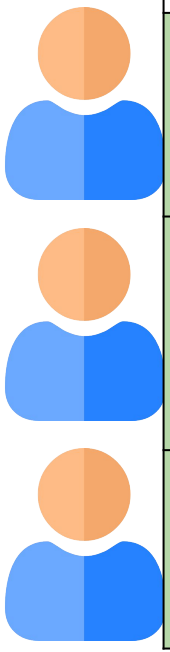
Human-in-the-Loop Resource Allocation in Restless Multi-Armed Bandits and Their Application to Public Health Interventions

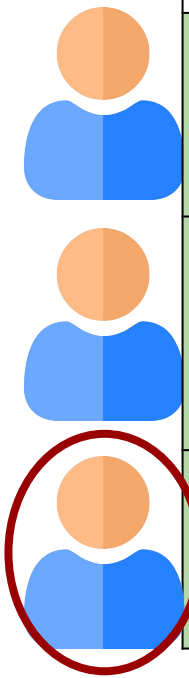
Aviva Prins

Motivating Example



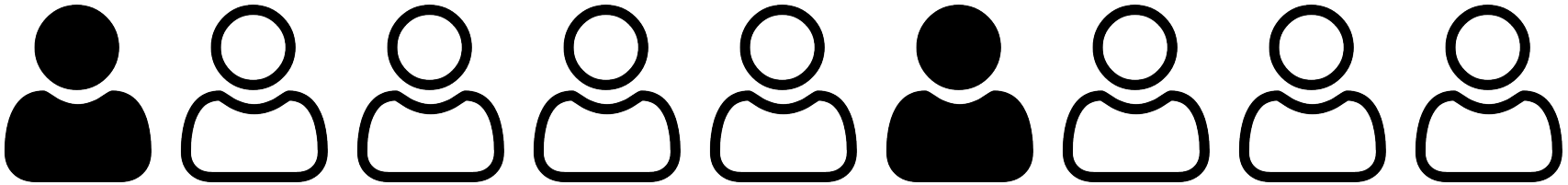
Day 1	2	3	4	5	6	7	8	9	10	11
✓	✓	✓	✓	✓	✓	✓	✓	?	?	
✓	✓	✓	✓	✓	✗	✓	✓	?	?	
✓	✓	✓	✓	✗	✓	?	?	?	?	

Motivating Example



Day 1	2	3	4	5	6	7	8	9	10	11
✓	✓	✓	✓	✓	✓	✓	✓	?	?	?
✓	✓	✓	✓	✓	✗	✓	✓	?	?	?
✓	✓	✓	✓	✗	✓	✗	✗	✗	✗	✓

Restless MAB Model



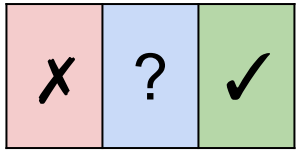
Choose *k* out of *N* people

at each timestep *t*

such that adherence is maximized

Restless MAB Model

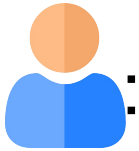
: Markov Decision Process

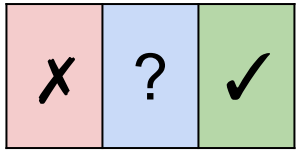


: State Space

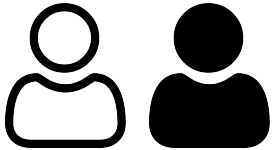
$$b := \Pr(s_{t+1} = 1 \mid s_t, a = 0)$$

Restless MAB Model

 : Markov Decision Process



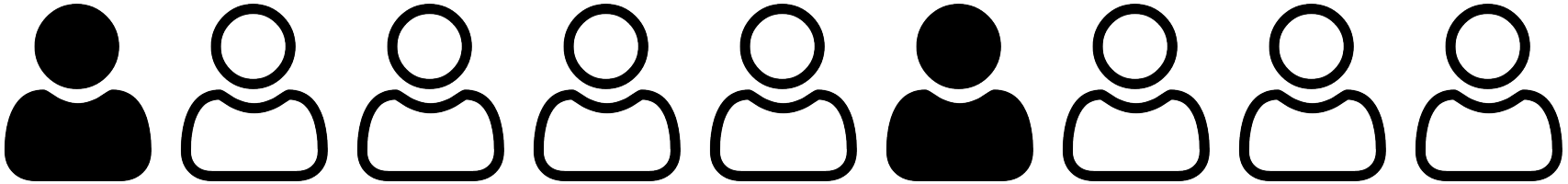
: State Space

 : Action Space

$P(s, s', a)$: Transition Matrix

$r(s) = s$: Reward Function

Restless MAB Model

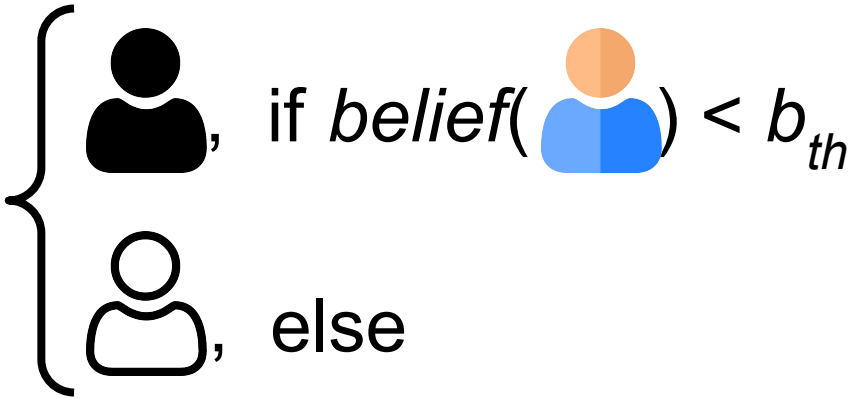


Choose *k* out of *N* arms

at each time step *t* such that:

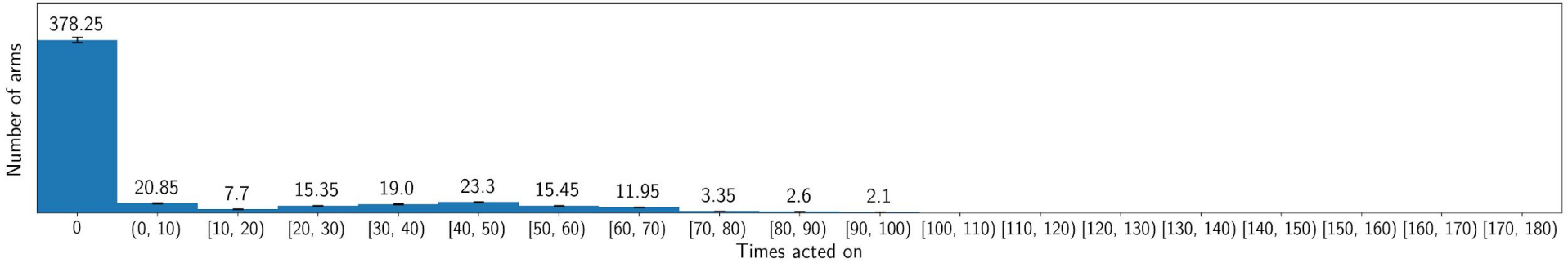
$$R(S) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it}$$ is maximized

Forward Threshold Optimal

optimal action = 

Threshold Whittle Algorithm

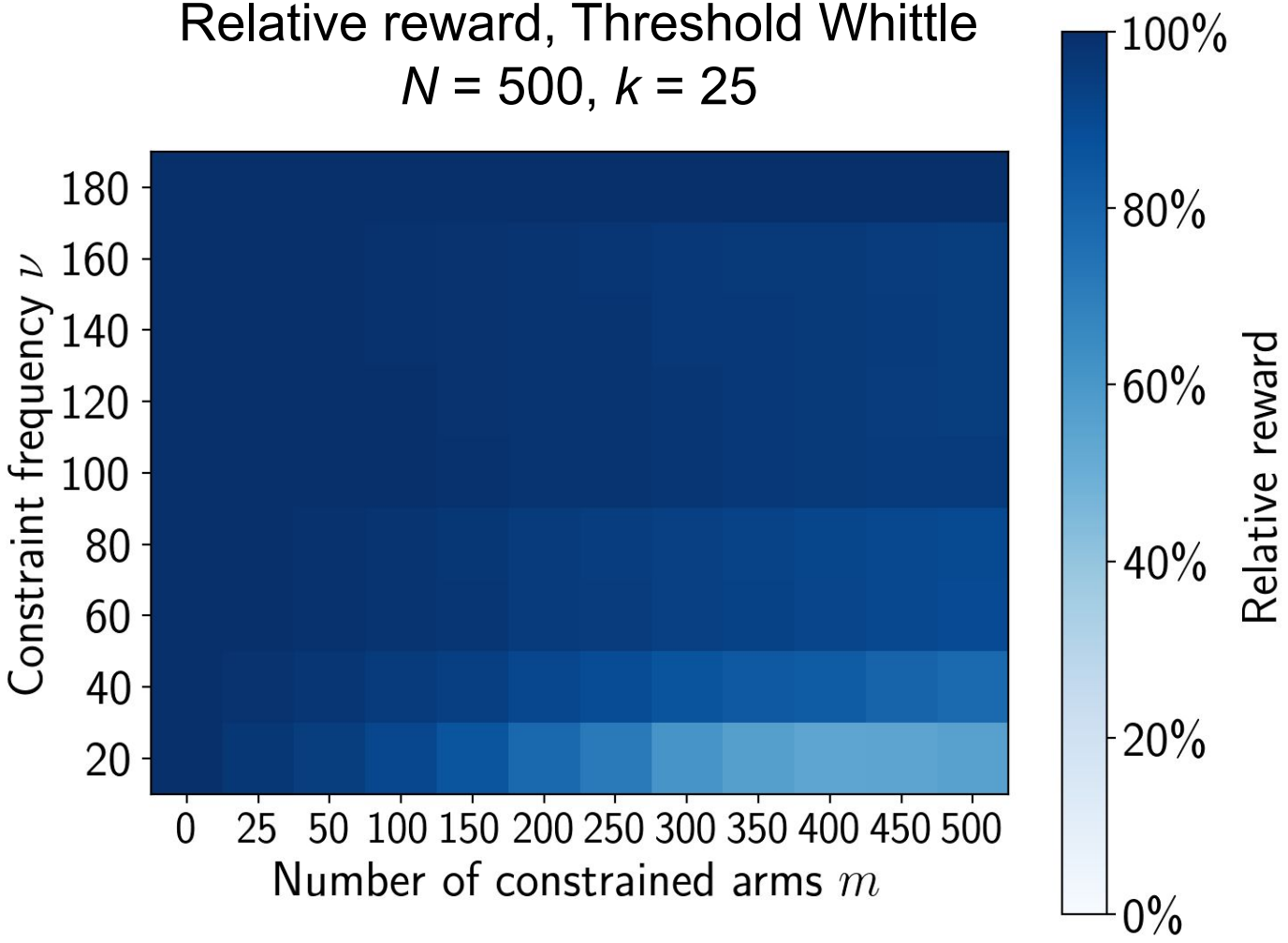
Histogram of actions
$N = 500$, $k = 25$

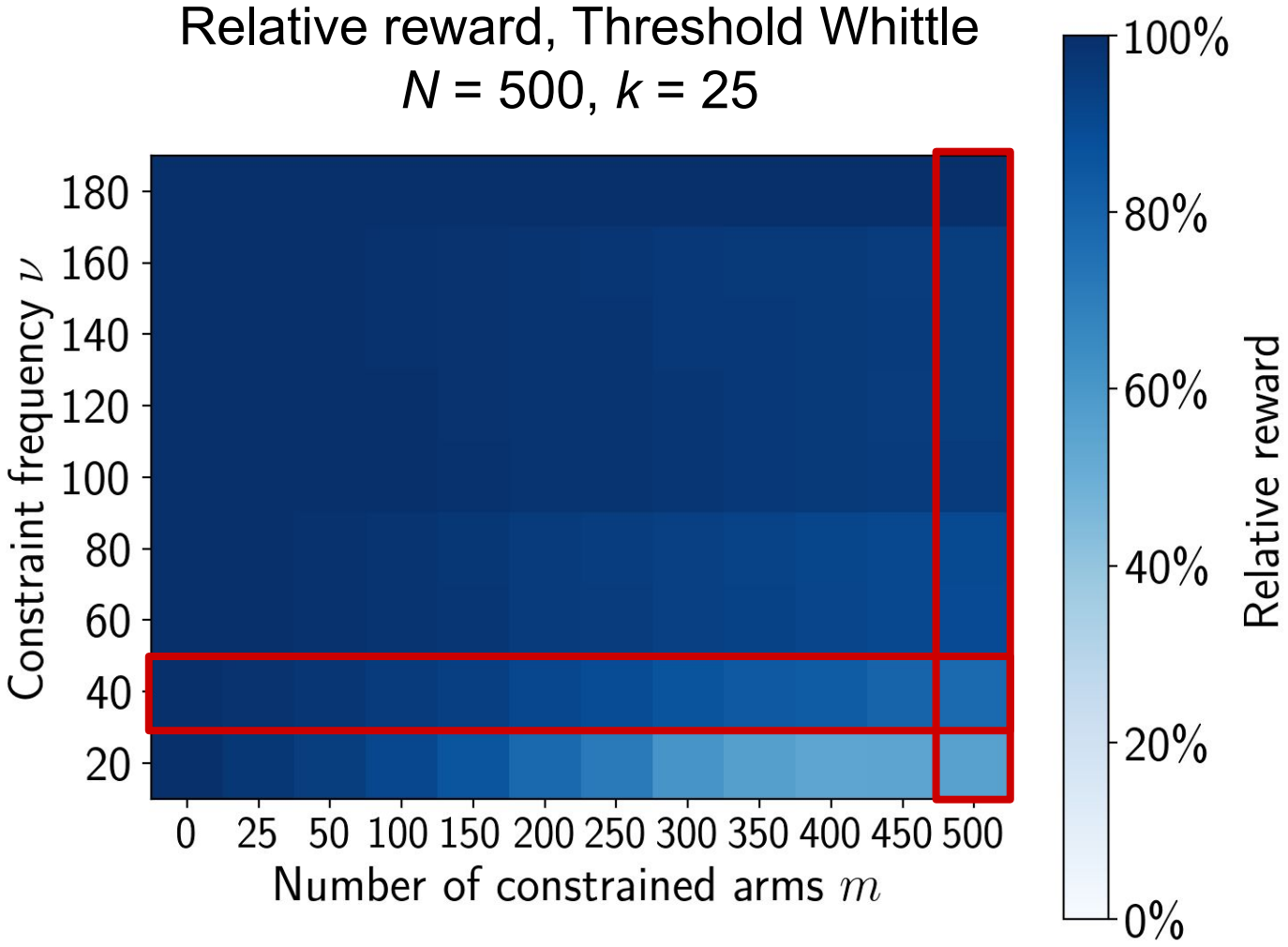


Human-in-the-Loop

Health worker applies a frequency constraint:

Ensure  is called at least every v days

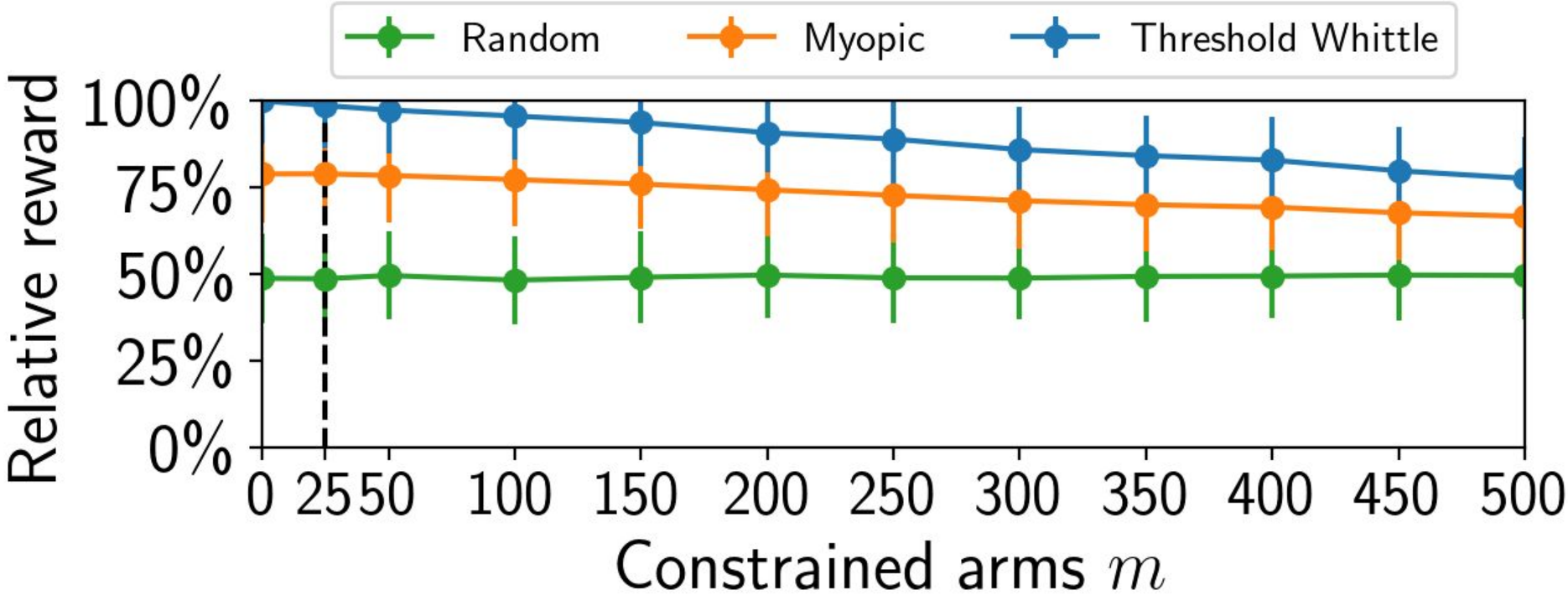

Relative reward, Threshold Whittle
N = 500, k = 25
Constraint frequency ν
180
160
140
120
100
80
60
40
20
0
25
50
100
150
200
250
300
350
400
450
500
Number of constrained arms m
100%
80%
60%
40%
20%
0%
Relative reward


Relative reward, Threshold Whittle
N = 500, k = 25
Constraint frequency ν
180
160
140
120
100
80
60
40
20
0
25
50
100
150
200
250
300
350
400
450
500
Number of constrained arms m
100%
80%
60%
40%
20%
0%
Relative reward

Relative reward

N = 500, k = 25

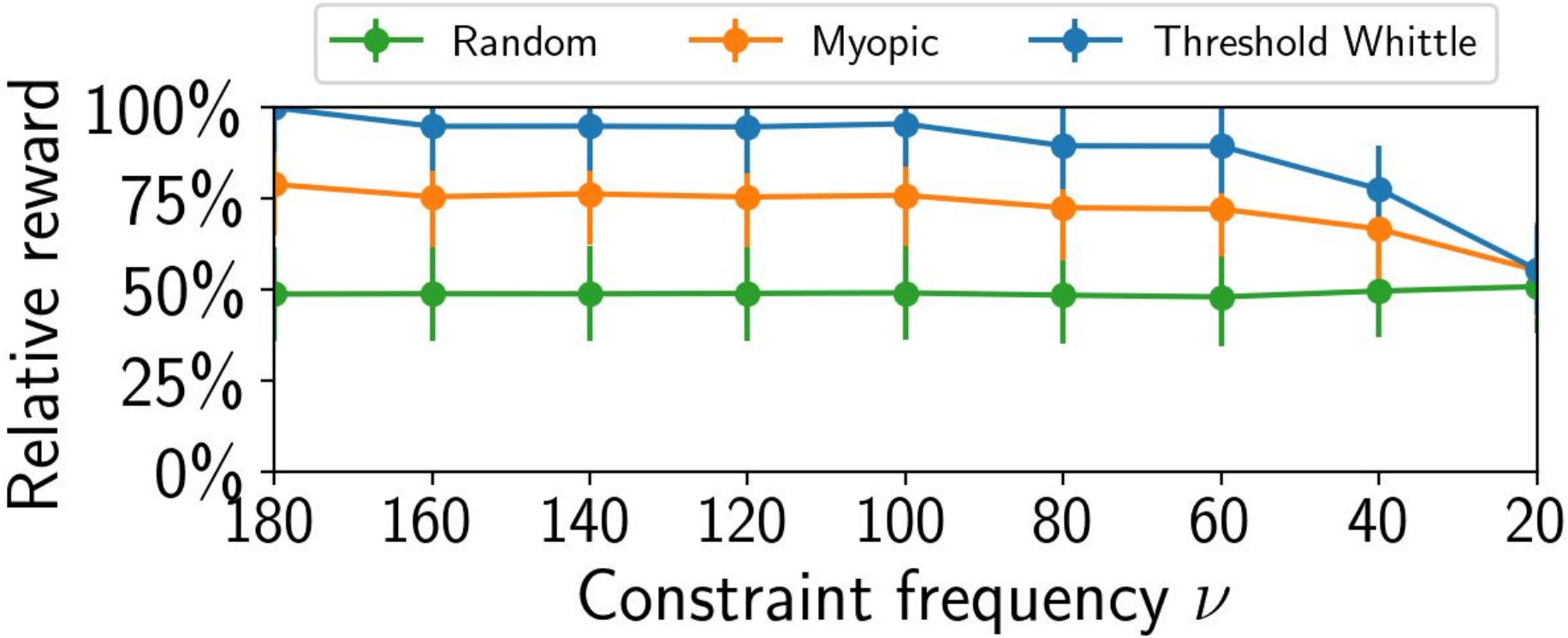
call at least every ν = 40 days



Relative reward

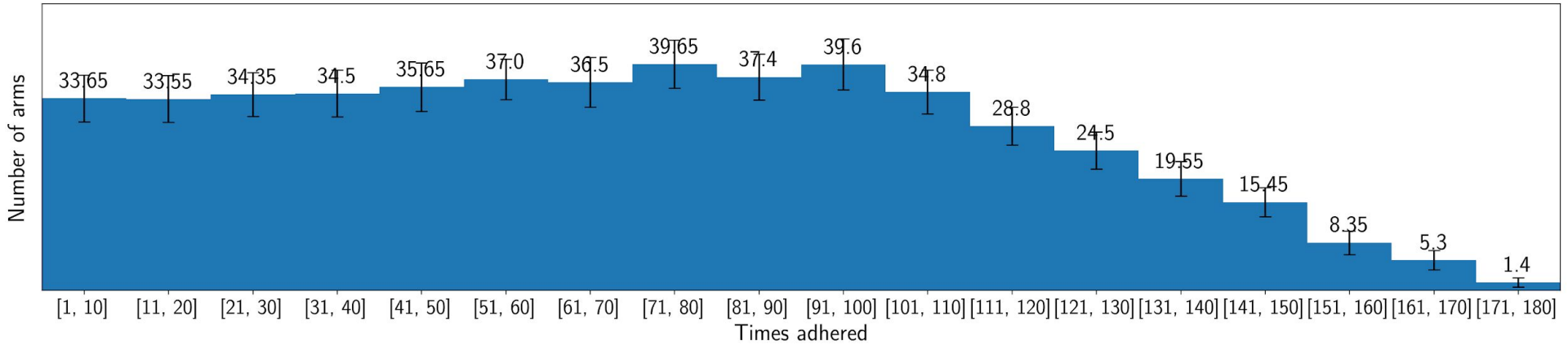
$N = 500$, $k = 25$

$m = 500$ constrained arms



Simulated Results

Histogram of Adherence
$N = 500$, $k = 25$
no constraints

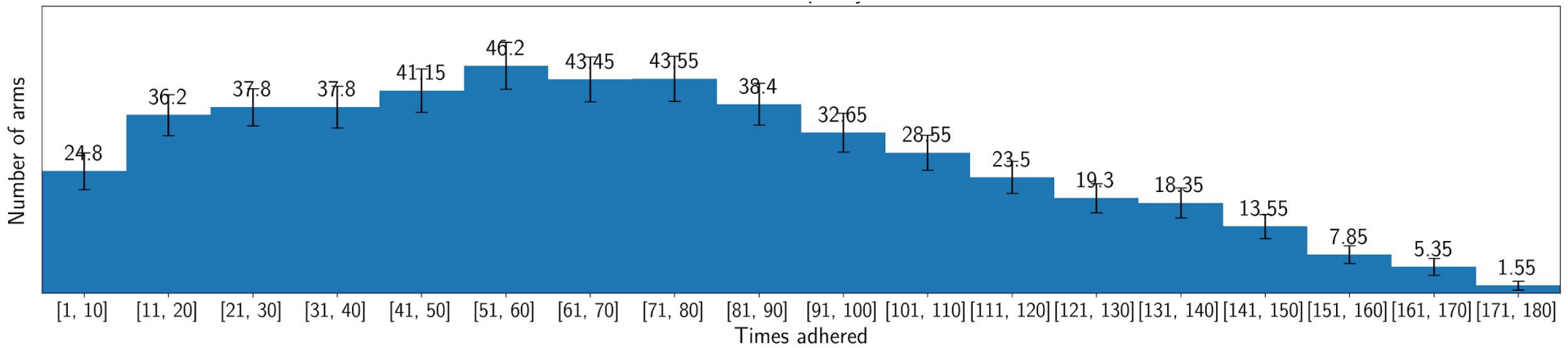


Simulated Results

Histogram of Adherence

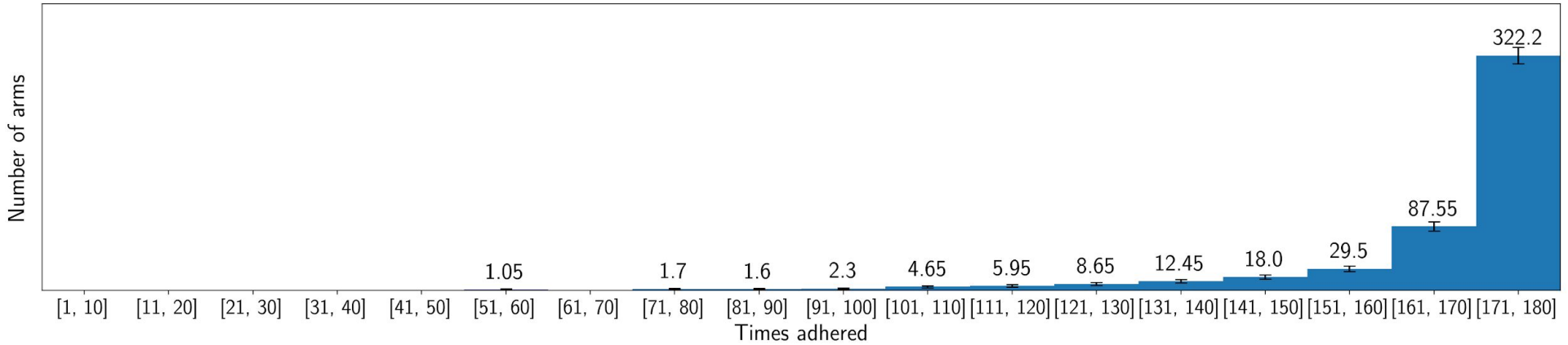
$N = 500$, $k = 25$, $m = 500$

call at least every $\nu = 40$ days



Simulated Results on Real Patient Data

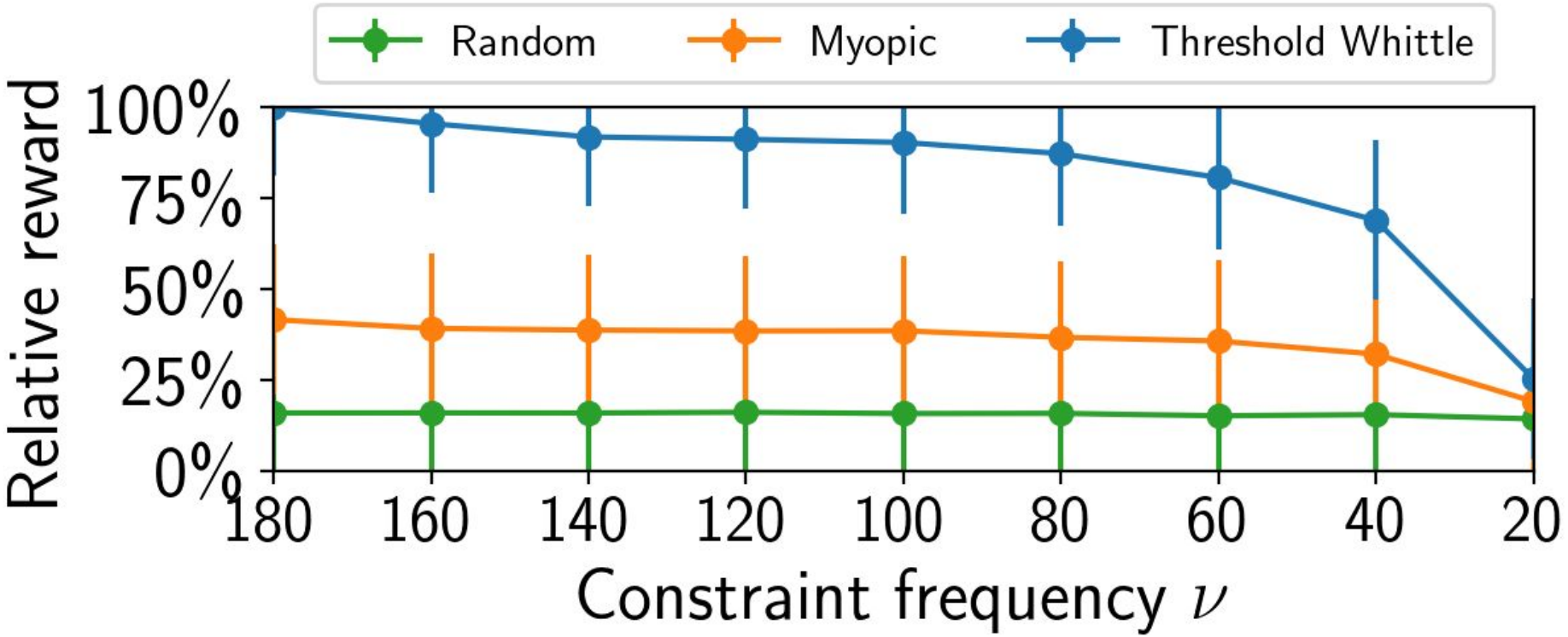
Histogram of Adherence
$N = 500$, $k = 25$
no constraints



Relative reward

$N = 500$, $k = 25$

$m = 500$ constrained arms



Conclusion

Our human-in-the-loop model is very effective:

- More diverse allocation of resources
- Elicits health worker feedback
- Versatile

Thank you

- CRCS Summer Fellowship
- Prof. Milind Tambe
- Dr. Rediet Abebe
- Aditya Mate
- Jackson Killian