

# DATA REQUIREMENTS FOR THE EVALUATION OF STATE WORKER PROFILING MODELS

## Introduction

The goal of this project is to measure and analyze the accuracy of state Worker Profiling and Reemployment Services (WPRS) models. Summarized claimant data will be used to identify, to the greatest extent possible given the available data and analysis results, factors correlated with good or poor model performance.

## Required Data

Since different WPRS models and systems are found throughout the country, the specific data elements submitted will be slightly different for each state. This study requires three general types of information for each claimant. The first is the set of **data elements used to predict benefit exhaustion** at the time of the initial claim. These will be used to replicate the probability assigned to the claimant at the time of the initial claim and form the basis of the predicted outcome. The second general type of data is a **measure of benefit receipt** that will form the basis for the actual outcome. The last general type of data is an **indication of whether the claimant was referred to services** under WPRS. The analysis will compare the actual and predicted outcomes and explore the reasons for different levels of accuracy.

## Data Elements Used To Predict Benefit Exhaustion

The data elements provided for this study should be those used in the state WPRS model to compute probabilities of exhaustion. For example, if your state model uses industry, education and wage replacement rate, you would provide these data elements for all claimants with benefit year begin dates in the specified time range. For variables requiring categorization, you may either provide the value before categorization or the binary categorical variables, whichever is easier. For computed variables such as wage replacement rate, the state may provide the computed variable or the data elements used to compute it and the computation method. Data for the variables in the model and computation methods should be consistent with the information provided on the WPRS Survey in questions 22 and 24.

Collection of these data elements is necessary because a critical part of the subsequent analysis will correlate the predictive variables or categories with model performance. However, if states can additionally provide the predicted value assigned by the model or screen at the time of the initial claim, this would be very helpful in validation of the replicated state model.

## **Data On Benefit Receipt By Claimants**

In order to help determine whether the initial prediction of benefit exhaustion was a good one, a measure of benefit utilization is necessary. There are several ways that the state could provide this information, but the preferred approach is to provide the maximum potential benefit amount and the amount of benefits paid over the benefit year for each claimant.

## **Data On Referral To Services**

The final data set should include a single variable that indicates whether each claimant was referred to mandatory employment services through the WPRS program. This should be a simple indicator variable and can be coded as “yes/no”, “0/1” or any other suitable system.

## **Selecting Claims And Claimants**

Some states profile all claims, while other states exclude certain claimants (e.g., job attached) from worker profiling entirely. The data provided for this project should contain information for only those claimants who would be assigned scores by the profiling model and be eligible for referral under WPRS. For example, if the state does not profile additional or transitional claims, then those data should be excluded. Similarly, if the state does not normally create profiling scores for job attached claimants, data for those individuals should be excluded as well. In general, the subset of claims requested should be in agreement with answers to questions 11 and 26 on the WPRS Survey.

States are requested to provide these data for all claimants who meet the above criteria and whose benefit years began in the period 01/01/2003 through 12/31/2003. To provide an analysis of a period with more robust growth, data are also requested for claimants who meet the above criteria and whose benefit years began in the period 01/01/1999 to 12/31/1999. If claimant data for 1999 is not readily available due to Y2K computer changes or archival processes, claimant data for 2000 may be substituted. Please submit data for each year as a separate file.

## **Data Format**

One record should be submitted for each claimant who meets the criteria described above. The record should include the variables used to compute profiling scores, the maximum benefit amount and the total amount of benefits paid over the course of the benefit year. The data should be submitted in ASCII format with a brief description of the field names and widths. Please note that no personal identifiers should be present in the data that are submitted.

Below is an example of the data file from a hypothetical state that uses wage replacement rate, industry categorized, and years of education categorized to predict benefit exhaustion.

Case	Maximum Benefit Amount	Benefits Drawn	WPRS score	Referred To Services	Years Of Educ.	NAICS code	Weekly Benefit Amount	Base Period Wage
1	6500	5000	0.56	Y	16	623546	250	29000
2	5850	5850	0.86	Y	12	243567	225	23400
3	5980	2760	0.36	N	15	734539	230	28000

In this example, the state has elected to provide the raw data for industry (NAICS code) and education (years of education) as opposed to the categorized forms. The state would have provided information in its survey response on how these variables are categorized; therefore, the categorical forms of the variables could easily be re-created.

The state has also chosen to provide the raw data for wage replacement rate rather than the computed variable. The survey response would show how the state computes each claimant’s wage replacement rate. In this example, the state computes wage replacement rate as weekly benefit amount divided by the base period wages divided by 52; therefore, the state has provided both of these variables.

To help determine the actual claimant outcomes, data are also provided for the maximum benefit amount in dollars and the amount of benefits paid to the claimant in dollars. The estimated probability (score) that the WPRS model calculated for each claimant is included to help verify that the re-created model is consistent with the state calculations. Similarly, a single variable, coded as “yes” or “no” is included that indicates whether each claimant was referred to services.

If the state elects to provide the categorized and computed variables, the data file might look like this:

Case	Maximum Benefit Amount	Benefits Drawn	WPRS score	Referred To Services	Categorized Education	Industry Category	Wage Replacement Rate
1	6500	5000	0.56	Y	4	6	0.448
2	5850	5850	0.86	Y	2	2	0.500
3	5980	2760	0.36	N	3	7	0.427

The survey asks detailed questions regarding how these data are converted into probabilities; therefore sufficient information should be present to allow an accurate replication of the state model.

If you have questions regarding this data request, please contact Scott Gibbons at [gibbons.scott@dol.gov](mailto:gibbons.scott@dol.gov) or 202-693-3008.