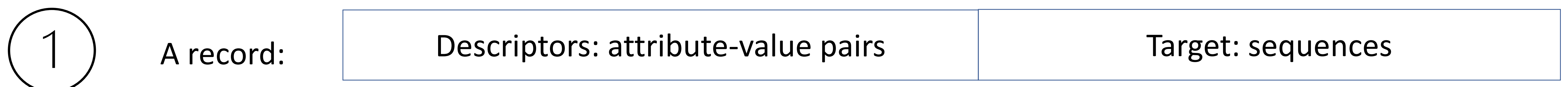


# Mining Sequences with Exceptional Transition Behaviour of Varying Order

COMING SOON

Rianne M. Schouten, Marcos L. de Paula Bueno, Wouter Duivesteijn, Mykola Pechenizkiy

We find **exceptional** subgroups with a Markov chain model with **more** (or less) parameters than the model estimated on the entire dataset, by using **quality measures** based on **information-theoretic** scoring functions.



2 Model class:



Figure 1 In a **first** order Markov chain, the **current** value predicts the next value.

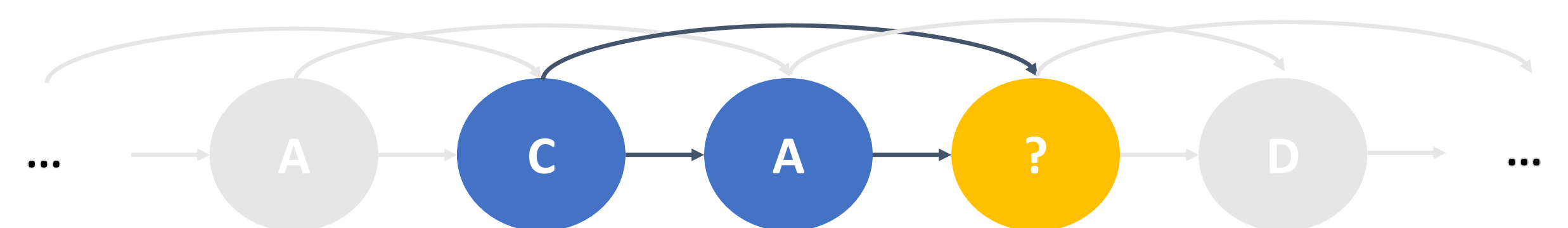


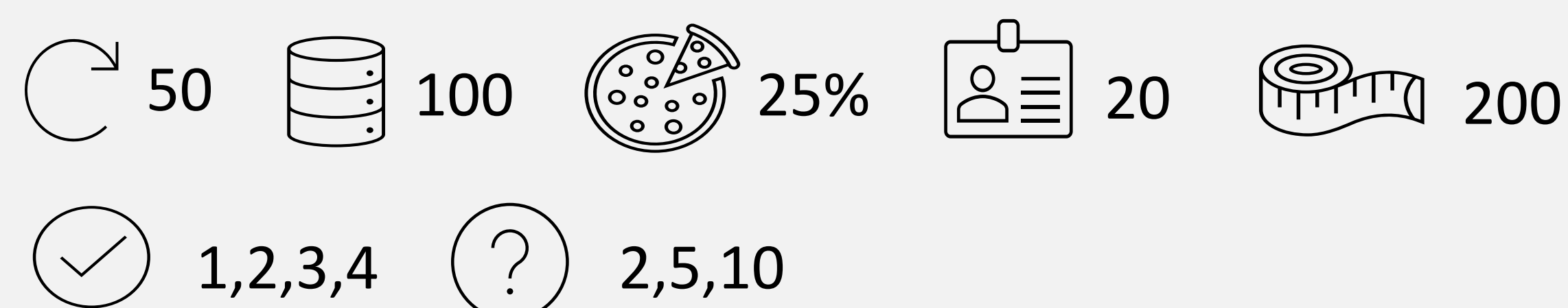
Figure 2 In a **second** order Markov chain, every combination of **two** values predicts the next value.

3 Quality measure:

$$\varphi_{EMM-AIC} = 2\mathcal{L}(P(SG|\theta^{SG})) - 2K^{SG} - 2\mathcal{L}(P(SG|\theta^{\Omega})) + 2K^{\Omega}$$

## Synthetic Data Experiment

**Simulation:** A ground truth subgroup of **varying order**. All other sequences a **first order** chain.



State-space 2 5 10

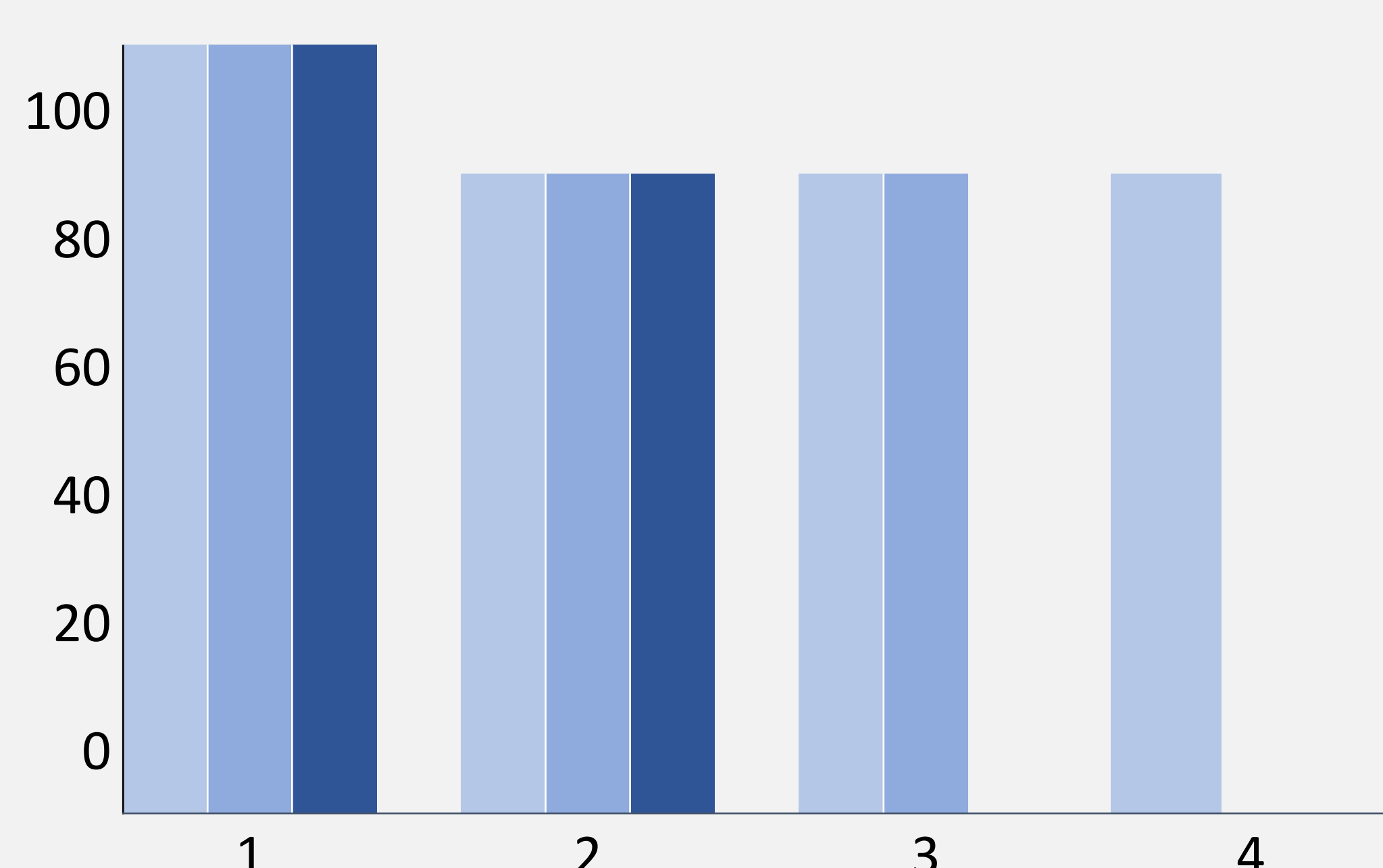


Figure 3 Percentage of repetitions where order of the true subgroup is found, for varying state-space.

## Real-World Data Experiment Blood glucose

**Overall data.** A **second order** Markov chain.

**Exceptional subgroup.** A **first order** Markov chain.

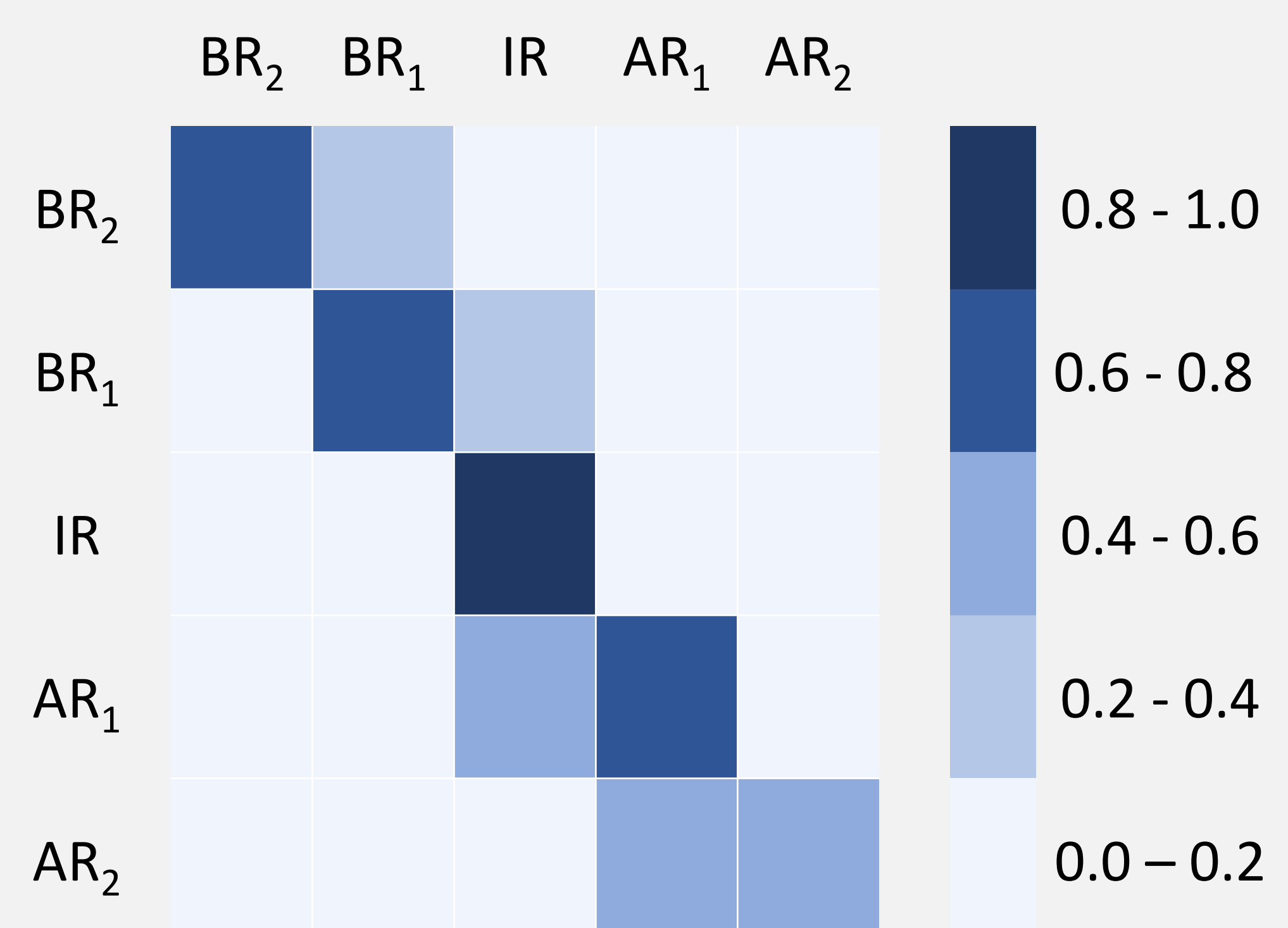


Figure 4 First order transition probability matrix of the exceptional subgroup.

Low HbA1c gives stable blood glucose sequences