Article:

Van Landeghem, B. orcid.org/0000-0001-6759-8893, Desiere, S. and Struyven, L. (2021) Statistical profiling of unemployed jobseekers : the increasing availability of big data allows for the profiling of unemployed jobseekers via statistical models. IZA World of Labor. 483. ISSN 2054-9571

10.15185/izawol.483

Part I ("one-pager")

Title: Statistical Profiling of Unemployed Jobseekers

Teaser: The increasing availability of big data allows us to profile unemployed jobseekers through the use of statistical models.

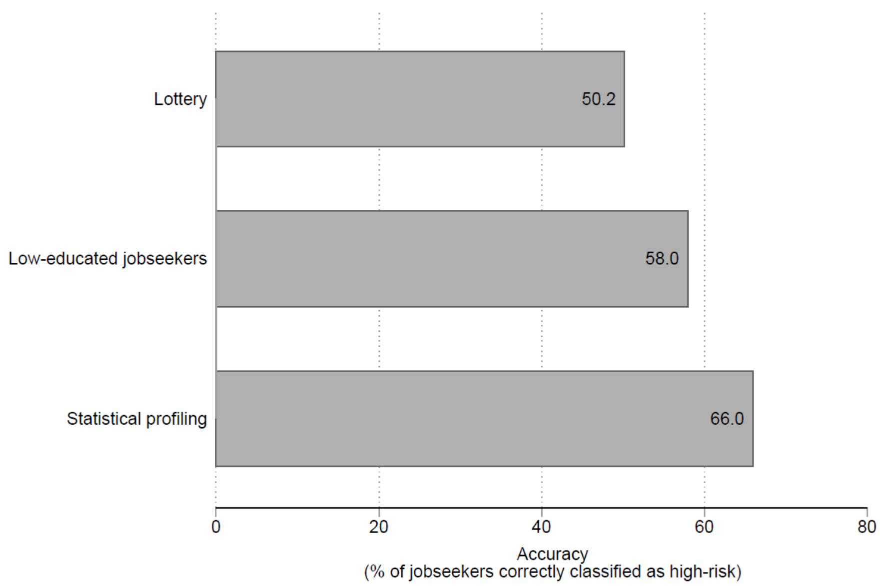
Keywords: Statistical Profiling; Long-term Unemployment; Benefit Exhaustion; Labour Market Discrimination.

Elevator pitch:

Statistical models can help public employment services to identify factors associated with long-term unemployment and to identify groups at risk. Statistical profiling models will probably even become more prominent as new machine learning techniques in combination with the increasing availability of big data will improve their predictive power. However, a policy maker cannot just define an outcome variable at the start of the project and walk away: a continuous dialogue between data analysts, policy makers and caseworkers is very important. Indeed, throughout the process, normative decisions are to be made: profiling practices misclassify many individuals. They can reinforce but also prevent existing patterns of discrimination.

Graphical abstract

Statistical profiling is more accurate in predicting long-term unemployment than a lottery or simple selection rules



Source: Desiere, S., B. Van Landeghem and L. Struyven (2019), Wat het beleid aanbiedt aan wie: een onderzoek bij Vlaamse werkzoekenden naar vraag en aanbod van activering. HIVA, KU Leuven.
Note: We compare three selection rules which all classify 33.8% of new jobseekers as high-risk (at risk of being unemployed for more than six months). The lottery randomly labels 33.8% of jobseekers as high-risk. The second rule labels low-educated jobseekers as high-risk, while jobseekers with higher educational levels are classified as low-risk. The selection rule based on the statistical profiling model, developed by the Flemish Public Employment Service, classifies all jobseekers with a predicted score of being employed after six months lower than 45% as high-risk jobseekers. Administrative data allow us to verify which individuals were correctly classified.

Key findings

Pros (max. 5)	**Cons** (max. 5)
+ Statistical models reveal systematic patterns in the association between socioeconomic and sociodemographic variables, and the outcome of interest.	- The improvement compared to a lottery is modest and many individuals tend to be misclassified.
+ They can direct future research on the causes why some groups are more at risk, an on how the gap can be closed.	- Statistical profiling risks to reinforce patterns of discrimination.
+ They can be used in the process of scrutinizing governments. (Have gaps between groups become smaller over time, are resources being directed to those most at risk?)	- Statistical profiling models predict outcomes, but do not tell us whether it is efficient to target high-risk individuals with certain programmes.
+ Statistical profiling models give us an indication of the potential duration of an unemployment spell. This can be important information for a caseworker but also for the jobseeker, as it is well-known that a jobseeker's perception about the chance of finding a suitable job within a given time window can have an impact on their job search strategy.	
+ Under some circumstances these models can also reduce existing patterns of discrimination.	

Author's main message

Statistical profiling can help us to identify individuals at risk of becoming long-term unemployed and variables predicting long-term unemployment. The models do however not unravel the mechanisms behind these relationships, and will hence not inform us directly about suitable policies to tackle long-term unemployment. It is also not straightforward to evaluate whether policies which target these groups as identified at risk are effective: statistical profiling does not necessarily help us to conduct causal analysis. Policymakers who consider using statistical profiling to target individuals (e.g. to define groups for a programme) should evaluate the ethical implications: individuals are often misclassified and statistical profiling can reinforce patterns of discrimination prevailing in a society.

Part II

The <u>whole paper</u> should be about 27,000 characters (incl. spaces) for an article with two figures (NOT including *references*). Part II should be approximately 24,600 characters (again excluding references). As a rough guide, a figure is worth approximately 1,800 characters so the total character count should be adjusted accordingly depending on the number of figures in your paper.

Please submit your paper as a Word file (.doc), and use 12pt Times New Roman and 1.5 line spacing.

Part II includes:

- **Motivation** (about 1,000 characters)
- **Discussion of pros and cons** (about 18,000 characters)
- **Background information**
- **Figures** (about two, approx. 3,600)
- **Limitations and gaps** (about 1,000 characters)
- **Summary and policy advice** (about 1,000 characters)

- **Acknowledgments**
- **Competing interests**
- **Reference list**

(For further information: see Author guidelines http://wol.iza.org/dms/documents/Guidelines-and-style-sheets/IZA_WoL_Author_Guidelines.pdf)

Motivation

While profiling of jobseekers is as old as employment activation itself, the methods of profiling have changed profoundly.

Traditionally, profiling of jobseekers by employment services has been rule-based, with some discretion for caseworkers, and related to large general groups, for example younger versus older jobseekers. More recently, however, governments are increasingly developing and implementing statistical profiling models based on administrative and/or survey data to predict whether a jobseeker will become long-term unemployed. This development is in line with a broader expectation of governments to conduct evidence-based policy making, to prevent a prolonged spell of joblessness, and to tailor services to the individual. Given the increasing popularity of statistical profiling and the increasing opportunities to build profiling models, it seems useful to us to review these practices, to discuss how these might inform policy making as well as the potential moral implications that we face when implementing these models to target the unemployed.

Background Information 1:

Key Principles of Statistical Profiling: The Statistical Model and Decision Rules

A statistical model will return an estimate of an individual's probability of exhausting benefits, an estimate for the number of months in unemployment/on benefits etc. It is then up to the researchers and policy makers to define decision rules or cut-off points. Let us assume a policy maker who is interested in identifying new benefit claimants who are at risk of exhausting their benefits. One might decide to categorize those as high-risk for whom the predicted chance of exhausting benefits is larger than 50%, and those individuals as low-risk for whom the latter chance is lower than 50%. But one could also use other cut-off values, such as 60% or 80%. Alternatively, one can adopt relative decision rules, which are dependent on the budget that is available to target these individuals that are identified as high-risk. The Kentucky Profiling model for example, gives each individual a score from 1 to 20, where score 1 means that one belongs to the 5% individuals with the lowest predicted risk, and score 20 that one belongs to the 5% individuals with the highest predicted risk.[1] Individuals with score 20 are then invited for a mandatory training programme, and if the budget allows those with score 19 are invited and so on, up to the point where the budget is exhausted.

Background Information 2: Evaluating Profiling Practices

Profiling, or a particular profiling practice, hinges upon the statistical profiling model, and the decision rule that uses the predictions of the model as inputs to sort people into groups (e.g. low-risk and high-risk). A natural way to evaluate a profiling practice is then to investigate how many people were correctly classified. One might encounter terms such as sensitivity (the percentage of those classified as high-risk who do eventually exhaust their benefits) and specificity (the percentage of those classified as low-risk who do eventually not exhaust benefits). Obviously we cannot look at sensitivity and specificity in isolation from each other, as there is a tradeoff between the two. If misclassification is less likely near the top and bottom of the ranking, making the high-risk pool smaller will increase the sensitivity, but will decrease the specificity and vice versa.

In general, one likes to observe that those with the highest profiling scores exhaust their benefits much more frequently than those with the lowest scores. A useful single-item metric is the Profiling Score Effectiveness Metric[2] (PSEM), which can be calculated as follows:

$$PSEM = 1 - (100-y)/(100-x)$$

Where x is the overall percentage of individuals who eventually exhaust their benefits, and y the x% individuals with the highest profiling score who eventually exhaust their benefits.

If the model does not do any better than a lottery, the metric will be zero. If we have the perfect model, the metric will be one. As an example, imagine that we are in a state where 20% of new benefit claimants will eventually exhaust their benefits. For the 20% of individuals with the highest profiling score (as predicted by a statistical model), 60% of individuals eventually exhaust their benefits. The PSEM for this model is then equal to: 1-(100-60)/(100-20)=0.5

Background Information 3: Profiling versus Targeting

It is important to make a distinction between profiling on the one hand, and targeting on the other hand. Statistical profiling aims to split the pool of jobseekers in homogeneous groups, groups of individuals who have the same chance of becoming long-term unemployed or to exhaust benefits. It is then up to stakeholders (the Public Employment Service, policymakers …) to decide how to use this information. One route is to not use it for operational purposes, but to look at the common characteristics that the model has identified for people who are more at risk than others. The model does not inform us about why this is the case, and how to effectively help these people at risk. But at least this information might spark a debate, might direct future research and eventually the development of policies that aim to support these people identified as being at risk. Another extreme is to actively use profiling to target individuals in a fully automated way. Those who are classified as high-risk are then directed to voluntary or mandatory Active Labour Market Programmes, are set to be contacted more frequently by caseworkers etc. But again, the statistical profiling model will be able to identify people at risk, but will not tell us whether the interventions are effective. As a middle ground, one can also use the classification produced by the profiling practice to inform caseworkers or the employment office, who can then use the classification as an additional source of information to make decisions regarding the coaching and monitoring they provide for an individual. We however lack evidence on whether and under what circumstances providing information to caseworkers that stems from profiling models will improve targeting, and this is definitely an avenue for future research.

Discussion of Pros and Cons

Predictors of Unemployment duration

Statistical profiling is often based on cross-tabulations or regression models.[2] But recently, machine-learning techniques have been used as well for this purpose. Indeed, some of these novel techniques have been implemented into standard statistical software packages. Moreover, computing power of personal computers and servers has increased significantly over the last decades, and the amount of administrative data available for (policy) research is on the rise in many countries. In addition, as job search and interactions between jobseekers and caseworkers now often happens online, tools on the websites of public employment services are able to gather valuable data from online CVs, and surveys can be administered at a low cost online from nearly all clients rather than from a sample. These recent developments have offered public employment services such as the Flemish Employment and Vocational Training Office the opportunity to develop statistical profiling models using modern data-hungry and computationally expensive machine learning techniques.[3] Machine learning techniques allow us to develop models that fit the data better than if we were to use standard regression techniques. Importantly, machine learning techniques take into account the problem of overfitting. Adding terms to a regression will always increase the predictive power, and in the extreme each observation is completely described by the regression equation. But this does not mean that this equation will have as much predictive power for next cohorts. It might well be that we have one woman in our sample aged 34 and 3 months, living in West street and having two children of 6 and 8 and that she becomes long-term unemployed. But in the next cohort we might have a woman with exactly the same combination of characteristics who finds a job immediately. Machine learning techniques will take this problem into account.

Academic research has certainly been an important source of inspiration for the development of statistical profiling. The academic literature in various disciplines across the social sciences has explored the determinants of unemployment duration or benefit exhaustion. These studies often provide very useful ideas for developing profiling models. Obviously, building such a model remains a matter of trial and error and making ad hoc decisions. Firstly, whether variables are good predictors will differ across countries. Secondly, in smaller-scale academic studies, one often uses variables (such as measures for soft skills) that are not always straightforward to collect at the population level if this requires filling out extensive surveys. Finally, apart from practical hurdles, legal restrictions related to privacy protection can prevent policymakers to collect, use or merge data for the development of profiling models.

In 2000, a case study to investigate predictors of unemployment duration was conducted in Minnesota.[4] Although small-scale (989 usable observations), it was rather influential as it e.g. inspired the Dutch Public Employment Service when developing their profiling model. The case study recruited Unemployment Insurance (UI) claimants who were then re-interviewed one year later to track their labour market history. Variables that were associated with lower reemployment success are being non-white, having worked with your previous employer for more than 5 years, being female and at the same time having children under the age of 18. Variables that are associated with higher reemployment success are measures of economic need (number of children under the age of 18 and economic hardship). Other variables that have explanatory power are one's region of residence and one's occupation. The explanatory power of macroeconomic variables has been widely documented now in policy reports discussing statistical profiling models based on administrative data. It is interesting that a large number of variables such as years of education, self-reported job search and conscientiousness, are not significant. This might be due to the small sample size, but there might also be several mechanisms at work that cancel out each other. For example, we know that a good labour market history, a curriculum vitae with few gaps and a short on-going unemployment spell are very important for being successful on the labour market. But as the results in this case study show, being employed with one employer for a long time predicts lower reemployment success, maybe because skills have become obsolete or because these individuals are not used to negotiate with employers and to be active on the job market. Having young children might create a need for higher income, and to accept a job more quickly and to increase search efforts, but as the above results suggest, opportunity costs can rise due to caring responsibilities. Finally, there can also be mediating effects, since variables are correlated with each other. For example, the authors find that conscientiousness predicts higher reemployment success once other variables are dropped from the model.

A similar project, but of a much larger scope, was set up by the Institute for Labor Economics (IZA) and is called the IZA Evaluation Dataset.[5] The IZA Evaluation Dataset is a German nationwide sample of 12 monthly cohorts of individuals who became unemployed between June 2007 and May 2008, and the dataset tracks these individuals for 2.5 years after inflow. Regression analyses on these data show that higher school-aged educational attainment and further (professional) training are related with less months spent in unemployment, as also are being younger, a strong employment record, having high numeracy skills, a high internal motivation or locus of control, being conscientious, being optimistic about finding a job and having a good labour market history. Again, macroeconomic conditions (the local unemployment rate) can have strong predictive power.

But it is interesting that in the literature, especially the economics literature, there are not many (influential) studies that report exercises that are very close to statistical profiling of the unemployed. There are two main reasons for this. Firstly, the international literature is often more interested in the general population, and less in a specific subsample of individuals who enter unemployment. One is interested in how education, cognitive and noncognitive skills are related to labour market success within the population, not necessarily of those who are currently unemployed. Secondly, and more importantly, the economics literature is not just interested in models that accurately predict an outcome. The literature pays more attention to the mechanisms that actually lead to these outcomes. When we discussed the Minnesota study, we already mentioned that some variables might have an effect through different channels. It is quite difficult to disentangle all these effects, and one needs a setting in which we can evaluate what happens if one variable changes, all else being equal. Therefore, studies tend to focus on one factor at the time, rather than presenting a comprehensive model that accurately predicts ex post outcomes. But statistical profiling models can offer us a prediction of the unemployment spell for each individual, which, according to economic research, is very relevant information for jobseekers themselves. It is well-known that jobseekers are often too optimistic about their employment chances. These biased perceptions imply that they take different decisions with regards to their job search strategy than if they would have had the correct information.[6]

Normative Implications of Statistical Profiling

Before asking data analysts to build and implement a statistical profiling model, policymakers can define the outcome variable which the statistical model should predict (e.g. exhausting benefits, becoming long-term unemployed) and decide how they want to use this information (see the background box 2). For example, they can decide to route high-risk individuals to a mandatory training programme, or to offer them subsidized training. However, this is not the endpoint of the policymaker's responsibility.

A first strand of concerns comes from the misclassifications produced by the model. Models are far from perfect. In the last large-scale exercise to investigate models from the profiling models used in different states of the USA, the highest Profiling Score Effectiveness Metrics were around 0.25, and the lowest were even below 0.1.[2] As models will never perfectly predict the ex-post outcome, the data analyst and policy maker will need to have a dialogue on which decision rule to implement, and how to trade sensitivity for specificity. Suppose that specific support is available for individuals who are categorized as high-risk. A government might be very worried about people in need not getting the appropriate support. This government is hence interested in a profiling practice with a high specificity: those who are categorized as low-risk should indeed be those who are not exhausting their benefits ex post. If the statistical model gives us predictions for the chance of exhausting benefits, one can then adapt the decision rule by choosing a lower cut-off value above which individuals are categorized as high-risk.

Conversely, if after an election we have a switch of regime, the new government might like to economize and might be worried about people having access to support who actually do not need it. The new government is hence interested in a profiling practice with a high sensitivity: those who are categorized as high-risk should indeed exhaust their benefits ex post. The new government could hence decide to increase the threshold value above which one is classified as high-risk. The pool of 'high risk' individuals will then become smaller. If misclassification is less likely near the tails of the ranking, the sensitivity will increase. The specificity, however, will decrease, as the fraction of those who eventually exhaust benefits will increase among the low-risk pool. It is hence clear that different regimes might prefer a different sensitivity-specificity combination, but that the optimal combination (given the model) will also depend on how one intends to use profiling. A mandatory programme for the 'high-risk' group might lead to a different conclusion than an opportunity for a voluntary training for the high-risk group.

Another very important normative implication is the potential reinforcement of stigmatization of minorities. Statistical profiling models aim to predict outcomes but do not really focus on causal relationships. A good example of this are the variables for one's employment history. These often tend to be good predictors for unemployment duration, but as discussed above, do not tell us much about the mechanisms. But when implementing these predictive models, it is still important to have an idea of the relationship between cause and consequence when it comes to understanding the normative implications of a profiling practice. A huge body of literature offers us very convincing evidence that there is a causal relationship between race and the job finding rate, or, in other words, that labour market discrimination exists.[7] As a consequence, race can be an important determinant of whether or not one exhausts benefits. But do we therefore necessarily want to invite jobseekers belonging to a minority more often to demanding mandatory training sessions? On the other hand, if profiling is primarily used to offer additional opportunities to vulnerable jobseekers (rather than to monitor job search), inviting proportionally more jobseekers belonging to a minority may pose less ethical questions, and may even be considered positive discrimination.[8]

If indeed this relationship is causal, and ethnic minorities suffer from discrimination, a policy maker might prefer not to reinforce the stigmatisation. It is, however, not straightforward to avoid such reinforcement by statistical profiling models. In practice, Public Employment Services complying with data privacy law like the GDPR have adopted rules making it unlawful to include contentious variables such as gender, age and ethnicity in the statistical model. However, these contentious variables are often correlated with other predictors, such as proxies for local labour market conditions. Indeed, minorities tend to be concentrated in certain areas. Recently, economists have proposed a simple methodology to partially mitigate the problem, and have applied their suggestion to the Worker Profiling and Reemployment Services model in the United States.[9] The main idea goes that the contentious variables are still included into the model, but at the prediction stage, the actual values of the latter variables are not used but replaced with the average for the population. This practice avoids that other explanatory variables pick up the effect of labour market discrimination. Obviously, the predictive power of the model decreases, but the reinforcement of stigmatization does as well: in their empirical example, the percentage of black jobseekers in the high-risk group decreases from 22% to 16%.

It is worth noting that after such adjustments, statistical profiling models might even prevent discrimination, especially when the alternative is discretionary power of caseworkers to allocate people to programmes. In the hiring context, for example, there is evidence that limiting discretionary power of hiring managers to overrule test results increases the quality of the pool of hired workers.[10] Algorithms are more transparent than the minds of decision-making humans, who often are prone to unconscious biases against certain groups.[11]

Gaps and limitations

Profiling models show us systematic patterns in unemployment duration, but a large part of the differences between people remains unexplained by the data. This means that these models do a bit better in classifying individuals ex ante than a lottery, but many individuals still tend to be misclassified ex post. Moreover, profiling models focus on predicting an outcome, and not on causal relationships. The latter implies that it can be difficult to interpret the results from a complex profiling model. Many variables in such models might also impact each other, e.g. the level of education can have an impact on the income that one previously earned. Including all these variables together in a model makes it difficult to interpret the coefficients, and, for example, to predict what would happen if a subsidized programme is able to increase the educational attainment of the jobseeker. While statistical profiling models might help us to a certain extent to identify people at risk, they do not reveal which policy programmes are effective for whom. In any case, statistical profiling models will support rather than replace caseworkers. Many Public Employment Services are currently experimenting with striking the right balance between automated decisions and human-based decision so that both approaches can reinforce each other. Finally, it is worth emphasizing that complex statistical profiling tools should ultimately improve the labour market outcomes of jobseekers. To the best of our knowledge, this has not yet been carefully evaluated in the academic literature, although two studies have shown that statistical profiling rules can outperform caseworkers with regard to assigning jobseekers to the optimal programme.[12,13]

Summary and policy advice

Matching survey data with administrative data can offer windows of opportunity to build more accurate statistical profiling models. While statistical profiling models are helpful to better identify jobseekers who are at risk of becoming long-term unemployed, there are some caveats when it comes to use them to better target unemployed individuals. Even if models contain a rich set of variables, a large number of individuals will be misclassified. Moreover, there is a risk of reinforcing the stigmatization of minority groups if used incorrectly. Policy makers need to maintain an intense dialogue with researchers to determine an ideal tradeoff between false positives and false negatives. Statistical profiling can be an additional source of information for caseworkers and public Employment Services. It does not inform us about causal relationships, but it can help us to ask ourselves the question why a certain group is more at risk than others, and can hence be a guide to develop research projects that investigate causal mechanisms. Policy makers may want to promote the conjunction of statistical profiling with causal inference methods such as large-scale randomized controlled trials in combination with machine learning. For example, one can identify whether indeed people at risk of becoming long-term unemployed do benefit more from certain programmes.[14] Or, one might investigate with (causal) machine learning techniques whether programmes have heterogeneous treatment effects.[12] Instead of profiling jobseekers with respect to the predicted unemployment duration, one could then profile them with respect to the predicted effectiveness of a programme.

Acknowledgments

We thank the editor and two anonymous referees for their valuable comments and suggestions.

Key References

[1] Black, D.A., J. Galdo and J.A. Smith (2007). Evaluating the Worker Profiling and Reemployment Services System using a regression discontinuity approach. *American Economic Review* 97(2), 104-107.

[2] Sullivan, W. F., L. Coffey, L. Kolovich, C.W. McGlew, D. Sanford and R. Sullivan. (2007). *Worker Profiling and Reemployment Services Evaluation of State Worker Profiling Models: Final Report.* ETA Occasional Paper(2007-15).

[3] Desiere, S., K. Langenbucher and L. Struyven (2019). *Statistical profiling in Public Employment Services.* OECD Working Paper, OECD.

[4] Wanberg, C.r., L.M. Hough and Z. Song (2002). Predictive validity of a multidisciplinary model of reemployment success. *Journal of Applied Psychology* 87(6), 1100-1120.

[5] Arni, P., M. Caliendo and K. Zimmermann (2014). The IZA Evaluation Dataset Survey: a scientific use file. *IZA Journal of European Labor Studies* 3, 1-20.

[6] Spinnewijn, J. (2015). Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs. *Journal of the European Economic Association* 13, 130-167.

[7] Bertrand, M. and S. Mullainathan (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94(4), 991-1013.

[8] Desiere, S. and L. Struyven (2020). Using Artificial Intelligence to classify Jobseekers: The Accuracy-Equity Trade-off. Journal of Social Policy, Forthcoming.

[9] Pope, D.g. and J.R. Sydnor (2011). Implementing Anti-Discrimination Policies in Statistical Profiling Models. *American Economic Journal: Economic Policy* 3, 206-231.

[10] Hoffman, M., L.B. Kahn and D. Li (2018). Discretion in Hiring. *Quarterly Journal of Economics* 133, 765-800.

[11] Kleinberg, J., J. Ludwig, S. Mullainathan and C.r. Sunstein (2018). Discrimination in the Age of Algorithms. *Journal of Legal Analysis* 10, 113-174.

[12] Cockx, B, M. Lechner and J. Bollens (2019). *Priority to Unemployed Immigrants? A Causal Machine Learning Evaluation of Training in Belgium.* IZA Discussion Papers, 12875.

[13] Lechner, M. and J. Smith (2007). What Is the Value Added by Caseworkers? *Labour Economics* 14, 135-151.

[14] Altmann, S., A. Falk, S. Jäger and F. Zimmermann (2018). Learning About Job Search: A Field Experiment with Job Seekers in Germany. *Journal of Public Economics* 164, 33-49.

Further reading:

Arni, P., M. Caliendo and R. Mahlstedt (2014). *Predicting the Risk of Long-Term Unemployment: What Can We Learn from Personality Traits, Beliefs and Other Behavioral Variables? Working Paper*.

Dohmen, T. and B. Van Landeghem (2019). *Numeracy and Unemployment Duration*. IZA Discussion Papers, 12531.

Farber, H.S., C.M. Herbst, D. Silverman and T. von Wachter (2019). Whom Do Employers Want? The Role of Recent Employment and Unemployment Status and Age. *Journal of Labor Economics* 37, 323-349.

Loxha, A. and M. Morgandi (2014). *Profiling the Unemployed: a Review of OECD Experiences and Implications for Emerging Economies. World Bank Working Paper*. Social Protection and Labor Discussion Paper, no. SP 1424.

Mueller, A., J. Spinnewijn and G. Topa. Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence and Bias. *American Economic Review*, forthcoming.