

# **Supplementary Material: Robot Causal Discovery Aided by Human Interaction**

Filip Edström, Thomas Hellström, Xavier de Luna

March 24, 2023

## **Contents**

|          |                |          |
|----------|----------------|----------|
| <b>1</b> | <b>Tables</b>  | <b>2</b> |
| <b>2</b> | <b>Figures</b> | <b>7</b> |

# 1 Tables

Table 1: Assoc and HC versions from both starting points for all DAGs.  $N$  is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and  $\Delta$  is the difference in Relative SHD-Em from Finish to Start.

| DAG    | $N$  | Version    | Start $\pm$ CI   | Finish $\pm$ CI  | $\Delta$ |
|--------|------|------------|------------------|------------------|----------|
| Asia   | 500  | MMHC Assoc | $0.46 \pm 0.007$ | $0.10 \pm 0.006$ | -0.36    |
| Asia   | 500  | MMHC HC    | $0.46 \pm 0.007$ | $0.00 \pm 0.001$ | -0.46    |
| Asia   | 500  | PC Assoc   | $0.73 \pm 0.008$ | $0.18 \pm 0.012$ | -0.55    |
| Asia   | 500  | PC HC      | $0.73 \pm 0.008$ | $0.01 \pm 0.002$ | -0.72    |
| Asia   | 1000 | MMHC Assoc | $0.42 \pm 0.007$ | $0.05 \pm 0.006$ | -0.36    |
| Asia   | 1000 | MMHC HC    | $0.42 \pm 0.007$ | $0.00 \pm 0.000$ | -0.42    |
| Asia   | 1000 | PC Assoc   | $0.68 \pm 0.008$ | $0.18 \pm 0.012$ | -0.50    |
| Asia   | 1000 | PC HC      | $0.68 \pm 0.008$ | $0.00 \pm 0.001$ | -0.67    |
| Asia   | 1500 | MMHC Assoc | $0.39 \pm 0.006$ | $0.03 \pm 0.005$ | -0.36    |
| Asia   | 1500 | MMHC HC    | $0.39 \pm 0.006$ | $0.00 \pm 0.000$ | -0.39    |
| Asia   | 1500 | PC Assoc   | $0.63 \pm 0.007$ | $0.14 \pm 0.012$ | -0.49    |
| Asia   | 1500 | PC HC      | $0.63 \pm 0.007$ | $0.00 \pm 0.001$ | -0.63    |
| Asia   | 3000 | MMHC Assoc | $0.37 \pm 0.004$ | $0.01 \pm 0.003$ | -0.36    |
| Asia   | 3000 | MMHC HC    | $0.37 \pm 0.004$ | $0.00 \pm 0.000$ | -0.37    |
| Asia   | 3000 | PC Assoc   | $0.59 \pm 0.005$ | $0.07 \pm 0.010$ | -0.52    |
| Asia   | 3000 | PC HC      | $0.59 \pm 0.005$ | $0.00 \pm 0.000$ | -0.59    |
| Sachs  | 500  | MMHC Assoc | $0.85 \pm 0.005$ | $0.54 \pm 0.009$ | -0.32    |
| Sachs  | 500  | MMHC HC    | $0.85 \pm 0.005$ | $0.15 \pm 0.003$ | -0.70    |
| Sachs  | 500  | PC Assoc   | $0.84 \pm 0.004$ | $0.23 \pm 0.009$ | -0.62    |
| Sachs  | 500  | PC HC      | $0.84 \pm 0.004$ | $0.13 \pm 0.003$ | -0.71    |
| Sachs  | 1000 | MMHC Assoc | $0.86 \pm 0.004$ | $0.65 \pm 0.006$ | -0.21    |
| Sachs  | 1000 | MMHC HC    | $0.86 \pm 0.004$ | $0.09 \pm 0.004$ | -0.77    |
| Sachs  | 1000 | PC Assoc   | $0.74 \pm 0.004$ | $0.44 \pm 0.011$ | -0.29    |
| Sachs  | 1000 | PC HC      | $0.74 \pm 0.004$ | $0.06 \pm 0.002$ | -0.67    |
| Sachs  | 1500 | MMHC Assoc | $0.86 \pm 0.003$ | $0.68 \pm 0.005$ | -0.18    |
| Sachs  | 1500 | MMHC HC    | $0.86 \pm 0.003$ | $0.02 \pm 0.004$ | -0.84    |
| Sachs  | 1500 | PC Assoc   | $0.80 \pm 0.004$ | $0.69 \pm 0.009$ | -0.11    |
| Sachs  | 1500 | PC HC      | $0.80 \pm 0.004$ | $0.09 \pm 0.005$ | -0.71    |
| Sachs  | 3000 | MMHC Assoc | $0.85 \pm 0.004$ | $0.69 \pm 0.004$ | -0.16    |
| Sachs  | 3000 | MMHC HC    | $0.85 \pm 0.004$ | $0.06 \pm 0.003$ | -0.79    |
| Sachs  | 3000 | PC Assoc   | $0.74 \pm 0.003$ | $0.61 \pm 0.008$ | -0.13    |
| Sachs  | 3000 | PC HC      | $0.74 \pm 0.003$ | $0.06 \pm 0.004$ | -0.68    |
| Survey | 500  | MMHC Assoc | $0.73 \pm 0.012$ | $0.00 \pm 0.000$ | -0.73    |
| Survey | 500  | MMHC HC    | $0.73 \pm 0.012$ | $0.00 \pm 0.000$ | -0.73    |
| Survey | 500  | PC Assoc   | $0.81 \pm 0.011$ | $0.00 \pm 0.000$ | -0.81    |
| Survey | 500  | PC HC      | $0.81 \pm 0.011$ | $0.00 \pm 0.000$ | -0.81    |
| Survey | 1000 | MMHC Assoc | $0.59 \pm 0.010$ | $0.00 \pm 0.000$ | -0.59    |

Table 1: Assoc and HC versions from both starting points for all DAGs.  $N$  is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and  $\Delta$  is the difference in Relative SHD-Em from Finish to Start.

| DAG       | $N$  | Version    | Start $\pm$ CI   | Finish $\pm$ CI  | $\Delta$ |
|-----------|------|------------|------------------|------------------|----------|
| Survey    | 1000 | MMHC HC    | $0.59 \pm 0.010$ | $0.00 \pm 0.000$ | -0.59    |
| Survey    | 1000 | PC Assoc   | $0.74 \pm 0.012$ | $0.00 \pm 0.000$ | -0.74    |
| Survey    | 1000 | PC HC      | $0.74 \pm 0.012$ | $0.00 \pm 0.000$ | -0.74    |
| Survey    | 1500 | MMHC Assoc | $0.54 \pm 0.009$ | $0.00 \pm 0.000$ | -0.54    |
| Survey    | 1500 | MMHC HC    | $0.54 \pm 0.009$ | $0.00 \pm 0.000$ | -0.54    |
| Survey    | 1500 | PC Assoc   | $0.69 \pm 0.013$ | $0.00 \pm 0.000$ | -0.69    |
| Survey    | 1500 | PC HC      | $0.69 \pm 0.013$ | $0.00 \pm 0.000$ | -0.69    |
| Survey    | 3000 | MMHC Assoc | $0.46 \pm 0.010$ | $0.00 \pm 0.000$ | -0.46    |
| Survey    | 3000 | MMHC HC    | $0.46 \pm 0.010$ | $0.00 \pm 0.000$ | -0.46    |
| Survey    | 3000 | PC Assoc   | $0.57 \pm 0.015$ | $0.00 \pm 0.000$ | -0.57    |
| Survey    | 3000 | PC HC      | $0.57 \pm 0.015$ | $0.00 \pm 0.000$ | -0.57    |
| Child     | 500  | MMHC Assoc | $0.60 \pm 0.004$ | $0.44 \pm 0.004$ | -0.16    |
| Child     | 500  | MMHC HC    | $0.60 \pm 0.004$ | $0.21 \pm 0.004$ | -0.39    |
| Child     | 500  | PC Assoc   | $0.75 \pm 0.004$ | $0.45 \pm 0.005$ | -0.29    |
| Child     | 500  | PC HC      | $0.75 \pm 0.004$ | $0.17 \pm 0.003$ | -0.57    |
| Child     | 1000 | MMHC Assoc | $0.53 \pm 0.004$ | $0.40 \pm 0.004$ | -0.13    |
| Child     | 1000 | MMHC HC    | $0.53 \pm 0.004$ | $0.12 \pm 0.004$ | -0.41    |
| Child     | 1000 | PC Assoc   | $0.66 \pm 0.005$ | $0.45 \pm 0.008$ | -0.21    |
| Child     | 1000 | PC HC      | $0.66 \pm 0.005$ | $0.11 \pm 0.004$ | -0.55    |
| Child     | 1500 | MMHC Assoc | $0.51 \pm 0.004$ | $0.41 \pm 0.006$ | -0.10    |
| Child     | 1500 | MMHC HC    | $0.51 \pm 0.004$ | $0.12 \pm 0.006$ | -0.39    |
| Child     | 1500 | PC Assoc   | $0.62 \pm 0.004$ | $0.48 \pm 0.004$ | -0.14    |
| Child     | 1500 | PC HC      | $0.62 \pm 0.004$ | $0.06 \pm 0.003$ | -0.57    |
| Child     | 3000 | MMHC Assoc | $0.46 \pm 0.004$ | $0.43 \pm 0.004$ | -0.04    |
| Child     | 3000 | MMHC HC    | $0.46 \pm 0.004$ | $0.12 \pm 0.009$ | -0.34    |
| Child     | 3000 | PC Assoc   | $0.59 \pm 0.006$ | $0.46 \pm 0.005$ | -0.13    |
| Child     | 3000 | PC HC      | $0.59 \pm 0.006$ | $0.01 \pm 0.002$ | -0.58    |
| Insurance | 500  | MMHC Assoc | $0.81 \pm 0.003$ | $0.63 \pm 0.003$ | -0.18    |
| Insurance | 500  | MMHC HC    | $0.81 \pm 0.003$ | $0.61 \pm 0.003$ | -0.20    |
| Insurance | 500  | PC Assoc   | $0.85 \pm 0.003$ | $0.63 \pm 0.003$ | -0.22    |
| Insurance | 500  | PC HC      | $0.85 \pm 0.003$ | $0.50 \pm 0.003$ | -0.34    |
| Insurance | 1000 | MMHC Assoc | $0.75 \pm 0.003$ | $0.58 \pm 0.003$ | -0.17    |
| Insurance | 1000 | MMHC HC    | $0.75 \pm 0.003$ | $0.53 \pm 0.004$ | -0.22    |
| Insurance | 1000 | PC Assoc   | $0.78 \pm 0.002$ | $0.57 \pm 0.002$ | -0.21    |
| Insurance | 1000 | PC HC      | $0.78 \pm 0.002$ | $0.42 \pm 0.004$ | -0.36    |
| Insurance | 1500 | MMHC Assoc | $0.71 \pm 0.003$ | $0.58 \pm 0.003$ | -0.13    |
| Insurance | 1500 | MMHC HC    | $0.71 \pm 0.003$ | $0.48 \pm 0.004$ | -0.23    |
| Insurance | 1500 | PC Assoc   | $0.74 \pm 0.004$ | $0.55 \pm 0.003$ | -0.19    |
| Insurance | 1500 | PC HC      | $0.74 \pm 0.004$ | $0.42 \pm 0.005$ | -0.33    |
| Insurance | 3000 | MMHC Assoc | $0.71 \pm 0.003$ | $0.59 \pm 0.002$ | -0.11    |

Table 1: Assoc and HC versions from both starting points for all DAGs.  $N$  is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and  $\Delta$  is the difference in Relative SHD-Em from Finish to Start.

| DAG       | $N$  | Version    | Start $\pm$ CI   | Finish $\pm$ CI  | $\Delta$ |
|-----------|------|------------|------------------|------------------|----------|
| Insurance | 3000 | MMHC HC    | $0.71 \pm 0.003$ | $0.50 \pm 0.003$ | -0.20    |
| Insurance | 3000 | PC Assoc   | $0.67 \pm 0.002$ | $0.49 \pm 0.002$ | -0.19    |
| Insurance | 3000 | PC HC      | $0.67 \pm 0.002$ | $0.28 \pm 0.003$ | -0.39    |
| Mildew    | 500  | MMHC Assoc | $0.91 \pm 0.001$ | $0.74 \pm 0.002$ | -0.17    |
| Mildew    | 500  | MMHC HC    | $0.91 \pm 0.001$ | $0.72 \pm 0.001$ | -0.19    |
| Mildew    | 500  | PC Assoc   | $0.88 \pm 0.002$ | $0.71 \pm 0.002$ | -0.18    |
| Mildew    | 500  | PC HC      | $0.88 \pm 0.002$ | $0.72 \pm 0.002$ | -0.16    |
| Mildew    | 1000 | MMHC Assoc | $0.85 \pm 0.000$ | $0.78 \pm 0.001$ | -0.07    |
| Mildew    | 1000 | MMHC HC    | $0.85 \pm 0.000$ | $0.68 \pm 0.001$ | -0.17    |
| Mildew    | 1000 | PC Assoc   | $0.87 \pm 0.001$ | $0.79 \pm 0.001$ | -0.09    |
| Mildew    | 1000 | PC HC      | $0.87 \pm 0.001$ | $0.63 \pm 0.001$ | -0.24    |
| Mildew    | 1500 | MMHC Assoc | $0.87 \pm 0.001$ | $0.82 \pm 0.001$ | -0.05    |
| Mildew    | 1500 | MMHC HC    | $0.87 \pm 0.001$ | $0.67 \pm 0.000$ | -0.19    |
| Mildew    | 1500 | PC Assoc   | $0.87 \pm 0.001$ | $0.78 \pm 0.001$ | -0.09    |
| Mildew    | 1500 | PC HC      | $0.87 \pm 0.001$ | $0.61 \pm 0.000$ | -0.27    |
| Mildew    | 3000 | MMHC Assoc | $0.87 \pm 0.001$ | $0.80 \pm 0.001$ | -0.07    |
| Mildew    | 3000 | MMHC HC    | $0.87 \pm 0.001$ | $0.65 \pm 0.001$ | -0.22    |
| Mildew    | 3000 | PC Assoc   | $0.87 \pm 0.001$ | $0.80 \pm 0.001$ | -0.07    |
| Mildew    | 3000 | PC HC      | $0.87 \pm 0.001$ | $0.57 \pm 0.001$ | -0.30    |
| Mildew    | 5000 | MMHC Assoc | $0.87 \pm 0.000$ | $0.85 \pm 0.000$ | -0.02    |
| Mildew    | 5000 | MMHC HC    | $0.87 \pm 0.000$ | $0.67 \pm 0.000$ | -0.20    |
| Mildew    | 5000 | PC Assoc   | $0.84 \pm 0.000$ | $0.79 \pm 0.000$ | -0.04    |
| Mildew    | 5000 | PC HC      | $0.84 \pm 0.000$ | $0.56 \pm 0.000$ | -0.27    |
| Alarm     | 500  | MMHC Assoc | $0.69 \pm 0.003$ | $0.63 \pm 0.003$ | -0.06    |
| Alarm     | 500  | MMHC HC    | $0.69 \pm 0.003$ | $0.43 \pm 0.006$ | -0.25    |
| Alarm     | 500  | PC Assoc   | $0.82 \pm 0.003$ | $0.71 \pm 0.003$ | -0.12    |
| Alarm     | 500  | PC HC      | $0.82 \pm 0.003$ | $0.46 \pm 0.004$ | -0.37    |
| Alarm     | 1000 | MMHC Assoc | $0.56 \pm 0.003$ | $0.52 \pm 0.003$ | -0.05    |
| Alarm     | 1000 | MMHC HC    | $0.56 \pm 0.003$ | $0.26 \pm 0.006$ | -0.31    |
| Alarm     | 1000 | PC Assoc   | $0.70 \pm 0.004$ | $0.63 \pm 0.004$ | -0.07    |
| Alarm     | 1000 | PC HC      | $0.70 \pm 0.004$ | $0.33 \pm 0.006$ | -0.37    |
| Alarm     | 1500 | MMHC Assoc | $0.51 \pm 0.003$ | $0.47 \pm 0.002$ | -0.04    |
| Alarm     | 1500 | MMHC HC    | $0.51 \pm 0.003$ | $0.17 \pm 0.004$ | -0.34    |
| Alarm     | 1500 | PC Assoc   | $0.63 \pm 0.004$ | $0.58 \pm 0.004$ | -0.05    |
| Alarm     | 1500 | PC HC      | $0.63 \pm 0.004$ | $0.31 \pm 0.005$ | -0.32    |
| Alarm     | 3000 | MMHC Assoc | $0.48 \pm 0.003$ | $0.47 \pm 0.003$ | -0.01    |
| Alarm     | 3000 | MMHC HC    | $0.48 \pm 0.003$ | $0.28 \pm 0.004$ | -0.20    |
| Alarm     | 3000 | PC Assoc   | $0.49 \pm 0.006$ | $0.47 \pm 0.006$ | -0.02    |
| Alarm     | 3000 | PC HC      | $0.49 \pm 0.006$ | $0.12 \pm 0.005$ | -0.37    |
| Alarm     | 5000 | MMHC Assoc | $0.40 \pm 0.003$ | $0.40 \pm 0.003$ | -0.00    |

Table 1: Assoc and HC versions from both starting points for all DAGs.  $N$  is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and  $\Delta$  is the difference in Relative SHD-Em from Finish to Start.

| DAG        | $N$  | Version    | Start $\pm$ CI   | Finish $\pm$ CI  | $\Delta$ |
|------------|------|------------|------------------|------------------|----------|
| Alarm      | 5000 | MMHC HC    | $0.40 \pm 0.003$ | $0.14 \pm 0.004$ | -0.26    |
| Alarm      | 5000 | PC Assoc   | $0.44 \pm 0.005$ | $0.43 \pm 0.004$ | -0.01    |
| Alarm      | 5000 | PC HC      | $0.44 \pm 0.005$ | $0.11 \pm 0.007$ | -0.32    |
| Barley     | 500  | MMHC Assoc | $0.88 \pm 0.001$ | $0.77 \pm 0.001$ | -0.11    |
| Barley     | 500  | MMHC HC    | $0.88 \pm 0.001$ | $0.76 \pm 0.001$ | -0.12    |
| Barley     | 500  | PC Assoc   | $0.99 \pm 0.001$ | $0.85 \pm 0.001$ | -0.14    |
| Barley     | 500  | PC HC      | $0.99 \pm 0.001$ | $0.77 \pm 0.001$ | -0.21    |
| Barley     | 1000 | MMHC Assoc | $0.85 \pm 0.001$ | $0.77 \pm 0.001$ | -0.08    |
| Barley     | 1000 | MMHC HC    | $0.85 \pm 0.001$ | $0.69 \pm 0.001$ | -0.16    |
| Barley     | 1000 | PC Assoc   | $0.97 \pm 0.001$ | $0.87 \pm 0.001$ | -0.10    |
| Barley     | 1000 | PC HC      | $0.97 \pm 0.001$ | $0.77 \pm 0.001$ | -0.20    |
| Barley     | 1500 | MMHC Assoc | $0.84 \pm 0.001$ | $0.78 \pm 0.001$ | -0.07    |
| Barley     | 1500 | MMHC HC    | $0.84 \pm 0.001$ | $0.69 \pm 0.001$ | -0.15    |
| Barley     | 1500 | PC Assoc   | $0.96 \pm 0.001$ | $0.88 \pm 0.001$ | -0.08    |
| Barley     | 1500 | PC HC      | $0.96 \pm 0.001$ | $0.77 \pm 0.001$ | -0.19    |
| Barley     | 3000 | MMHC Assoc | $0.78 \pm 0.001$ | $0.73 \pm 0.001$ | -0.05    |
| Barley     | 3000 | MMHC HC    | $0.78 \pm 0.001$ | $0.58 \pm 0.002$ | -0.21    |
| Barley     | 3000 | PC Assoc   | $0.86 \pm 0.002$ | $0.80 \pm 0.002$ | -0.06    |
| Barley     | 3000 | PC HC      | $0.86 \pm 0.002$ | $0.67 \pm 0.002$ | -0.19    |
| Barley     | 5000 | MMHC Assoc | $0.74 \pm 0.001$ | $0.70 \pm 0.001$ | -0.04    |
| Barley     | 5000 | MMHC HC    | $0.74 \pm 0.001$ | $0.51 \pm 0.002$ | -0.23    |
| Barley     | 5000 | PC Assoc   | $0.86 \pm 0.001$ | $0.79 \pm 0.001$ | -0.07    |
| Barley     | 5000 | PC HC      | $0.86 \pm 0.001$ | $0.66 \pm 0.001$ | -0.20    |
| Hailfinder | 500  | MMHC Assoc | $0.65 \pm 0.001$ | $0.41 \pm 0.002$ | -0.24    |
| Hailfinder | 500  | MMHC HC    | $0.65 \pm 0.001$ | $0.37 \pm 0.002$ | -0.28    |
| Hailfinder | 500  | PC Assoc   | $0.71 \pm 0.002$ | $0.53 \pm 0.003$ | -0.19    |
| Hailfinder | 500  | PC HC      | $0.71 \pm 0.002$ | $0.41 \pm 0.002$ | -0.30    |
| Hailfinder | 1000 | MMHC Assoc | $0.60 \pm 0.001$ | $0.50 \pm 0.001$ | -0.10    |
| Hailfinder | 1000 | MMHC HC    | $0.60 \pm 0.001$ | $0.30 \pm 0.002$ | -0.30    |
| Hailfinder | 1000 | PC Assoc   | $0.65 \pm 0.002$ | $0.53 \pm 0.002$ | -0.12    |
| Hailfinder | 1000 | PC HC      | $0.65 \pm 0.002$ | $0.32 \pm 0.002$ | -0.33    |
| Hailfinder | 1500 | MMHC Assoc | $0.56 \pm 0.001$ | $0.47 \pm 0.001$ | -0.10    |
| Hailfinder | 1500 | MMHC HC    | $0.56 \pm 0.001$ | $0.25 \pm 0.002$ | -0.32    |
| Hailfinder | 1500 | PC Assoc   | $0.63 \pm 0.003$ | $0.51 \pm 0.002$ | -0.12    |
| Hailfinder | 1500 | PC HC      | $0.63 \pm 0.003$ | $0.30 \pm 0.003$ | -0.33    |
| Hailfinder | 3000 | MMHC Assoc | $0.54 \pm 0.001$ | $0.41 \pm 0.001$ | -0.13    |
| Hailfinder | 3000 | MMHC HC    | $0.54 \pm 0.001$ | $0.19 \pm 0.002$ | -0.35    |
| Hailfinder | 3000 | PC Assoc   | $0.61 \pm 0.003$ | $0.48 \pm 0.003$ | -0.13    |
| Hailfinder | 3000 | PC HC      | $0.61 \pm 0.003$ | $0.26 \pm 0.003$ | -0.35    |
| Hailfinder | 5000 | MMHC Assoc | $0.52 \pm 0.001$ | $0.41 \pm 0.001$ | -0.12    |

Table 1: Assoc and HC versions from both starting points for all DAGs.  $N$  is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and  $\Delta$  is the difference in Relative SHD-Em from Finish to Start.

| DAG        | $N$  | Version  | Start $\pm$ CI   | Finish $\pm$ CI  | $\Delta$ |
|------------|------|----------|------------------|------------------|----------|
| Hailfinder | 5000 | MMHC HC  | $0.52 \pm 0.001$ | $0.15 \pm 0.001$ | -0.37    |
| Hailfinder | 5000 | PC Assoc | $0.61 \pm 0.003$ | $0.48 \pm 0.003$ | -0.12    |
| Hailfinder | 5000 | PC HC    | $0.61 \pm 0.003$ | $0.25 \pm 0.003$ | -0.35    |

## 2 Figures

OptSingle sometimes doesn't ask any questions, e.g. Figure 1. This is due to not all 500 starting PDAGs having an undirected edge.

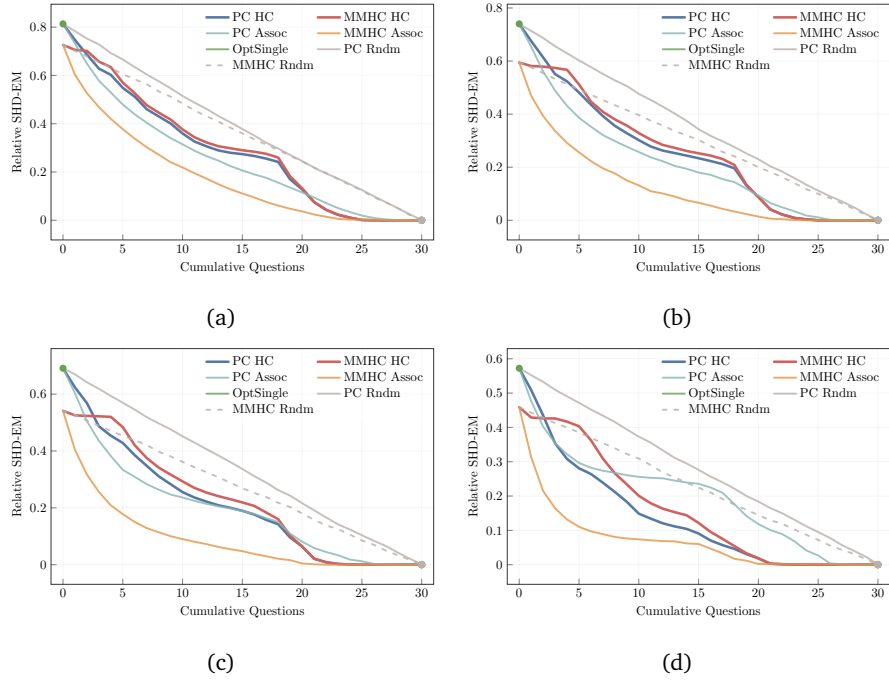


Figure 1: SHD-EM averaged over 500 replicates on the Survey DAG (6 nodes, 6 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ .

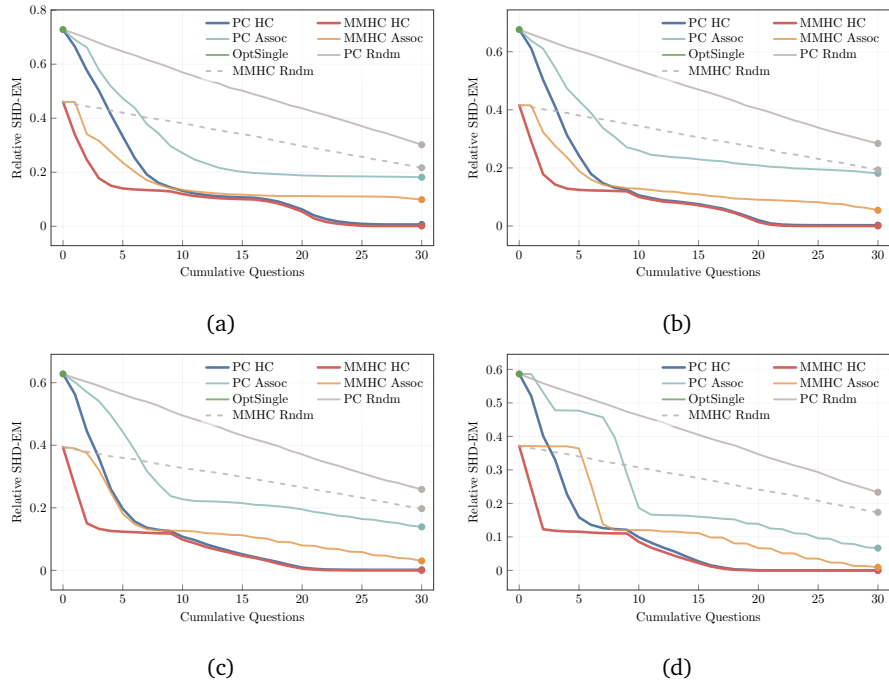


Figure 2: SHD-EM averaged over 500 replicates on the Asia DAG (8 nodes, 8 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ .



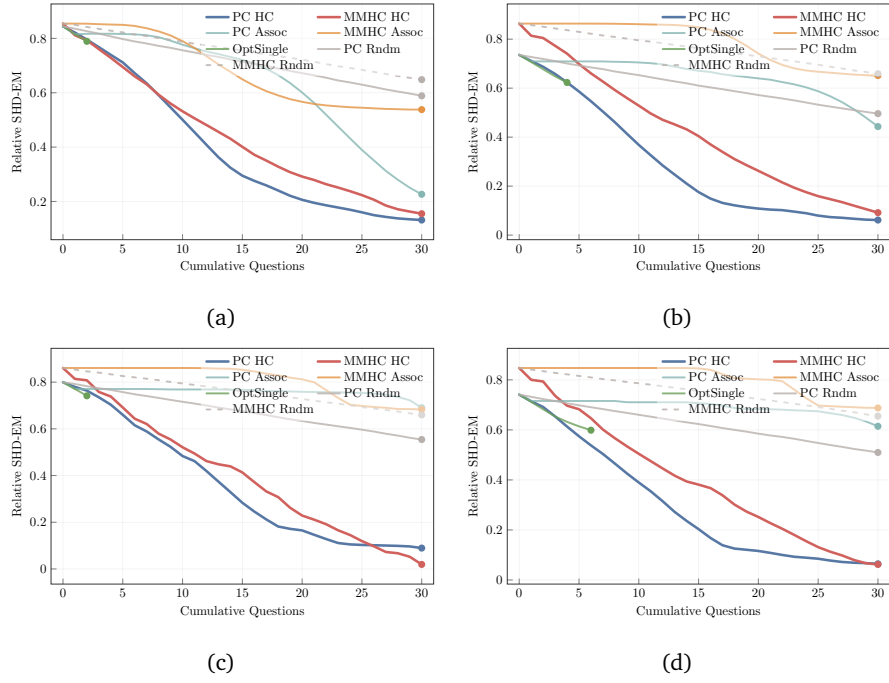


Figure 3: SHD-EM averaged over 500 replicates on the Sachs DAG (11 nodes, 17 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ .

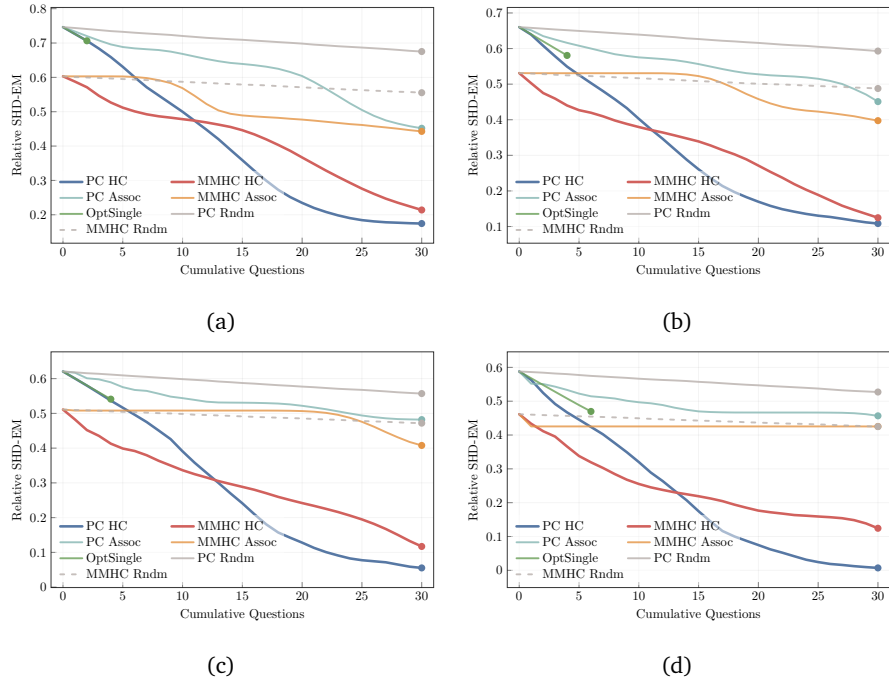


Figure 4: SHD-EM averaged over 500 replicates on the Child DAG (20 nodes, 25 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ .

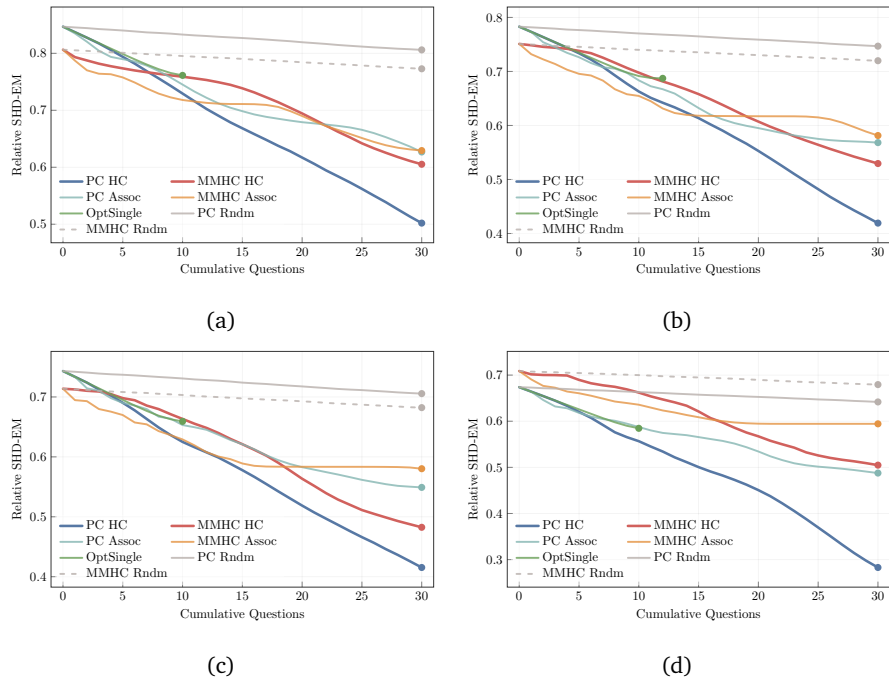


Figure 5: SHD-EM averaged over 500 replicates on the Insurance DAG (27 nodes, 52 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ .

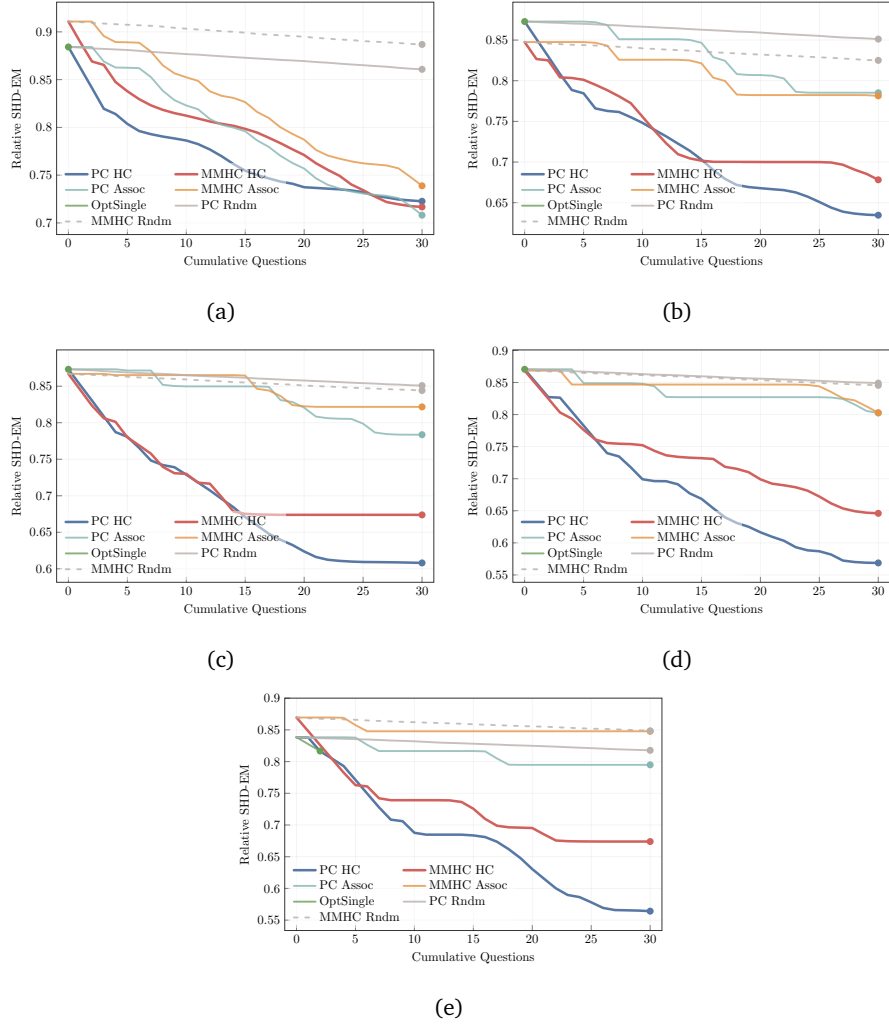


Figure 6: SHD-EM averaged over 500 replicates on the Mildew DAG (35 nodes, 46 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ , (e)  $N = 5000$ .

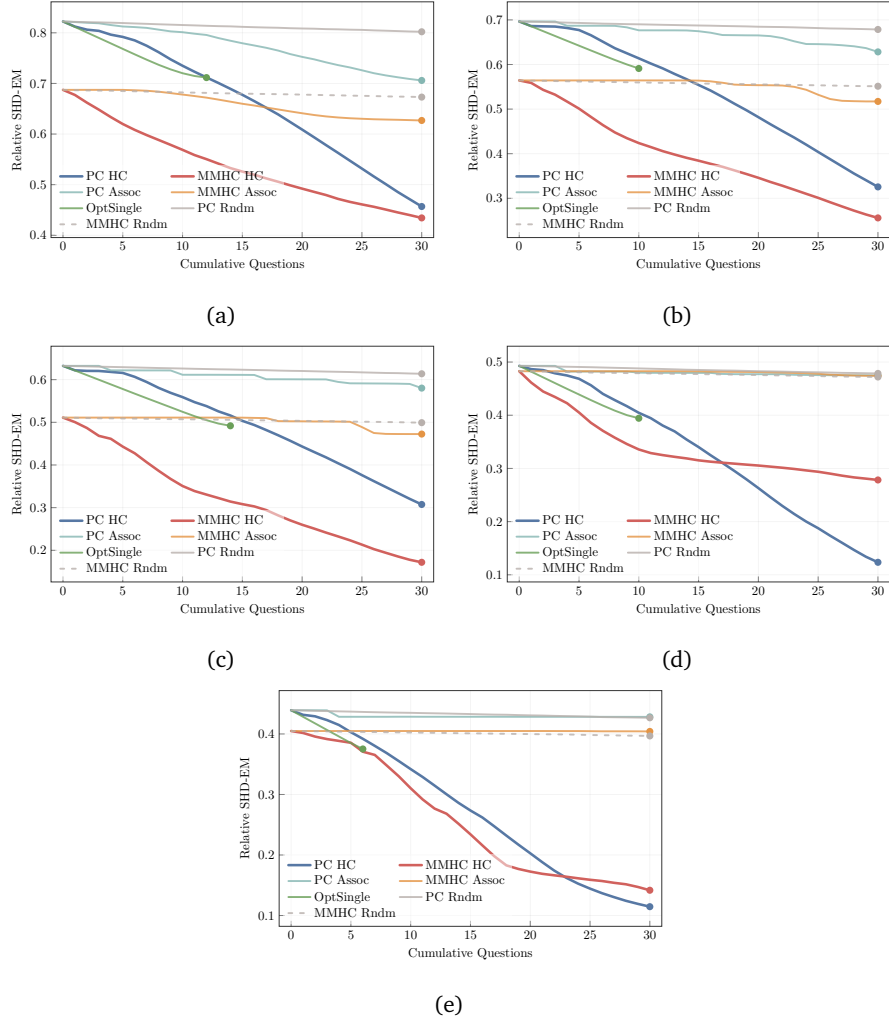


Figure 7: SHD-EM averaged over 500 replicates on the Alarm DAG, (37 nodes, 46 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ , (e)  $N = 5000$ .

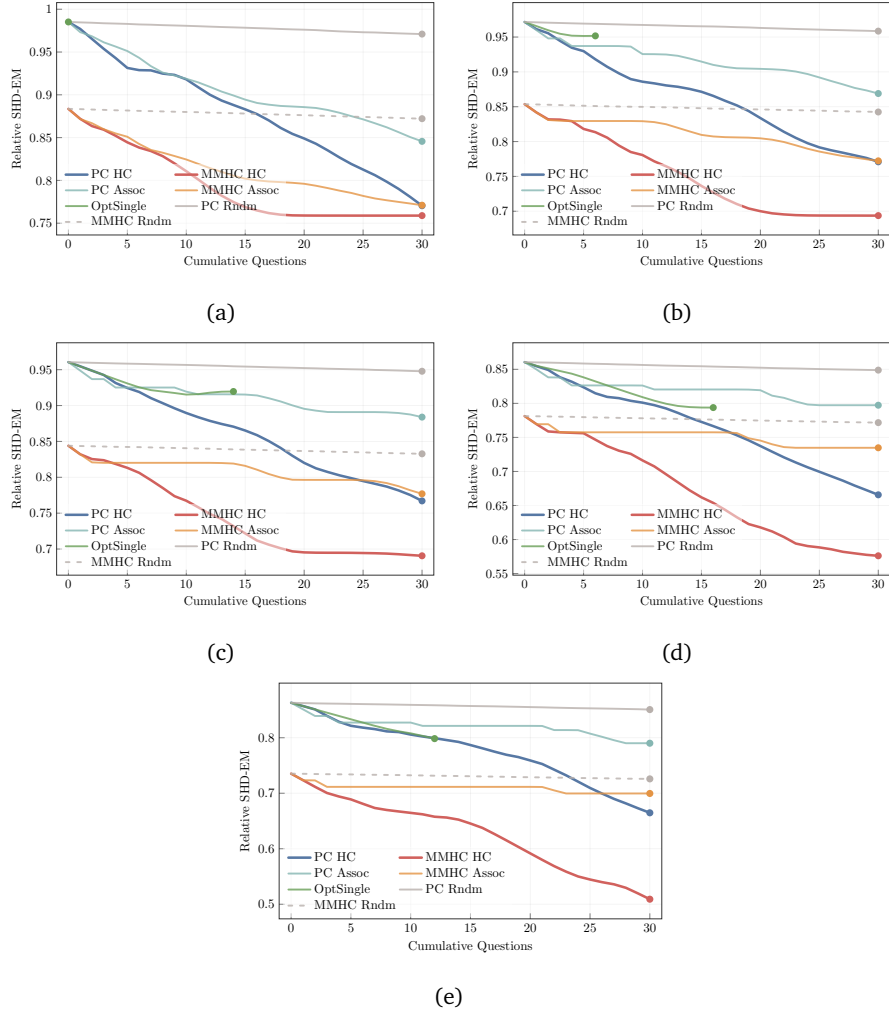


Figure 8: SHD-EM averaged over 500 replicates on the Barley DAG (48 nodes, 84 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ , (e)  $N = 5000$ .

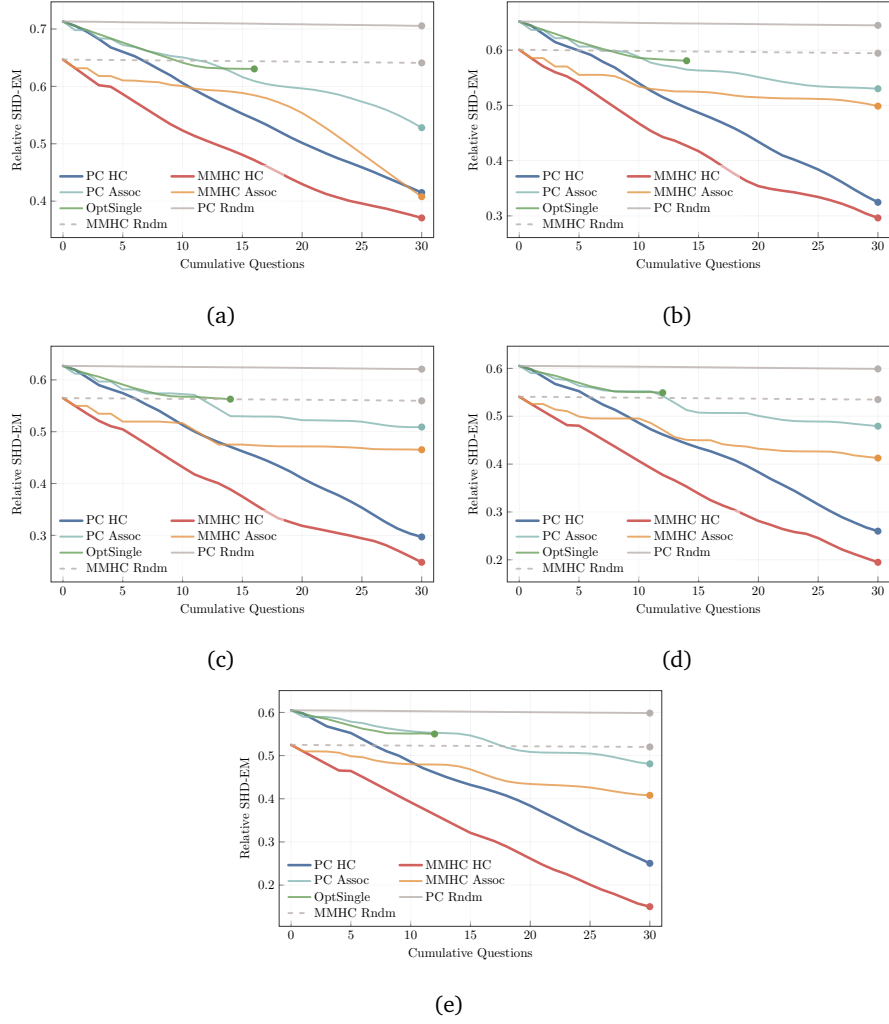


Figure 9: SHD-EM averaged over 500 replicates on the Hailfinder DAG (56 nodes, 66 edges) for sample sizes (a)  $N = 500$ , (b)  $N = 1000$ , (c)  $N = 1500$ , (d)  $N = 3000$ , (e)  $N = 5000$ .