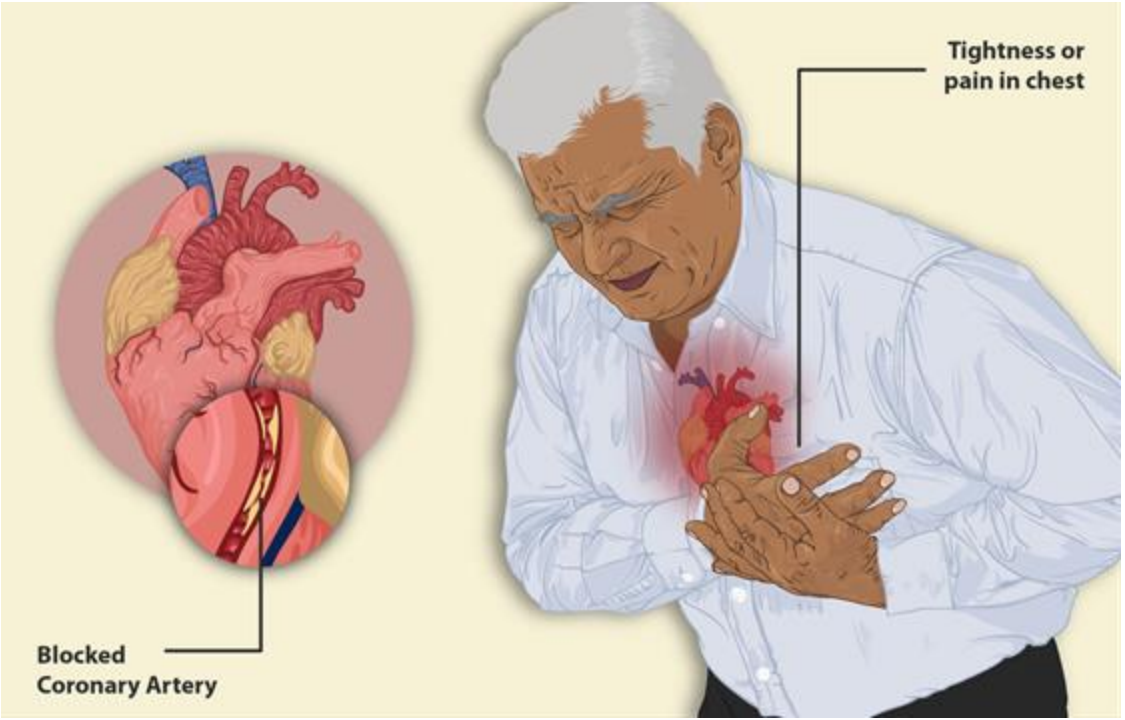
But What if AIs Could Impact Life-or-Death Decisions?



Heart Attack? Not Easy to Determine

- Patient arrives at emergency
- Some cases are obvious (e.g., severe chest pain)
- In other cases, life-threatening blockages can have subtle symptoms:
  - a subtle squeezing sensation in the chest, shortness of breath, nausea
  - can be easily attributed to other causes (e.g., acid reflux, viral infection, pinched nerve in the back)
  - key decision: whether to conduct an imaging test for heart attack

The Decision: Test or No Test

Benefit of testing

- More information on condition → reduces false negatives (should treat but didn't)

Cost of testing

- Tests are expensive and risky to patients
  - thousands USD + monitoring/hospitalization
  - single highest dose of ionizing radiation (long-term cancer risks)
- Treatment, Cardiac Catheterization, also very costly and risky for patients
  - tens of thousands USD
  - very invasive, radiation + high risk of kidney failure, arterial damage, and stroke

Prediction and Judgment: Humans

Today, doctors receive information about a patient, make a prediction of the health risk in their heads, and then decide whether to test.

Prediction and Judgment: Humans

Risk Ventile	Yield (1
1	0.017
2	0.022
3	0.034
4	0.049
5	0.063
6	0.082
7	0.075
8	0.076
9	0.092
10	0.094
11	0.114
12	0.124
13	0.145
14	0.143
15	0.158
16	0.193
17	0.199
18	0.206
19	0.254
20	0.351

But there is a lot of waste

Researchers ranked patients into twenty bins with bin #1 being the lowest

They tested a random sample within each bin and Yield = probability of treatment among the tested

only 1.7% of those tested in the 1st bin needed treatment

Prediction and Judgment: Humans

Risk Ventile	Yield (SE) (1)		Cost, $ (2)
1	0.017	(0.003)	650,838
2	0.022	(0.003)	587,572
3	0.034	(0.004)	366,289
4	0.049	(0.005)	270,292
5	0.063	(0.005)	222,940
6	0.082	(0.006)	178,145
7	0.075	(0.006)	181,552
8	0.076	(0.006)	203,132
9	0.092	(0.007)	165,491
10	0.094	(0.007)	171,460
11	0.114	(0.007)	140,606
12	0.124	(0.008)	140,064
13	0.145	(0.008)	119,263
14	0.143	(0.008)	131,469
15	0.158	(0.008)	121,253
16	0.193	(0.009)	105,463
17	0.199	(0.009)	102,103
18	0.206	(0.009)	103,568
19	0.254	(0.010)	90,504
20	0.351	(0.011)	74,739

But there is a lot of waste

Researchers ranked patients into twenty bins with bin #1 being the lowest

They tested a random sample within each bin and Yield = probability of treatment among the tested

only 1.7% of those tested in the 1st bin needed treatment

cost = for each life year saved, how much was spent on testing

if you test 100 patients but only treat 1.7, then the cost per life-year saved is very high ($650K USD!) –> “low-value care”

Prediction and Judgment: Humans

Risk Ventile	Yield (SE) (1)		Cost, $ (2)	Test rate (SE) (3)	
1	0.017	(0.003)	650,838	0.015	(0.000)
2	0.022	(0.003)	587,572	0.024	(0.000)
3	0.034	(0.004)	366,289	0.030	(0.001)
4	0.049	(0.005)	270,292	0.036	(0.001)
5	0.063	(0.005)	222,940	0.042	(0.001)
6	0.082	(0.006)	178,145	0.043	(0.001)
7	0.075	(0.006)	181,552	0.048	(0.001)
8	0.076	(0.006)	203,132	0.048	(0.001)
9	0.092	(0.007)	165,491	0.052	(0.001)
10	0.094	(0.007)	171,460	0.053	(0.001)
11	0.114	(0.007)	140,606	0.056	(0.001)
12	0.124	(0.008)	140,064	0.061	(0.001)
13	0.145	(0.008)	119,263	0.064	(0.001)
14	0.143	(0.008)	131,469	0.064	(0.001)
15	0.158	(0.008)	121,253	0.070	(0.002)
16	0.193	(0.009)	105,463	0.075	(0.002)
17	0.199	(0.009)	102,103	0.079	(0.002)
18	0.206	(0.009)	103,568	0.089	(0.002)
19	0.254	(0.010)	90,504	0.103	(0.002)
20	0.351	(0.011)	74,739	0.127	(0.003)

Doctors' decisions to test are, overall, not bad: They tested 1.5% of patients in the first bin and 12.7% in the last bin. But it could have been a lot better.

Prediction and Judgment: Humans

Risk Ventile	Yield (SE) (1)		Cost, $ (2)	Test rate (SE) (3)	
1	0.017	(0.003)	650,838	0.015	(0.000)
2	0.022	(0.003)	587,572	0.024	(0.000)
3	0.034	(0.004)	366,289	0.030	(0.001)
4	0.049	(0.005)	270,292	0.036	(0.001)
5	0.063	(0.005)	222,940	0.042	(0.001)
6	0.082	(0.006)	178,145	0.043	(0.001)
7	0.075	(0.006)	181,552	0.048	(0.001)
8	0.076	(0.006)	203,132	0.048	(0.001)
9	0.092	(0.007)	165,491	0.052	(0.001)
10	0.094	(0.007)	171,460	0.053	(0.001)
11	0.114	(0.007)	140,606	0.056	(0.001)
12	0.124	(0.008)	140,064	0.061	(0.001)
13	0.145	(0.008)	119,263	0.064	(0.001)
14	0.143	(0.008)	131,469	0.064	(0.001)
15	0.158	(0.008)	121,253	0.070	(0.002)
16	0.193	(0.009)	105,463	0.075	(0.002)
17	0.199	(0.009)	102,103	0.079	(0.002)
18	0.206	(0.009)	103,568	0.089	(0.002)
19	0.254	(0.010)	90,504	0.103	(0.002)
20	0.351	(0.011)	74,739	0.127	(0.003)

Doctors' decisions to test are, overall, not bad: They tested 1.5% of patients in the first bin and 12.7% in the last bin. But it could have been a lot better.

The highest dollar amount any society puts on saving a life-year is about USD 150K. This means that we should test all patients below the line and not test any patient above the line

53% of existing tests caused more harm than benefit; while many more tests should have been conducted but didn't.

Source: https://www.nber.org/system/files/working_papers/w26168/revisions/w26168.rev1.pdf?mod=article_inline *("A Machine Learning Approach to Low Value Health Care: Wasted Tests, Missed Heart Attacks, and Mis-Predictions" by Sendhil Mullainathan and Ziad Obermeyer, 2019)*

System Thinking in AI and Heart Attack Testing

Physicians have poor incentives. They receive some private benefit from doing more tests, in the form of extra revenue or protection from malpractice risk. Also, they are not that accurate at predicting risk. Moral hazard

- Task level: AIs could help reduce low value tests and increase missed high value tests.
- System level: if AI makes a prediction that doctors, hospitals, insurers, and patients can see, rather than a system where the prediction is private in the doctor's mind. Then, the doctor must submit an explanation if they decide to administer a test when the AI predicts the risk is below a threshold.