Legend
- **All features** (red square)
- **Arguably causal features** (gray diamond)
- **Causal features** (light blue circle)
- **Constant** (dark cross)
- **Same performance** (dashed line)

Left panel — y-axis: Out-of-domain accuracy; x-axis: Tasks

Right panel — y-axis: Shift gap (higher is better); x-axis: Tasks

Tasks (left to right): ASSISTments, Hospital Readmission, Hypertension, ICU Length of Stay, Utilization, PublicCoverage, Poverty, Income, Food Stamps, Diabetes, Voting, College Scorecard, Hospital Mortality, Childhood Lead, Sepsis, Unemployment