon the estimated terms. To account for this, we use bootstrapped standard errors. Of primary concern is how to parameterize both the main model of interest $f(T^{*}|X, R = 1)$ and the measurement error terms. The data we have are measured measured monthly. While certainly it is possible to use an underlying continuous distribution (such as the Weibul), estimates will ultimately be based upon a discretization of that distribution. For simplicity, we have chosen to use discrete distributions such as the Poisson and Negative Binomial.

# 4 Estimation of Measurement Error Terms

We next turn to the estimation of each of the five terms in the measurement error model. The main restriction in estimation of these terms will be the small sample sizes encountered. While the formal derivation above requires conditioning on all variables in the main model, if these variables have little or no predictive power, more parsimonious models can be estimated. We focus on two groups of variables in our models below. Age, gender and marital status have been shown to affect response behavior for receipt of programs and surveys in general. Hence, we examine these variables as predictors. We also note that our validation sample differs from the full BHPS sample along three lines: respondents are more likely to have lower educational qualifications, are less likely to own their own home and, at least for the Job Seekers Allowance, are less likely to live in London or Southeast England. The final model of program duration we will estimate includes only two additional variables, the number of children and the regional unemployment rate. In other results not reported here, these variables were not found to have significant predictive power. For parsimony, we choose to include regressors which are either significant predictors (large effects) or statistically significant.

A second issue is how to handle censoring of the true spell. In the likelihood function for the estimation of the primary model, censoring of the true spell is not an issue. However, in estimating terms conditional upon the true spell (such as $\Pr[S = 1|T^*, X, R = 1]$ or $\int_0^\infty f(T|X, S = 1, R = 1, T^*)$) we may need to account for censoring of the conditioning variable. Two approaches are quite obvious. If we believe that we have a correctly specified model, then we should simply use the observations for which no censoring of the true spell occurs: censoring on a conditioning variable should not affect estimation. A second approach is to include an indicator variable for censoring. Censoring in the survey response only effects estimation of $f(T|X, S = 1, R = 1, T^*)$ and $f(T|X, S = 1, R = 0)$ and can be handled using standard approaches.

A third important issue is the choice of specification. For the terms involving estimation of probability functions (probabilities of false positive and false negative spells), we choose simple probit models. While the probit model does involve some distributional assumptions, it is well known to fit data well and to provide predictions which are relatively robust. Specification of models for the conditional durations, the duration of a reported event given a true spell and the duration of false positives, we have chosen the Poisson distribution. This distribution has two desirable properties: first and foremost it is discrete which matches the way the data were collected. This improves efficiency, and allows for simpler programming in the second stage where the distributions are used (rather than using discrete approximations to continuous functions). Secondly, the distribution is simple, which enhances efficiency in the face of small samples. This is a particular concern for the false positive spells.

## 4.1 Estimation of False Positive Probability

The first term of interest is the estimation of the probability that a true spell is associated with reported spell (or the probability that a reported spell is a false negative): $\Pr[R = 1|X, S = 1]$. The last line of table 3 presents simple estimates of this rate. The goal of estimation here is primarily prediction. We focus on probit models, as they have desirable properties for prediction, specifically that all predictions are bounded in the unit interval. Qualitatively results do not appear to differ in meaningful ways for other models. Estimates of the marginal effects (at the mean of the primary sample) are presented in table 4. The small samples make it difficult to include all regressors of interest. We base the estimates we present on two potential issues. First, we note that prior research has established differences in response behavior for men and woman, and often by age and marital status. Column 1 for each program presents models including these three demographic variables. The validation sample differs from the full survey sample along three dimensions: the level of qualifications, home ownership rates, and the percentage living in London or Southeast England. Column 2 for each program presents models including these three variables. Finally, based upon both the predictive and statistical significance of the previous two models, a combined model was estimated and is reported here.

In general, it appears that little is related to the probability of a true spell being associated with a survey report of Income Support. Only age was marginally significant, and it's impact is relatively modest. While we include this variable in the final specification reported in column three, it is also quite reasonable to assume that a simple marginal probability may suffice for correcting for measurement error in this program. This is in sharp contrast to both the Working Family Tax credit and the Job Seekers Allowance. In both of these programs, gender is highly predictive, however with opposite signs. In general in measurement error literature, men are thought to have higher reporting errors. While this is supported for the Working Family Tax Credit, it is apparently not true for the Job Seekers Allowance. We note also that marital status and no-qualifications are highly predictive for Job Seekers Allowance. Men with no qualifications are more likely to have a false positive than men with any other qualifications. Similarly, those who are married or cohabiting are more likely to have a false positive report.

## 4.2 Estimation of Probability of Reporting a Spell

The second term of interest appears in the denominator of the ratio in the summation and measures the probability of reporting a spell when a true spell occurs: $\Pr[S = 1|R = 1, X]$. The second to last line of table three presents a simple estimate of this rate, unconditional on any covariates. Similar to the section above, we focus on probit models and present estimates of the marginal effects in table 5. We again focus on two sets of variables: those found by other researchers to affect response behavior and those which differ between the validation sample and the primary sample.

In general, very little appears to be related to the probability of reporting a spell (the probability of a false negative). Only home ownership is significant in the Income Support model and only age is significant in the working family tax credit model. No variables were significant in the Job Seekers Allowance models. In both the Working Family Tax Credit and in the Job Seekers Allowance models, gender had a large estimated impact, but was imprecisely estimated leading to statistical insignificance. In the final model, we include gender. Overall, we find that home owners are much less likely to report income support when they receive it, the marginal effect of this coefficient was a very large -0.436. Especially given the higher home ownership in the primary sample, this appears an important effect. In the Working Family Tax Credit program, older workers were less likely to report receiving this credit. Nothing was statistically significantly related to reporting Job Seekers Allowance. This is somewhat surprising in contrast to the results in the previous section. The sample size here is larger than in the previous section. The final model estimated included gender.

## 4.3 Probability of Reporting a Spell Conditional on True Spell Length

The third term of interest is the probability of reporting a true spell condition upon the true spell's length: $\Pr[S = 1|R = 1, X, T^*]$. Of considerable interest here is whether the length of the spell effects this probability. If the length of the spell does not affect this probability, then the ratio $\frac{\Pr[S=1|T^*,X,R=1]}{\Pr[S=1|X,R=1]}$ will be identically 1, and this term may be dropped from the likelihood function. Of concern is that the relationship may be highly non-linear, and that indeed, seem effects may dominate this term. Although not reported here, a number of models with a variety of non-linear and step effects were estimated. We report the three types of models which appeared to be most predictive, while preserving parsimony: models only including the length of the true spell entering linearly, models with quadratic length functions, and models with an indicator for spell length a year or longer. We also repeat the type of analysis from the previous two sections, examining including variables from the main model.

Table 6 presents models only including the length of spell, and no other regressors. These models are comparable in parsimony to the simple probabilities presented in table 3 for the first two models of response. The first column in each program presents a model estimated only using observations with completed administrative spells. While the second column presents models using all spells, but includes an indicator that the spell was a complete administrative spell (not right censored). In general we find little difference between the models fit using all observations and the models fit using only complete administrative spells. However, in the Income Support model the indicator for a complete spell is typically statistically significant, and in the Job Seekers Allowance models it is insignificant but large. For the Income Support and Working Family Tax Credit models, a quadratic function in the length of spell appears to have the most predictive power, while for the Job Seekers Allowance a linear model appears to perform well. In all three models, there is reason to believe that the length of the administrative spell affects the likelihood of reporting that spell, with longer spells more likely to be reported.

Table 7 includes the demographic variables into the specifications. In the first two models, only the linear term for the length of spell is included. In the final model for Income Support and Working Family Tax Credit programs, the quadratic terms for length of spell are re-introduced. Comparing these results to those in both table five (the model without length of spell) and table 6 (the model with only length of spell) demonstrates that the length models and demographic models in those table are largely unaffected by combining terms. Of particular interest is the fact that the statistically significant effects appear to be the most stable across specifications, while the predictively large effects are more likely to change with inclusion of additional variables. It should be noted that this does not imply the previous models are in any sense wrong, but rather sheds insight into interactions on response behavior. The third column presents a combination of the parsimonious models in the third and fourth column of tables 6 and 7.

## 4.4 Distribution of Reported Spells Conditional on Length of True Spell

Investigating the distribution of reported spells conditional on the length of the true spell presents a number of difficulties. Perhaps the most important challenge is that the data only allow the spell to measured at the level of the month. This makes continuous models somewhat suspect, especially for reporting behavior. Also, as noted in table 3, except for Job Seekers Allowance, there were few matched spells which were not censored. However, in both the Income Supplement and in the Working Family Tax Credit, there is substantial agreement on censoring: either both measures (administrative and survey) or neither measure was censored. There appears to be a great amount of agreement between true and reported spells. This is particularly true for cases where both spells are censored.

A number of possible models could be used for the distribution of reported spell length. The data are, as collected and compiled, discrete in nature: individuals are asked in which months they received income from the particular source. While one can argue that a continuous process underlies the data, even that is questionable, as the eligibility of participants for these programs are not continuously evaluated, but rather are typically evaluated at a monthly, or in the case of Job Seekers Allowance, weekly intervals. For these reasons, we focus upon discrete distributions in our choice of modelling. The Poisson distribution has a number of appealing features, although it does have the disadvantage of the strong assumption that mean and variance are equal. An alternative to the Poisson is the negative binomial distribution, which allows for overdispersion. We have estimated both but present the Poisson estimates here. We noted that while the negative binomial models typically indicated overdispersion relative to the Poisson, the Poisson appear to predict probabilities closer to those observed empirically, in particular for agreement between the survey and the administrative, as well as those differences that are 1 or 2 periods different. Hence, for the purposes of estimation, the Poisson appears to be preferred.

Table 8 presents estimates using only the true length of the spell as a regressor and a Poisson model. As one would expect, there is a strong positive relationship between the true length of the spell and the reported length of the spell. The first column in each program uses only observations where the record spell is not censored. The cost is that these estimates are based on a smaller number of observations. This is particularly true for income support and for working family tax credits. The second column uses the same specification but includes all observations and an indicator for observations where the true spell length is censored. All specifications adjust for censoring in the dependent variable using standard adjustments to the likelihood function. The third and fourth columns include non-linear approaches: quadratic in the length of the true spell and an indicator for spells lasting longer than 1 year.

The results in table 9 indicate a number of conclusions. First, the censoring of the true spell length does not appear to have dramatic effects on the estimation. The positive coefficient on the indicator for the true spell being censored is quite sensible, indicating that the reported length would be longer, were the spell not censored is certainly consistent with the positive coefficient on the length of the spell. In all three programs, the quadratic model does not appear to be in any meaningful way better than the simple linear model. While the indicator for spells longer than a year is significant in both the Income support and the Job Seekers Allowance program, it is not clear that this fits the data better, in particular for prediction. In the Job Seekers Allowance, in particular, it is based on a small number of cases, and so is somewhat suspect. While not presented here, we estimated negative binomial models. However, the Negative Binomial model significantly under predicts agreement between the reported and true spell and places too high a probability on larger deviations.

Table 9 presents expanded models including various regressors in an approach parallel to that taken in previous sections. In table 7, the estimates suggested that few of these coefficients were important in prediction of the probability of reporting, however, they do appear to be highly predictive of the spell length reported. Across all three programs both gender and the level of schooling are highly predictive, with men under-reporting Income Supplement and Working Family Tax Credit duration, but over-reporting Job Seekers Allowance duration. Those with low educational attainment tend to under-report Income

Supplement duration but over-report both Working Family Tax Credit and Job Seekers Allowance Duration. Married people tend to under-report duration of Income Support and may over-report duration of Job Seekers Allowance, while older people tend to over-report duration of Working Family Tax Credits. The negative binomial model, finds qualitatively similar results, but again, prediction from the negative binomial models tends to lead to lower agreement probabilities than appear in the data.

In both tables 8 and 9 we report not only statistical significance of the coefficient on log-spell length, but also evidence as to whether it is statistically significantly different than 1. We find in all cases that the estimate reject the hypothesis that the coefficient is 1. This provides evidence that the simple linear model of response error does not hold. Previous authors have considered models which would require that coefficient to be one. We find evidence rejecting that assumption.

## 4.5 Estimation of Spell Duration for False Positive Spells

Table 3 demonstrates that the sample sizes for estimation of duration of false positive spells are small. Income Support has the most with 24, while Job Seekers Allowance has only 7 false positive observations. We estimate models for both Income Support and the Working Family Tax Credit program, following the approach above. We are hesitant to conclude much from this exercise except that these spells do not appear to be similar to the actual spells, and hence it appears crucially important to control for this fact. We do note that, different from models of non-response and conditional duration above, the demographic variables appear to be somewhat more predictive. One interpretation of this is that these models reflect duration on some program, and may indicate program confusion. We hope to explore this issue further in future work. The results are presented in Table 10.

# 5 Estimation of Duration Models

The underlying model we estimate is a relatively simple specification. We use the negative binomial specification. A number of authors have noted that the Poisson distribution does not fit data well, because of apparent overdispersion of the data. Some authors (Cheshire et al, 2002) have noted that measurement error may cause overdispersion. This specification allows us to examine that claim. Noting that the claim holds only in the face of strong assumptions, we will see that a variety of patterns exist. Thus the underlying $f(T^*|X, R=1)$ is assumed to be negative binomial. Equation 1 then becomes summations over values of $T^*$ rather than integrals, simplifying programming. For X variables, we choose a set of variables which consist of standard demographic variables and two variables which vary with location: Age, education, gender, number of children, ****. We focus upon the coefficients on education and unemployment, as these have important policy implications. The final likelihood function uses the estimated probabilities from the five terms estimated above

$$\begin{aligned} L(\beta|T,X,) &= \left(\sum_{T^*=0}^{80} f(T|X,S=\widehat{1,}R=1,T^*)\frac{\Pr\left[S=1|\widehat{T^*},X,R=1\right]}{\Pr\left[S=\widehat{1|}X,R=1\right]} * f\left(T^*|X,R=1:\beta\right)\right)\Pr\left[R=\widehat{1|}\right. \\ &\quad + f\left(T|X,\widehat{S=}1,R=0\right) * \Pr\left[R=\widehat{0|}X,S=1\right]. \end{aligned}$$

Note that $\beta$ only enters the distribution of $T^*$. The remaining terms are functions of the estimated parameters from the previous section and the data. Note that this implies the sampling variance from the first stage estimates must be accounted for in the second stage estimates. In order to accomplish this, we use bootstrapping with the boostrap stratified to reproduce the ratio of administrative matches to the full sample. This is an important feature for two reasons. First, given the small sample sizes in some of the categories in the validation data, simple random sampling could result in a sample with no false positives (for example). Hence from a practical reason it is important to carefully stratify the samples across the administrative matches. Secondly, the first order taylor approximation will be a function of the ratios of validation to estimation sample (see, for example, Bollinger and David, 1996). We also argue that due to small sample sizes, the bootstrap is preferred to an asymptotic approximation. We report significance based on a symmetric 95% empirical confidence interval. We used 200 replicates for each model.

The results are presented in table 11. The first column reports baseline estimates where no correction for measurement error was taken. They provide perspective on the implications of the correction. In all cases we use a negative binomial as the baseline model. In all cases the overdispersion parameter (alpha) is highly positive and significant. The highest overdispersion appears to be in the Income Support program while overdispersion in the Working Family Tax Credit program is quite modest. In the income support program, only age is statistically significant in the baseline model, with older participants having longer spells. In the Working Family Tax Credit program, Age is again positive and significant as is the number of children. Men, married or cohabiting couples, and workers with either low or high qualifications experience significantly shorter spells. The spell length is also shorter when the unemployment rate is higher. This result may indicate that those who qualify for

WFTC, may be more likely to lose their job as the unemployment rate rises. Like the Income Support program, few of the demographic variables are statistically significant in the baseline estimate for the Job Seekers Allowance program. More children increases the duration while owning your own home significantly decreases duration. The coefficients on education are consistent with what one might suspect: higher qualified individuals have shorter durations of Job Seeking, while the lowest qualifications have longer. Similarly, the unemployment rate positively effects duration, although none of these variables are statistically significant. The intercept terms, all statistically significant, are important in duration models, particularly when X variables are insignificant. As might be expected, the intercept for Income support is the largest, indicating the longest average duration while intercept for the Job Seekers Allowance is the smallest, indicating the shortest duration.

The second column presents estimates which use the simplest corrections possible where no X variables are included in the measurement error terms of the likelihood. For example, the first term - the probability of a true spell being associated with an observed spell - are simply the rates given in table 3. The only conditioning variables in any of the five measurement error terms included in the likelihood are those associated with the duration of the true spell. In the income support model and the Working Family Tax Credit model, this leads to higher estimates of the dispersion parameter. In the case of the WFTC program it more than doubles. Only in the Job Seekers Allowance program does the overdispersion parameter decrease, however, it decreases significantly from 1.5 to 0.2. The intercept terms fall somewhat for both the IS and WFTC programs. This is as we might expect, since a significant issue is false negatives. In the income support program, none of the variables are statistically significant in the corrected model. Most coefficients decreased in magnitude, with the exception of the unemployment rate which becomes positive but remains statistically insignificant. In the WFTC program, the coefficients on statistically significant variables all rose in magnitude with the exception of the unemployment rate. However, many of the previously significant coefficients lost significance: age, both qualifications variables and the unemployment rate. In part this is to be expected as the first stage models induce significant sampling variation into the estimates here. Of particular interest is the fact that gender remains statistically significant and rises substantially in magnitude. Indicating that men (or perhaps more accurately male headed households where the male applies for the program) have shorter durations on this program. In the JSA program, the coefficient on higher qualifications becomes statistically significant, but is slightly smaller in magnitude. The coefficient on Own House remains significant although it too falls in magnitude. The number of children, which was statistically significant and positive, falls substantially in magnitude and becomes statistically insignificant. In general, coefficients in this model fell in magnitude.

The final column in table 11 presents the full correction using the preferred models from the previous sections. In all cases, the overdispersion parameter is now substantially smaller than the baseline estimate. Only the JSA program is it higher than the simple model of the second column, however it is still substantially smaller than the original estimate and is now statistically insignificant. This is consistent with the previously literature finding that correcting for measurement error reduces overdispersion. In all three programs, standard errors rose quite substantially reflecting the small samples used in the first stage. In the IS model, the model with the smallest percentage of false negatives, the intercept term is only slightly larger than the estimate in the baseline model. In the WFTC model, the intercept has risen dramatically over both the baseline model and the simple correction. In the JSA model, the intercept is slightly lower than the simple correction but still larger than the baseline correction, although it is now statistically insignificant. The pattern is somewhat puzzling in that we might expect the intercept to fall the largest for the program with the highest false negative rate. However, the correction is complicated. One might argue that the preferred final model would have lower dispersion and a higher intercept. For the JSA model this would be the second column, the simpler correction. This is also consistent with the findings that few if any X variables predicted the measurement error terms in the JSA model.

In the income support model, none of the explanatory variables are statistically significant. There is no particular pattern discernible with respect to size or sign changes. We think this reflects perhaps poor prediction of these variables in general. In the WFTC model, the full correction statistical significance is preserved for marital status, the number of kids and the unemployment rate. The coefficient on gender has dramatically changed from negative and statistically significant in both previous models to very large positive but statistically insignificant. Age as well has reversed sign but become insignificant. . In the JSA final model, statistical significance is lost for all X variables. However, with exceptions of male and No qualifications, sign and magnitude is quite comparable. Similar to the WFTC, the coefficient on male dramatically reverses sign, but remains insignificant. The sign reversal on no-qualification is not as dramatic, but is still somewhat puzzling.

# 6 Conclusion

It is difficult to draw strong conclusions about particular parameters in any of these models. There does seem to be evidence that some of the findings in the original models may be more closely associated with the measurement error process, but no obvious pattern is clear. Perhaps the most obvious conclusion is that better validation studies and more comprehensive approaches to estimation are important. We suspect that in the JSA case, the simple correction model is sufficient. However, in the WFTC case and possibly in the IS case, it may be necessary to consider more complicated models of measurement error correction. In all cases, the inclusion of a correction for false negatives and the inclusion of a correction for false positives in the WFTC case may be highly warranted.

# References

[1] Abrevaya, J. and J.A. Hausman (1999), "Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells" Annals of Economics and Statistics, no. 55/56, pp. 243-275.

[2] Augustin, T. and J. Wolff, (2004) "A Bias Analysis of Weibull Models Under Heaped Data," Statistical Papers, 45, 211-229.

[3] Bollinger, C.R. and M.H. David, (1997) "Modeling Discrete Choice with Response Error: Food Stamp Participation" (with Martin H. David), Journal of the American Statistical Association, September 1997, vol. 92, no. 439.

[4] Bollinger, C.R. and M.H. David (2001) "Response error in duration of spells of Unemployment Compensation: Implications for modeling duration" (with Martin H. David) Proceedings of the Survey Methods Section of the American Statistical Association.

[5] Boudreau, C.(2003) Duration Data Analysis in Longitudinal Surveys.University of Waterloo: Department of Statistics and Actuarial Science.

[6] Cheshire, A., M.G.B. Dumangane and R.J. Smith (2002), Duration Response Measurement Error",Journal of Econometrics, 2002, 111, 169-194.

[7] Hill, D. H. (1994) 'The Relative Empirical Validity of Dependent and Independent Data Collection in a Panel Survey', Journal of Official Statistics, 10(4): 359-380.

[8] Holt, D., McDonald, J. W. and Skinner, C. J. (2004) The Effect of Measurement Error on Event History Analysis, in Measurement Errors in Surveys (eds P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman), John Wiley & Sons, Inc., Hoboken, NJ, USA.

[9] Jäckle, A. and Lynn, P. (2008) 'Respondent Incentives in a Multi-Mode Panel Survey: Cumulative Effects on Non-Response and Bias'. Survey Methodology, 34(1), 105-117.

[10] Jäckle, A., Lynn, P., Sinibaldi, J. and Tipping, S. (2013) 'The Effect of Interviewer Experience, Attitudes, Personality and Skills on Respondent Co-operation with Face-to-Face Surveys', Survey Research Methods, 7(1), 1-15.

[11] Lynn, P.,Jäckle, A., Jenkins, S.P. and Sala, E. (2012) 'The Impact of Interviewing Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation Study'. Journal of the Royal Statistical Society, Series A, 175(1), 289-308.

[12] Marquis, K. H., Moore, J. C. and Huggins, V. J. (1990) 'Implications of SIPP Record Check Results for Measurement Principles and Practice', Proceedings of the Survey Research Methods Section, Alexandria, VA, American Statistical Association, 564-569.

[13] Meyer, B.D. and R.M. Goerge, (2011) "Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation" working paper.

[14] Paggiaro, A. and Torelli, N. (2004) "The Effect of Classification Errors in Survival Data Analysis" Statistical Methods and Applications, 13(2), 213-225.

[15] Romeo, C.J. (1997) "Measuring Information Loss Due to Inconsistencies in Duration Data from Longitudinal Surveys," Journal of Econometrics, 78(1), pp. 159-177.

[16] Skinner, C.J. and Humphreys, K. (1999) Weibull regression for lifetimes measured with error Lifetime data analysis, 5 (1). 23-37. ISSN 1380-7870

[17] Torelli,N. and U. Trivellato (1993) Modelling inaccuracies in job-search duration data, Journal of Econometrics, Volume 59, Issues 1–2, Pages 187-211. 827-835.

Table 1: Descriptive Statistics for the Primary Sample (BHPS)

| | IS | WFTC | JSA |
|---|---|---|---|
| Total Spells | 386 | 436 | 320 |
| Right Censored Spells | 47.4% | 46.3% | 9.3% |
| Mean Spell Length | 14.6 months | 13.6 months | 5.5 months |
| Mean Spell Length Complete | 8.5 months | 8.4 months | 4.7 months |
| Age | 47.2 | 32.9 | 33.4 |
| Male | 35.5% | 32.8% | 60.0% |
| Married or cohabiting | 42.0% | 75.9% | 52.2% |
| Number of kids under 16 | 0.72 | 1.77 | 0.568 |
| No qualifications | 45.6% | 12.6% | 17.2% |
| Mid level aualifications | 33.3% | 52.8% | 50.9% |
| High aualifications | 21.0% | 34.6% | 31.9% |
| Own house | 28.5% | 55.5% | 57.5% |
| London and SE England | 17.6% | 16.7% | 23.4% |

Table 2: Descriptive statistics for the Validation Subsample

| | IS | WFTC | JSA |
|---|---|---|---|
| Total Observations | 77 | 76 | 64 |
| Matched Spells | 40 | 37 | 23 |
| Record Only Spells | 13 | 19 | 34 |
| Survey Only Spells | 24 | 20 | 7 |
| Right Censored Record Spells | 60.4% | 39.3% | 12.2% |
| Right Censored Survey Spells | 50.0% | 40.3% | 16.7% |
| Mean Survey Spell Length | 12.8 months | 16.0 months | 8.0 months |
| Mean Record Spell Length | 14.7 months | 20.0 month | 6.6 months |
| Age | 48.1 | 34.4 | 32.2 |
| Male | 33.8% | 29.7% | 60.3% |
| Married or cohabiting | 40.3% | 68.9% | 44.4% |
| Number of kids under 16 | 0.766 | 1.95 | 0.746 |
| No qualifications | 48.1% | 14.8% | 20.6% |
| Mid level qualifications | 37.6% | 59.5% | 52.4% |
| High qualificaations | 14.3% | 25.7% | 27.0% |
| Own house | 9.1% | 31.1% | 42.8% |
| London or SE England | 18.1% | 17.6% | 17.5% |

Table 3: Match and Censoring Rates in Validation Sample

| | IS | WFTC | JSA |
|---|---|---|---|
| All Spells | 77 | 76 | 64 |
| Matched Spells | 40 | 36 | 23 |
| Neither Censored | 12 | 17 | 20 |
| Both Censored | 19 | 14 | 1 |
| Only Survey Censored | 1 | 2 | 1 |
| Only Record Censored | 8 | 3 | 1 |
| Record Only (false negative) | 13 | 18 | 29 |
| Record Censored | 5 | 5 | 5 |
| Survey Only (false positive) | 24 | 20 | 7 |
| Survey Censored | 12 | 7 | 3 |
| Response Rates | | | |
| True Spells Reported | 75.5% | 66.7% | 44.2% |
| Reported Spells with True Spell | 62.5% | 64.3% | 76.7% |

Table 4a: Probit Marginal Effect Estimates
Probability of a True Spell given an Observed Spell (Part 1)

| | Income Support | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| Age | -0.007** | | -0.005* |
| Male | -.111 | | |
| Married or cohabiting | -.176 | | |
| No qualifications | | -0.067 | |
| Home Owner | | 0.035 | |
| London or SE | | 0.136 | |
| Log-Likelihood | -39.2 | -41.7 | –40.7 |

Table 4b: Probit Marginal Effect Estimates of
Probability of a True Spell given an Observed Spell (Part 1)

| | Working Family Tax Credit | | |
|---|---|---|---|
| | Model 1 | Model2 | Final Model |
| Age | -0.003 | | |
| Male | -0.748** | | -0.724** |
| Married or cohabiting | 0.083 | | |
| No qualifications | | 0.078 | |
| Home Owner | | -0.129 | |
| London or SE | | 0.103 | |
| Log-Likelihood | -20.3 | -35.7 | -20.5 |

Table 4c: Probit Marginal Effect Estimates
Probability of a True Spell given an Observed Spell (Part 1)

| | Job Seekers Allowance | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| Age | -0.011 | | |
| Male | 0.372* | | 0.356* |
| Married or cohabiting | -0.261 | | -0.253 |
| No qualifications | | -0.426* | -0.565** |
| Home Owner | | 0.053 | |
| London or SE | | -0.154 | |
| Log-Likelihood | -13.1 | -13.2 | -10.9 |

Table 5a: Probit Marginal Effect Estimates of Probability of a Suvery Spell given a True Spell (Part 2)

| | Income Support | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| Age | 0.002 | | |
| Male | -0.0002 | | |
| Married or cohabiting | -0.044 | | |
| No qualifications | | -0.067 | |
| Home Owner | | -0.481** | -0.475** |
| London or SE | | 0.091 | |
| Log-Likelihood | -29.2 | -26.6 | -26.8 |

Table 5b: Probit Marginal Effect Estimates
Probability of a Suvery Spell given a True Spell (Part 2)

| | Working Family Tax Credit | | |
|---|---|---|---|
| | Model 1 | Model2 | Final Model |
| Age | -0.016* | | -0.015* |
| Male | 0.154 | | |
| Married or cohabiting | 0.227 | | 0.272** |
| No qualifications | | -0.120 | |
| Home Owner | | 0.141 | |
| London or SE | | -0.150 | |
| Log-Likelihood | -30.5 | -33.1 | -30.8 |

Table 5c: Probit Marginal Effect Estimates
Probability of a Suvery Spell given a True Spell (Part 2)

| | Job Seekers Allowance | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| Age | -0.004 | | |
| Male | 0.141 | | 0.123 |
| Married or cohabiting | -0.021 | | |
| No qualifications | | -0.116 | |
| Home Owner | | -0.042 | |
| London or SE | | -0.045 | |
| Log-Likelihood | -37.1 | -37.7 | -37.4 |

Table 6a: Probit Marginal Effects Estimates
Probability a True Spell is Reported with Length of Spell Only (Part 3)

| | Income Support | | | |
|---|---|---|---|---|
| | Complete Spells | All Spells | Quadratic | Year Shift |
| True Spell Length | 0.003 | 0.002 | 0.030 | -0.008 |
| Complete Record Spell | | -0.209 | -0.247* | -0.209 |
| True Spell Length Squared | | | -0.001 | |
| True Spell >=11 | | | | 0.264 |
| Log-Likelihood | -13.9 | -27.8 | -26.8 | -27.0 |

Table 6b: Probit Marginal Effects Estimates of
Probability a True Spell is Reported with Length of Spell Only (Part 3)

| | Working Family Tax Credit | | | |
|---|---|---|---|---|
| | Complete Spells | All Spells | Quadratic | Year Shift |
| True Spell Length | 0.0009 | -0.002 | 0.091** | -0.023** |
| Complete Record Spell | | -0.219 | -0.300 | -0.234 |
| True Spell Length Squared | | | -0.002** | |
| True Spell >=11 | | | | 0.629** |
| Log-Likelihood | -23.0 | -34.8 | -35.9 | -31.3 |

Table 6c: Probit Marginal Effects Estimates
Probability a True Spell is Reported with Length of Spell Only (Part 3)

| | Job Seekers Allowance | | | |
|---|---|---|---|---|
| | Complete Spells | All Spells | Quadratic | Year Shift |
| True Spell Length | 0.027* | 0.040** | -0.092 | 0.030 |
| Complete Record Spell | | 0.241 | 0.493 | 0.247 |
| True Spell Length Squared | | | 0.008 | |
| True Spell >=11 | | | | 0.159 |
| Log-Likelihood | -32.3 | -33.9 | -32.2 | -33.8 |

Table 7a: Probit Marginal Effects Estimates
Probability of True Spell with Length of Spell&X (Part 3)

| | Income Support | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| True Spell Length | 0.004 | 0.006 | 0.048* |
| True Spell Squared | | | -0.001* |
| Complete Spell | -0.229 | -0.269 | -0.246* |
| Age | -0.002 | | |
| Male | -0.037 | | |
| Married or Cohabiting | 0.140 | | |
| No Qualifications | | -0.184 | |
| Own House | | -0.499** | -0.541** |
| London or SE | | -0.059 | |
| Log-Likelihood | -27.4 | -24.7 | -23.8 |

Table 7b: Probit Marginal Effects Estimates
Probability of True Spell with Length of Spell&X (Part 3)

| | Working Family Tax Credit | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| True Spell Length | 0.0003 | 0.003 | 0.022** |
| True Spell Squared | | | -0.0005** |
| Complete Spell | -0.159 | -0.179 | |
| Age | -0.019* | | -0.005** |
| Male | 0.182 | | |
| Married or Cohabiting | 0.277* | | 0.087* |
| No Qualifications | | -0.074 | |
| Own House | | 0.151 | |
| London or SE | | -0.158 | |
| Log-Likelihood | -29.3 | -32.3 | -24.4 |

Table 7c: Probit Marginal Effects Estimates
Probability of True Spell with Length of Spell&X (Part 3)

| | Job Seekers Allowance | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| True Spell Length | 0.042** | 0.039** | 0.039** |
| True Spell Squared | | | |
| Complete Spell | 0.247 | 0.251 | 0.248 |
| Age | -0.006 | | |
| Male | 0.108 | | |
| Married or Cohabiting | -0.014 | | 0.090 |
| No Qualifications | | -0.187 | |
| Own House | | -0.040 | |
| London or SE | | 0.048 | |
| Log-Likelihood | -32.8 | -33.2 | -33.4 |

Table 8a: Poisson Model Estimates
Reported Spell Duration on log True Spell Length (Part 4)

| | Income Supplement | | |
|---|---|---|---|
| | Complete Spells | All Spells | Year Shift |
| Intercept | 2.023$^{**}$ | 1.521$^{**}$ | 1.274$^{**}$ |
| Log True Spell Length | 0.121$^{✠}$ | 0.350$^{✠✠}$ | 0.522$^{✠✠}$ |
| Censored True Spell | | 0.318$^{**}$ | 0.383$^{**}$ |
| True Spell $>$ year | | | -0.373$^{*}$ |
| Log Likelihood | -70.7 | -182.7 | -181.3 |

Table 8b: Poisson Model Estimates
Reported Spell Duration on log True Spell Length (Part 4)

| | Working Family Tax Credit | | |
|---|---|---|---|
| | Complete Spells | All Spells | Year Shift |
| Intercept | 0.315 | 0.182 | -0.099 |
| Log True Spell Length | 0.757✠✠ | 0.806✠✠ | 0.985** |
| Censored True Spell | | 0.536** | 0.511** |
| True Spell > year | | | -0.307 |
| Log Likelihood | -87.5 | -119.7 | -119.1 |

Table 8c: Poisson Model Estimates
Reported Spell Duration on log True Spell Length (Part 4)

| | Job Seekers Allowance | | |
|---|---|---|---|
| | Complete Spells | All Spells | Year Shift |
| Intercept | -1.918$^{**}$ | -2.043$^{**}$ | -2.813 |
| Log True Spell Length | 1.512$^{✠✠}$ | 1.559$^{✠✠}$ | 2.019$^{✠✠}$ |
| Censored True Spell | | 0.073 | 0.254 |
| True Spell > year | | | -0.791$^{**}$ |
| Log Likelihood | -44.6 | -50.6 | -48.4 |

$^{✠}$indicates statistically significantly different than 1 at the 95% confidence level; $^{✠✠}$indicates statistically significantly different than 1 and 0 at the 95% confidence level; $^{*}$indicates statistically significantly different than 0 at the 90% confidence level; $^{**}$ indicates statistically significantly different than 0 at the 95% confidence level.

Table 9a: Poisson Model Estimates
Reported Spell Duration on True Spell Length and X's (Part 4)

| | Income Support | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| Intercept | 2.128** | 1.547** | 2.053** |
| Log True Spell Length | 0.273$^{\maltese\maltese}$ | 0.347$^{\maltese\maltese}$ | 0.285$^{\maltese\maltese}$ |
| Censored True Spell | 0.328** | 0.379** | 0.342** |
| Age | -0.003 | | |
| Male | -0.673** | | -0.618** |
| Married or Cohabiting | -0.297** | | -0.332** |
| No Qualifications | | -0.160 | -0.209* |
| Own House | | na | |
| London or SE | | 0.074 | |
| Log-Likelihood | -162.0 | -181.2 | -160.6 |

Table 9b: Poisson Model Estimates
Reported Spell Duration on True Spell Length and X's (Part 4)

| | Working Family Tax Credit | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| Intercept | -1.810** | 0.206 | -1.513 |
| Log True Spell Length | 0.848✠✠ | 0.739✠✠ | 0.775✠✠ |
| Censored True Spell | 0.596** | 0.707** | 0.739** |
| Age | 0.058** | | 0.053** |
| Male | -0.883** | | -0.797** |
| Married or Cohabiting | 0.067 | | |
| No Qualifications | | 0.704** | 0.498** |
| Own House | | 0.106 | |
| London or SE | | -0.174 | |
| Log-Likelihood | -95.3 | -108.3 | -91.5 |

Table 9c: Poisson Model Estimates
Reported Spell Duration on True Spell Length and X's (Part 4)

| | Job Seekers Allowance | | |
|---|---|---|---|
| | Model 1 | Model 2 | Final Model |
| Intercept | -1.541** | -1.586** | -1.526** |
| Log True Spell Length | 1.422** | 1.298** | 1.314** |
| Censored True Spell | 0.182 | 0.398 | 0.333 |
| Age | -0.040** | | -0.028** |
| Male | 1.020** | | 0.983** |
| Married or Cohabiting | 0.387 | | |
| No Qualifications | | 0.639** | 0.279 |
| Own House | | 0.018 | |
| London or SE | | -0.312 | |
| Log-Likelihood | –40.1 | -47.2 | -40.7 |

Table 10: Poisson Model of False Positive Spells

| | Income Support | | | Working Family Tax Credit | | |
|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Final Model | Model 1 | Model 2 | Final Model |
| Intercept | 1.605** | 2.394** | 1.791** | 2.083** | 2.349** | 2.054** |
| Age | 0.017** | | 0.012** | 0.046** | | 0.039** |
| Male | -0.220 | | | -1.094** | | -1.062** |
| Married or Cohabiting | -0.052 | | | -0.240 | | |
| No Qualifications | | 0.352** | 0.122 | | 0.400* | -0.014 |
| Own House | | -0.314 | | | 0.335** | 0.148 |
| London or SE | | -0.972** | -0.806** | | 0.109 | |
| Log-Likelihood | -91.2 | -91.7 | -88.1 | -43.7 | -55.5 | -43.7 |

| Table 11a: Income Support Primary Model Estimates | | | |
|---|---|---|---|
| Negative Binomial Base Specification | | | |
| | Base | No X's | Preferred X's |
| Age | 0.022* | 0.011 | -0.006 |
| Male | -0.382 | -0.254 | -0.599 |
| Marr/Cohab | 0.411 | 0.272 | -0.076 |
| N. Kids | -0.098 | -0.069 | 0.0002 |
| No Quals | -0.028 | -0.009 | -0.031 |
| High Quals | -0.476 | -0.328 | -0.316 |
| Own House | -0.326 | -0.205 | -0.254 |
| London/SE | -0.230 | -0.021 | 0.524 |
| Urate | -.039 | 0.044 | -0.003 |
| Intercept | 3.044* | 2.598* | 3.359* |
| Alpha | 1.906* | 2.263* | 0.728* |
| n | 386 | 386 | 386 |

| Table 11b: Working Family Tax Credit Primary Model | | | |
|---|---|---|---|
| Negative Binomial specification | | | |
| | Base | No X's | Preferred Models |
| Age | 0.027* | 0.042 | -0.018 |
| Male | -0.372* | -0.545* | 0.724 |
| Marr/Cohab | -0.342* | -0.341 | -0.298* |
| N. Kids | 0.215* | 0.230* | 0.123* |
| No Quals | -0.550* | -0.640 | -0.836 |
| High Quals | -0.296* | -0.314 | -0.030 |
| Own House | -0.160 | -0.161 | -0.204 |
| London/SE | 0.165 | 0.195 | 0.128 |
| Urate | -0.157* | -0.107 | -0.96* |
| Intercept | 2.934* | 2.519* | 3.489* |
| Alpha | 0.827* | 2.060* | 0.184* |
| n | 436 | 436 | 436 |

Primary Model Estimates: Job Seekers Allowance

Table 11c: Job Seekers Allowance Primary Model

Negative Binomial Specification

| | Base | No X's | Preferred |
|---|---|---|---|
| Age | 0.004 | 0.007 | 0.031 |
| Male | 0.144 | 0.118 | -0.511 |
| Marr/Cohab | -0.343 | -0.136 | -0.359 |
| N. Kids | 0.250* | 0.108 | 0.173 |
| No Quals | 0.408 | 0.183 | -0.125 |
| High Quals | -0.343 | -0.312* | -0.331 |
| Own House | -0.555* | -0.357* | -0.471 |
| London/SE | 0.119 | 0.094 | 0.081 |
| Urate | 0.101 | 0.045 | 0.061 |
| Intercept | 1.315 | 1.987* | 1.808 |
| Alpha | 1.495* | 0.205* | 0.731 |
| n | 320 | 320 | 320 |