Intelligent Interactive Systems

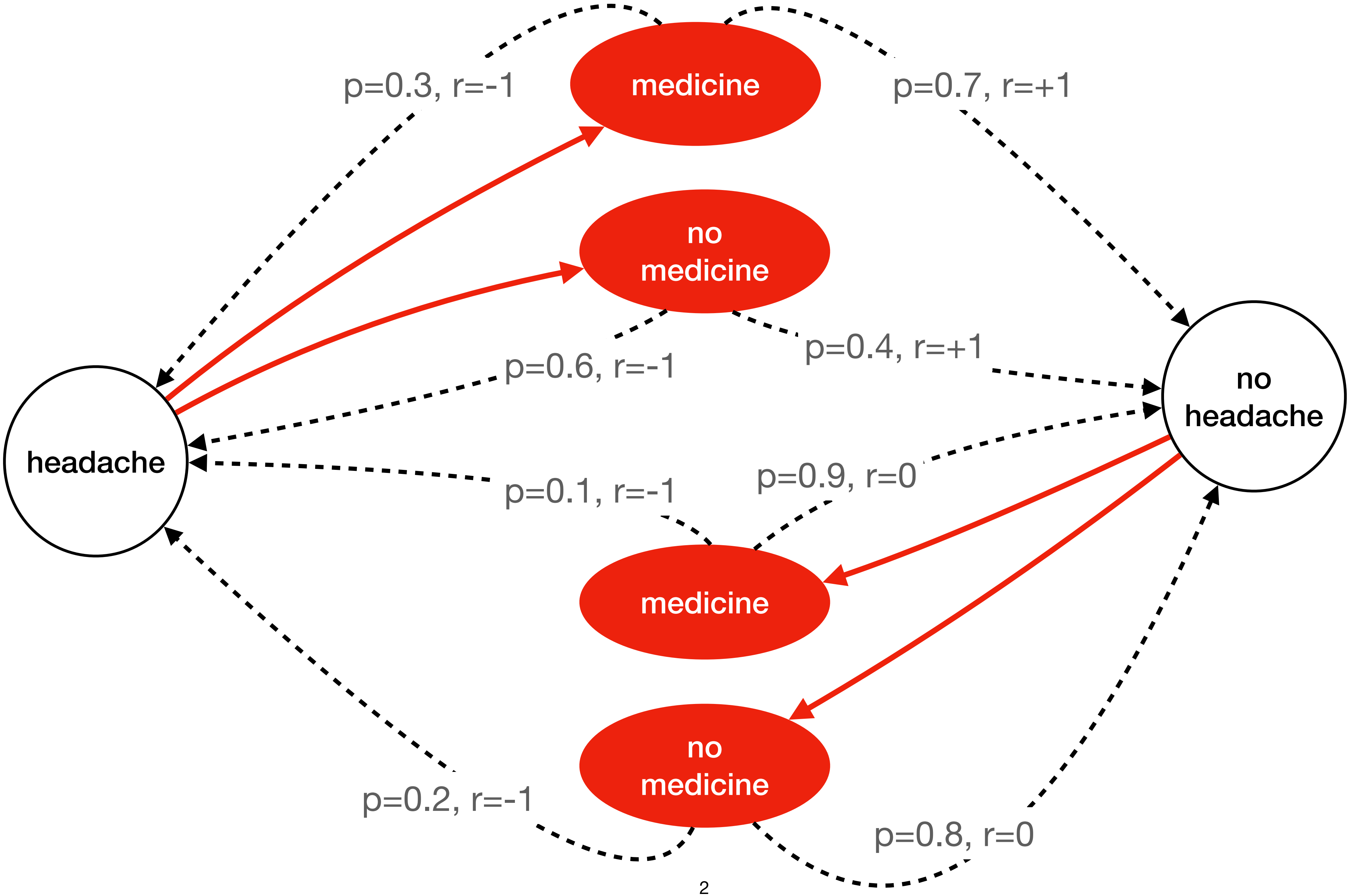
MDP applications

Mark Lee, University of Birmingham

Original slides by

Andrew Howes, University of Birmingham and Aalto University

MDP: Rewards



medicine
p=0.3, r=-1
p=0.7, r=+1
no medicine
p=0.6, r=-1
p=0.4, r=+1
headache
no headache
p=0.1, r=-1
p=0.9, r=0
medicine
no medicine
p=0.2, r=-1
p=0.8, r=0

Markov Decision Process

- S is a set of states.
- A is a set of actions
- State transition function. The new state is s' with probability conditioned on the previous state s and the action a:
    - $P(s' \mid s, a)$
- Rewards. A scalar reward is received depending on the transition from s to s':
    - $r = R(s', s)$

The Bellman Equation for MDPs

$$V^{\pi*}(s) = \max_a \left\{ R(s,a) + \gamma \sum_{s'} P(s'|s,a) V^{\pi*}(s') \right\}.$$

- $V^{\pi^*}(s)$ is the value of state s assuming the optimal policy π.
- It is defined as the maximum of the expected values of all actions a from state s.
- The expected value of an action has two parts, the reward defined by $R(s,a)$ and the discounted γ sum of all possible expected outcome values $V^{\pi^*}(s')$ if a is chosen assuming that the optimal policy continues to determine future actions selections.

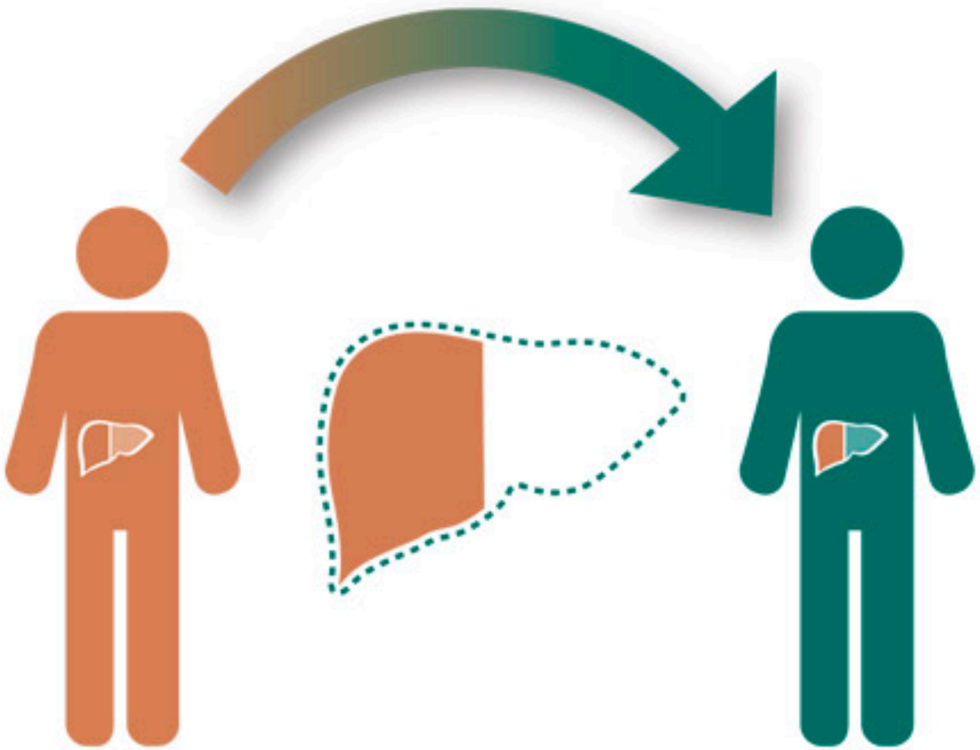
Applications of MDPs

- responding to an epidemic (Lefevre)
- drug infusion plans (Hu)
- heart disease treatment (Hauskrecht and Fraser)
- optimal time to iniative HIV treatment
- whether to accept/reject a liver donation.
- medical needle steering
- breast cancer screen policies
- statin therapy onset for diabetes
- etc.

Example 1

How it Works

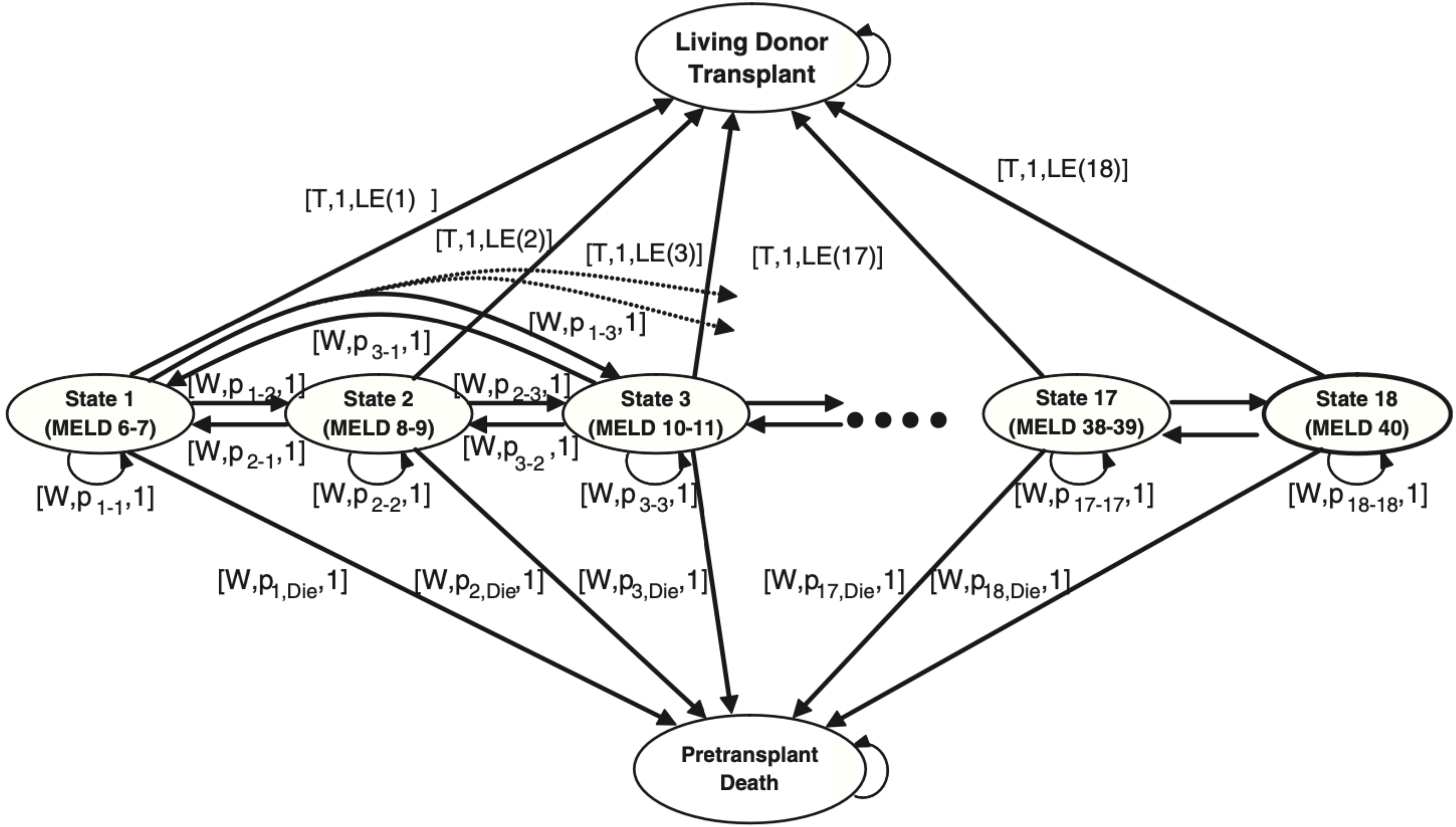
40-65% of a healthy donor's liver is removed and transplanted into the sick patient.



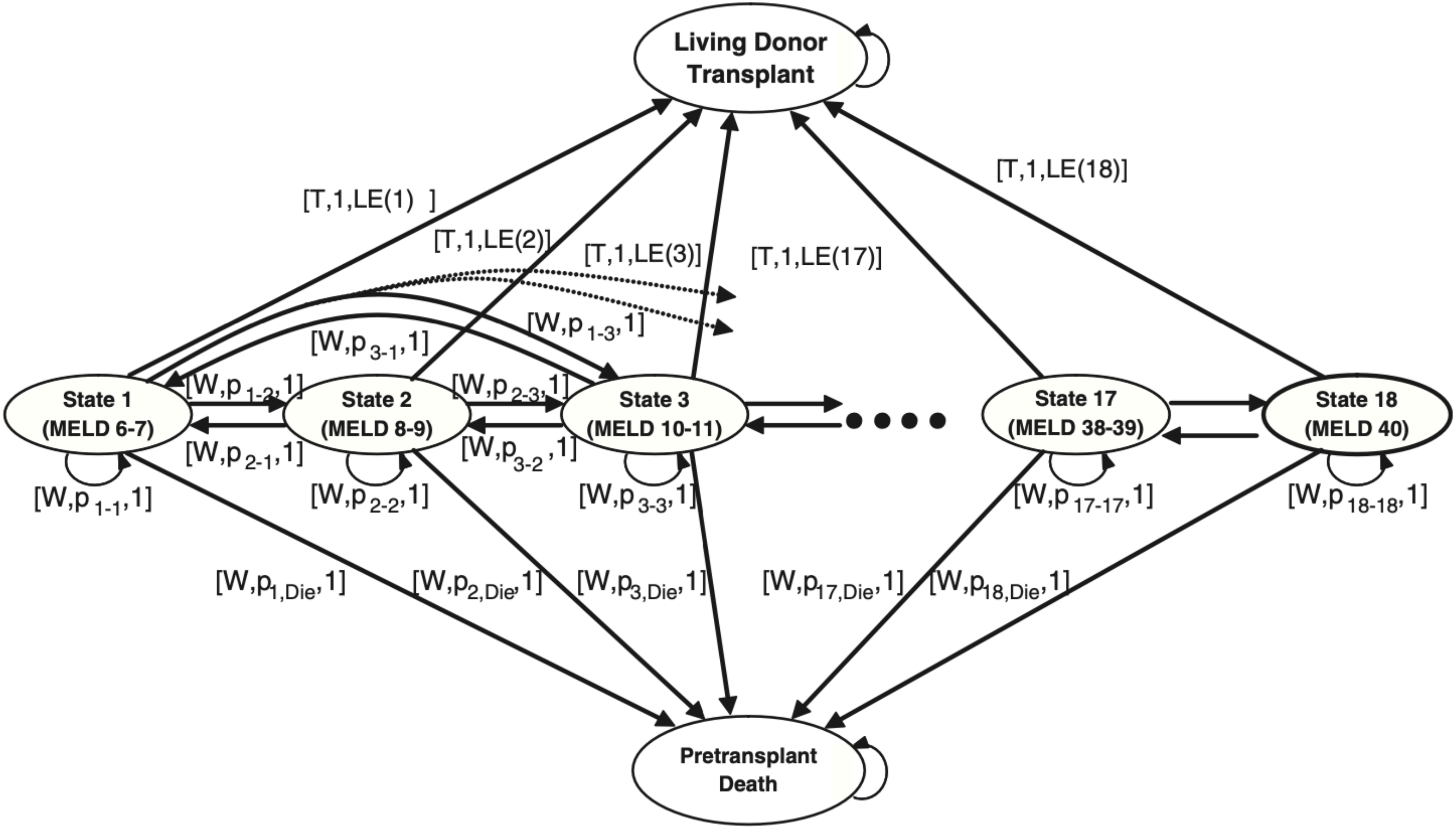
The transplanted portion of liver grows to meet the recipient's needs, and regrowth in the donor's liver happens in just a few months.

organ accept/reject recommendation

- Trade-offs
  - because post transplant survival only 10 to 20 years.
  - but also transplant success depends on health.
- Complexity
  - 14 types of liver offer
  - x 18 health states
  - x 30 rankings of quality
  - = 18x14x30 = 7560 nodes in a decision tree.



Alagoz, O., Hsu, H., Schaefer, A. J., & Roberts, M. S. (2010). Markov decision processes: a tool for sequential decision making under uncertainty. *Medical Decision Making*, *30*(4), 474-483.



Alagoz, O., Hsu, H., Schaefer, A. J., & Roberts, M. S. (2010). Markov decision processes: a tool for sequential decision making under uncertainty. *Medical Decision Making*, *30*(4), 474-483.

Discussion

- Unfortunately Alagoz et al. did not have the data to test their model thoroughly.
- But, subsequent work, using a similar methodology, does provide some evidence…

Example 2

MDP Diabetes recommendation

State space

- **Chronic** complications (retinopathy, cataracts etc.)
- **Acute** complications (e.g. heart failure)
- Severe **risk** = 1, otherwise 0
- Period - years elpased since occurence of diabetes.

$$S^t = \left(S^t_{Chronic}, S^t_{Acute}, S^t_{Risk}, S^t_{Period}, S^t_{FPG}\right) \quad \mathrm{t} = 1, \ldots, \mathrm{T}$$

Action space

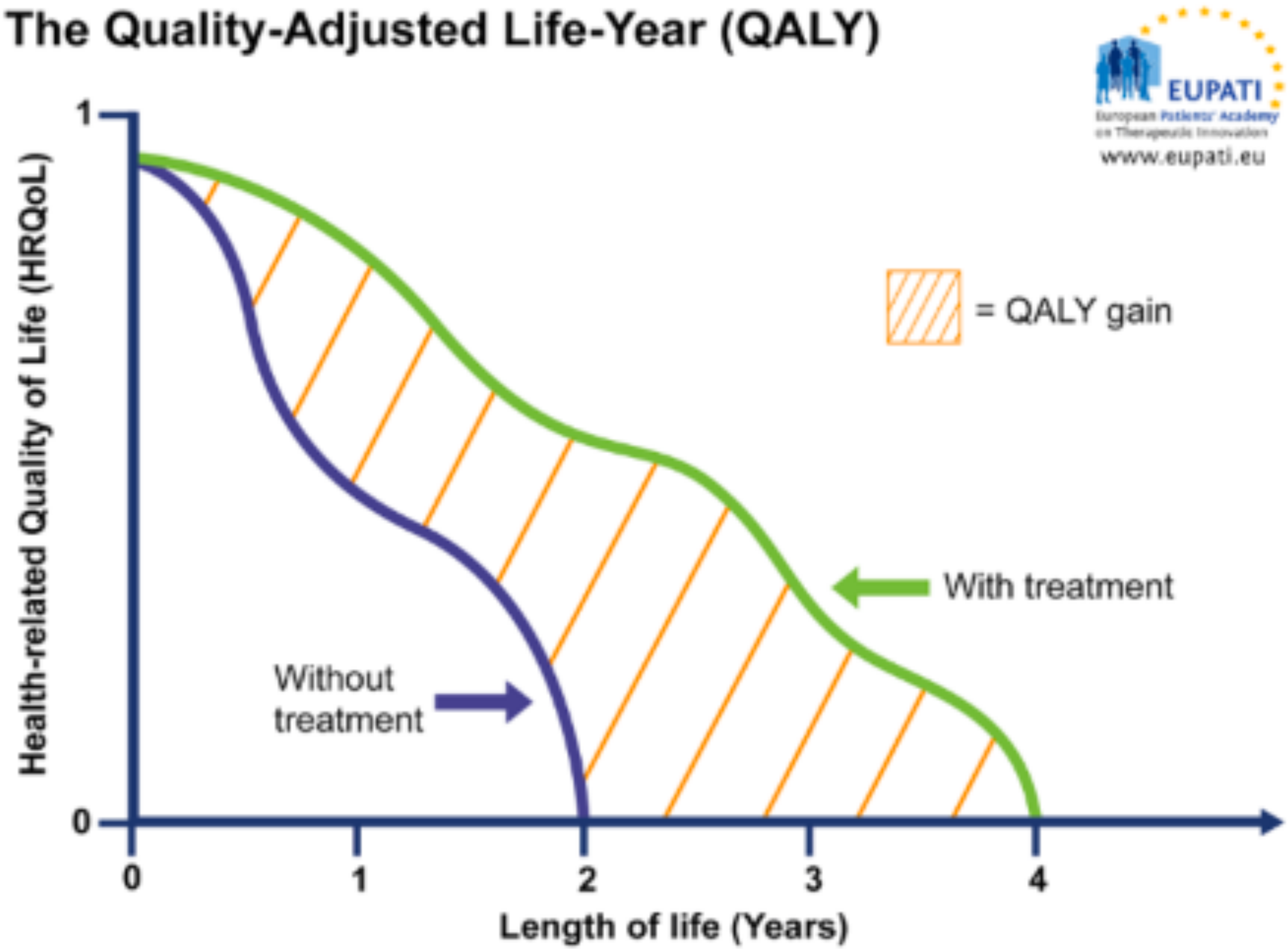
Type	Action no	Description
Monotherapy	A1	Metformin (MET)
	A2	Sulfonylureas (SUL)
	A3	Others
Dual therapy	A4	Metformin + sulfonylureas (MET + SUL)
	A5	Metformin + DPP-4 inhibitors (MET + DPP4I)
	A6	Others
Triple therapy	A7	Metformin + sulfonylureas + alpha-glucose inhibitor (MET
	A8	Others

State Transition Function

- Determined from the data.
- Counted the number of occurrences of state s to s' transitions after having taken each type of medicine (an action).
- Then normalised.

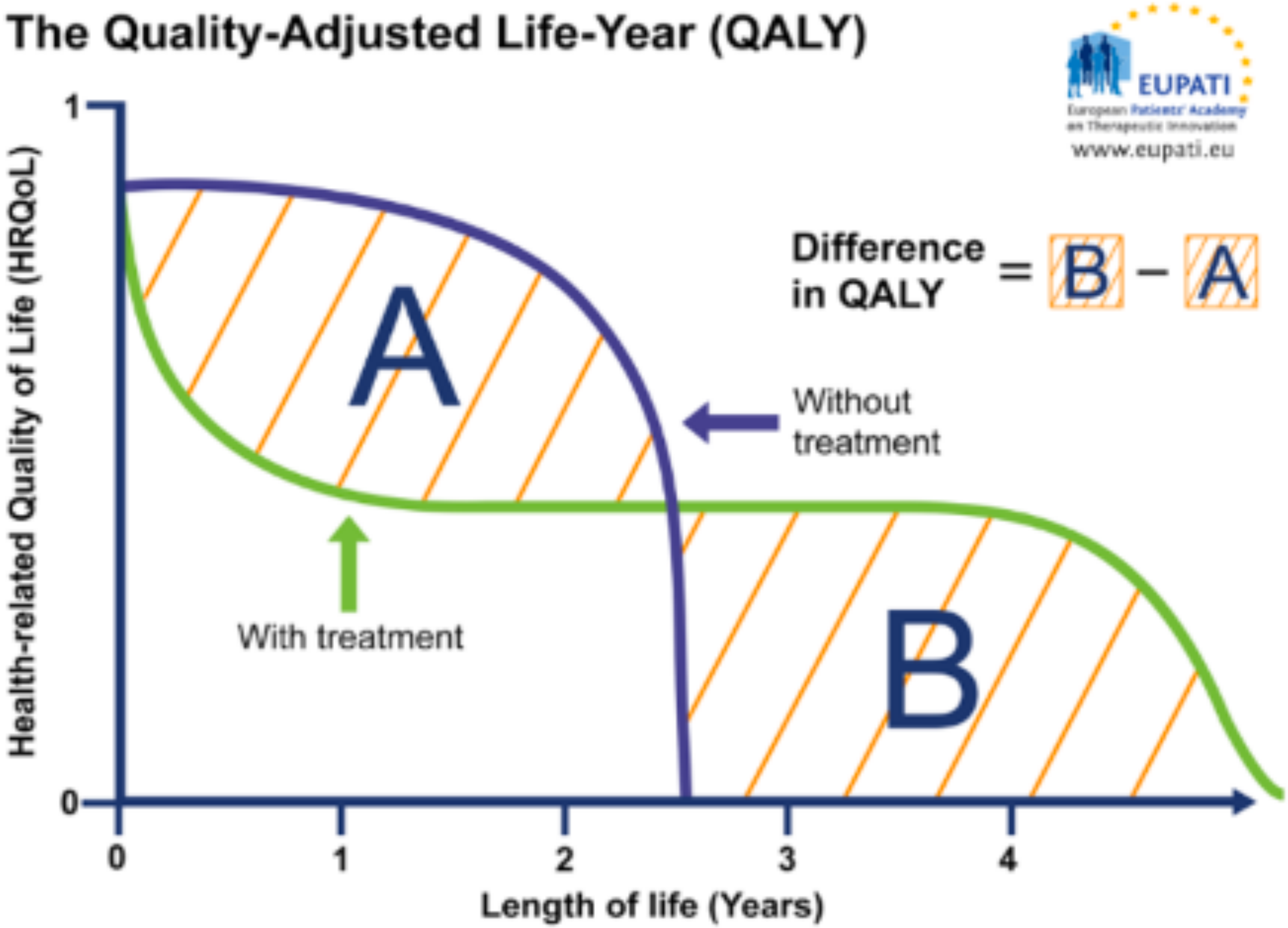
$$\mathrm{T(s,s',a)} \text{ for } s, s' \in S, a \in A.$$

Reward



https://learning.eupati.eu/mod/book/view.php?id=510&chapterid=472

Reward



https://learning.eupati.eu/mod/book/view.php?id=510&chapterid=472

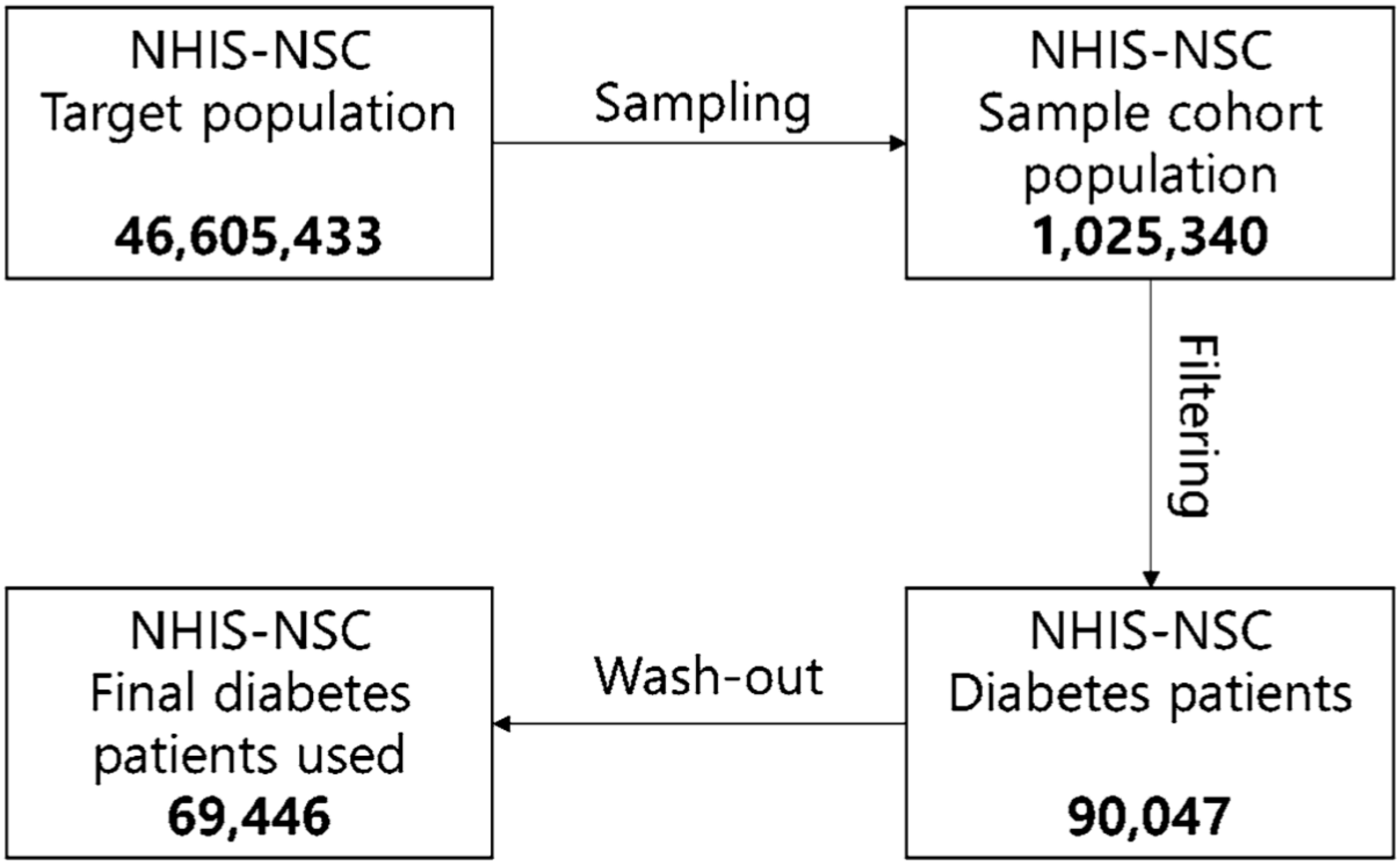
Reward

- QALYs quantify the quality of a year of life with the discomfort due to medical interventions.
- Treatment will offer some reward to the patient, such as a potentially longer life.
- However, there are costs
  - reduction in quality of life,
  - side effects due to medications,
  - or a financial cost (medication or hospitalization).

$$R(a, s') = R^{WTP}\left[(1 - d^{Chronic}(s'))(1 - d^{Acute}(s'))(1 - d^{Risk}(s'))(1 - d^{Period}(s'))\right] - C^{MED}(a)$$

Data cleansing

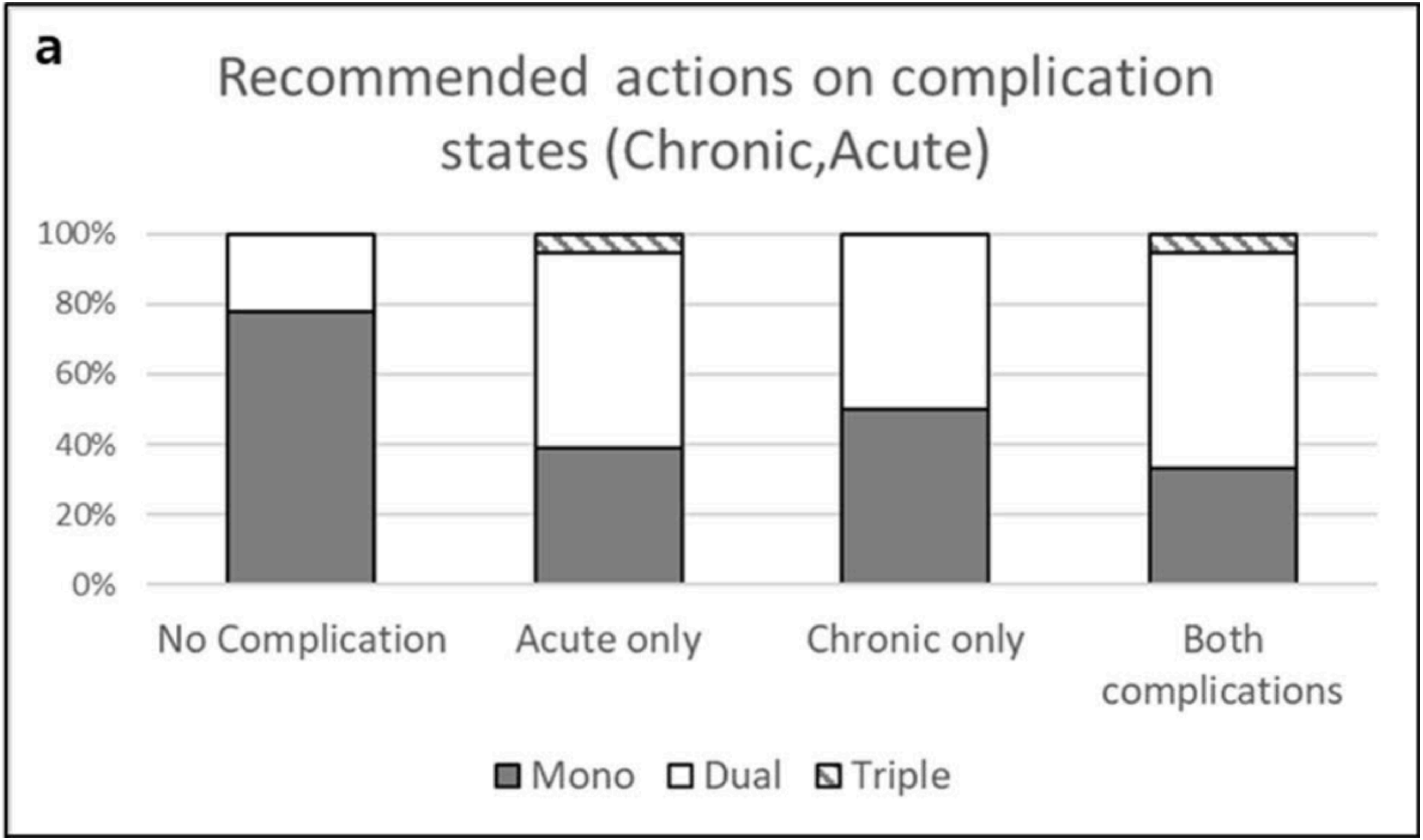
- 46 million patient records sampled reduced to 69 thousand used.



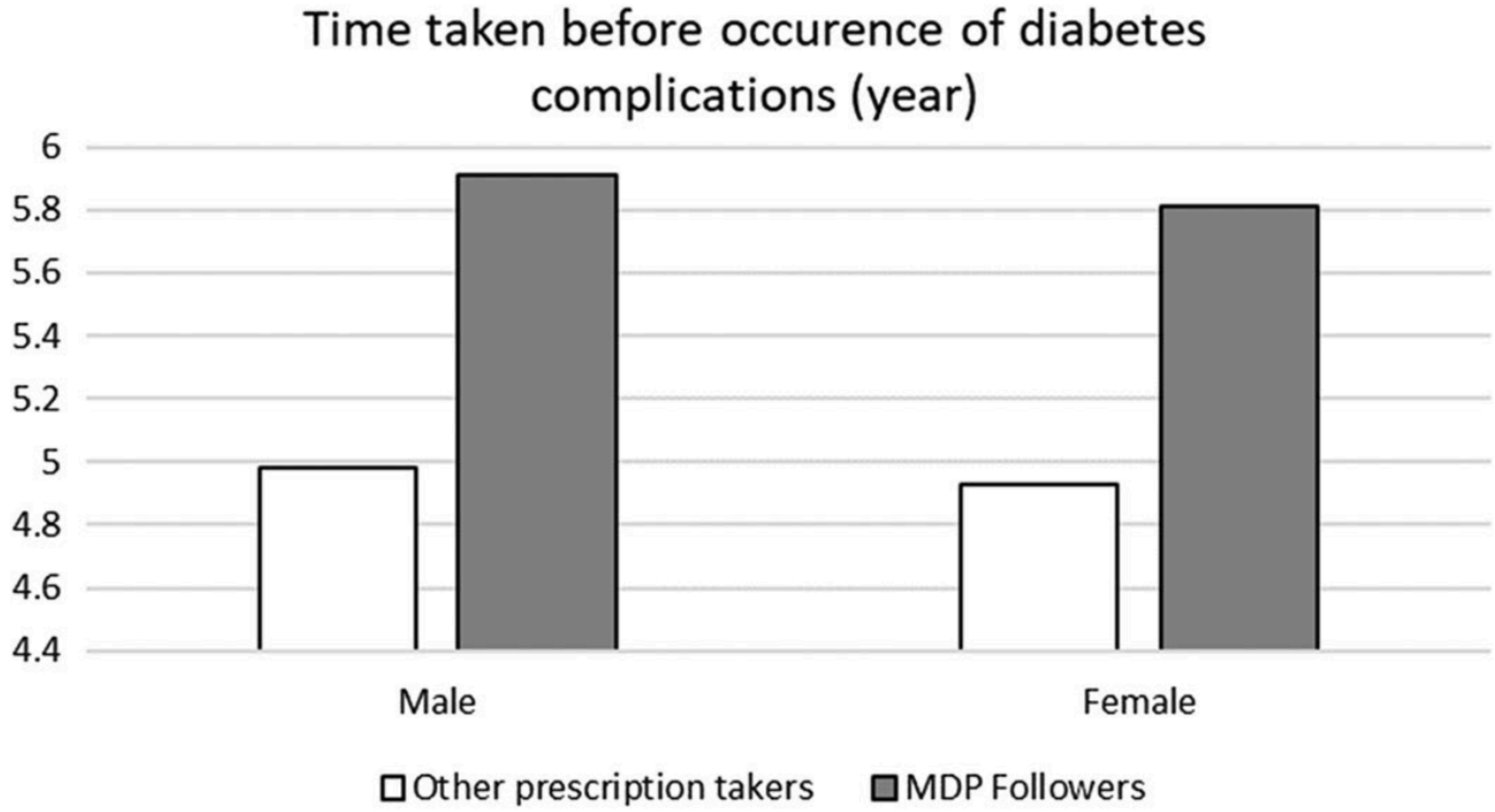
Statistics of filtered patients

	Male	Female
Sex (%)	54	46
Age, mean (SD)	58.4 (25.1)	61.7 (25.4)
Period of having diabetes (years), mean (SD)	7.5 (2.7)	8.6 (2.5)
BMI (kg/m^2), mean (SD)	25.7 (2.7)	25.4 (2.5)
FPG (mg/dL), mean (SD)	142.9 (58.1)	145.8 (58.5)
Total cholesterol (mg/dL), mean (SD)	188.5 (45.1)	192.4 (45.7)
SBP (mmHg), mean (SD)	129.2 (24.6)	134.1 (24.8)
DBP (mmHg), mean (SD)	79.6 (14.5)	81.3 (14.7)
Smoker (%)	65	48

Results - MDP recommendations



Results - Quality of life



Summary

- Both Alagoz et al. and Oh et al. have proposed MDP-based treatment recommendation systems for medical conditions.
- Oh et al. designed the states, actions, reward functions, and transition probability matrices of an MDP for diabetes management.
- The model was tested against a national medical records database requiring fitting of reward and state transition functions.
- The results show that diabetes medication recommended by their MDP system is realistic as it correctly recommends changing treatment patterns.

Exercises

- MDPs provide generative models of the decision making process involved in solving real-world problems.
    - These generative models can be used as the basis of an “MDP recommender system”.
    - What would the advantages/disadvantages of such a system be?
    - Would MDPs provide a better or worse basis for explanation than other types of recommender system?

Reading

- Alagoz, O., Hsu, H., Schaefer, A. J., & Roberts, M. S. (2010). Markov decision processes: a tool for sequential decision making under uncertainty. Medical Decision Making, 30(4), 474-483.

- Oh, S. H., Lee, S. J., Noh, J., & Mo, J. (2021). Optimal treatment recommendations for diabetes patients using the Markov decision process along with the South Korean electronic health records. Scientific reports, 11(1), 6920.