Statistical profiling

Lessons from OECD countries

Nicola Duell and Gael Fostier de Moraes

April 2020

THE WORLD BANK

Table of contents

Introduction

1. Challenges adressed

In OECD countries:

- Reduce long-term unemployment
- Reduce spending on unemployment benefits and minimum income schemes
- Efficient use of resources
- Avoid conflict between equity and efficiency (e.g. creaming)

2. Objectives of the statistical profiling tool

- For the Public Employment Services
  - → **Segment** jobseekers into groups that do not need specific job-search support and groups who need it To better **target employment policies,** in-depth counselling and ALMP
  - → Increasing **efficiency**
  - → **Increasing equity** of services provided PES and subcontracted private agencies

3. What is a statistical profiling tool?

- It helps to segment jobseekers by their **difficulty / likelihood** to find a job, their **distance** to the labour market, their **degree of disadvantage**, their likelihood to exhaust unemployment benefits (different notions are used in the different countries)
- It aims to **predict** the probability a jobseeker becomes long-term unemployed
- It helps to assigning PES clients at the initial stage of their **journey** into broad groups:

  → **Activation streams:** e.g. self-help or intensive follow-up by counsellors, access to specific programmes; coordination with social service providers

3. What is a statistical profiling tool? (cont.)

- It helps with **diagnostic**; understanding which reasons lead to unemployment
- Usually IT-based

A multitude of underlying concepts for classifying individuals

- Segmentation by the degree of "**employability**"
- Segmentation by the "**distance from the labour market**"
- Segmentation by degrees of "**disadvantage**" (see e.g. Ireland's "Disadvantage to the Labour Market Model")

→ Mostly classification into 3 to 4 groups: e.g.

-Employment-ready no further assistance needed

-Needs support

-Needs intensive support

Advantage of statistical profiling

- Takes into account that there is considerable **heterogeneity** across individuals belonging to an at-risk group in how disadvantaged they are

- It allows for **early intervention** for those in need

- **Avoid 'cream-skimming',** which caseworkers might have incentives to do, objective resource allocation

- It may **save time of staff** which can be used for in-depth counselling for those who need it most

- It provides **assistance to the decision-making** of PES counsellors

Experience in selected OECD countries

- **US,** Australia: since the 1990s


- Ireland, the Netherlands, Denmark since the years 2000s



- In the years 2010: Italy, Austria, Spain, Belgium




Part I

Overview of key elements of statistical profiling models used in 12 OECD countries

4. Defining output variables

Driven by **employment policy objectives** (e.g. rapid vs sustainable labour market integration; reduce benefit dependency), mostly:

- **Length of unemployment** period (e.g. leaving unemployment in general; becoming long-term unemployed > 6 months, **mostly > 12 months**)
- **Exit to employment** within 12 months (e.g. Ireland)
- Outflows of unemployment benefit registry, or more generally being **off benefit**
- Outcome variables are mainly to be processed in a **binary** way; few cases with a continuum

Examples

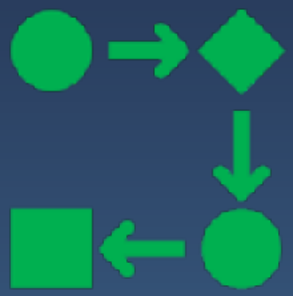
	Outcome (probability of)	Data derived from
Australia	Long-term unemployed (12 months)	Personal interview; online trial ongoing
Austria	Labour market integration probability[b]	Administrative data
Belgium (Flanders)	Long-term (>6 months) unemployed	Administrative data; "click" data
Denmark	Long-term (>26 weeks) unemployed	Online questionnaire; Administrative data
Ireland	Probability of exit to employment within 12 months	Questionnaire as part of benefit claim process, administrative data
Italy	Long-term unemployed (12 months)	Administrative data
Latvia	Long term unemployed (12 month)	Personal (individual) interview, questionnaire, administrative data
Netherlands	Long-term unemployed (12 months)	Online questionnaire
New Zealand	Lifetime income support costs (LET), change in lifetime income support and staff costs from receiving a case management service (SEM)	SEM/LET are based on administrative data
Sweden	LTU (6 months)	Administrative data
US	Exhausting the 26-week entitlement to UI benefits	Online questionnaire; Administrative data
Spain (model Fedea)	**Probability of leaving unemployment**	**Administrative data (+questionnaire)**

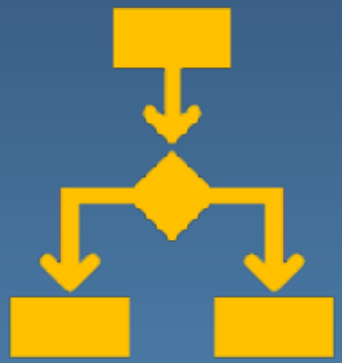
b. Outcome variable in Austria: Labor market integration probability measured in short and long term (short term: 3 months of unsubsidized employment within 7 months; long term: 6 months of unsubsidized employment within 24 months).

Sources: Desiere et al., 2019; and 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)

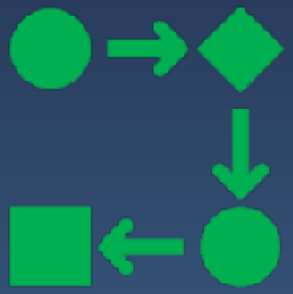
5. Defining explanatory variables

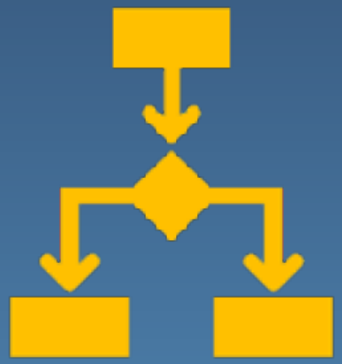
- **Socio-demographic factors** (age, gender,)
- **Labor market history (in most countries)**: prior periods of employment, unemployment, inactivity, occupations held, sectors worked in
- **Formal qualification**
- Mainly "**hard**" skills; few examples of "**soft**" skills
- **Motivation**, jobseeker behavior (most countries)
- **In several countries** (e.g. Austria): **Household factors**: e.g. caring obligations, marital status (e.g. Ireland)
- **In several countries** (strong focus in NL): **Health conditions**
- **Less often other individual factors** (such as addictions, ex-offenders, etc.)





5. Defining explanatory variables (cont.)





- Large variety in the number of variables or “items” considered (up to 500 items in initial phases of research, reduced to mostly 20 to 100) for calibration of the model
- Number of variables retained in the questionnaires are smaller, vary between 15 to 50 variables
- Large variations with regard to taking **local labor market conditions** into account

Examples

	Type of data sources (in addition to socioeconomic info)				
	Job readiness			Motivation	Opportunities
	Labour market history[a]	Hard skills	Soft skills	Jobseekers' behaviour	Regional labour market info
Australia	Yes	Yes	No	No	No
Austria	Yes	Yes	No	No	Yes
Belgium (Flanders)	Yes	Yes	No	Yes	No
Denmark	Yes	Yes	Yes	Yes	No
Ireland	Yes	Yes	No	Yes	Yes
Italy	Yes	Yes	No	No	Yes
Latvia	Yes	Yes	Yes	Yes	Yes
Netherlands		Yes	Yes	Yes	
New Zealand	Yes	Yes	No	No	No
Sweden	Yes	Yes	No	No	Yes
US		Yes	No	No	Yes

Source: Desiere et al., 2019

Explanatory variables collected through administrative data

Include for example:

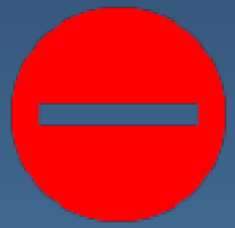
- Age
- Gender
- Family status (otherwise asked through questionnaire)
- Educational level (otherwise asked through questionnaire)
- Eligibility to unemployment benefit / social assistance
- Degree of disability / eligibility to disability benefits
- Duration of unemployment
- Job-search behaviour (e.g. “clicks” on job-search portal, see e.g. Belgium/Flanders)

Examples of explanatory variables that are collected through questionnaires

- Literacy / numeracy skills
- Languages skills, migration background
- Location/transportation
- Subjective view on health, employability, employment barriers
- Subjective views on chances to return to the labour market
- Motivation
- Job search behaviour
- Reservation wages
- Desired job/occupation(aspirations (e.g. Catalonia in Spain)
- Time and risk preferences (being developed in Denmark)
- Socio-emotional skills, personality traits difficult to grasp, mostly excluded

Taking regional/local dimension into account

- **Pro:** is more accurate, as the regional labour market context on the supply side and demand side varies greatly
- **Cons:** Adds to the complexity of the model
- Depends on data availability and data quality at local level



6. Data collection

- **Administrative data**
  - Limits: Administrative data are often **incomplete** for the constructing a profiling tool:
  - E.g. lack info on life circumstances, motivation, some socio-demographic characteristics, level and type of skills and type of work experience, …
- **Questionnaire**
  - Collect new data when jobseeker register, which are relevant for segmentation.
  - Longitudinal (follow-up employment status for up to 18 months)

  Specialist assessment
  - In few case include results of **interviews** with case manager
  - Assessments by specialists (psychologists, occupational doctors, …)
- **Combination** of different sources

Who is involved in development of questionnaires?

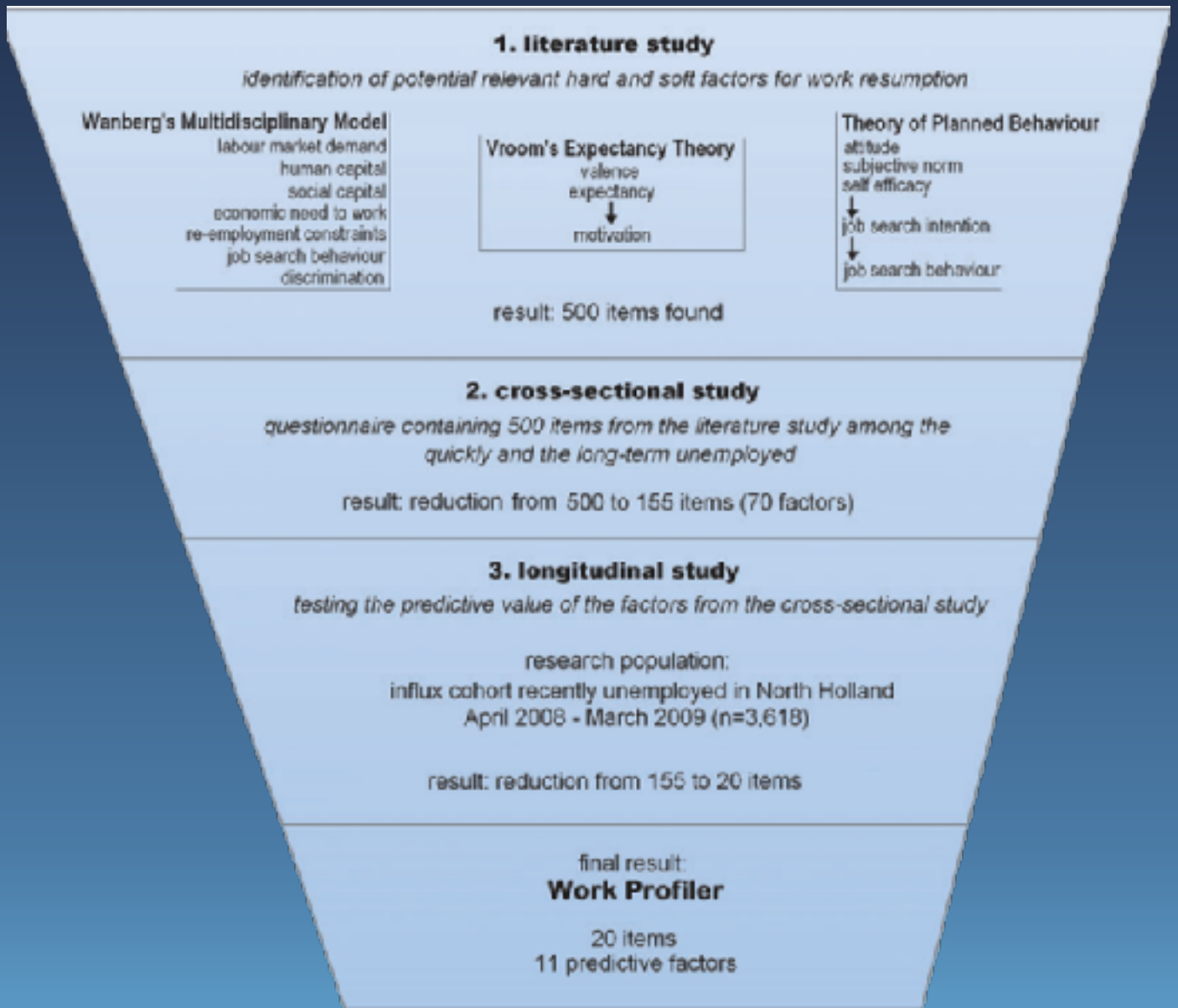
- Initiated by Ministry of Labour, or central PES
- External consultants / researchers to develop questionnaires and carry out the statistical modelling (mainly logit, probit regressions)
- Expert consultation
- In some cases social partners have been involved
- Involving PES counsellor is an advantage

  - To build on their experience (e.g. for developing the questionnaire)
  - Increase acceptance of the tool
  - Implementing “modular” – diagnostics departing from a row statistical segmentation

Developing the questionnaire

- **Literature review** on employment barriers
- **Quantitative methods** on the basis of: the labor force survey, household surveys, to identify employment barriers
- **Qualitative methods**: e.g. interviews with counsellors, experts
- **Pilots**: testing responses to pilot questionnaires, pilots may analyze responses to questionnaire depending on whether or not jobseeker received help
- **Issues, e.g. legal issues, quality of responses**: A number of factors are affected by privacy and discrimination legislation; for example, ex-offenders and disability

Example: Netherlands
(with a focus on workability as explanatory variable)



Implementation of the questionnaire

- Mostly it is **compulsory** to fill in the questionnaire during registration at PES
- Out of 11 country example collected:
  - **Compulsory for jobseekers** (9 countries); voluntary for all only in Denmark; voluntary if online (Latvia)
  - **Compulsory for counsellors** (9 countries); voluntary in Denmark and Sweden
- Via telephone, face-to-face, online (increasingly)

(Desiere et al. 2019), countries include: Australia, Austria, Belgium (Flanders), Denmark, Ireland, Italy, Latvia, Netherlands, New Zealand, Sweden, US

7. Statistical methods / modeling

Estimating individual probabilities to become long-term unemployed

- **Mostly used**:
  - **Logistic regression** (e.g. Australia, Austria, Italy, Netherlands, Sweden, US)
  - **Probit models**; estimating separate probit models by gender (e.g. Ireland)
- **Linear multiple regressions** (5 States in the US, multivariate NL)
- **Duration models** (based on administrative data)
- **Factor analysis** (e.g. Spain, Latvia)
- **Random forest model** (new Zealand, Belgium/Flanders)
- **Big data** (Denmark, Belgium/Flanders)

7. Statistical methods / modeling (cont.)

Estimating individual probabilities to become long-term unemployed

- In a number of countries, the quality of predictability of the models is assessed to be quite good (varies between 60% and 86% in countries where this was tested, see Desiere et al 2019)

- Recent improvements: e.g. controlling for endogeneity (because those with higher employment barriers will be referred to more intense employment services and follow-up or ALMP)



Weighting of explanatory variables

Example: Australia

(a) Through statistical model

- Weight the answers to predict level of disadvantage, translate weights into point system, adding up to a score
- The statistical model estimates the relative weights of various socio-economic factors and personal characteristics, as indicators of labor market disadvantage. These are then translated into points
- Factors tested in the survey have formed the basis of the Job Seeker Classification Instrument JSCI
- The number of scores show to which stream the person should be allocated to

Weighting of explanatory variables

(b) Other ways to attributing points:

Not all variables are captured by a longitudinal survey. Therefore, need to develop a point system, examples (Australia):

- **Geographic location** (regional grouping): pointing system based on unemployment rates and employment growth rates for the region. (see also example of Austria on geographical cluster)
- **Country of birth:** Pointing system-based unemployment rates in Australia (Census)
- **Age:** based on the survey data, of labor market perception of age as a barrier for employment and on LTU rates among age tranches. Points were awarded to young people to reflect the lack of experience
- **Language and literacy**: self-assessment
- **Disability/Medical condition:** not based on survey but rather on professional expertise
- **Disadvantage resulting from personal factors requiring professional or specialist judgement (secondary classification):** not based on survey

8. Who uses the tool and how?

Segmentation

- **PES staff: Assigning clients** at the initial stage of their journey to broad **groups**, automatically (without counsellor involvement)
- **Based on scoring, mostly segmentation of jobseekers into 3-6 segments**
- **Segmentation to assign to different streams of service provision** (e.g. Australia, US)
- **Few examples: assigning already "integration pathways"**
- Automated referral **to more intensive services (in house or out-house;** then counsellor decides kind of intervention**)**

Governance

- **Performance management** (in-house, of external providers)
- **Payment structure for external providers**, in case employment services are outsourced (e.g. Australia)

7. Who uses the tool and how? (cont.)

Use by counsellor for service provision

- Used by counsellor as a tool helping with diagnostic
- In addition, PES counsellor use other tools for screening: skills/competences profiling, advice of a health specialists, advice of a psychologist; case workers' experience and in-depth interviews following qualitative guidelines
- Assistance for decision-making of counsellor, different services and ALMP are then offered to these groups after meeting with a counsellor
- Both: first automated assignment to stream, then in-depth screening of jobseekers (kind of multi-step approach)

Example Ireland

- Compulsory questionnaire: 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**

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

Referral to active labour market programmes

Balancing use of a statistical profiling tool and automated referral and **discretion** of counsellor:

- Use of counselling and ALMP to reduce the risk of unemployment remains a **black box** (example: contracting out in Australia)
- At the discretion of the case worker
- Discussion about pros and cons of caseworker autonomy
- Different political priorities at local/regional level, mix of central and local/regional ALMPs
- An evaluation in Switzerland showed that use of statistical profiling tool: led to adjustments in the counselling strategies of the counsellor
- Belgium (Flanders): proposal of intervention pathways, as a suggestion to jobseeker

Use by the jobseeker

- **Potential use for empowering jobseekers** (automatic e-counselling to jobseekers)
  - Helps to define own job-search and upskilling strategies
- **Limits**
  - Needs to be well designed, as it may otherwise be demotivating
  - Complexity of multiple employment barriers as well a s combination of barriers and potentials might be not well grasped
  - How to avoid exclusion?
  - Risk of people not being reached out?
  - For the harder-to-place it may be relevant to develop an individual action plan, allow for failures, re-adjust action plan: this can hardly be automatized
  - Experiences with blended counselling: still interaction with the counsellor

Use by government / ministry of labour

Policy development

- To get a more precise understanding of the characteristics of unemployment and non-employment and its development
- To develop and target active labour market programs.

8. Who uses the tool and how? (cont.)

- Several countries have abandoned their statistical profiling model (e.g. France), forth and back in Denmark
- Problems: case workers did not use it, because:
  - PES counsellors might see their role undermined
  - Experience of case workers valued as being more correct. This depends on the experience and qualification of staff as well as caseload (e.g. many years of experience, low staff turn-over, continuous training, and a low unemployed/caseworker ratio increase effectiveness of caseworkers)
  - The tool was too complex
  - Added-value of the tool was not clear

9. Which lessons learned?

- **Political objectives** / activation strategies: outcome indicator
- **Tension:**
  - **"Objectivity" of the tool** (e.g. if based on self-assessments of jobseekers; quality of the data?; choice of statistic/econometric model?) and
  - **"Experience"/subjectivity of case worker**
- Tension: between
  - **Accuracy** *vs*
  - **Low complexity** of the tool (short or long questionnaire, taking into account local labour market context factors)

9. Which lessons learned? (cont.)

Role of counsellor:

- **Dialogue**: important to involve stakeholders and counsellors to develop the tool
- Counsellors need to be trained to make use of a statistical profiling tool
- Changed job profile: time gained allows to focus on in-depth diagnostic and developing individual pathways

9. Which lessons learned? (cont.)

Need to balance between automated self-help e-services and individualized face-to-face counselling

Multistep approaches:

- Combining statistical profiling for segmentation and IT-based deeper diagnosis of (i) labor market barriers; (ii) potentials
- New approaches: Potentials for IT-based assistance to job counsellors to decide on activation strategy and referral to ALMPs (including to intense counselling)
- In addition: PES counsellor use other tools for screening: skills/ competences profiling, advice of a health specialists, advice of a psychologist; case workers' won experience and in-depth interviews following qualitative guidelines

9. Which lessons learned (cont.)?

- **Development takes time**
- Development of questionnaires, implementing them in order to obtain sufficient results to conduct the statistical model takes time.
- Try and error in the implementation, needs to adapt the tool over time

Part II
Examples of Australia, Ireland, Netherlands, US

Australia 

Job Seeker Classification Instrument (JSCI)

- **Outcome variable:** Long-term unemployed (12 months)
- **Model:** Logistic regression
- **Main use:** JSCI scores are used to determine a variety of actions to be taken by the employment services in Australia. It defines whether the jobseeker:
  - Needs further evaluation
  - If they are eligible to receive intensive assistance
  - Allocate jobseekers into the different existing streams of assistance

Australia 

- **Methodological issues:**
  - JSCI was based on 8 principles
    - Classification, minimum number of factors, accuracy, reliance on survey factors, net impact, transparency, consistency and acceptance
  - The instrument was developed on three steps
    - **Formal research:** an extensive survey was conducted on jobseekers, and administrative data compiled. Preliminary factors were drawn from analysis then tested for accuracy
    - **Expert judgement:** a Classification Working Group was created to make recommendations regarding existing factors and new ones to be added and other recommendations regarding the JSCI instrument
    - **Wider consultations:** main stakeholders such as companies, employment service companies and other organizations were consulted to give feedback and further considerations

Australia 

- **Explanatory variables:**

  - Age
  - Gender
  - Recency of work experience (*)
  - Job seeker history
  - Educational attainment (*)
  - Vocational qualifications
  - English proficiency (*)
  - Country of birth
  - Indigenous status

  *JSA Triggers (JSCI Supplementary Assessment)

  - Indigenous location
  - Geographic location
  - Proximity to labor market
  - Access to transport
  - Phone contactability
  - Disabilities or medical conditions (*)
  - Stability of residence (*)
  - Living circumstances (*)
  - Criminal convictions (Disclosed ex-offender) (*)
  - Other personal factors (*)

Netherlands

Work Profiler

- **Outcome variable:** Long-term unemployed (12 months)
- **Model:** Logistic regression
- **Main use:** the tool generates two outcomes
  - The first is a percentage of likelihood that the job-seeker will reenter the market before 12 months. This is used to determine is the user will be directed to face-to-face or to online based services
  - The second is a quick diagnosis, indicating strong points and weaknesses from the user's profile. This is used by the UWV (NL PES) to offer tailored services to the job-seeker

Netherlands

- **Methodological issues:**
  - An elaborate study took place before the creation of the Work Profiler. The study 'Predictors of Work Resumption' was carried out by the UWV Centre for Knowledge (Kenniscentrum UWV) and the School of Medical Sciences of the University Medical Centre Groningen (UMCG) between 2006 and 2011
  - **The study was conducted in 3 phases:**
    - **Literature study** - Wanberg Multidisciplinary model, Vroom's Expectancy Theory and Theory of Planned Behavior). 500 factors were found
    - **A cross-sectional study** – questionnaire containing 500 items comparing the answers of the long-termed unemployed with those of whom resumed work quickly. Results narrowed 500 factors to 155
    - **A longitudinal study –** questionnaire with the 155 items was applied to groups of unemployed in 2008 and one year later in 2009 again on the same people. The conclusions allowed the team to keep only 20 variables
    - **New study in 2018**

Netherlands 

- **Explanatory variables:**

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

United States

Worker Profiling and Reemployment Services (WPRS)

- **Differences in the model across States**
- **Outcome variable:** Exhausting entitlement to unemployment benefits
- **Model:** mainly logistic regressions
- **Main use:** List eligible candidates for referral to reemployment services
- 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
- Candidates are sorted by their probability of exhaustion

United States

Explanatory variables vary by State

Variables used in most States:

- Unemployment benefit exhaustion
- Education
- Industry (change)
- Occupation (change)
- Tenure
- Local unemployment rate

Additional variables found in some States

- Wage replacement rate
- Potential duration
- Time from employment separation to the date the claim is filed
- Maximum benefit
- Weekly benefit amount
- Number of employers in the base period

United States

Example Kentucky (1996)

- claimant's previous wage,
- Unemployment benefit parameters
- reservation wage
- pensions
- assistance receipt
- prior UI receipt
- job tenure
- work experience
- reason for separation

Independent variables:

- industry growth
- occupation growth
- county unemployment rate
- county employment growth

United States

Methodological issues

- Using State-specific models increases success of the model
- States need to continually update their models to reflect recent changes in the economy, e.g., growth or decline of occupations and industries
- It is necessary to control for the effect of reemployment services when developing new profiling models
- 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 SICs to NAICS.
- More than half have never revised their models

Ireland 

Probability of Exit (PEX, 2009), Labour Market Disadvantage Model (2014)

- **Outcome variable:** Exit long-term unemployment
- **Model:** probit
- **Main use:** Jobseekers are assigned to to a "**participation path**" before the first meeting with the employment consultant according to their **"risk category**"
- **With this information and a more in-depth diagnostic interview**, the consultant can decide, based on his knowledge and experience, what precise combination of services and measures are offered to the job applicant

Ireland 

Probability of Exit (PEX, 2009)

(questionnaire), explanatory variables:

- 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
  - on benefits
  - on community work over the previous five years
- Location

Ireland 

Labour Market Disadvantage Mododel Explanatory variables (administrative data only)

- Age
- nationality
- education
- previous occupation
- household characteristics (including marital status, spousal earnings)
- unemployment benefit or unemployment allowance
- Location
- **No variables on employment history**

Ireland 

- **Methodological issues**
  - Labour Market Disadvantage model 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
  - **Probit model:** separate regressions by gender