

Supplementary Materials for “Bayesian Active Questionnaire Design for Cause-of-Death Assignment Using Verbal Autopsies”

Toshiya Yoshida

University of California Santa Cruz, USA

TOYOSHID@UCSC.EDU

Trinity Shuxian Fan

University of Washington, USA

FANSX@UW.EDU

Tyler McCormick

University of Washington, USA

TYLERMC@UW.EDU

Zhenke Wu

University of Michigan, USA

ZHENKEWU@UMICH.EDU

Zehang Richard Li

University of California Santa Cruz, USA

LIZEHANG@UCSC.EDU

1. Additional results for the synthetic data examples

Tables 1 and 2 contain more extensive results on the classification accuracy and questionnaire length for the synthetic data under the correctly specified model and misspecified models, respectively.

2. Additional results for the PHMRC data example

2.1. Descriptive statistics of the PHMRC dataset

Table 3 summarizes the sample size of each cause of death in the PHMRC data.

2.2. More results of different stopping rules

Figures 1 to 5 show more detailed results for the proportion of correctly classified deaths among deaths due to each cause of death using different stopping criteria.

Table 1: Classification accuracy and questionnaire length for the synthetic data under the correctly specified model. Median and 5th and 95th percentiles of the questionnaire length are shown for each stopping rule conditions.

p_{1st}	d	Stopping Rule	Acc	Median	Lower	Upper
0.8	0.75	Point Est	0.84	5	3	14
		Pred $r = 0.5$	0.92	10	4	50
		Pred $r = 0.7$	0.95	13	6	50
0.8	0.5	Point Est	0.85	5	3	14
		Pred $r = 0.5$	0.92	10	4	50
		Pred $r = 0.7$	0.95	14	6	50
0.8	0.25	Point Est	0.88	6	3	16
		Pred $r = 0.5$	0.94	10	4	50
		Pred $r = 0.7$	0.96	15	6	50
0.8	0	Point Est	0.96	8	5	35
		Pred $r = 0.5$	0.96	13	6	50
		Pred $r = 0.7$	0.96	20	7	50
0.9	0.75	Point Est	0.95	7	5	17
		Pred $r = 0.5$	0.95	12	6	50
		Pred $r = 0.7$	0.96	17	7	50
0.9	0.5	Point Est	0.95	7	5	18
		Pred $r = 0.5$	0.95	12	6	50
		Pred $r = 0.7$	0.96	18	7	50
0.9	0.25	Point Est	0.95	7	5	24
		Pred $r = 0.5$	0.96	12	6	50
		Pred $r = 0.7$	0.96	18	7	50
0.9	0	Point Est	0.97	10	7	50
		Pred $r = 0.5$	0.96	16	8	50
		Pred $r = 0.7$	0.96	23	8	50
0.95	0.75	Point Est	0.97	9	6	24
		Pred $r = 0.5$	0.96	13	7	50
		Pred $r = 0.7$	0.96	18	8	50
0.95	0.5	Point Est	0.97	9	6	35
		Pred $r = 0.5$	0.96	13	7	50
		Pred $r = 0.7$	0.96	20	8	50
0.95	0.25	Point Est	0.97	10	6	50
		Pred $r = 0.5$	0.96	15	7	50
		Pred $r = 0.7$	0.97	22	8	50
0.95	0	Point Est	0.98	12	8	50
		Pred $r = 0.5$	0.96	20	8	50
		Pred $r = 0.7$	0.97	26	9	50

Table 2: Classification accuracy and questionnaire length for the synthetic data under the misspecified model. Median and 5th and 95th percentiles of the questionnaire length are shown for each stopping rule conditions.

p_{1st}	d	Stopping Rule	Acc	Median	Lower	Upper
0.8	0.75	Point Est	0.86	4	3	11
		Pred $r = 0.5$	0.88	4	3	18
		Pred $r = 0.7$	0.94	5	3	25
0.8	0.5	Point Est	0.88	5	3	11
		Pred $r = 0.5$	0.88	4	3	18
		Pred $r = 0.7$	0.94	5	3	27
0.8	0.25	Point Est	0.90	5	3	11
		Pred $r = 0.5$	0.90	4	3	22
		Pred $r = 0.7$	0.94	6	3	30
0.8	0	Point Est	0.96	6	5	17
		Pred $r = 0.5$	0.94	6	4	30
		Pred $r = 0.7$	0.99	8	5	50
0.9	0.75	Point Est	0.94	6	4	12
		Pred $r = 0.5$	0.96	6	4	23
		Pred $r = 0.7$	0.98	7	4	50
0.9	0.5	Point Est	0.94	6	4	13
		Pred $r = 0.5$	0.96	6	4	23
		Pred $r = 0.7$	0.98	7	4	50
0.9	0.25	Point Est	0.94	6	4	16
		Pred $r = 0.5$	0.96	6	4	23
		Pred $r = 0.7$	0.99	7	4	50
0.9	0	Point Est	0.96	7	6	20
		Pred $r = 0.5$	0.98	8	5	42
		Pred $r = 0.7$	0.99	9	5	50
0.95	0.75	Point Est	0.96	7	5	16
		Pred $r = 0.5$	0.98	7	5	33
		Pred $r = 0.7$	0.98	8	5	50
0.95	0.5	Point Est	0.96	7	5	17
		Pred $r = 0.5$	0.98	7	5	33
		Pred $r = 0.7$	0.99	9	5	50
0.95	0.25	Point Est	0.96	7	5	17
		Pred $r = 0.5$	0.98	8	5	34
		Pred $r = 0.7$	0.99	9	5	50
0.95	0	Point Est	0.98	8	6	25
		Pred $r = 0.5$	0.99	8	6	50
		Pred $r = 0.7$	0.99	10	6	50

Cause of death	Sample Size
Stroke	630
Other Non-communicable Diseases (NCD)	599
Pneumonia	540
AIDS	502
Maternal	468
Other Cardiovascular Diseases	416
Renal Failure	416
Diabetes	414
Acute Myocardial Infarction	400
Cirrhosis	313
TB	276
Other Infectious Diseases	263
Diarrhea/Dysentery	228
Road Traffic	202
Breast Cancer	195
Falls	173
COPD	171
Homicide	167
Leukemia/Lymphomas	156
Cervical Cancer	155
Suicide	124
Fires	122
Drowning	106
Lung Cancer	106
Other Injuries	103
Malaria	100
Colorectal Cancer	99
Poisonings	86
Bite of Venomous Animal	66
Stomach Cancer	62
Epilepsy	48
Prostate Cancer	48
Asthma	47
Esophageal Cancer	40

Table 3: Sample size of each cause of death in the PHMRC dataset.

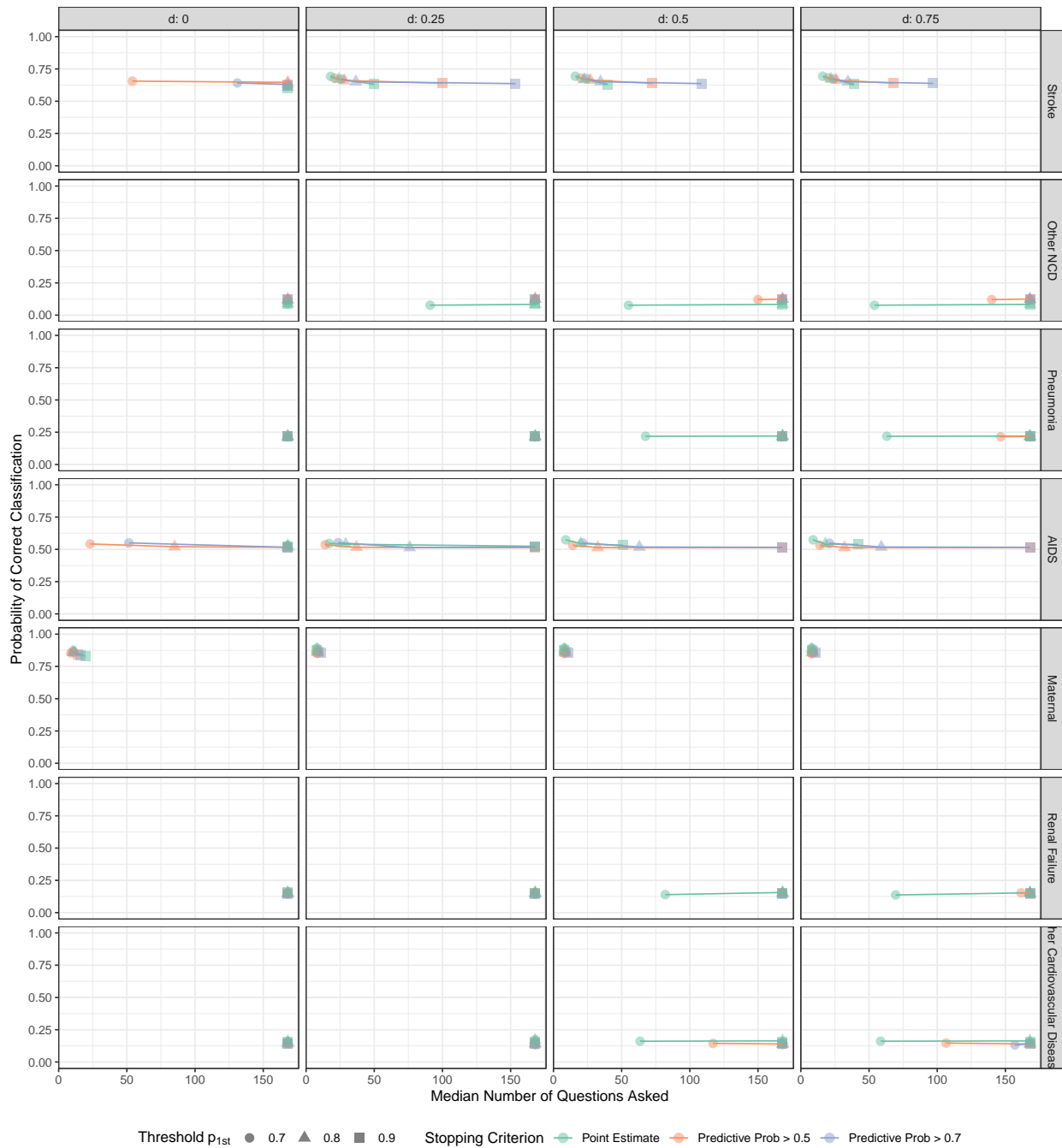


Figure 1: Proportion of correctly classified deaths among deaths due to each cause using different stopping criteria. The causes are ordered by their sample size in the PHMRC data.

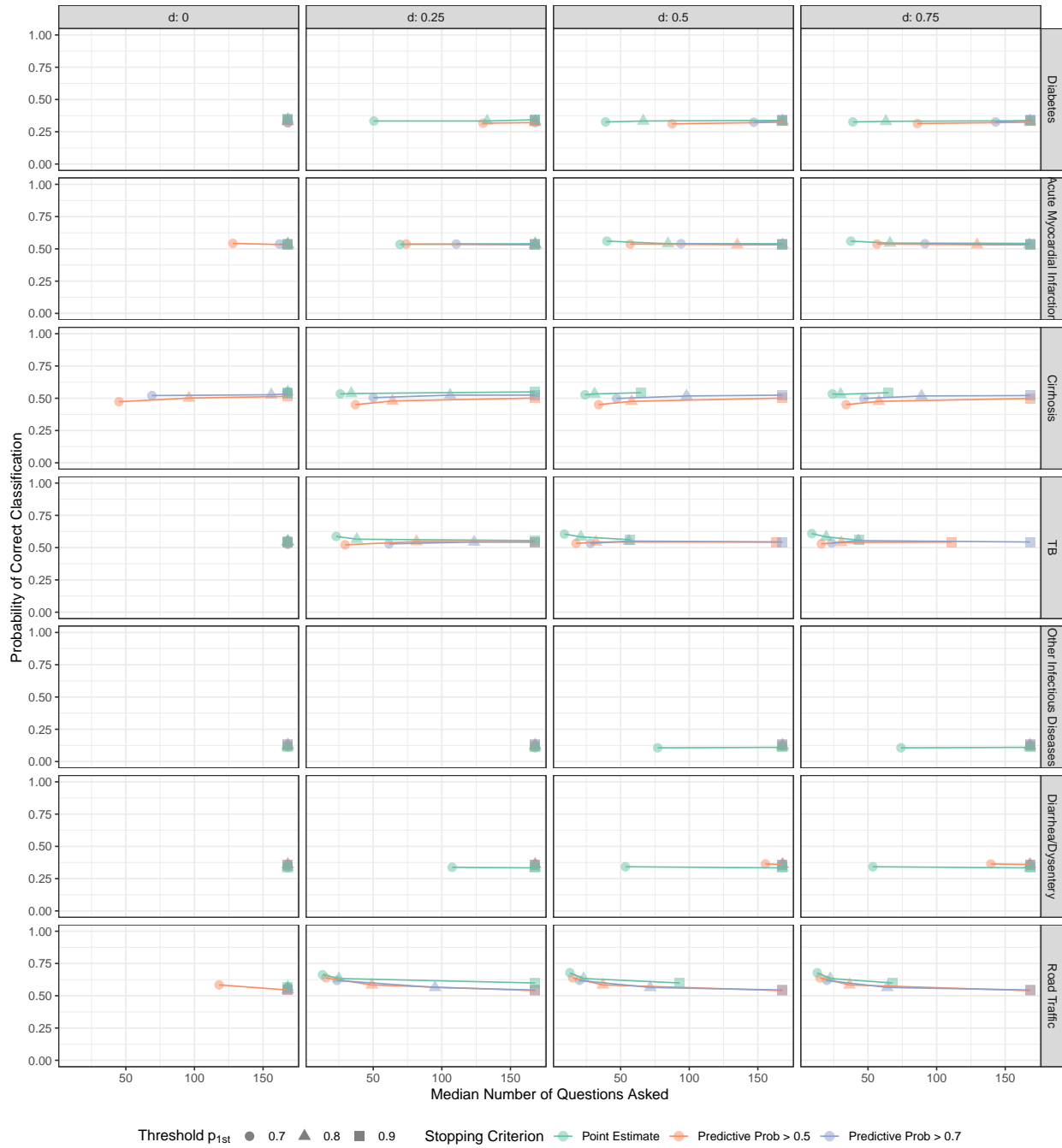


Figure 2: (Continued) Proportion of correctly classified deaths among deaths due to each cause using different stopping criteria. The causes are ordered by their sample size in the PHMRC data.

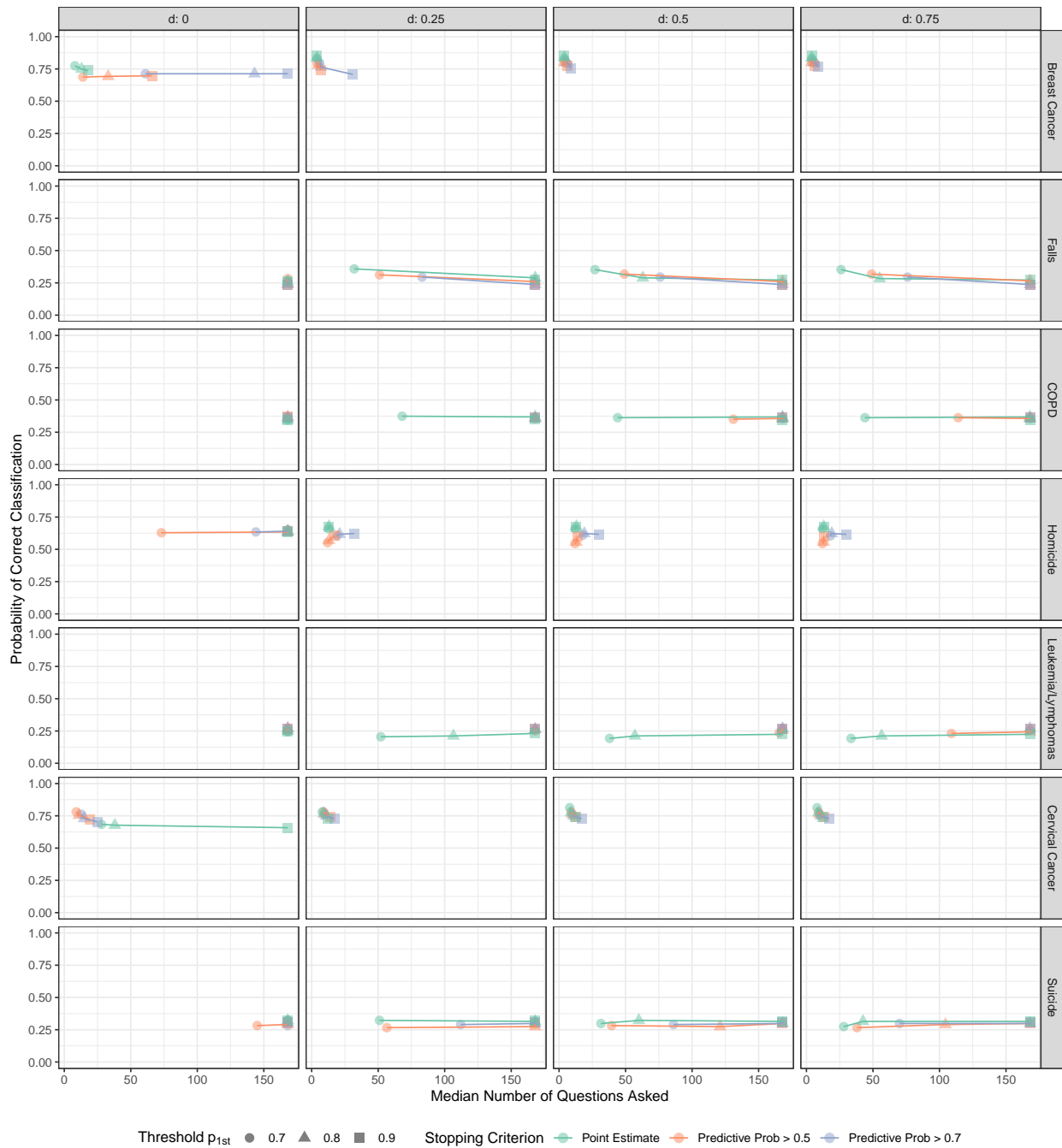


Figure 3: (Continued) Proportion of correctly classified deaths among deaths due to each cause using different stopping criteria. The causes are ordered by their sample size in the PHMRC data.

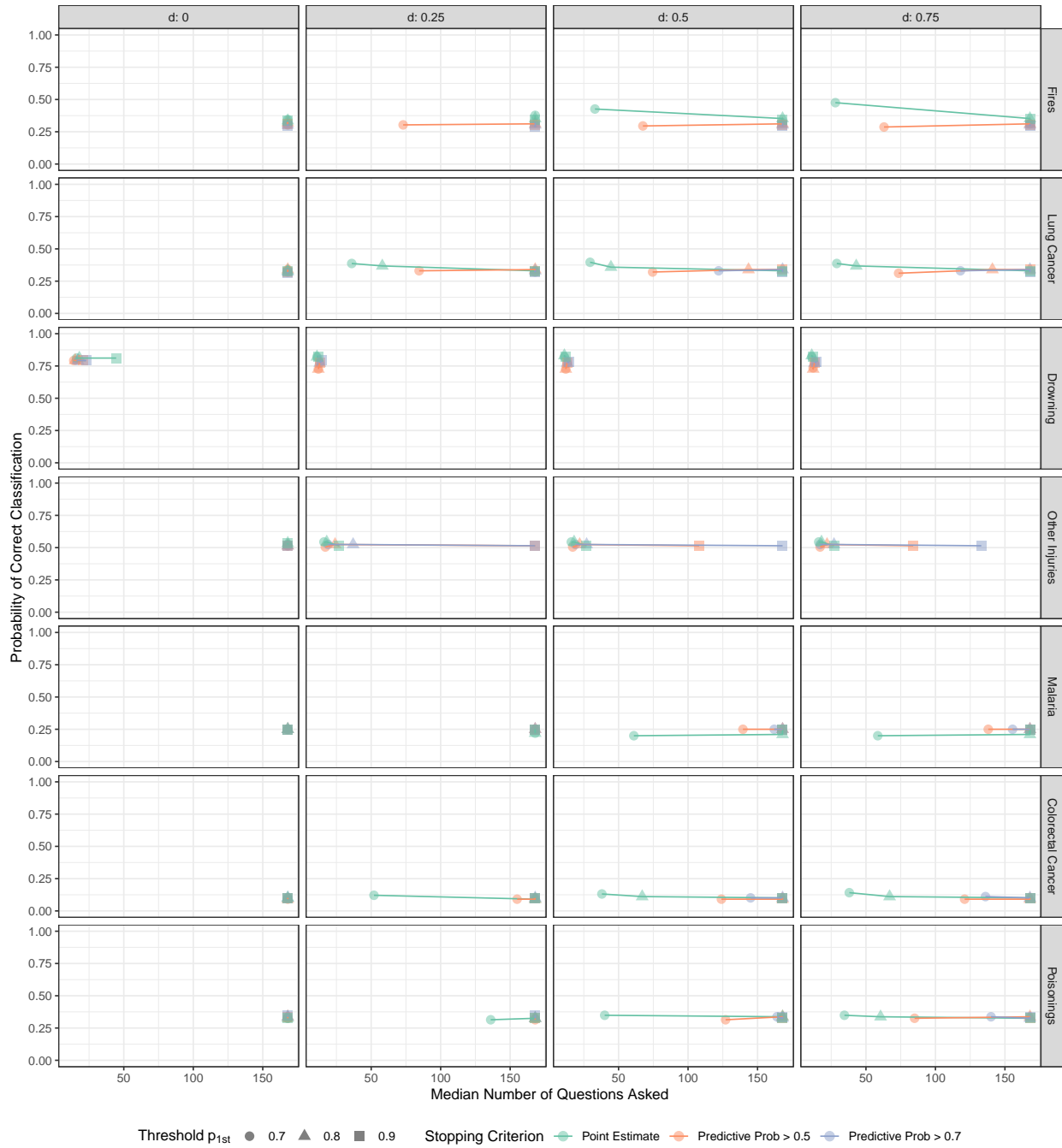


Figure 4: (Continued) Proportion of correctly classified deaths among deaths due to each cause using different stopping criteria. The causes are ordered by their sample size in the PHMRC data.

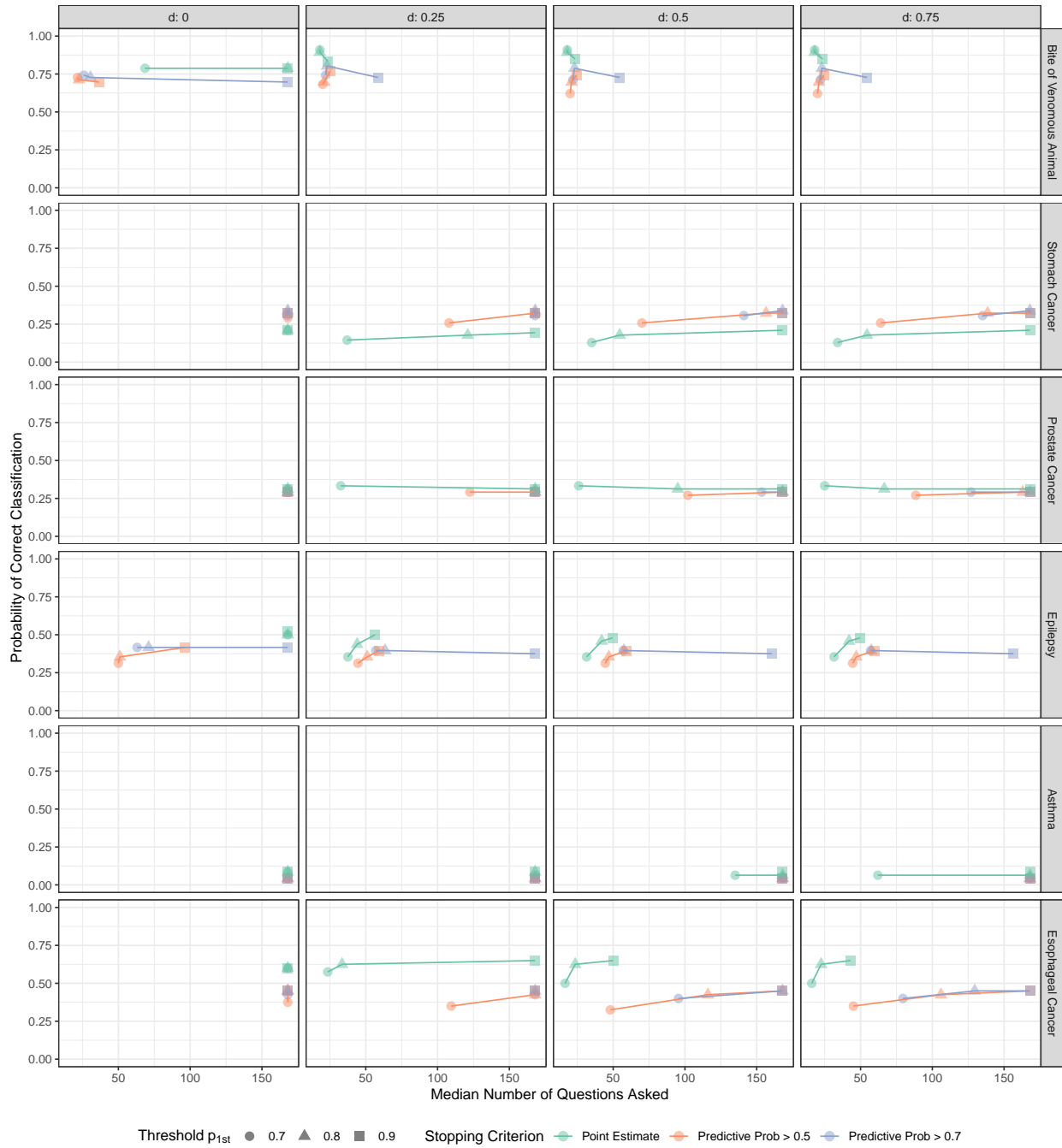


Figure 5: (Continued) Proportion of correctly classified deaths among deaths due to each cause using different stopping criteria. The causes are ordered by their sample size in the PHMRC data.