Review of statistical profiling

Nicola Duell and
Gael de Moraes

CONTENTS

Australia

1. Background

Australia has a long history of recognizing the needs of disadvantaged job seekers. Back in the 1993–94 budget, there was a shift from the target group approach to the use of 'risk-based criteria' for identifying and assessing disadvantaged job seekers. Previously, characteristics such as Aboriginal and Torres Strait Islander status, migrant status, age, disability, and single-parent status were used to identify 'at greatest risk' job seekers for early or preferential access to labor market assistance, thus defining target groups. This approach was not considered as sufficiently accurate (Lipp 2005).

Another key driver for the development of a statistical profiling in Australia was the reorganization of employment service delivery and active labor market programs (ALMPs) to job seekers. Reforms involved the creation of Centrelink, a single delivery point for accessing employment services, and the development of a contestable market for publicly funded employment placement services, the Job Network (Lipp 2005). Thus, employment services in Australia, in contrast with the majority of other countries, are delivered by private contractors, since 1998. The government provides resources for such private agencies and nongovernmental organizations (NGOs) to perform their services.

Under the first Employment Services Contract, over 200 providers delivered Job Matching (a placement service), 112 delivered Job Search Training (a two-week program delivered around the fourth to sixth month of unemployment), and 125 delivered Intensive Assistance (case management for the long-term unemployed and other disadvantaged job seekers). Centrelink staff referred job seekers to these services. On a voluntary basis, job seekers could register with multiple Job Matching services. In the first two years, the Job Network achieved a similar number of placements as the previous arrangements but at a lower cost. From 2000, the Department of Employment and Workplace Relations began to regularly publish star ratings of provider performance at over 1,400 individual sites, based on how many paid placements and paid outcomes were achieved, with regression adjustments considering client characteristics and the state of the local labor market (OECD 2012).

To better allocate its resources, the government uses a statistical profiling system called Job Seeker Classification Instrument (JSCI), to identify the quantity and difficulty of job seekers. Centrelink is responsible for the implementation and operation of the JSCI.

2. Databases used

Australia uses existing administrative information of the applicant as data input into its system, along with information provided by an online questionnaire and from personal interviews.[1] Information on the disability/medical conditions factor may be taken from a job seeker's Employment Services Assessment (ESAt) or Job Capacity Assessment.[2]

3. Output variable

Long-term unemployed, that is, 12 months out of work.[3]

4. Explanatory variables

The JSCIs contain the following 18 factors:

a) Age
b) Gender
c) Recency of work experience*
d) Job seeker history
e) Educational attainment*
f) Vocational qualifications
g) English proficiency (language and literacy)*
h) Country of birth
i) Indigenous status
j) Indigenous location
k) Geographic location
l) Proximity to labor market
m) Access to transport
o) Phone contactability
p) Disabilities or medical conditions*
q) Stability of residence*
q) Living circumstances (family status)
r) Criminal convictions (disclosed ex offender)*
s) Other personal factors[4] (personal characteristics requiring professional or specialist judgment)*

The questionnaire contains 49 questions. Items marked in red trigger JSCI Supplementary Assessment (JSA). JSA "is conducted when the JSCI identifies a job seeker as having particular severe or multiple barriers to employment or when the nature of the employment barrier requires professional and/or specialist advice or assistance."[5]

1 Desiere, Langenbucher, and Struyven 2019, 12 and 22.
2 https://docs.employment.gov.au/system/files/doc/other/fact_sheet_the_role_of_the_job_seeker_classification_instrument.pdf.
3 Desiere, Langenbucher, and Struyven 2019, 12.
4 Loxha and Morgandi 2014, 29; Australian Government, JSCI.
5 Lipp 2005, 7.

Job seekers can be identified as requiring professional or specialist assessment when the following characteristics are observed or disclosed:

- Low motivation/self-confidence/self-esteem
- Personal presentation which may adversely affect their ability to secure employment
- Psychological problems (for example, aggressive behavior, depression, anxiety, grief, family/relationship difficulties)
- Substance abuse problems
- Experience of torture or trauma.[6]

The variables are weighted according to the assessed difficulty that the factor brings into placing the job seeker in the job market.[7] A more detailed description of how the weight of each variable was constructed can be found below (5. B. Development of the pointing system).

5. Statistical/econometric model used

Logistic regression.[8]

6. Development of a questionnaire for implementation of the profiling tool

To better assess the applicants, upon registration, a questionnaire is given to each of them.[9] It consists of 49 questions overall, of which the individual must fill at least 18. The questions were previously only asked and answered through the phone or personally. Recently, the Australian government started switching to an online questionnaire.[10]

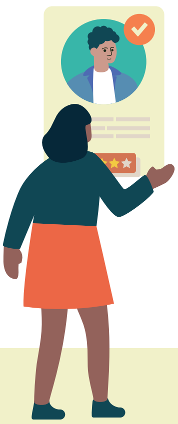
a) Methodology used to develop the questionnaires

The JSCI is based on the current Job Seeker Screening Instrument and the Client Classification Levels Questionnaire used for case management.[11]

The Department of Employment and Workplace Relations' experience with these instruments has been considered in the development of the JSCI, particularly in the formulation of the principles guiding its development.

- **Classification.** The JSCI is a classification instrument and not a detailed assessment of an individual job seeker's needs.
- **Minimum number of factors.** The JSCI should encompass the minimum number of factors necessary to form an acceptable instrument.
- **Accuracy.** There is no perfectly accurate method of determining the precise employment needs of each job seeker. On this basis, the JSCI is a relative and not an absolute measure of the circumstances of one job seeker against another.

6 Lipp 2005, 25.
7 Commonwealth of Australia 1998, 1.
8 Desiere, Langenbucher, and Struyven 2019, 12.
9 "Services Australia or employment services providers conduct the JSCI with job seekers. Job seekers participating in the Online JSCI Trial can complete their JSCI using the Job Seeker Snapshot on the jobactive website. In general, job seekers participating in jobactive, Disability Employment Services (DES), Community Development Program (CDP), Transition to Work, Time to Work and ParentsNext employment services need a JSCI." Australian Government, JSCI. https://www.employment.gov.au/job-seeker-classification-instrument.
10 Desiere, Langenbucher, and Struyven 2019, 22.
11 Commonwealth of Australia 1998, 2.



- **Reliance on survey factors.** Well-developed statistical screening instruments can have important advantages over other alternative methods of identifying disadvantaged clients, in terms of objectivity, reliability, and accuracy.
- **Net impact.** Survey factors should be considered for their net impact on the ability of a job seeker to gain employment.
- **Transparency.** The purpose and operation of the JSCI should be easily understood by the community, Job Network members, job seekers, and Centrelink staff.
- **Consistency.** The instrument will need to operate consistently across different groups of job seekers, different labor markets, special and disadvantaged groups, and intensive assistance levels.
- **Acceptance.** The JSCI should be accepted by Job Network members and have credibility as a classification instrument.

The JSCI has been developed through the following:

Formal research. An extensive survey of job seekers was undertaken by the department, with data collected since late 1995. The survey and administrative data have been analyzed to identify the risk factors that are associated with prolonged unemployment. The analysis determined estimates of the average effect of these factors. Factors tested in the survey form the basis of the JSCI.

Expert judgement. A Classification Working Group was established to make recommendations on additional factors that could not be tested in the survey (such as homelessness and disclosed ex offender status) but that also contribute to labor market disadvantage (LMD).

Wider consultations. Major stakeholders, apex organizations, and the employment services industry were consulted. The consultations were very supportive of the Classification Working Group proposals. Highly positive feedback was received about the JSCI. Representatives from case management organizations and apex bodies commented on the comprehensive nature of the instrument and how it is a significant improvement on the previous screening and classification instruments.

Formal research

The JSCI has been developed through formal research—from an extensive job seeker survey, the 1997 JSCI Survey, undertaken by the department—which correlates various personal characteristics, such as age, educational attainment, and family status, with the incidence of long-term unemployment.

1. The first stage of the survey was to gather system-based data from the Commonwealth Employment Service JOBSYSTEM (CES JOBSYSTEM) database on job seekers newly registering or already on the CES register in late 1995 and early 1996.
2. The second stage, in March–May 1997, had a follow-up survey questionnaire (and telephone contact, where necessary and possible) mailed to these job seekers, primarily to determine their labor market status. The survey questionnaire also asked clients about a number of characteristics that are either recorded in a different way or not at all recorded by JOBSYSTEM. These are their main activity in the labor market in the past five years before registration; any disability they have that may affect their chances of finding employment, household type/family status, and English language and literacy skills; and whether or not they possess any work-related qualifications.

a. The 12–15-month period between the first and second stages allowed time to distinguish between clients who remained unemployed for long periods and those who were able to find work. This enabled the identification of the factors that contribute to the difficulty of placing people into employment and increase the likelihood of becoming or remaining long-term unemployed.

3. Statistical analysis was undertaken to identify factors that increase the likelihood of becoming or remaining long-term unemployed. Respondents to the survey were matched against the department's records on labor market program access. Those job seekers who were in full-time programs such as JobSkills, the Landcare and Environment Action Programme, and New Work Opportunities were excluded from the analysis.
4. The statistical model estimates the relative weights or 'points' of various socioeconomic factors and personal characteristics, as indicators of LMD. The weights represent the independent effects of each factor (say age) on the predicted probability of becoming long-term unemployed after controlling for the influence of all other factors (such as educational attainment and duration of unemployment) that have been included in the model. Factors tested in the survey have formed the basis of the JSCI.[12]

Parameters

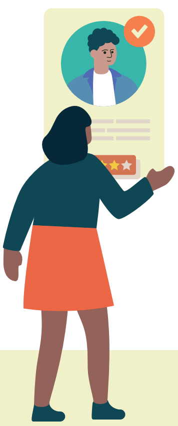
Resources. Staff will be skilled in administering the instrument in as consistent and accurate a way as possible rather than skilled in objectively determining a job seeker's level of disadvantage. Most Centrelink customer service officers are not professional assessors and are not required to have the qualifications necessary to administer an instrument involving a substantial level of specialist or professional judgment.

It should be possible to administer the JSCI with the resources available to Centrelink. The major cost of administering the JSCI within Centrelink will be the staff time to interview each client to gather the necessary information. Therefore, the instrument should be simple for both job seekers and staff to use and should incorporate the minimum number of necessary factors.

In addition, the development of any secondary classification process should consider the availability of specialist staff to exercise some professional judgment, for example, occupational psychologists.[13]

Systems. The JSCI is a system-based device that will be part of the department's Integrated Employment System (IES). The IES can include a number of factors and sub-factors and simple secondary processes. It may be possible to build complex processes into the system, but they may not necessarily deliver a better outcome.

Time frame. Development of the JSCI has been undertaken in a time frame that accommodates system deadlines, the timing of the tender assessment process (for probity reasons, the Classification Working Group could meet only after tender submissions closed), and the time needed for classification of job seekers before the opening of the new market on May 1, 1998.



The JSCI was revised several times, for example, with the introduction of new activation strategies. Thus, with the introduction of the 'Participation Model' in 2003, the number

12 Commonwealth of Australia 1998, 9.
13 Commonwealth of Australia 1998, 12.

of factors included was revised and reduced from 18 to 14 at that time. In addition to duration of unemployment, the factors omitted were transport, proximity to labor markets, and small community dynamic (Lipp 2005). Later changes increased the number of factors to 18 (OECD 2014, Connecting People to Jobs).

Legal issues. A number of factors are affected by privacy and discrimination legislation, for example, ex offenders and disability. The department has obtained legal advice for developing the JSCI, which has been considered by the Classification Working Group in its examination of possible JSCI factors. The department has also consulted with the Privacy Commissioner while finalizing the JSCI.[14]

Consultation. In addition to utilizing the industry expertise of the Classification Working Group, the draft JSCI developed with the Working Group was the subject of wider consultations with

- Key stakeholders, including other Commonwealth departments and agencies;
- Apex organizations;
- The employment services Industry Reference Group (the recommendations of the Working Group were referred to the Industry Reference Group for comment); and
- Case managers and providers (drawn from all states and territories) in focus group settings in Sydney, Adelaide, and Melbourne.

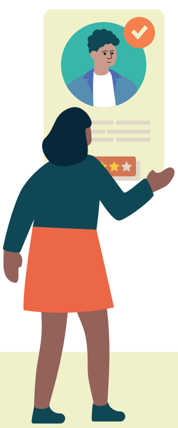
b) Development of the pointing system

- **Geographic location (regional grouping).** Pointing system based on unemployment rates (average 12 months) and employment growth rates (last eight quarters) for the region. The more the unemployment, the higher the pointing.[15]
- **Country of birth.** Pointing system-based unemployment rates in Australia experienced by migrants from different countries using the 1996 Census data and survey. Higher unemployment rates are translated into more points. Also, the pointing system tries to capture lack of recognition of qualifications earned overseas, lack of knowledge of and attachment to the Australian labor market, cultural differences, and English language problems. For example,
  - Category A - (Australia and low unemployment rat–s - 5.0 percent to 9.77 percent);
  - Category B - (medium unemployment rates - 9.8 percent to 17.1 percent); and
  - Category C - (high unemployment rates - 17.7 percent to 54.1 percent).
- **Age.** Based on the survey data, labor market perception of age as a barrier for employment, and Long-Term Unemployment (LTU) rates among age tranches. Points were awarded to young people to reflect the lack of experience.
- **Education**. Level of education as related to probability of unemployment. Based on the survey.
- **Vocational qualifications.** Based on the survey findings regarding vocational skills and also on recognized qualifications. “This is to clarify the position of migrants, including those with professional qualifications, who should not be assumed to have an advantage in the labour market if their overseas qualifications are not recognized in Australia.”[16] It also tries to capture the value added of “educational attainment which has a vocational orientation (e.g. a nursing or education degree in comparison with a general arts degree)” and “non-educational qualifications which are necessary

14 Commonwealth of Australia 1998, 13.
15 Commonwealth of Australia 1998, 19.
16 Commonwealth of Australia 1998, 26.

pre-requisites for entry to particular occupations (e.g. trade or other special licenses, tickets and other technical qualifications)."[17]

- **Duration of unemployment (omitted after review).**[18] This is based on the survey. It tries to capture the degradation of motivation to look for job, assuming that a higher duration of unemployment means less motivation; the degradation of network and contact advantage in finding jobs; and years worked before unemployment.
- **Recency of work experience.** This is based on the survey. It tries to capture the impact of absence from employment in skills degradation, the lack of networking and contacts, and the type of work experience (part-time, full-time, seasonal, and so on).
- **Family status.** This is based on the survey. It tries to capture one person's responsibility for others (financial or otherwise) and, on the other hand, personal support received from others. For example, "Survey findings show that those living alone are more likely to become long term unemployed than other job seekers" and "Job seekers who are sole parents generally face significant disadvantage in the labour market."[19] Indirectly, the variable captures lack of suitable childcare arrangements, the number and ages of children, the number of hours available to work, and so on.
- **Transport (removed).**[20] This is not based on survey. It refers to the 'job seeker's physical ability to attend interviews and travel daily to work. Key sub-factors relate to access to adequate private or public transport'.
- **Contactability.** This is not based on survey. It refers to the job seeker's ability to contact and be contacted by the employers.
- **Proximity to the labor market (removed).**[21] This is not based on survey. This variable tries to capture the physical distance of the job seeker from labor, for example, people living in rural areas and taking over 90 minutes to get to the city center.
- **Language and literacy.** Due to the difficulty of assessing the level of English and literacy of someone without referring to an expert's opinion, the following tool was developed: "Job seekers will be asked to self-assess their language ability. In addition, language ability will be assessed by the Centrelink officer's judgement based on structured observation during their interaction with the job seeker."[22]
- **Disability/medical condition.** This is not based on survey but rather on professional expertise. It tries to capture the workability of job seekers with disabilities and difficult medical conditions, for example, "ability to understand work instructions, ability to lift and move objects."[23] To better assess this, the Department of Family and Community Services (DFaCS) developed the Work Ability Tables (WAT). The WAT generate an index of workability ranging from 0 to 99.[24] The WAT dimensions are
  - Ability to report regularly for work;
  - Ability to persist at work tasks;
  - Ability to understand and follow instructions;
  - Ability to communicate with others in the workplace;
  - Ability to travel to and from and move at work;
  - Ability to manipulate objects for work;
  - Work behavior;



17 Commonwealth of Australia 1998, 26.
18 Lipp 2005, 8.
19 Commonwealth of Australia 1998, 28.
20 "Statistical analysis showed that the 'access to transport' factor did not significantly assist the identification of the most disadvantaged job seekers." (Lipp 2005, 8).
21 Now accounted for under the geographic factor. (Lipp 2005, 8).
22 Commonwealth of Australia 1998, 36.
23 Commonwealth of Australia 1998, 39.
24 "Only selected job seekers with a disability (i.e. those streamed to secondary classification) will be assessed with the WAT to determine the impact of the disability on their ability to work." (Commonwealth of Australia 1998, 40).

  - Ability to learn and undertake a variety of tasks; and
  - Ability to lift, carry, and move objects at work.
- **Stability of residence.** This is not based on survey. It tries to capture the degree of stability of housing (stable, temporary, or homeless) and the consequences of this such as, lack of access to running water and electricity (which could result in hygiene problems), lack of stable support networks, and so on.
- **Disclosed ex offenders.** This is not based on survey. The factor tries to capture the possible reluctance of employers to employ ex offenders because of a criminal record, the restriction on availability for referral, and the availability to work at all (restrictions related to the sentence).
- **Disadvantage resulting from personal factors requiring professional or specialist judgment (secondary classification).** This is not based on survey. "Secondary classification will occur in circumstances where a job seeker has a particularly severe barrier or where the nature of the employment barrier requires professional and/or specialist advice."[25] The Classification Working Group identified the following personal characteristics that may be identified for secondary classification:
  - Poor motivation, poor self-confidence, and low self-esteem
  - Poor presentation
  - Psychological problems
  - Substance abuse
  - Torture and trauma experience (including victims of domestic violence and childhood sex abuse).

Staff implementing the JSCI are trained to recognize symptoms and appearances related to those characteristics. Upon identification of such symptoms, the staff can refer the job seeker to professional assistance.

The questionnaire was changed several times. For example, research in 2002 for the department's quality assurance program recommended

- Implementing a consistent verbatim approach, through training, to the administration of the JSCI by Centrelink to encourage increased consistency and accuracy in collection of job seeker data and
- Improving the JSCI tool to reduce question ambiguity, improving the structure and flow of the JSCI interview, removing unnecessary questions, and improving the ease of understanding questions for job seekers and Centrelink staff.

In response to these recommendations the revised JSCI comprises fewer interview questions, down from a maximum of 60 questions to 30 questions, and the flow of the questions has been improved. The revised JSCI questions and supporting information technology (IT) were tested extensively in usability laboratory trials. The overall results of the testing indicated that the new questions, question flow, and the IT tool are a significant improvement on the previous version that should ensure a more efficient and effective interview process (Lipp 2005).

7. Additional observations

Early studies and research that evaluated the JSCI suggest that job seekers are less likely to reveal sensitive information in the Centrelink interview than in telephone interviews because they are concerned about the potential ramifications on their eligibility for

25 Commonwealth of Australia 1998, 46.

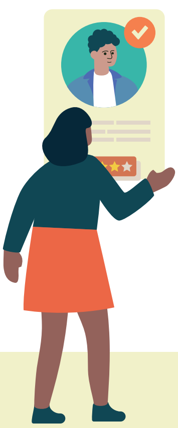
income support (a function performed by Centrelink) and employment prospects (Lipp 2005).

A study conducted in May 2004 examined the accuracy of the JSCI factors as recorded by Centrelink. In-depth interviews with a sample of those having a JSCI score were used to check the accuracy of the score. Centrelink's application of the JSCI was found to be 90.3 percent accurate in identifying the appropriate funding level. This represents a significant improvement over the 82.6 percent accuracy found when a similar exercise was conducted in 2002 before the latest changes to the Centrelink interview (Lipp 2005).

According to Lipp (2005), "the diagnostic statistics for the logistic regression model that underpins the JSCI tend to indicate that the model has low to moderate predictive power. One point to consider is that logistic regressions performed on cross-sectional unit-record data usually have low predictive power according to diagnostic tests and in this context, the JSCI model is not unique. However, when one considers the fact that some factors have been moderated, in terms of points allocated, and that there are simply some determinants that are unable (or at least very difficult) to be quantified, the JSCI is quite effective at predicting the exit rates from allowance across all allowees." Further analysis has been undertaken to determine the effectiveness of the JSCI in predicting exit probabilities for specific classes of allowance recipients/ pensioners such as Disability Support Pensioners, Parenting Payment Recipients, and Youth Allowance Recipients.

These are lessons learned by Australia, based on the experiences and on the development of the JSCI:[26]

- Classification is relative and not absolute. There is no perfectly accurate method of determining the precise employment needs of each job seeker. While job seekers can be placed on a continuum of relative job placement difficulty, there are no obvious criteria that separate one assistance level from another.
- Care needs to be taken to minimize overlap between factors so that 'double counting' is avoided.
- The wording of the questions used to gather information from job seekers needs to be simple and straightforward to ensure that responses are accurate.
- A statistical weights approach established by a survey is more defensible because of its objectivity. Well-developed statistical screening instruments have advantages over alternative methods of identifying disadvantaged clients.
- Instruments should be simple and easy to understand and operate. The additive numerical approach of the Job Seeker Screening Instrument is simpler and more easily understood than the Client Classification Levels, which uses a complicated algorithm behind the scenes to determine a rating.
- Regarding the Client Classification Levels, analysis of data suggests that factors regarded as 'personal barriers' are more important than 'employment strengths' in classifying job seekers.
- It is important to ensure that Job Network members have a clear understanding of the purpose of classification and hence the distinction between classification and detailed assessment of individual job seeker's needs. This avoids misconceptions and unrealistic expectations about the outcomes of the classification process.
- The operational processes used to gather information from clients need to be



26 Commonwealth of Australia 1998, 8.

rigorous and sensitive to clients, to gather accurate and useful data. In particular, it is generally considered that a face-to-face interview is the most reliable means of achieving this.
- All relevant factors should be considered. Some personal characteristics (such as poor motivation) that contribute to a job seeker's disadvantage in the labor market are not addressed directly in the current instruments, although some factors (such as duration of unemployment) may act as proxies for them.[27]

8. Use by the Public Employment Services (PES) and by counsellors/caseworkers

The JSCI has been used to measure job seeker disadvantage and to differentiate outcome fees, which are paid when a job seeker has been in a job for 13 weeks with a second payment after 26 weeks. It is implemented through a questionnaire addressed to the job seeker by Centrelink, with points allocated on the basis of answers to the questionnaire, together with administrative information (OECD 2012).[28]

It is compulsory for caseworkers in Australia, from all employment services awarded by the government, to use the profiling system.[29] The JSCI, also called the 'Job Seeker Snapshot', is a set of questions a job seeker is asked in an interview with the Department of Human Services (DHS) or their employment services provider. If eligible, job seekers may also complete their JSCI online using the Job Seeker Snapshot as part of the Online JSCI Trial. seeker himself/herself through his/her Jobactive account.

The JSCI is used to determine the best level of employment servicing for the job seeker. For example, under Jobactive, the JSCI is mainly used to determine a job seeker's eligibility for Streams A and B and to identify job seekers who have complex or multiple barriers to employment that may require further assessment through an ESAt.

The JSCI is also used to identify job seekers

- Who may benefit from a referral to a DHS social worker (for example, with disclosed domestic violence, family grief, or trauma);
- With lower English language, literacy, and numeracy skills who may benefit from the Skills for Education and Employment or Adult Migrant English Programs; and
- With unrecognized overseas qualifications who may require further information on Assessment Subsidy for Overseas Trained Professionals.

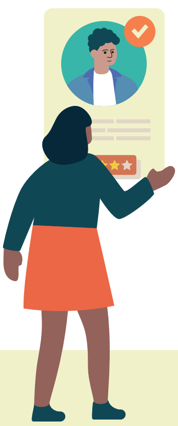
It is necessary to have a JSCI for job seekers participating in the following ALMPs: Jobactive, Disability Employment Services (DES), Community Development Programme (CDP), Transition to Work (TtW), Time to Work Employment Services (TWES), and ParentsNext.

The counsellors (of the subcontracted providers) are free to implement the strategy they consider as the most useful to bring the job seeker into work (black box approach) (OECD 2012).

27 Commonwealth of Australia 1998, 8.
28 OECD 2012.
29 Desiere, Langenbucher, and Struyven 2019; OECD 2012, 12.

JSCI issues an individual score to applicants,[30] who are then directed to different segments of employment services according to their predicted difficulty to reenter the job market.

The JSCI was initially used to stream job seekers to different Job Network services and to different funding levels for more intensive assistance based on the level of disadvantage. Those not assessed as ‘at risk’ were referred to job search assistance. Those job seekers assessed as ‘at risk’ either at registration or subsequently if they continued to remain unemployed for long periods were referred to more intensive assistance. Initially, there were three funding levels for more intensive assistance, which were reduced to two levels in 2000. *The Active Participation Model* was introduced in July 2003 for the delivery of employment services through the Job Network. Unlike the previous model, access to more intensive assistance is now based not only on an early intervention strategy but also on duration of unemployment. In the past, duration of unemployment was a factor that contributed to the JSCI score but did not necessarily guarantee access to more intensive assistance for those who were unemployed for 12 months or more (Lipp 2005). Those assessed by Centrelink at registration as being at high risk of long-term unemployment have immediate access to Intensive Support customized assistance, tailored to their individual needs and available for a six-month period.



30 “Each JSCI factor is given a numerical ‘weight’ or points which indicate the average contribution that factor makes to the job seeker’s difficulty in finding and maintaining employment. The points are added together to calculate the JSCI score which reflects a job seeker’s relative level of disadvantage in the labour market. A higher score indicates a higher likelihood of the job seeker remaining unemployed for at least another year.” Australian Government, JSCI.

The Netherlands

1. Background

"Due to government-imposed cost cuts, clients on unemployment benefits in the Netherlands will increasingly receive computerized services instead of face-to-face services. The instrument, called the 'Work Profiler', enables the selection of job seekers towards face-to-face or computerized services by estimating the clients' chances of finding work within a year."[31]

2. Databases used

Online questionnaire

3. Output variable

Long-term unemployed (12 months). It shows the client's chance of resuming work within one year.

4. Explanatory variables

For a short version of the questionnaire for counsellors, 11 factors have been retained (developed by 2011 and implemented in 2015. A new study of 2018 proposes 18 factors):

- Age
- Years employed in last job
- Problems understanding Dutch (listening, writing, reading, and speaking)
- Views on return to work
- Feeling too ill to work
- Job search behavior (contact with employers)
- Job search intention
- External variable attribution

31 Wijnhoven and Havinga 2014, 1.

- General work ability
- Physical work ability
- Mental work ability.[32]

Of the final 11 factors, 6 enhance and 5 reduce a person's probabilities for work resumption.[33]

a) Other variables used

"Since the Work Profiler was developed by UWV to be used for selection and quick diagnosis, its short questionnaire only focuses on the most important predictors for work resumption. Hence, the Work Profiler does not offer a total personal profile with in-depth insights into the competencies and capacities for return to work. To this aim, UWV has more extensive tests done by its specialized 'Competencies Centres'."[34]

5. Statistical/econometric model used

Logistic regression.[35]

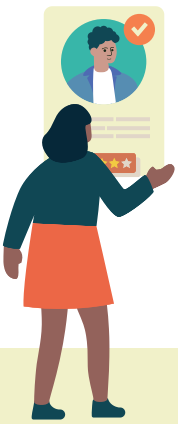
A percentage reflecting the probability to resume work is calculated according to questionnaire responses.[36] The method of calculation does not seem to be disclosed.

6. Development of a questionnaire for implementation of the profiling tool

a) Methodology used to develop the questionnaires

A study 'Predictors of Work Resumption' was carried out to identify which factors are predictive for return to work at the start of unemployment. The research project was a collaboration between the Uitvoeringsinstituut Werknemersverzekeringen Center for Knowledge (Kenniscentrum UWV) and the School of Medical Sciences of the University Medical Cente[37]

The research project 'Predictors of Work Resumption' was carried out in three subsequent stages: a literature study, followed by a cross-sectional study, and finally a longitudinal study.[38] Through this process, it aimed to reduce a large list of possible variables to only a few relevant ones, which bore more impact on the likelihood of reemployment.



32 Heijnen and Dekenga 2014, 14.
33 Wijnhoven and Havinga 2014, 8.
34 Wijnhoven and Havinga 2014, 5.
35 Desiere, Langenbucher, and Struyven 2019.
36 Wijnhoven and Havinga 2014, 3.
37 Brouwer et al. 2011; Wijnhoven and Havinga 2014, 5.
38 Wijnhoven and Havinga 2014, 6.

Figure 1. Methodology used for developing questionnaires

1 LITERATURE STUDY } *identification of potential relevant hard and soft factors for work resumption*

Wanberg's Multidisciplinary Model
- labour market demand
- human capital
- social capital
- economic need to work
- re-employment constraints
- job search behaviour
- discrimination

Vroom's Expectancy Theory
- valence
- expectancy

↓
- motivation

Theory of Planned Behaviour
- attitude
- subjective norm
- self efficacy

↓
- job search intention

↓
- job search behaviour

result: 500 items found

2 CROSS-SECTIONAL STUDY } *questionnaire containing 500 items from the literature study among the quickly and the long-term unemployed*

result: reduction from 500 to 155 items (70 factors)

3 LONGITUDINAL STUDY } *testing the predictive value of the factors from the cross-sectional study*

- research population:
- influx cohort recently unemployed in North Holland
- April 2008 - March 2009 (n=3,618)

result: reduction from 155 to 20 items

final result: WORK PROFILLER

- 20 items
- 11 predictive factors

1. The initial extensive literature study produced many hard and soft factors that may influence work resumption. From the literature, three theoretical models that encompassed these factors were selected: the Wanberg model (Wanberg et al. 2002), the Theory of Planned Behaviour (Ajzen 1991), and the Expectancy Theory of Motivation (Vroom 1964). A total of 500 items that were found relevant were used during the next stage.
2. The 500 items were processed into a questionnaire for the cross-sectional study, which compared the answers of the long-term unemployed with those of the ones who resumed work quickly. By this process, the number of items was reduced from 500 to 155. The nature of cross-sectional research itself did not allow establishing which factors are predictive; it could only demonstrate the di erences between the two populations.

The results of the cross-sectional research showed that the contribution to predicting the chance of job recovery was the largest for questionnaire A and the smallest for questionnaire C.

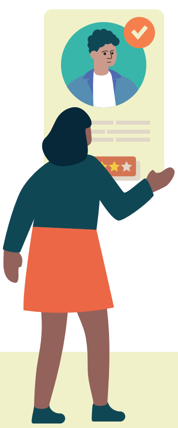
Further reduction of the list of questions is made on the basis of the distinctiveness of items. So, items for which long-term unemployed and quick resumers show no clear (significant) distinctions have been removed for the longitudinal study. Items are also eliminated that in a logistic regression analysis, in which the likelihood of returning to work fluctuated, did not significantly contribute to the probability prediction. Questionnaire C focused on competencies and the Big Five (personality traits).

3. The true predictive value of factors could only be established through a longitudinal study, which was done during the last stage of the research process. In this phase, a large group of recently unemployed people were asked to fill in a questionnaire containing the now reduced list of 155 items. A year later, a follow-up was done with the same people in the administration of the unemployment o ce to see whether they did or did not resume work. The longitudinal study was carried out in the Province of North Holland from April 2008 to March 2009 among all those who became unemployed (about 33,000 questionnaires were sent). The total research population consisted of 3,618 respondents, of which 58 percent found work within a year and 42 percent still had an unemployment benefit after one year. Job seekers whose unemployment benefits ended before 12 months for causes other than work resumption, such as reaching the maximum duration of their benefit, were excluded from the analysis. The longitudinal study allowed for a further reduction of items from 155 to 20.[39]

The differences of explanatory variables in their predictive power between work resumers and non-work resumers were analyzed. The average values and percentages of hard and soft characteristics were looked at separately for those who do not return to work and those who return to work calculated. T-tests and Chi-square tests have been calculated to determine the significance, to identify differences found between the two groups. The following 'hard' and 'soft' factors were considered (Brouwer et al. 2011):

Hard features. Sex, age, ethnicity, problems understanding the Dutch language, educational attainment, training direction, household composition, age of children living at home, number of persons in the household, number of persons with a fixed income in the household, occupation, nature of the profession, industry, number of years worked, number of years in last appeal, number of employers in the last five years, resigned several times in the past five years, work pattern for unemployment (in hours/week), nature of the last employment, unemployment benefit only, level of unemployment benefit, remaining unemployment benefit duration (in days), hours of paid work/week in addition to unemployment benefit, and hours of volunteer work/week in addition to unemployment benefit.

Soft features. For example, vision of return to work, feeling too sick to work, financial necessity, job search behavior around job application preparations, active job search behavior (among which approaching potential employers, sending letters of application after vacancy, approaching acquaintances and open applications), belief in one's own



39 Wijnhoven and Havinga 2014, 6.

ability (self-efficacy), believe in one's own ability when applying for, belief in one's own ability when approaching companies, making an impression, job search attitude (seeking benefit and pleasantness of work and seeking the usefulness and necessity of work), subjective norm (close family and partner, friends and acquaintances, people in the neighborhood, and social support), job search intent (job search hours the next month), perceived barriers to getting work (among which physical health problems, mental health problems, lack of affordable childcare, care for family members, availability for different working hours, financial problems, incorrect training, too little or not the right work experience, foreign background, and so on), able to work several hours a week, work motivation, disadvantages of not working, general confidence, willingness to accept work (variable and/or different working hours, dirty work, moving with compensation/long travel time to work, compensation, adaptation to questionable work (more difficult/temporary work, work outside the box, full-time job, and undesirable work), wishes regarding job, versatility (variation/learning opportunities), social contacts, good working conditions, health problems, and general work ability (physical work ability and psychological work ability).

Predictors of return to work within 12 months

The multivariate logistic regression analysis included 24 factors. As a result of these regressions, 10 factors remained, of which age is the most important predictor of return to work within 12 months.

To determine which hard and soft factors influence return to work, logistic regressions were performed. Before performing a multivariate logistic regression analysis, the independent variables involved in the analysis were checked as to whether they are not too strong to be correlated (multicollinearity). The economic crisis may have an impact on the job resumption of the unemployed.

There was an attempt to also control for the effect of the economic crisis that occurred during the data collection period of the questionnaire.

Development of a short questionnaire

The Knowledge Center UWV had decided in consultation with UWV WERKbedrijf to develop a short questionnaire that can be used by the work coaches. The instrument consisted of 9 out of 10 factors that remained as predictors of job recovery within 12 months. The factor 'perceived health' had not been included in this abbreviated list because this consisted of a questionnaire of 13 items. After reanalysis by means of multivariate logistic regression (where perceived health was left out of the regression analyses), it turned out that psychological work ability was still characterized as a significant predictor. Therefore, at the request of the work coaches, physical work ability also should be included in the new instrument. Ultimately, the abbreviated list consists of 11 factors with 19 items.

7. Additional observations

Tests on the Work Profiler show that it predicts with a certainty of 70 percent. This means that during the first few months of unemployment, the Work Profiler is able to determine correctly for 7 out of 10 job seekers who will resume work within a year.[40]

40 Wijnhoven and Havinga 2014, 8.

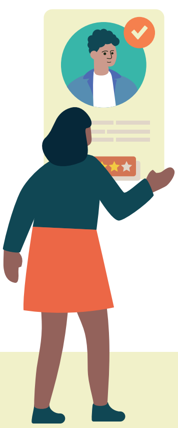
Interesting conclusions for the Dutch labor market context: the older unemployed are at a higher risk of long-term unemployment (Brouwer et al. 2011). It can also be observed that people with *many years of work in the last job* are less inclined to accept an *adaptation-seeking job* (deviating in degree of difficulty, know their own subject, or get a job with a temporary contract) and a job that scores less high on *varied work and many social contacts.* The lower *working search intent* is associated with lower *self-confidence when applying* and a *more negative attitude about the usefulness and need* to look for work. People with *a negative view of returning to work* generally turn out to be ones having lower *self-confidence with regard to job application preparation* (score low to find good vacancies, write good cover letters, be able to list strengths and weaknesses) and with regard to *impressing potential employers* when applying. Furthermore, these respondents see the *usefulness and necessity of job search* (attitude) as less important and consequently exhibit less *job search intention* and *job search behavior* (approaching potential employers). They also indicate that they have little *social support* in seeking and finding one job. This negative *view of returning to work* also includes an older *age*, one lower *work ability*, and more *barriers due to physical* (especially musculoskeletal complaints and complaints related to shortness of breath and pain in chest or heart area) and *mental health* problems and as a result of not doing so *be available for different working hours* (night shift, evening shift, shift work, irregular shift). The *motivation for work* and the *willingness to accept questioning work* (outside own field or more difficult work, work on a temporary basis contract) are lower and the *attribution of their unemployment to external (variable and stable) factors* is also greater.

The instrument was implemented nationwide at the end of 2015. A new research[41] was conducted addressing the following main research questions:

1. What are predictive personal and situation characteristics for resuming work and duration of the unemployment from unemployment benefits?
2. How can the research results in question 1 be translated into an objective scoring system (specifications) for an instrument that is directly applicable in the service from UWV?
3. What are the psychometric properties of the instrument?

Main results of this most recent research. A model with 18 factors predicts better the risk of not resuming work within 12 months. The logit regression model correctly predicts 70.3 percent of resumption of work. Of the 18 factors, there are 6 hard factors stemming from administrative data of UWV; the others are collected through the questionnaires applied, among which 8 are soft factors.

The study also looked separately at a predictive model for unemployment duration (in other words, time to return to work). This was done with the Cox regression model. The factors in this model largely corresponded to the factors in the most sparse, interpretable logistic regression model (17 out of 20 were equal). The predictive properties of the Cox regression model were comparable to the logistic regression model.



The factors in the Work Explorer are sufficiently reliable for the purpose for which they will be deployed (Chapter 7). When determining the psychometric properties, we looked at reliability measures based on inter-item relationships (Cronbach's alpha, Guttman's lambda) of the soft factors measured by multiple items in the extensive Work

41 Dusseldorp, Hofstetter, and Sonke 2018.

Explorer questionnaire used in this research (1.0+). The minimum required reliability is for diagnosis and guidance 0.70; all factors that occur in Work Explorer 2.0 meet this requirement (both in the total group, as in the groups of work resumers and non-resumers). Also, the test-retest correlations of most factors are more than sufficient ($r \geq 0.70$).

8. Use by counsellors/caseworkers

The Work Profiler will be used by the Dutch PES (UWV) for two purposes:

1. **Selection.** The client's chance of work resumption within a year will be employed to determine whether the job seeker will be o ered computerized or face-to-face services. The Work Profiler indicates the chances of work resumption through a percentage, with 0 percent being the lowest and 100 percent being the highest probability for return to work.
2. **Quick diagnosis.** UWV utilizes the outcome of the Work Profiler—which provides a quick diagnosis by illustrating which of the 11 predictive factors for work resumption need to be positively influenced to increase the client's chances of returning to work—to o er tailored services that will increase a client's chances of reemployment. The scores on the individual factors indicate the client's strong points and those that may need improvement.

Individual results of the questionnaire can be used to calculate the probability of resuming work. However, there is uncertainty and inaccuracy as not all predictive factors are considered. The higher the number of variables considered, the higher the accuracy (Brouwer et al. 2011).

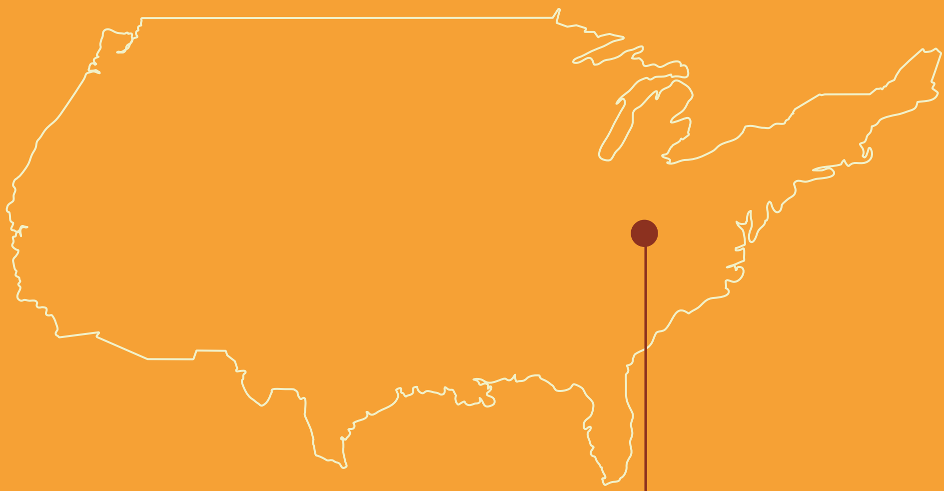
In the model, resumption of work within 12 months is coded as 1 and non-resumption of work in the follow-up period as 0. This implies that as the outcome of the probability formula is higher, the chance of returning to work is greater. Therefore, it is important to choose the cutoff point in such a way that the chance of making a wrong decision the least is where costs are the highest. Brouwer et al. (2011) showed that with a cutoff point of 0.50 69 percent are correctly classified (based on the 11 factors mentioned).

For the diagnosis, the customer's score (between 0 and 100) is compared for each factor with the average scores of the reference group formed by the group of work resumers. The customer is classified per factor into one of the five categories: far above average, above average, average, below average, and far below average (Brouwer et al. 2011).

The indicator is used to select who will have a face-to-face conversation with a counsellor (limited spots), to select what types of services can be offered to a job seeker and from which skills improvements he/she could benefit the most.[42]

The implementation of the tool was tested in PES in 3 locations in 2011 among 9 job coaches and 50 job seekers. The results of this pilot were positive. The working coaches indicated that the results of the Personal Explorer helped them in their qualitative assessment of whether they should offer services to the customer (Brouwer et al. 2011).

42 Wijnhoven and Havinga 2014, 2.



United States

1. Background

The Worker Profiling and Reemployment Services (WPRS) system, mandated by Public Law 103-152 of the Unemployment Compensation Amendments of 1993, **is designed to identify and rank or score unemployment insurance (UI) claimants by their potential for exhausting their benefits for referral to appropriate reemployment services** (Sullivan 2007).

State Work Force Agencies (SWAs), in the different states, have the freedom to implement the models that suit their needs the most.

2. Database

Administrative data and questionnaire

3. Outcome indicator

Benefit exhaustion (50 SWAs).

This approach did not change markedly until 2016 when the Reemployment and Eligibility Assessment (REA) was introduced. Rather than focusing exclusively on the claimants most likely to exhaust their benefits, states can now serve job seekers within a range of scores and align the services to the claimants' characteristics. Evaluations of REA are currently ongoing (Desiere, Langenbucher, and Struyven 2019).

Experiences with different outcome indicators:

- Specific benefit duration - 1 SWA (Kentucky, since the 1990s)
- Proportion of total benefits paid - 1 SWA.

4. Explanatory variables

Overview of the variables

Dependent variables

- Benefit exhaustion - 50 SWAs
- Specific benefit duration - 1 SWA
- Proportion of total benefits paid - 1 SWA
- Exhaustion of benefits and long-term unemployed

Independent variables

- Industry (39 SWAs)
- Occupation (30 SWAs)
- Education (39 SWAs)
- Job tenure (40 SWAs)
- Local unemployment rate (24 SWAs).

Additional variables found

- Wage replacement rate (15 SWAs)
- Time from employment separation to the date the claim is filed, known as delay in filing (15 SWAs)
- Number of employers in the base period (8 SWAs)
- Potential duration (7 SWAs)
- Maximum benefit (10 SWAs)
- Weekly benefit amount (16 SWAs).

Variables used in statistical models vary widely. The majority of SWAs still use the variables recommended by Employment and Training Administration (ETA) when WPRS became law. These are as follows:

- Industry (39 SWAs)
  - The most common method to verify employment and industry classification is a crossmatch against the UI wage record files; 48 SWAs use the crossmatch method.
  - 5 base the industry classification on the initial claim interview.
  - Even if the industry classification is not used in the model, it is collected for other purposes.
- Occupation (30 SWAs)
  - 12 SWAs use the Dictionary of Occupational Titles (DOT) codes as their occupational classification system.
  - 11 SWAs use the O*NET system
    Some directly
    Some based on feedback from the One-Stop.
  - 7 use the Standard Occupational Classification (SOC) classification system.
- Education (39 SWAs)
- Job tenure (40 SWAs)
- Local unemployment rate (24 SWAs).

These variables are in general entered into the models directly. Other SWAs may collect these variables and not use them in their models or use these variables to create other variables that are in the models, such as industry unemployment rate.

Extended number of variables (example Kentucky)

Kentucky used the following explanatory variables (Hawkins et al. 1996): "the (Kentucky) model contained a large number of explanatory variables, including those related to a

- claimant's previous wage,
- UI benefit parameters,
- reservation wage,
- pensions,
- assistance receipt,
- prior UI receipt,
- industry growth,
- occupation growth,
- job tenure,
- work experience,
- reason for separation,
- county unemployment rate,
- county employment growth."

5. Model

The majority of SWAs use logistic regressions (logit):

- Out of 46 SWAs, 43 use a statistical model.
  - Of these, 38 use logistic regression (logit) as the functional form.
  - 5 use linear multiple regression.
  - 1 uses neural network.
  - 1 uses Tobit and 1 uses discriminant analysis.

6. Additional observations

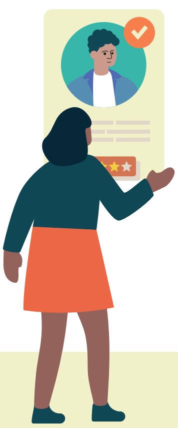
Updates

The major reason for updates has been to convert the occupational classification system from DOT to SOC or O*Net and industry classification system from Standard Industrial Classification (SICs) to North American Industry Classification System (NAICS).

More than half (29) have never revised their models.

Assessing effectiveness

Study of Sullivan (2007) shows the following:



- Development of a metric that demonstrates the effectiveness of various profiling scores. The metric is a statistic that demonstrates the effectiveness of a profiling score. Normally, the metric ranges from 0 to 1. If a profiling score is as effective as a random number generator, then the metric will be insignificantly different from 0. If a metric is a perfect predictor of UI benefit exhaustion, then it will take a value

of 1. A metric of 0.100, means that, for individuals with high scores, the profiling score selects exhaustees 10 percent better than a random number. For the metric, the authors also calculate a standard error. For SWAs, the standard error allows comparison of multiple profiling models for statistically significant improvements.
- **Control for endogeneity.** Because profiling and referral affect observed benefit exhaustion, it is necessary to control for the effect of reemployment services when developing new profiling models.

Results from former evaluation showed (Dickinson, Decker, and Kreutzer 2002, in Sullivan et al. 2007)

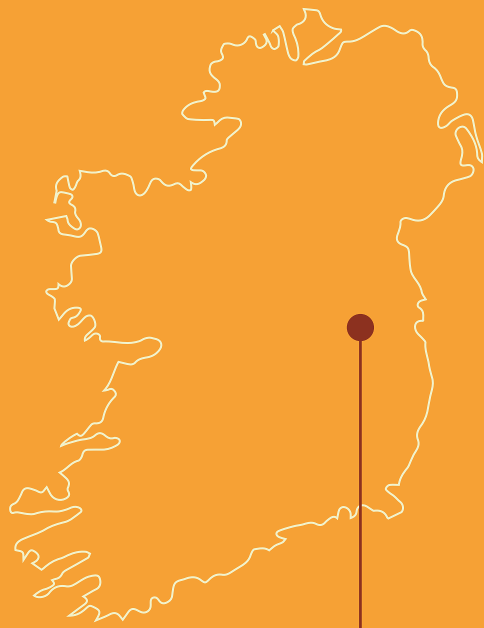
- The states that were using national coefficients provided by the Department of Labor were not as successful as those that had developed state-specific models and
- States need to continually update their models to reflect recent changes in the economy, for example, growth or decline of occupations and industries.

7. Use

- List eligible candidates for referral to reemployment services. Thus, these participants were receiving more services than claimants who were not referred.
- In 50 SWAs, lists of candidates are either mailed or sent electronically to the reemployment services provider.
- In most SWAs, the lists go directly to workshop/orientation staff, while in a few, they go to local management personnel.
- In three SWAs, the lists are sent to administrative staff for review before being sent to the local service provider.
- The two most important determinants of the number of candidates to be served are staff availability and space. Most of the decisions on the number to be served are made locally.
- In 6 SWAs, the number of claimants to be selected and referred is determined by central office personnel and/or negotiation between central and local office personnel.
- In all SWAs (except for the one SWA that does not calculate a score) that use the statistical model, candidates are sorted by their probability of exhaustion.

Implementation of the listing

- The list of eligible candidates for referral to appropriate reemployment services is produced when the model is run.
- 49 SWAs run the model against the claimant's first payment file. The remaining 4 run it against the initial claim file.
- 42 SWAs run the model weekly. The remaining 11 run the model daily.



Ireland

1. Background

The statistical profiling model, called Probability of Exit (PEX) model, was rolled out as part of the Government's new integrated employment and support service, Intreo, which was introduced in October 2012.

In 2014, a revised statistical profiling model was elaborated (McGuiness, Kelly, and Walsh 2014).[43] The objective is to use the result of the model to characterize the labor market prospects of persons who are already long-term unemployed (and who have not been administered the PEX questionnaire), such that those most in need can be prioritized in service provision.

2. Database

a) Database and model development of previous model

The first model developed in 2009 was based on administrative data and a questionnaire (O'Connell et al. 2009[44]).

To develop the model, a unique questionnaire was administered to a large number of (newly registered) job seekers who were then followed up (for 18 months) in the unemployment registers. The individuals who were considered reemployed were those who left the register toward employment and have not returned to employment within the 52 weeks (after registration) or who had exited toward employment for at least 6 weeks earlier.

43 Predicting the Probability of Long-Term Unemployment in Ireland Using Administrative Data, ESRI Survey and Statistical Report Series Number 51.

44 O'Connell et al. 2009.

Due to the lack of information contained in the unemployment registers at the time, a large number of variables had to be collected via questionnaires including education, health, literacy/numeracy levels, and employment/unemployment history (including the number of months spent in employment, on benefits, on community work over the previous five years).

b) Database in the revised PEX model, the LMD model

The new study is based on administrative data only. As a consequence, the range of explanatory variables included within the model developed in the new study is more restricted (McGuiness, Kelly, and Walsh 2014). The administrative data had become richer.

There is a *Job Seekers Longitudinal Dataset*, although the different data sets behind it do not contain the same information. The data sets for recipients of the different out-of-work benefits were combined (instead of calculating two models for the two kinds of benefits and data sets).

3. Output variable

Exit long-term unemployment

4. Explanatory variables

Explanatory variables of the PEX model include

- Age;
- Household indicators (marital status, spouse earnings);
- Education (including also literacy/numeracy levels);
- Health;
- Employment/unemployment history including
  - The number of months spent in employment,
  - Benefits,
  - Community work over the previous five years, and
  - Location.

Explanatory variables of the new model with a limited set of variables (McGuiness, Kelly, and Walsh 2014) but no information on employment/unemployment histories:

- Age
- Nationality
- Education
- Previous occupation
- Household characteristics (including marital status and spousal earnings)
- Unemployment benefit or unemployment allowance
- Location.

The dependent variable is binary and indicates that the claim has closed to employment and has remained closed for at least six weeks.

5. Model

Probit model.

To develop the previous statistical model, separate probit models were estimated for males and females to predict the probability to exit to employment within 12 months. The models were found to have good predictive capabilities, as 83 percent and 85 percent of male and female job seekers, respectively, were correctly classified (when the cutoff point was raised to 0.8 probability of LTU).

More recently, an LMD model was developed, which predicts the probability of exit to employment (McGuinness, Kelly, and Walsh 2014). The new study was applied to those who had already reached 12 months duration on the Live Register. This means that the score can no longer be interpreted as a predicted probability of becoming long-term unemployed. Instead, the score can be viewed as the predicted probability that the individual should have already left the Live Register given their characteristics. Thus, the score is more a measure of relative LMD as opposed to an expected probability of future exit from the Live Register. The result for each characteristic tells the impact that the particular attribute will have on an unemployed person's likelihood of exiting to employment at 12 months.

For the new model, anonymized data were received for 63,795 individuals who made claims for Job seekers Benefit (JB) or Job seekers Allowance (JA) during the relevant period and who subsequently had their claims approved. Basic Live Register information on gender, age, previous occupation, marital status, spousal income, nationality, and geographical information is held for all claimants. Additional information on a range of educational and labor market attributes is held in the *Job Seekers Longitudinal Dataset* for approximately 63 percent of the sample. (McGuiness, Kelly, and Walsh 2014).

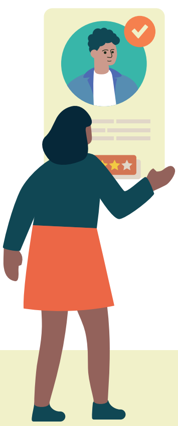
6. Additional observation

The model has fairly good predictive power, with 66 percent of cases correctly classified (at a cutoff of 0.5 probability).[45] The proportion of correctly predicted observations is only marginally below that of the previous PEX study which predicted 69 percent of cases and outperforms existing country profiling systems on which information is available.

7. Use

In Ireland, **initial registration in the unemployment system** includes **a compulsory requirement to complete a questionnaire to access benefits and services,** which serves as a basis to estimate the probability of remaining unemployed for more than 12 months.

Once the responses are processed, the system calculates the **"risk category"** of the person seeking employment, which is then assigned to a **"participation path" before the first meeting with the employment consultant.**



45 It is interesting to note that the previous PEX only had marginally better predictive power at the same cutoff, with 68 percent of cases correctly classified.

With this information and a more in-depth diagnostic interview, the consultant can decide, based on his/her knowledge and experience, what precise combination of services and measures are offered to the job applicant.

Austria

1. Background

The Austrian PES introduced its first statistical profiling model in November 2018 and has tested its performance and evaluated the acceptance by its caseworkers in 2019 (Desiere, Langenbucher, and Struyven 2019).

2. Database

Existing administrative data sources.

Administrative data of the PES as well as of the employment subject to social security database.

3. Outcome indicator

- The short-term function assesses the probability of moving into unsubsidized employment for at least 3 months in the first 7 months after the start of unemployment.
- The long-term function estimates the probability of moving into unsubsidized employment for at least 6 months over 24 months.

4. Explanatory variables

Focus on **socioeconomic variables and labor market history information**

- Gender
- Age
- Nationality
- Education
- Health limitations
- Care responsibilities
- Work experience (type and intensity)
- Frequency and duration of unemployment
- Participation in active labor market programs
- Opportunities (regional labor market development).

Regional labor market data are captured through **segmenting regional PES offices into five clusters**, based on supply and demand for labor in each region.

The full labor market history is available for about two-thirds of all new client inflows. The history is typically incomplete for youth, individuals with longer periods outside the labor market, and migrants (Desiere, Langenbucher, and Struyven 2019).

The varying availability of labor market history information requires different models to be estimated for various subgroups: youth, migrants, and people with a fragmented labor market participation.

5. Model

Logistic regression model.

AMS-Arbeitsmarktchancen-Modell (the labor market chances model of the PES).[46] The model is measuring the 'labor market integration probability.'

The registered unemployed are segmented into groups, depending on their labor market integration probability.

Criteria that were used for the segmentation (Holl et al. 2018)

1. For those with a high labor market probability. At least 66 percent of probability to be employed in the short term (80 days of employment within the following 7 months)
2. For those with low probability. Less than 25 percent of probability of being employed for 180 days over the next 2 years.

6. Additional observation

The model achieves a very high level of accuracy.

7. Use

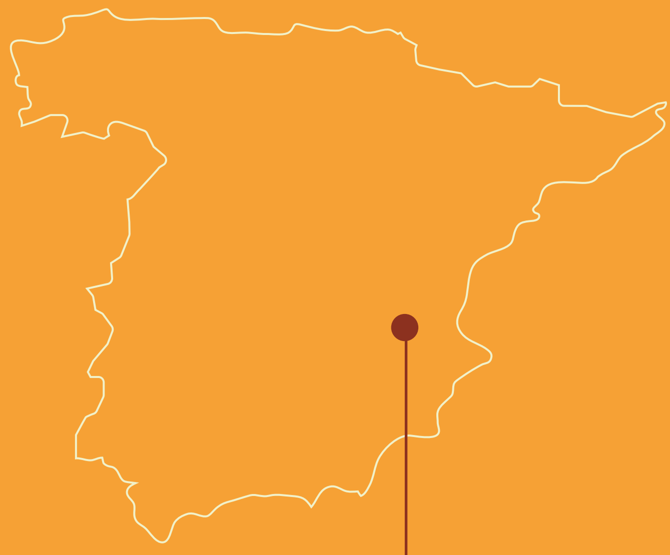
The statistical profiling model is the basis for labor market integration probability - assistance model (Arbeitsmarktchancen-Assistenzsystem AMAS), which is a computer-based model to assign job seekers into 3 groups: low, medium, and high labor market integration probabilities. The final decision for assigning a job seeker to a segment is done by the counsellor.[47]

The results of profiling are discussed between the counsellor and the job seeker.

The tool shall help promote early intervention; improve diagnostics, counseling, and matching; and make a more efficient use of budget through better targeting.

Note that soft profiling instruments are also being developed and used in Austria.

46 Jürgen Holl, Günter Kernbeiß, Michael Wagner-Pinter, 2018, Das AMS-Arbeitsmarktchancen-ModellDokumentation.
47 Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (2019), Arbeitsmarktchancen-Assistenzsystem (AMAS).



Spain

1. Background

Quite advanced work, implemented in some regions, in progress in other regions and at the federal level. Different steps have been undertaken at the federal level to set up a statistical profiling system

2. Development of a questionnaire

To obtain information required for profiling job seekers, the Vocational Guidance Service uses different sources, including an Employability Questionnaire. This questionnaire was designed by the Servicio Publico de Empleo Estatal (SEPE) in response to a request of the Autonomous Communities, who wanted to have a tool for profiling job seekers.

3. Statistical profiling model

Fedea (Foundation for Applied Economics Studies) has developed a profiling model at the request of the SEPE (the Spanish PES at the federal level).

4. Statistical profiling of intervention pathways

Based on the **statistical and logistic regression analysis** carried out by Fedea **and based also on the experience of the Employability Questionnaire,** the SEPE has been focusing the **analysis on profiling itineraries/pathways**, that is, on calculating probabilities of services/combinations of services of increasing the employability of job seekers (*not so much profiling persons, but profiling itineraries*). The objective is to rigorously be able to recommend and offer the most adequate combination of services that boost the employability of job seekers (information from 2018).

Work in progress. Neither of the two tools (questionnaire and profiling model) has gone further and has yet not linked the different segments or profiles to the combination of employment services most adequate to the needs of the segment/profile.

There is the plan to develop the model further and to develop an IT-based tool.

There is not yet one common statistical profiling tool in Spain. A model for statistical profiling is being contracted to an external research/consultancy institution.

In parallel, some *Communidades Autonomas* are more advanced and use their own tool.

5. Database

Administrative data

For developing the questionnaire and for Fedea's statistical profiling model, the main information source has been the Sistema de Información de los Servicios Públicos de Empleo **(SISPE) database with administrative information on the unemployed registered with PES,** thus neglecting other groups (inactive persons and other persons not registered).

6. Outcome variables

Profiling model of Fedea

Measuring the probability of a person to go out of unemployment

7. Variables in the questionnaire (descriptive variables)

The PES at the federal level, called SEPE, developed the Employability Questionnaire in 2014, based on the information managed by its internal statistics department (General Sub-directorate of Statistics and Information), which considers three blocks of individual factors and one of external factors. The blocks of individual factors are

1. **Technical and professional competencies,** with a weight of 40 percent
   - Qualification
   - Experience
2. **Transversal competencies**, with a weight of 30 percent
   - Communicational,
   - Digital,
   - Organizational,
   - Social
3. **Personal factors,** with a weight of 30 percent
   - Availability,
   - Motivation,
   - Objectives,
   - Search activity,
   - Attitude and knowledge of the labor market.

8. External factors - variables

To take account of **local labor market conditions**, in particular, **hiring by occupations**, with information from the SEPE occupational observatory. Depending on the dynamism of the local labor market, it adopts a multiplier value of 0.25, 0.5, or 1 (less, medium, and high dynamic, respectively).

a) Questionnaire: number of questions

The Employability Questionnaire consists of **33 questions** and assigns a **final quantitative value to the employability of the job seeker desegregated by the abovementioned blocks and in total:** low, medium, or high.

9. Statistical profiling model

A statistical profiling model has been developed based on the SISPE data provided by SEPE.[48] Using the information available in the administrative records, the model produces a basic diagnosis of the situation of each individual in relation to the labor market, which is summarized in an employability indicator that captures the likelihood of a job seeker moving from unemployment to employment before a certain period, considering his/her sociodemographic characteristics, work history, competencies, and the evolution of the labor market.

Some limits of this exercise, as the authors point out, are information used comes exclusively from administrative databases and would need to be complemented with additional relevant variables. Moreover, it would be necessary that at least one regional PES implements it to include this additional necessary information for a second stage of construction, which would include ad hoc questionnaires specially designed to deepen the cognitive and noncognitive abilities of the job seeker. Additionally, demand-side information should also be added, possibly coming from job portals, the occupational observatories, and social networks.

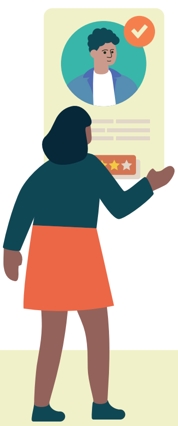
10. Use

The questionnaire is used by part of the **Autonomous Communities** of the common information system[49] as a tool for profiling job seekers; some of them have designed/customized it and use their own questionnaires.

The new project described above shall also aim to build the computer tool for profiling job seekers, to be used by employment counsellors, which will be available for all the Autonomous Communities.

11. Experiences in selected regions

Aragón and Catalonia have carried out statistical analyses of the relevance of the variables included in the questionnaire and have adapted it to their needs.



48 Felgueroso, F. García-Pérez, J. I., Jiménez-Martín, S. (Coord.) (2018). Perfilado estadístico: un método para diseñar políticas activas de empleo. Fundación de Estudios de Economía Aplicada (Fedea).
49 The information of the National Employment System (which includes the national and regional PES) in Spain is integrated in the so-called SISPE. The SISPE integrates the information on active employment policies and unemployment benefits from both the national and the regional PES. Within the SISPE model, two types of systems coexist: those Autonomous Communities using the system and information centrally managed by the SEPE (used by Aragón, Asturias, Cantabria, Castilla-La Mancha, Extremadura, Baleares, La Rioja, Madrid, Murcia, and Navarre, the so-called CEUS—*cesión de uso*—regions) and those with their own information systems that transfer the information to the SEPE database (Andalusia, Basque Country, Canarias, Castilla y León, Catalonia, Galicia, and Valencia).

In 2012, **Catalonia** implemented its own Employability Questionnaire, the Q-Questionnaire (*Questionnaire for the Improvement in Employability and Better Results in the Labour Insertion of the Job Seekers, Cuestionario-Q*), in all employment offices. This is the tool used for segmentation by groups of users of the regional employment offices according to structural, personal, and competencies factors related to both their employability and criticality.

The questionnaire is integrated in the initial interview with the job seeker that takes places in employment offices and is structured along different blocks of information, including personal data, preferences, training, professional information, jobs requested, availability to work, employment search skills, and administrative data. It is aimed at people who want to receive training, counseling, orientation, or intermediation services, as well as at those who want to claim unemployment benefits.

The main objective of Q-Questionnaire is to make an initial diagnosis of the needs, socioeconomic aspects, competencies, and preferences of the users, to achieve greater precision and efficiency in the referral of services. The result of the segmentation obtained from the questionnaire is the basis for the classification of users according to

- The collective (a total of 7), as the result of the combination of the insertion index according to the professional objectives and criticality of the sectors in which the user is seeking for a job;
- The sub-collective (13), which specifies the position of the professional objectives in relation to the labor market and the sectors in which the user is focused and to detect whether there is a need to improve the experience and/or the training;
- The level, which indicates whether there are improvement needs associated with any or all of the factors related to the job search, the basic competencies, and/or the transversal competences; and
- The degree of priority, such as the combination of criticality factors including age, gender, time spent in unemployment, the entitlement to benefits, and the degree of disability.

Once the segmentation has been obtained, an itinerary of the actions to be carried out is defined to improve the employability of the beneficiary, which will also be monitored.

The Q-Questionnaire is continuously evolving to improve the attention to the users and the results of the active labor measures, increase its analytical capacity, and achieve a unique labor record of the person.

It has contributed to changing the attention model in employment offices toward a closer relationship with job seekers; homogenizing the information gathered and the diagnoses carried out by counsellors; and systematizing the attention provided.[50]

Likewise, in 2016, **Aragón** customized the Employability Questionnaire designed by the SEPE,[51] by modifying many questions, including new ones and altering the order of the blocks of information. Thus, the Aragón Employability Questionnaire is currently accessed from the website of the regional employment service with username and password. It is the first step in the diagnosis and the basis for planning the personalized pathway and the intervention. It includes 36 questions about factors affecting employability, related to the attitude of the unemployed when looking for a job, and

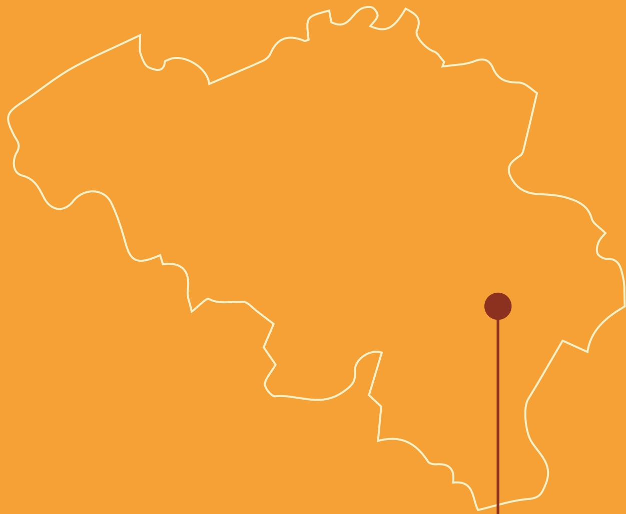
50 More information at https://serveiocupacio.gencat.cat/ca/inici
51 The questionnaire can be accessed at http://www.educationalpaths.com/es/cuestionarios/inaem.

identifies his/her skills, training, experience, interests, and personal and family situation, providing a quantitative value of employability. In 2017, technical assistance services were hired to improve the tool.

Andalusia is developing a system by which the calculation of the level of employability is based on the data collected in the 'employment demand' (*demanda de empleo*) submitted by the job seeker, including, among others, the coherence of the data between the requested occupations, professional experience, and training, as well as the availability of the job seeker. With this information and weighting each one of the items, a numeric value is calculated, which allows to classify the demand into a concrete value on which specific actions can be applied.[52]

In **Galicia**, a very tiny NGO Amigos de Galicia has implemented the project More and Better Jobs (*Máis e Mellor Traballo*), which uses new profiling tools to predict the probability of finding employment with certain characteristics, facilitate the diagnosis of the unemployed, and create personalized itineraries. The project is oriented to the labor market insertion of people with difficulties finding employment because of age, gender, family responsibilities, or lack of training and is supported by the Municipality and the Province of Pontevedra.

52 Social Inclusion Network (*Red de Inclusión Social*, RIS) 2017–2020. Working Group on Information Exchange between Social and Employment Services. Results Report, November 2017.



Examples of other quantitative methods - using machine learning in Belgium (Flanders)

1. Background

The statistical profiling model is part of a new contact strategy that was rolled out in October 2018. The strategy aims to reach and screen all new job seekers within six weeks after registration at the PES, but priority will be given to high-risk job seekers as predicted by the profiling model (Desiere, Langenbucher, and Struyven 2019).

In 2014, the Flemish PES founded an innovation lab that focuses on developing new apps and 'big data' analytics. Using a random forest model with hundreds of variables (called features), this lab has developed a statistical profiling model, called 'Next Steps', that estimates the probability of becoming long-term (> 6 months) unemployed. The model is built in a flexible way so that it can be updated regularly to remain accurate. It was implemented in October 2018.

2. Outcome indicator

Becoming long-term (>6 months) unemployed: it estimates, after 35 days of unemployment, that a job seeker will find work within 180 days.

3. Explanatory variables and database

- **Administrative data:**
  - Job seekers' socioeconomic characteristics
  - Job seekers' labor market history.
- Information collected **by caseworkers** during previous and current unemployment spells is also used in the model.
- An additional innovation is the use of '**click data**', which monitors job seekers' activity on the website of the PES, including clicking on job vacancies. This is considered a **proxy for job search behavior and motivation**.
- Research based on all job seekers newly registered in 2016 (and followed until Jan 2018).
- N = 288,765.
- **Linked to 4 data sets** on programs and services.

Three further data collection exercises are currently being examined.

a. Adding more behavioral information to the model using a short online questionnaire to capture job seekers' motivation and self-reliance
b. Developing a tool that visualizes barriers to employment
c. Developing a tool that suggests specific (online) programs to a job seeker and caseworkers based on the job seeker's profile and the experiences of other job seekers with a similar profile. Just as job seekers can be placed on a continuum, it is also possible to place the measures of the Flemish labor market policy on a continuum. The function of the target group's distance from the labor market ranks according to the distance from the labor market of the effectively reached group allows to compare the reached target group with the intended target group.

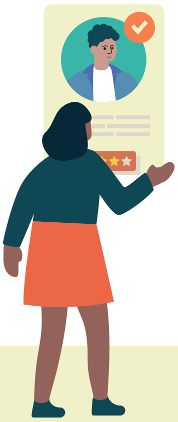
4. Model

Profiling scores from the "Vlaamse Dienst voor Arbeidsbemiddeling en Beroepsopleiding" (VDAB) models in Dutch, which translates to "Flemish Service for Employment and Vocational Training 'Next Steps'

5. Use

The model is meant to **assist caseworkers in decision-making, not to impose it**.

Priority for screening job seekers will be given to high-risk job seekers as predicted by the profiling model.



Examples of other quantitative methods: Denmark

1. Background

The Danish Agency for Labor Market and Recruitment (STAR) developed a profiling model (profilafklaringsværktøjet) using **machine learning techniques, more specifically decision tree classification**.

The results of statistical models were one of the main inputs into the profiling process used in Denmark between 2004 and 2008 (Loxha and Morgandi 2014).

The Job Barometer was integrated as part of an overall **Employability Profiling Process**, which consisted of three phases:

1. Preparation phase where the caseworker would receive initial CV from the job seeker, pull the information from the client's customer public assistance account, and run a **statistical test through the Job Barometer**
2. Interview phase based on the Dialogue Manual covering the critical areas such as job seekers' perspectives of their own job prospects for reemployment, vocational qualifications and experience, personal and social skills, financial situation, and health condition
3. Assessment phase to make the overall assessment that would lead to five employability potential categories.

This profiling system was abandoned and **replaced by a caseworker-based (and data-assisted) procedure, which did not incorporate statistical profiling.** The prime reasons were that

a. The system was heavy on documentation,
b. The system was too complicated for caseworkers in the sense that it differed by benefit category and region,
c. There were objections that the system would lead to 'discrimination' since the same job seeker might be assigned a different profiling category based on the region of residence, and
d. The difference between the five different 'profiles' was not meaningful enough.

The statistical profiling model (Rosholm, Svarer, and Hammer 2004) **was solely based on administrative data** that were directly assessable through the Danish Labor Market Authority.

Researchers used **duration analysis**.

A separate model was estimated for 120 subpopulations defined by benefit eligibility (UI versus welfare benefits), gender, age, and region. Thus, a separate Cox proportional hazard model with arbitrary baseline hazards was estimated for 120 subpopulations. Following estimation, the probability of survival in unemployment beyond 26 weeks (after the initial interview) was estimated.

The model performed relatively well, as 66 percent of observations were correctly classified (note that a separate cutoff point was chosen for each subpopulation), despite the fact that some important variables (such as education) were not included in the register data.[53] Researchers attribute this to having access to detailed labor market history variables (number of unemployment spells in each of the two previous years and the proportion of the year's spend on different types of benefits) and to subsampling.

New approaches in Denmark use a decision tree.

2. Outcome indicator

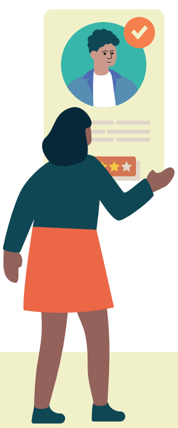
Becoming long-term (>26 weeks) unemployed

The decision tree identifies nine paths that predict the likelihood of becoming long-term unemployed (Desiere, Langenbucher, and Struyven 2019).

3. Data source

The model combines data from

- Administrative records and
- An online survey that gathers behavioral information.



53 The success rate was 82 percent for actual short-term unemployed, while it was only 55 percent for actual long-term unemployed.

In collaboration with the University of Copenhagen, a new survey instrument is currently being developed that aims to capture structural personality traits such as time and risk preferences.

4. Use

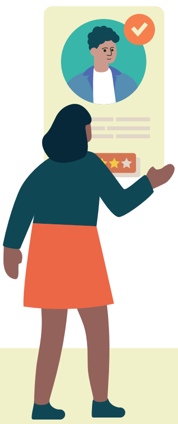
- The system is voluntary to use for job seekers, but if they use it, they get full access to the model's results.
- The system does not automatically refer job seekers to ALMPs but only supports caseworkers who keep full discretionary responsibility.

Example for other quantitative methods: latent class analysis for segmenting job seekers

A new approach to identify risk groups or potential target groups in relation to their employment barriers for designing and targeting employment services and ALMPs has been proposed by the OECD and World Bank (Fernandez et al. 2016; Sundaram et al. 2014).[54] They use a latent class analysis, a statistical method for segmentation. This method has previously been used in fields like medical profession and behavioral sciences, to identify hidden and underlying subgroups of individuals that share observable characteristics within a population of interest.

A major difference between the analysis made by OECD and World Bank and the statistical profiling tools used at PES is the population of interest. While the statistical profiling tools have been applied to registered unemployment and/or welfare benefit recipients, the approach chosen by Fernandez et al. (2016) focuses on working-age individuals who are entirely out of work (either actively searching for a job or inactive) or whose labor market attachment is 'weak'. 'Weak' labor market attachment can include individuals with unstable jobs working only sporadically, those working persistently with restricted working hours, and those with very low earnings (due to, for example, being partially unpaid or working informally) (Fernandez et al. 2016; OECD 2017). Sundaram et al. (2014) look at different categories of people out of work.

These studies begin with an analysis of characteristics of employment barriers and key characteristics for working-age population with different levels of labor market attachment on the basis of household survey data (European Union Statistics of Income and Living Conditions [EU-SILC] for European countries and Household, Income and Labour Dynamics in Australia [HILDA] for Australia). Duell et al. (2016) also analyze employment barriers of nonworking population in working age with different levels of labor market attachment on the basis of two surveys, the Labor Force Survey (LFS) data, and the European household survey data (EU-SILC) for 28 European countries; however, this latter study does not, as a second step, apply statistical cluster methods to classify groups at risks.



Both data sets, household data survey and LFS, have their advantages and disadvantages. Household data contain information on variables like household income, receipt of welfare benefit, and household composition. Employment characteristics as well as

54 Estonia, Ireland, Italy, Lithuania, Portugal, and Spain (OECD) as well as Bulgaria, Croatia, Greece, Hungary, Poland, and Romania (World Bank).

out-of-work status, job availability, and job search conform to the usual International Labour Organization (ILO) standard of unemployment when using LFS data, inactivity, and employment definitions are used. Also, both data sets contain different variables relevant for analyzing employment barriers. While it is possible to use both sources to get an understanding of characteristics of predefined groups, they cannot be mixed for defining segments (classes or groups) with the latent class analysis (which has to be understood as a bottom-up approach to identify groups that share the same characteristics).

Implementation of the latent class analysis

As explained in Fernandez et al. (2016), the statistical clustering method separates the highly heterogeneous population of individuals with no or weak attachment into groups (clusters) that are homogeneous with respect to the types of employment barriers that they face. The latent class algorithm identifies the most frequent response patterns in the data and then estimates probabilities of class membership based on the similarity of the response patterns with the frequent ones. These probabilities are model based and are therefore called posterior probabilities (Fernandez et al. 2016).

Fernandez et al. (2016) used four steps to identify the optimal number of latent classes. Models with different number of classes are estimated sequentially, and the optimal model is chosen based on a series of statistical criteria. In the first step, a baseline model is defined, which is then estimated for an increasing number of classes. The baseline model only uses employment barrier indicators, which were previously identified. In the second step, goodness-of-fit statistics and classification-error statistics are computed. In the third step, potential misspecification issues are analyzed. In the fourth step, the model is further refined by including the so-called active covariate. The main role of active covariates is to describe the latent classes; they can also interfere with the actual definition of the latent groups driven by the employment barrier indicators. Each group can then be further analyzed along with using the inactive covariates that do not affect the fit of the model but allow to get a more detailed profile (Sundaram et al. 2014).

Sundaram et al. (2014) used the following variables (with some variation across the six European countries analyzed):

1. Two or three sets of variables were selected as input variables for the clustering exercise to show the extent of labor market distance and to capture some of the main employment barriers at the supply side and the demand side: labor supply conditions (household-level incentives to work and physical ability to work), work experience, and human capital.
2. Active covariates include demographic variables that are normally used to disaggregate labor market outcomes; age, gender, highest educational level achieved, and degree of urbanization are typically active covariates.
3. Inactive covariates were used that indicate household welfare conditions, such as income quintile, labor market status of a partner, and household ownership, and variables for household demographics such as household size and whether the household has children under six years old, among others. The resulting groups are then labeled according to their main distinguishing characteristics.

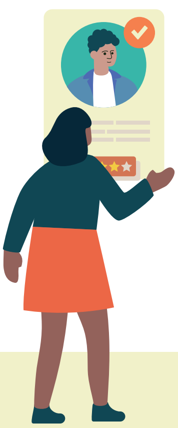
Fernandez et al. (2016) used, in the case of Spain for example, the following variables:

1. Input indicators such as low skills, care responsibilities, health limitations, no recent work experience, having never worked in the past, low overall work experience, high non-labor income, high earning replacements (benefits), and low job opportunities (limited hirings).
2. Variables used to describe the latent class included living with children, number of adults, women, age group, average age, average years of paid work experience, highest educational attainment, average years of schooling, migrant, severe health limitations, main activity status (employed, short-term unemployed, long-term unemployed, retired, unfit to work, domestic tasks, other inactive, and actively looking and available for work) residence in rural area, at risk of poverty, income distribution (5 quintiles), and average disposable income.
3. Further, the number of employment barriers for each class are shown.

OECD (2017) used roughly the same input indicators for Australia (with thresholds adapted to the Australian case) as for the above example of Spain. The variables used for describing the detailed characteristics of groups of individuals with no or weak labor market attachment include the following variables: three age groups, average age, four groups of highest educational level as well as average number of years of education, four categories of health status, migrant, indigenous, household type (single, couple without children, couple with children, lone parents, two or more adults, multifamily households, children older or younger than 15 years, age of youngest child, hours of non-parental childcare, households with other working household members), had any work activity (three different lengths over the past three years), type of labor market attachment during the reference period (unstable jobs < three months, restricted working hours < 10 hours, employees with zero or near zero earnings), main activity during reference period (working, unemployed, inactive), main reasons for restricting working hours (illness or disabilities, care or family responsibilities, education, cannot find full-time work, prefer working part-time, prefer current job, other reasons), main activity at time of household survey HILDA interview, type of employment (employee, self-employed, family business), share of employees with 'casual' jobs, length of unemployment spell (months), years of paid work experience (average), equivalized disposable household income, position in the income distribution, material deprivation and At Risk of Poverty or Social Exclusion (AROPE) indicator to measure poverty, average amounts of benefits (sickness and disability, unemployment benefits, social assistance, and family-related benefits), residence in rural area, and area of residence (major cities, remote areas, and so on).

The latent class analysis was carried out on the basis of household survey data (EU-SILC in the case of European countries and HILDA in the case of Australia; see Fernandez et al. 2016; OECD 2017; Sundaram et al. 2014).

Using the results of the latent class analysis



The clustering results are used for an assessment and a discussion on labor market and social policies in place and their potential to help overcoming these barriers. In contrast to the above-described statistical profiling tools designed for individual long-term unemployment risk assessment by the PES counsellor (see sections 2 to 4 above), these clustering approaches are designed as a 'policymaking-oriented profiling' tool. They depart from the analysis of employment barriers and the combination of different employment barriers which adds to the complexity of the individual case (Fernandez et al. 2016; and a series of country policy papers that will be published in 2017).

Summary and lessons learned

1. Background

- **Older experiences:** Mainly in the context of outsourcing of employment services.
- Increasingly, the objective is to allocate job seekers to a relevant ALMP or pathway.
- Research on labor market barriers/reasons for disadvantages.
- **Objective:** Increase efficiency in service provision.
- **Newer approaches:** To guide counsellors' decision.

2. Development of method

- Takes years.
- Updates of methodology over time in some cases.
- Updates of models, and in particular of variables and correlations, as explanatory power of context factors changes.
- The choice of variables depends on the labor market context; therefore, adaptations need to be made over time; variables for one country may not be relevant for another country.
- Choice of variables depends on activation models and policy targets.
- **Improvements:** For example, controlling endogeneity (because those with higher employment barriers will be referred to more intense employment services and follow-up or ALMP).

3. Model

- Mostly logit models are used. In some cases, probit models are also used (more recently).
- In a few cases, other methodologies were used: duration models and using machine learning as an additional tool to the logit/probit model.
- If not scoring for an employability level or distance to the labor market level but segmenting into groups according to (multiple) employment barriers, the latent class analysis or other statistical segmentation tools may be useful. These are useful at the macro level: for policy design and for assigning job seekers to well-defined 'integration or intervention pathways'. It is more complex and for advanced use.

4. Variables

- **Outcome indicators:** Mainly long-term unemployment.
- Mainly to be processed in a binary way; few cases with a continuum.

5. Databases

There are two types of databases that, in general terms, use statistical profiling models (Felgueroso et al. 2018):

1. Administrative databases, which come from the registers of the social security systems and/or of applicants for unemployment benefits. They usually include employment information and are generally accurate and easily accessible to public employment offices. This requires that administrative data are complete, contain detailed information, and are cleaned and structured. Administrative data are collected during registration as a job seeker or benefit claims.
2. Databases created specifically for the design and use of statistical profiling models by employment offices. These are data from questionnaires, developed specifically to build the model, that enrich it and improve its predictability.

There are several options for implementing these questionnaires, including face-to-face interviews with employment consultants, telephone interviews, or online surveys. Although the face-to-face survey could be considered the most accurate, it has been found that job applicants may be more willing to answer some sensitive questions over the phone (Lipp 2005). On the other hand, as job applicants become more familiar with the use of the internet, online questionnaires are more widely used.

- In some countries, a combination of administrative data and survey-based data is used.
- In some countries, the use of questionnaires in addition to administrative data serves to compile characteristics such as behavior, motivation, and socio-emotional competences.
- Using LFS or household survey data is a way to identify risk of being unemployed or remaining unemployed (also confirmed by recent experiences made with designing profiling tools in Chile and Peru/BID, Duell et al. 2016). This may serve to improve collection of administrative data and to design additional questionnaires..

6. Use

- Mostly segmentation of job seekers into 3–6 segments
- Sometimes assigning already 'integration pathways'
- Automated referral to more intensive services (in-house or external; then counsellor decides what exactly)
- Automated referral to service providers (external, according to streams)
- Assistance for decision-making of counsellor
- Both: first automated assignment to stream and then in-depth screening of job seekers (a modularized approach)
- E-services and blended services (for example, in the Netherlands (NL)
- Performance management (in-house, of external providers)
- Payment structure for external providers, in case employment services are outsourced.

In addition, PES counsellors use other tools for screening: skills'/competences' profiling, and so on; advice of a health specialist; advice of a psychologist; case workers' own experience; and in-depth interviews following qualitative guidelines.

References

Ajzen. I. 1991.*The Theory of Planned Behavior*, Organizational Behavior and Human Decision Processes 50, 179-211.

Brouwer, S., J. M. H. Schellekens, H. Havinga, et al. 2011. *Voorspellers van Werkhervatting: Een onderzoek onder werklozen in Noord-Holland.* Groningen/Amsterdam: Rijksuniversiteit Groningen UMCG/UWV Kenniscentrum.

Commonwealth of Australia. 1998. *Job Seeker Classification Instrument*. Canberra: Commonwealth Department of Employment, Workplace Relations and Small Business.

Csillag, M. , Duell, N. 2017. *Lessons from statistical profiling methods*, Consultation on the design of diagnostic mechanism about the users of the labor intermediation system in Chile, IDB, not published.

Desiere, S., K. Langenbucher, and L. Struyven. 2019. "Statistical Profiling in Public Employment Services: An International Comparison." *OECD Social, Employment and Migration Working Papers,* No. 224, OECD Publishing, Paris. https://doi.org/10.1787/b5e5f16e-en.

Dickinson, Decker, and Kreutzer. 2002. "Evaluation of WPRS Systems," in Randall W. Eberts, Christopher J. O'Leary, Stephen A. Wandner, eds. Targeting Employment Services. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Duell, N., Thurau, L., Vetter, T. 2016. Long-term unemployment in the EU. Trends and Policies. Bertelsmann Stiftung (Ed), Gütersloh.

Dusseldorp, E., H. Hofstetter, and C. Sonke. 2018. *Landelijke doorontwikkeling van de Werkverkenner.* TNO, NOA & VUmc, Leiden & Amsterdam. https://publications.tno.nl/publication/34626555/VsGrTh/TNO-2018-R10279bijlage.pdf

Felgueroso, F., García-Pérez, J. I, Jímenez-Martin, S.. 2018. *Perfilado estadístico: un método para diseñar políticas activas de empleo*, Fundación Ramón Areces, fedea

Fernandez et al. 2016, *"Faces of joblessness. Understanding employment barriers to inform policy", OECD Social, Employment and Migration Working Papers,* No. 192, OECD Publishing, Paris, https://doi.org/10.1787/5jlwvz47xptj-en.

Hawkins et al. 1996. Evaluation of Worker Profiling and Reemployment Systems: Interim Report. UI Occasional Paper 96-1. Washington, DC: U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Service.

Heijnen, K. and M. Dekenga (2017), "Experience from the Netherlands", Learning Exchange on "Service differentiation – tailoring online services for the unemployed", Mutual Learning Programme: DG Employment, Social Affairs and Inclusion, Manchester, United Kingdom,4-5 December. https://ec.europa.eu/social/main.jsp?langId=en&catId=89&newsId=2854&furtherNews=yes

Holl, J. Kernbeiß, Wagner-Pinter, M. 2018. *Das AMS-Arbeitsmarktchancen-Modell, Arbeitsmarktservice Österreich*, Wien.

Lipp, R. 2005. "Job Seeker Profiling: The Australian Experience." Keynote EU-Profiling

seminar. Nuremberg, January 12–14, 2005.

Loxha, Artan, and Matteo Morgandi. 2014. "Profiling the Unemployed: A Review of OECD Experiences and Implications for Emerging Economies." SP 1424, World Bank Group, Washington, DC.McGuiness, Kelly, and Walsh. 2014. *Predicting the Probability of Long-Term Unemployment in Ireland Using Administrative Data,* ESRI Survey and Statistical Report Series Number 51.

O'Connell, P. J., S. McGuinness, E. Kelly, and J Walsh. 2009. "National Profiling of the Unemployed in Ireland." ESRI Research Series Number 10, Economic and Social Research Institute, Dublin.

OECD (Organisation for Economic Co-operation and Development). 2012. Activating Jobseekers: *How Australia Does It*. Paris: OECD Publishing. https://doi.org/10.1787/9789264185920-en.

OECD. 2014. *Connecting People with Jobs: Activation Policies in the United Kingdom*, Connecting People with Jobs, OECD Publishing, Paris, https://doi.org/10.1787/9789264217188-en.

OECD. 2017. *Connecting People with Jobs: Key Issues for Raising Labour Market Participation in Australia,* OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264269637-en

Rosholm, M. Svarer, M. and Hammer, B.. 2004. *A Danish Profiling System*. IZA Discussion Papers 1418, Institute for the Study of Labor (IZA)

Sullivan, W.F, Coffey, L. Kolovich, L., McGlew, C.H., Sanford, D. y Sullivan, R. 2007: *Worker Profiling and reemployment services evaluation of state workers profiling models,* final report, Coffey Communications, LLC Bethes-da, Maryland

Sundaram, R., U. Hoerning, N. De Andrade Falcao, N. Millan, C. Tokman and M. Zini 2014. *Portraits of Labour Market Exclusion*, World Bank, Washington, D.C..

Vroom, V. H. Workandmotivation. NewYork: Wiley,1964. 331p. Carnegie Institute of Technology, Pittsburgh. PA

Wanberg et al. 2002. *Predictive Validity of a Multidisciplinary Model of Reemployment Success*, Journal of Applied psychology, Vol. 87, issue 6

Wijnhoven, Martijn, and Harriët Havinga. 2014. "The Work Profiler: A Digital Instrument for Selection and Diagnosis of the Unemployed." *Local Economy* 29 (6–7). https://doi.org/10.1177/0269094214545045.

Websites

Australian Government, JSCI. https://www.dewr.gov.au/job-seeker-assessment-framework/resources/online-job-seeker-classification-instrument-trial-evaluation-report#:~:text=The%20Job%20Seeker%20Classification%20Instrument,to%20a%20series%20of%20questions.