



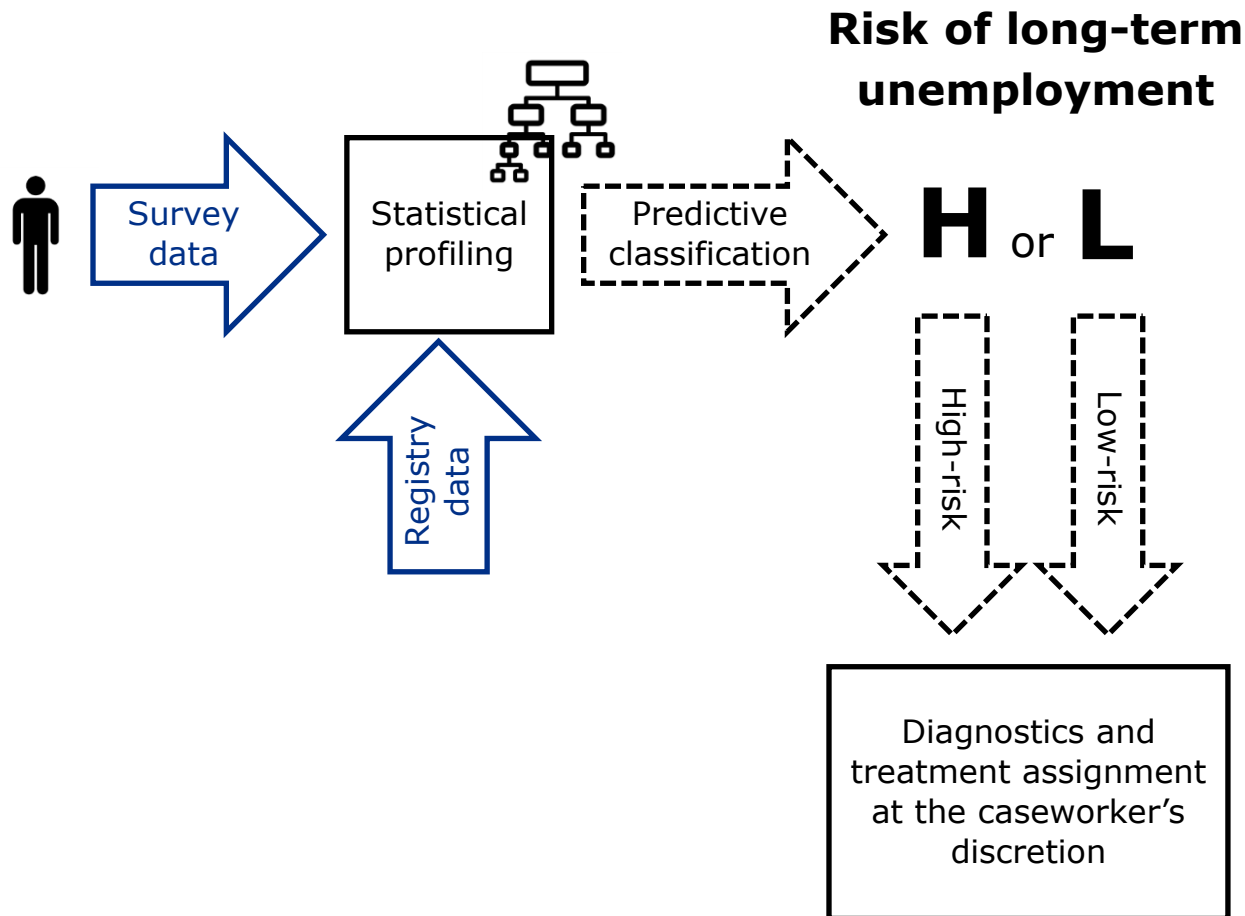
# **Predictive Profiling of UI Benefit Recipients in DK**

OECD Technical Workshop:  
Profiling tools and their use in active labour market policies

*Session 2—Design and methodology: Moving towards holistic profiling?*

# Predictive Profiling

## Newly Unemployed UI Benefit Recipients



### Principles

- Voluntary to use
- Support the assessment of the caseworker—not replace it
- Full transparency—caseworker and citizen gets the same results
- Intuitive methodology

### Advantages

- Objective and standardized assessment
- Identify high-risk jobseekers early
- Filter out low-risk jobseekers—optimize public resources
- Person-specific risk classification

### Disadvantages

- Assumes a stable environment
- Consumes a lot of data

# Data Sources

## Survey data

- Digital survey when unemployed citizens apply for benefits
- Voluntary to use
- Approx. 55,000 individuals every year (since 2014)
- 10 questions
- Examples include:
  - *"How fast do you think you will get a job?"* (7 categories + don't know)
  - *"Is there anything that makes it hard for you to get a job?"* (yes/no + don't know)

...

## Registry data

- Automatically included iff. survey is answered
- Variables that are relevant for labour market participation
- Examples of variables included:
  - Age
  - Gender
  - Education
  - Unemployment length
  - Number of unemployment spells
  - Former industry

...

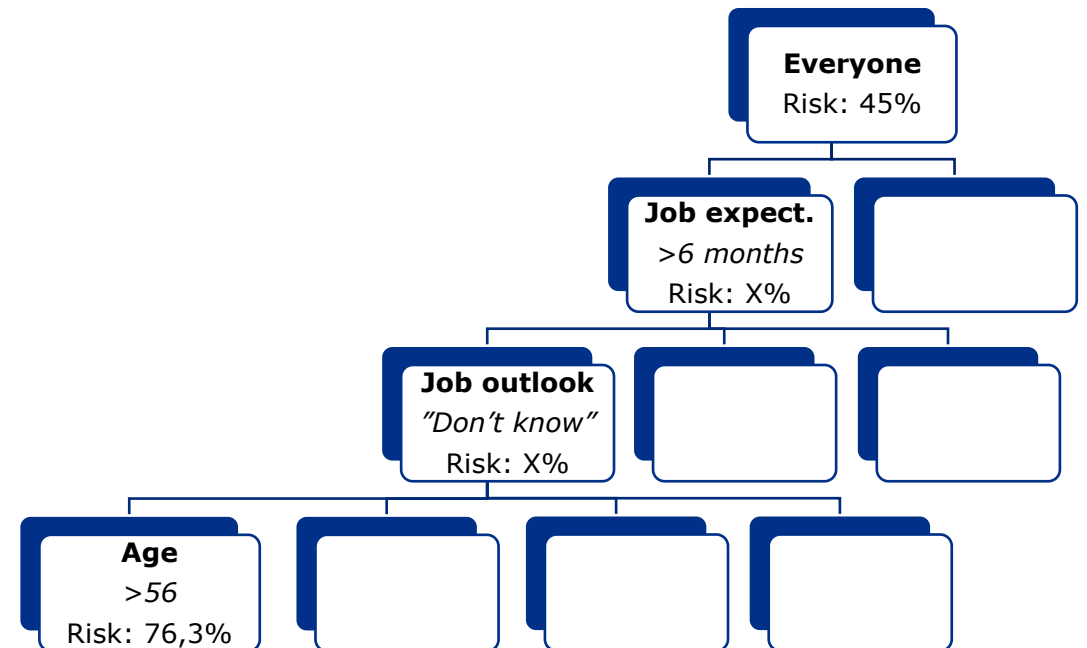
# Method

## Decision Tree Classification

### Description

- Algorithm that identify the variables that are most important for predicting “in risk of long-term (>26 weeks) unemployment”
- Maximizes the information gained (i.e., Kullback-Leibler divergence) by adding a new variable
- “Sorting” starts with the variable that gives the greatest information gain
- Variables that provide little information gain are removed
- Prediction precision >60% prob. of long-term unemployment
- 9 paths through the decision trees produce this accuracy

### An example of a path



# Imminent Next Step

## Update survey data and upgrade algorithm

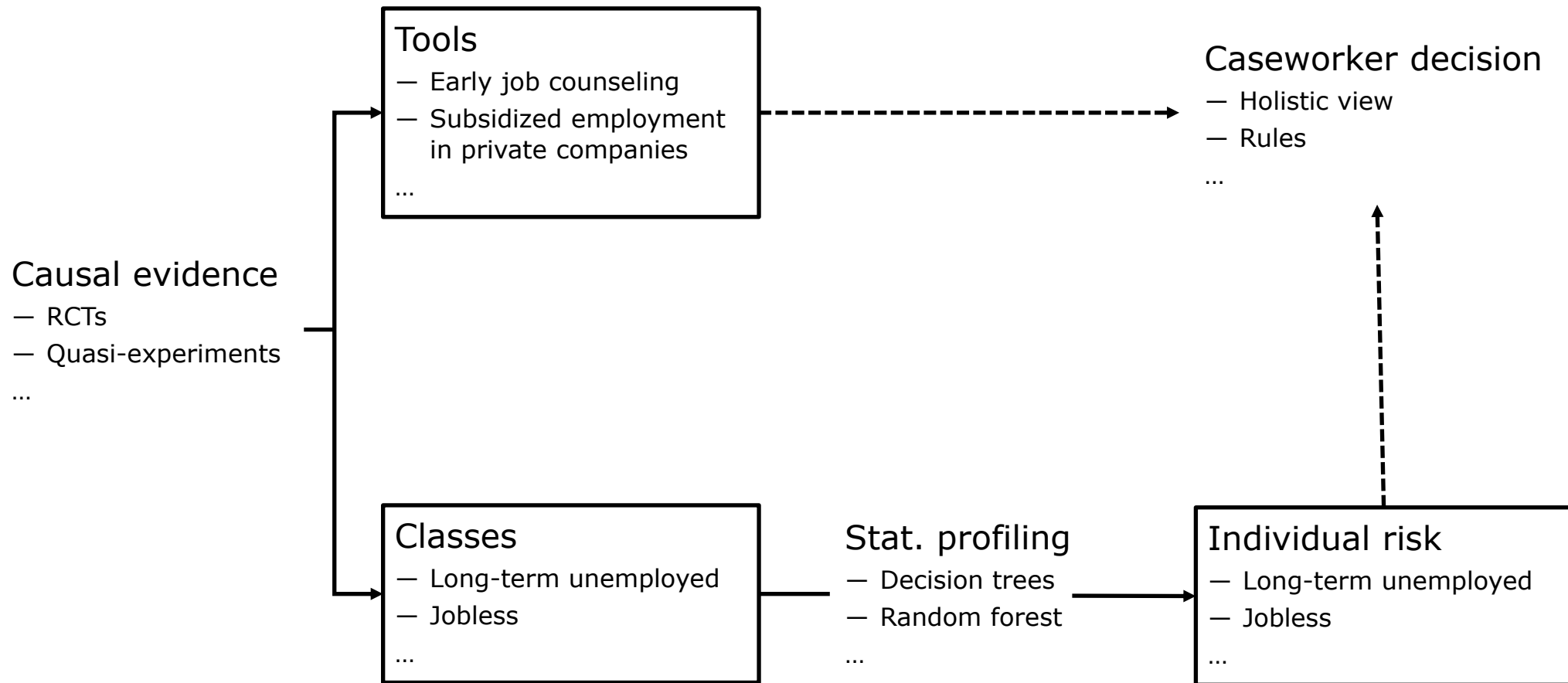
### Update survey data

- New questionnaire that aims to measure “deep” structural parameters relevant for job search
- Still voluntary to use
- Structural parameters to be measured:
  - Time preferences
  - Risk preferences
  - Social preferences
  - Locus of control
  - Beliefs about job prospects
- Questionnaire will be validate through relevant laboratory experiments

### Random forest algorithm

- Problems with decision trees
  - Tend to overfit = bad out of sample predictions
  - Splits are locally greedy = instable sorting
- Random forest algorithm:
  - Ensemble classifier that consists of many decision trees
  - Encourage diversity among the decision trees = avoid overfitting
  - Outcome decision by majority vote among all leafs reached in each tree = stability

# A Future Scenario for Predictive Profiling in DK



**Thank you!**