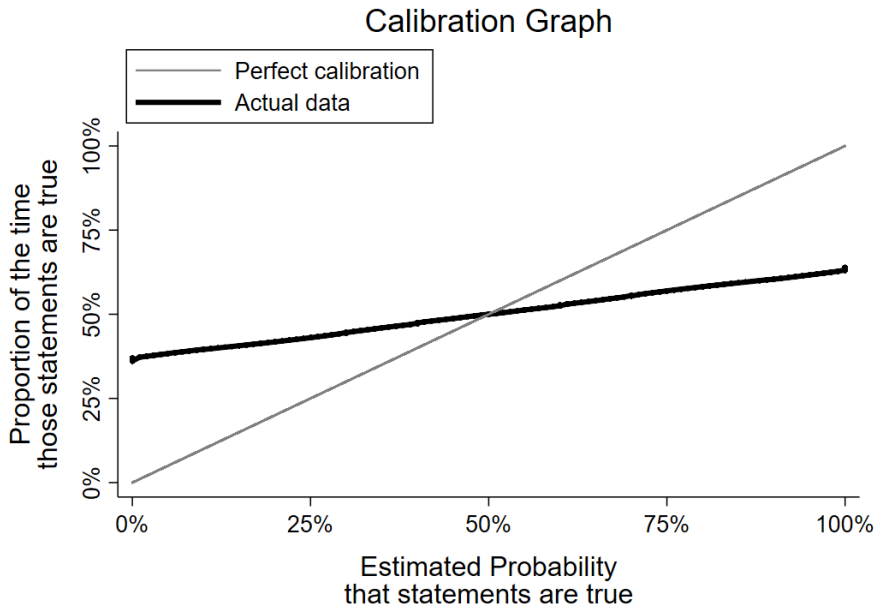
Probability and Decision Making
NDC Class Feedback (19 February 2021)

Prof. Jeffrey A. Friedman
Dartmouth College
jeffrey.a.friedman@dartmouth.edu

Thank you for taking the time to complete the survey. This document describes performance for the class as a whole and then for individual participants. As mentioned in the lecture, my colleagues and I welcome feedback on this material. The goal of our research is to support sound national security analysis and decision making, and we hope you found this exercise to be useful.



Calibration

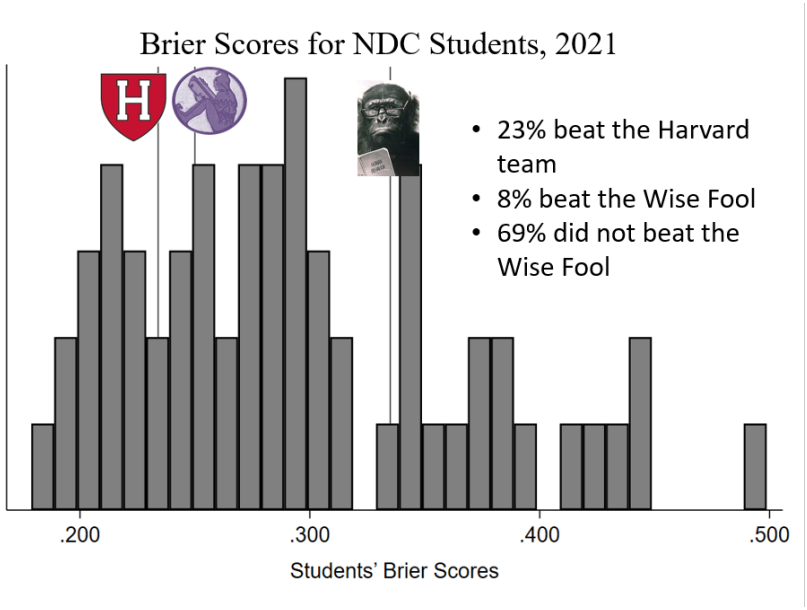
The "calibration graph" presented above shows how students' assessments of uncertainty compared to the actual chances that statements in the survey were true. The good news is that students were quite good at determining which statements were more likely to be true than others. That is not a trivial fact: it means that students consistently drew meaningful distinctions among different levels of certainty. The bad news is that students were overconfident when assigning certainty to their judgments. For instance, when students assigned a 100 percent probability to statements being true, those statements were actually true about 70 percent of the time.

Brier Scores

Another for evaluating students' probability assessments is to calculate Brier Scores. The Brier Score is a measurement of judgmental error: it is the squared difference between the probability estimate that you made, and the probability estimate that you *would* have made if you knew the answer for certain. As in golf, lower scores are better. We evaluated students' Brier Scores against several baselines:

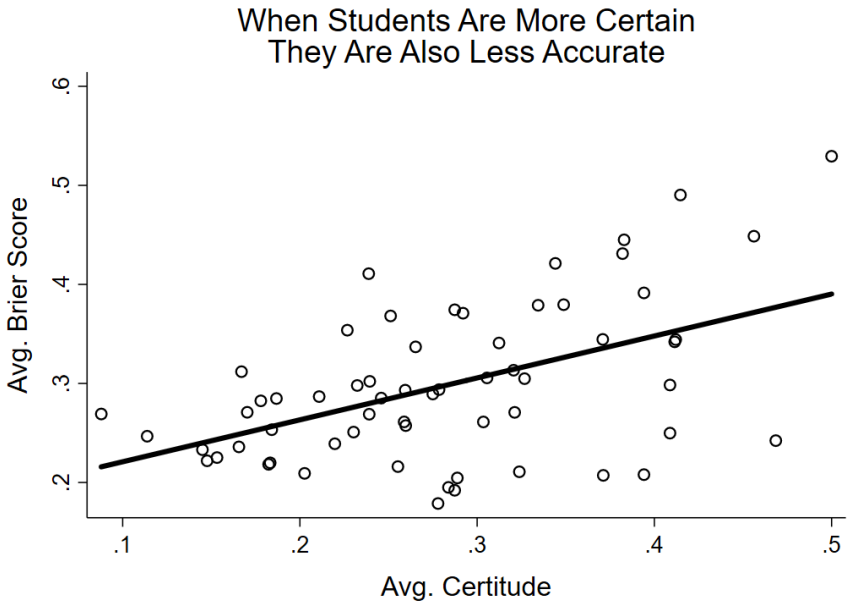
- A group of Harvard teaching fellows, who are intelligent and who have been trained in probabilistic reasoning, but who have no special knowledge of foreign policy and national security. Their average Brier Score was 0.234.
- The Wise Fool, who guesses 50% for every question we pose. The Wise Fool provides no useful information, but he never suffers from overconfidence. His Brier Score is always 0.250.
- The Chimp, who guesses answers at random. The Chimp's expected Brier Score is 0.335.

Among the 63 NDC students who took this year's survey, 23 percent beat the Harvard assessors and another 8 percent others beat the Wise Fool. That meant 69 percent of the class would have done better by saying they simply had no idea when responding to every question that we posed.



These results are not surprising. In almost every context we know of, individuals who have not previously received explicit feedback on their assessments of uncertainty prove to be substantially overconfident in their capabilities. The following graph shows that there is indeed a clear relationship between the certainty that NDC students assigned to their judgments and how well they performed on the probability assessment exercise on the whole.[1]

[1] The horizontal axis in this graph represents "certitude," which is the average difference between a student's probability estimates and 50 percent. Thus, an estimate of 30 percent and an estimate of 70 percent each have a "certitude" of 0.20 – they are equally-distant from an ignorant guess.



Prospects for improvement

Almost any major decision we make is surrounded by uncertainty. Any time that you state an opinion that you do not know as a fact to be true, you are making a probability assessment, however explicitly you say what these chances are. Yet most people do not receive structured feedback at any point in their lives on how well they perform this task. This creates what decision theorists call the "illusion of validity" – almost all of us believe we are better at assessing uncertainty than we really are.

The good news is that once people see these evaluations, they generally improve immediately and substantially. In courses at Harvard and Dartmouth (where we usually run such surveys twice), we find that students' performance typically improves by about 20 percent the second time around. A major research project at the University of Pennsylvania, funded by the Intelligence Advanced Research Projects Activity, found that just one hour of probability assessment training improved people's capabilities by about 15 percent. Those training sessions had noticeable impacts on performance as long as *four years* out (and presumably further: four years is as long as the project has been running). If this feedback provides you with only a small fraction of this common improvement, then we hope you believe that this exercise was worthwhile.

Additional references

Philip Tetlock and Daniel Gardner, *Superforecasting* (Crown, 2015). State-of-the-art insights on political forecasting based on a four-year Intelligence Community-backed research project that involved thousands of participants.

Daniel Kahneman, *Thinking, Fast and Slow* (FSG, 2011). A Nobel Prize-winning economist's lucid overview of common biases in decision making.

Nate Silver, *The Signal and the Noise: Why Most Predictions Fail – And Some Don't* (Penguin, 2012). An overview of how people make predictions in many different professions, focusing especially on the need to balance analysis and intuition when making important decisions.

Gerd Gigerenzer, *Calculated Risks: How to Know When Numbers Deceive You* (Simon & Schuster, 2002). Explores common problems that emerge when making decisions based on uncertainty, and makes practical, evidence-based suggestions for how to address that challenge.

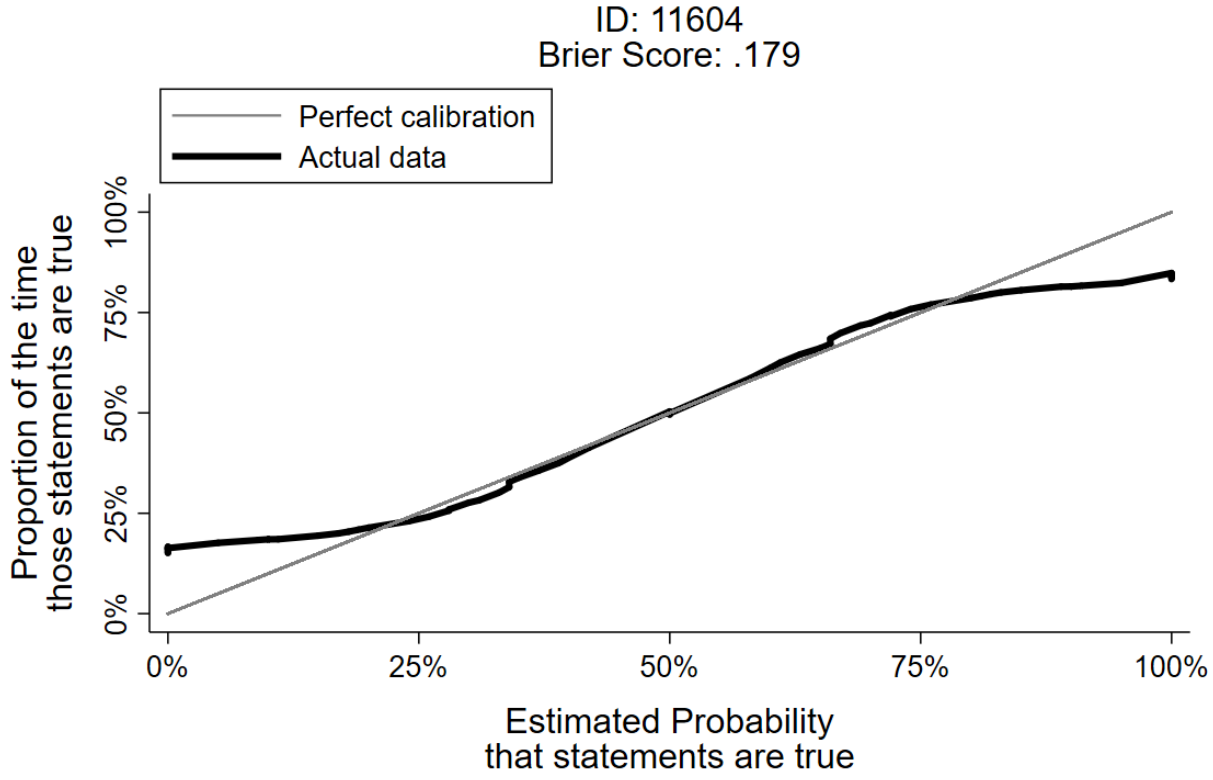
Jeffrey A. Friedman, *War and Chance: Assessing Uncertainty in International Politics* (Oxford, 2019). Analyzes the logic, psychology, and politics of assessing uncertainty in a national security context.

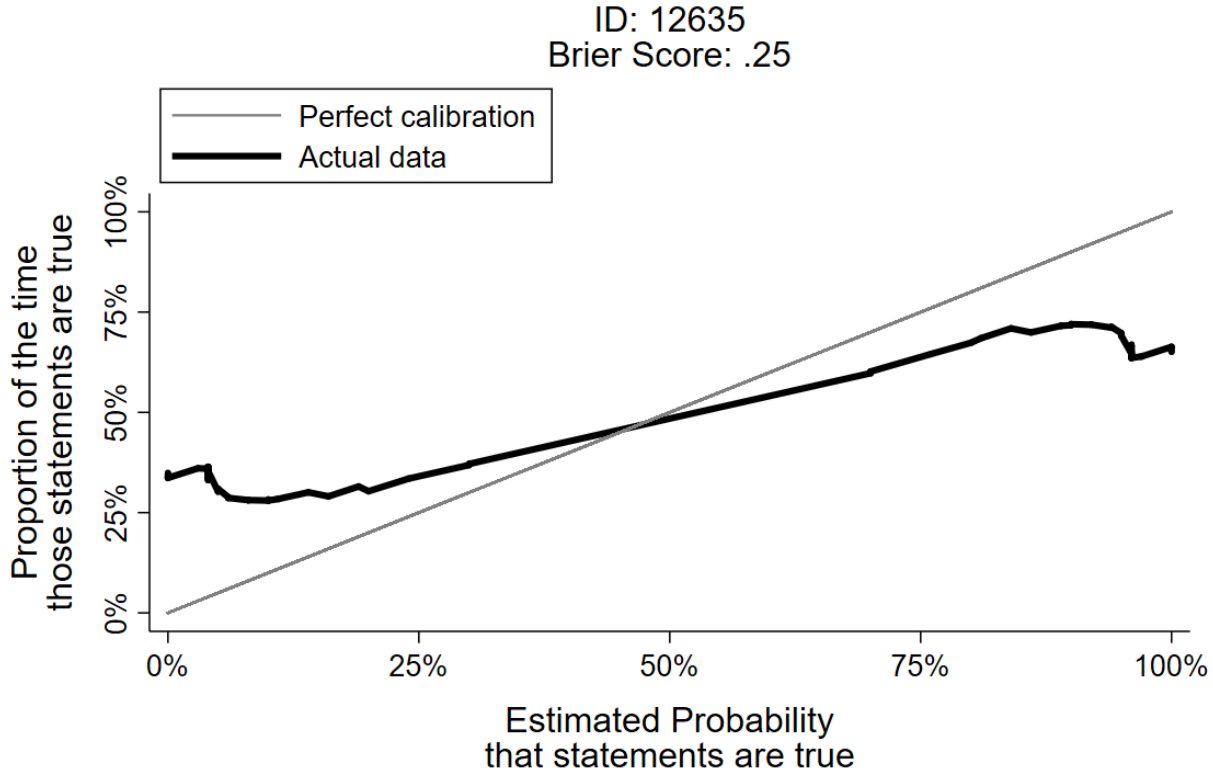
Individual performance

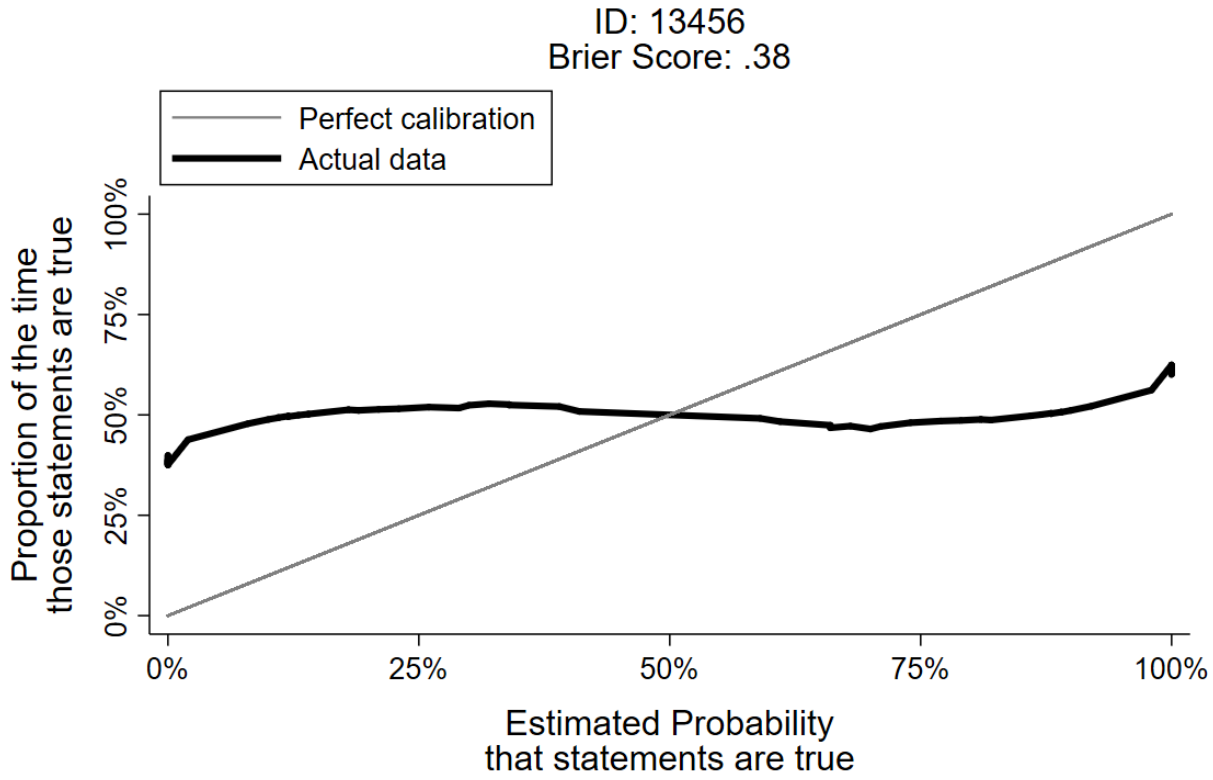
The remainder of this document presents individual students' calibration graphs, in ascending order of respondent ID. We processed every survey received by the end of 17 February. If you do not see your results here, please contact jeffrey.a.friedman@dartmouth.edu with your ID number. Even if you took the survey after the appointed time, we can usually track down students' results. If you forgot your ID number, it is usually possible to recover your graph if you can recall a few answers that you gave on the survey.

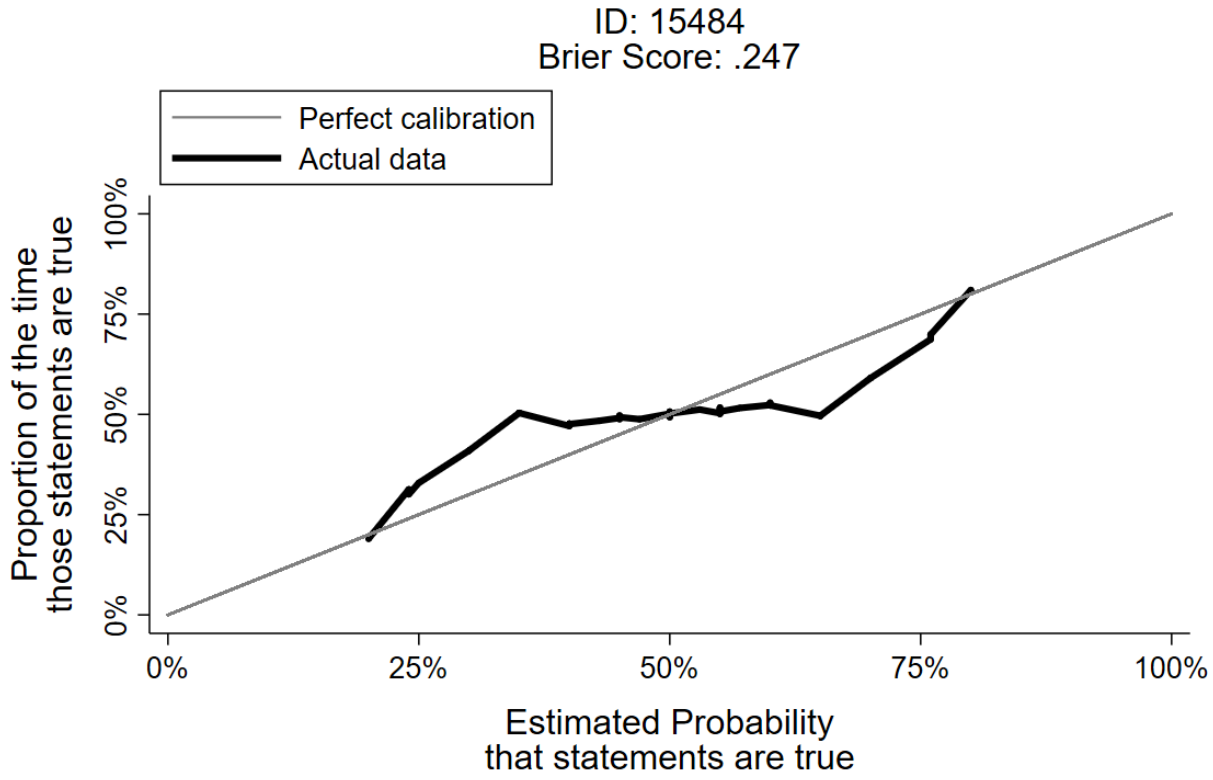
Each graph contains your Brier Score. Across NDC students as a whole, the average Brier Score was 0.298. Here are the percentiles for Brier Scores across the group:

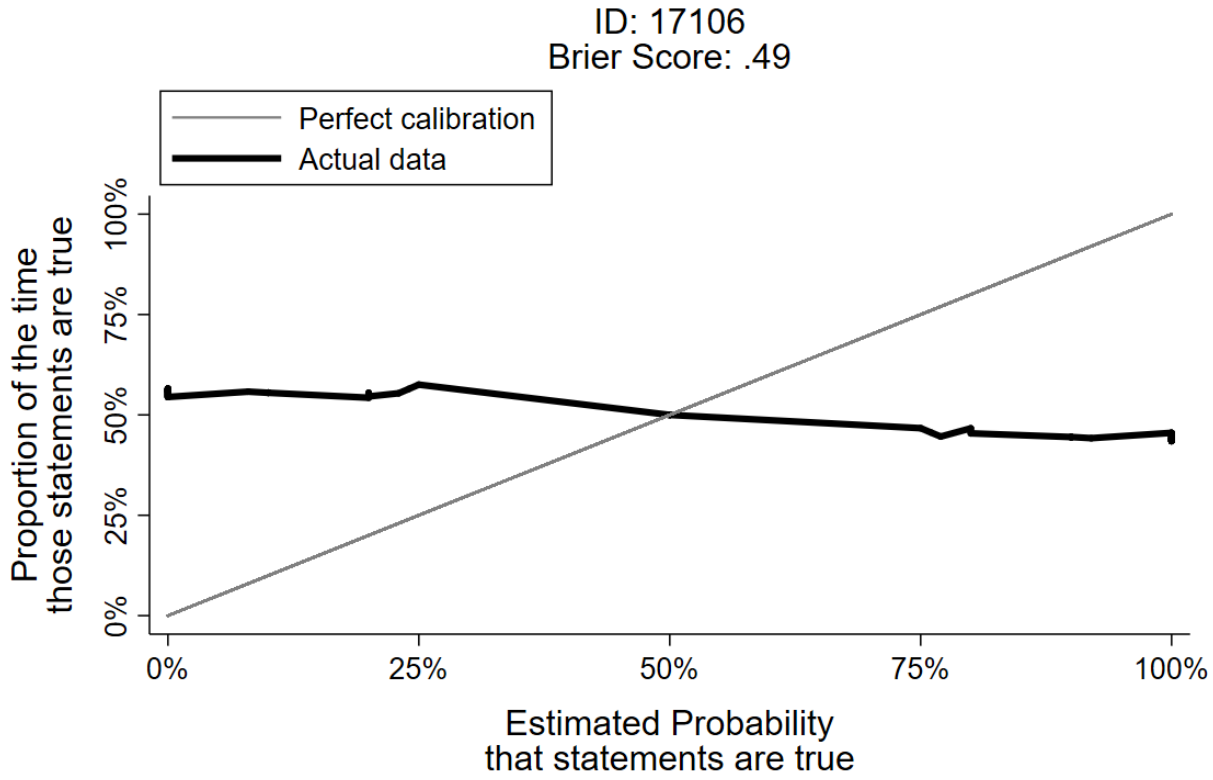
100th (best)	0.179
90th:	0.205
75th:	0.239
50th (median):	0.285
25th:	0.344
10th:	0.445

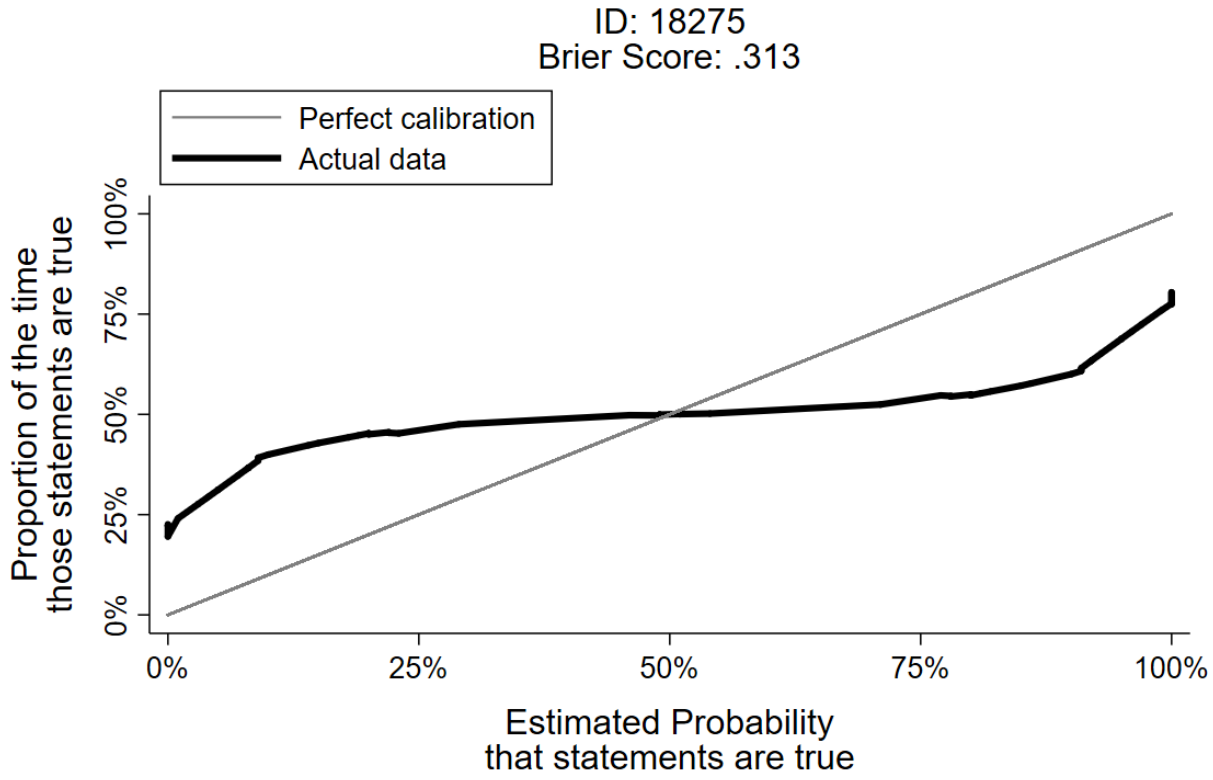

ID: 11604
Brier Score: .179
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

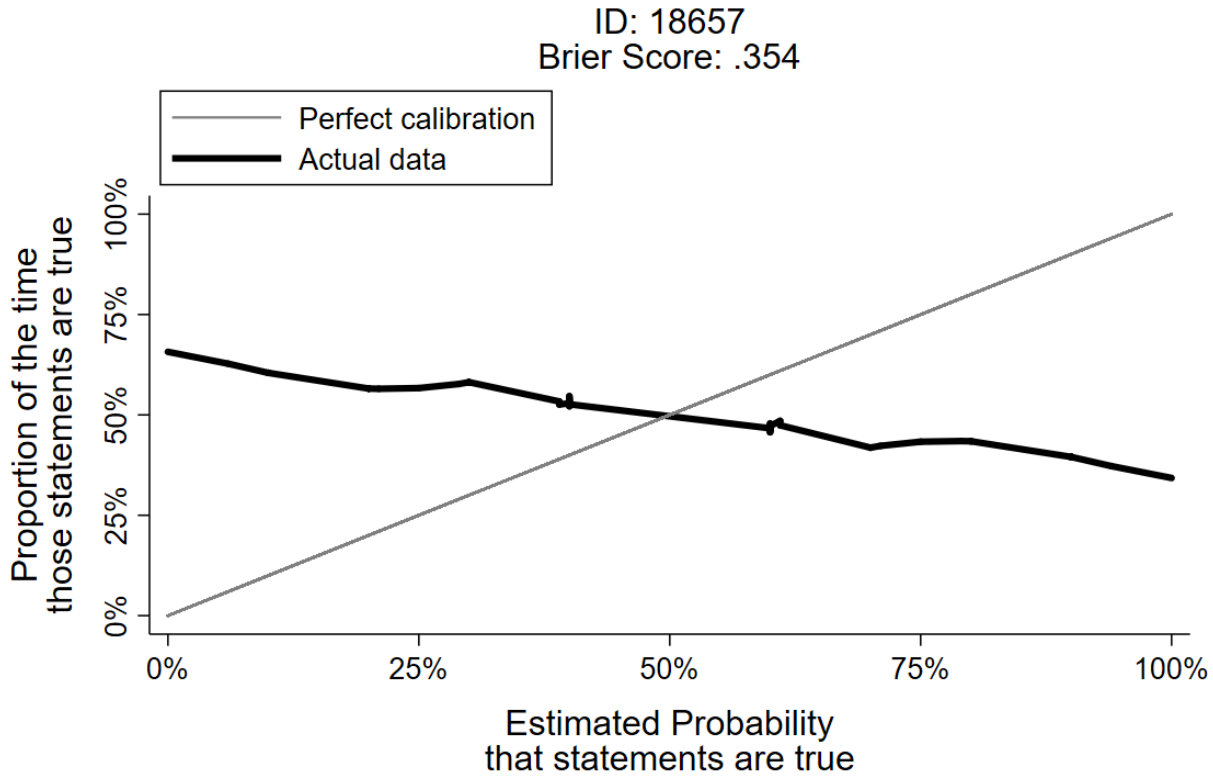

ID: 12635
Brier Score: .25
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

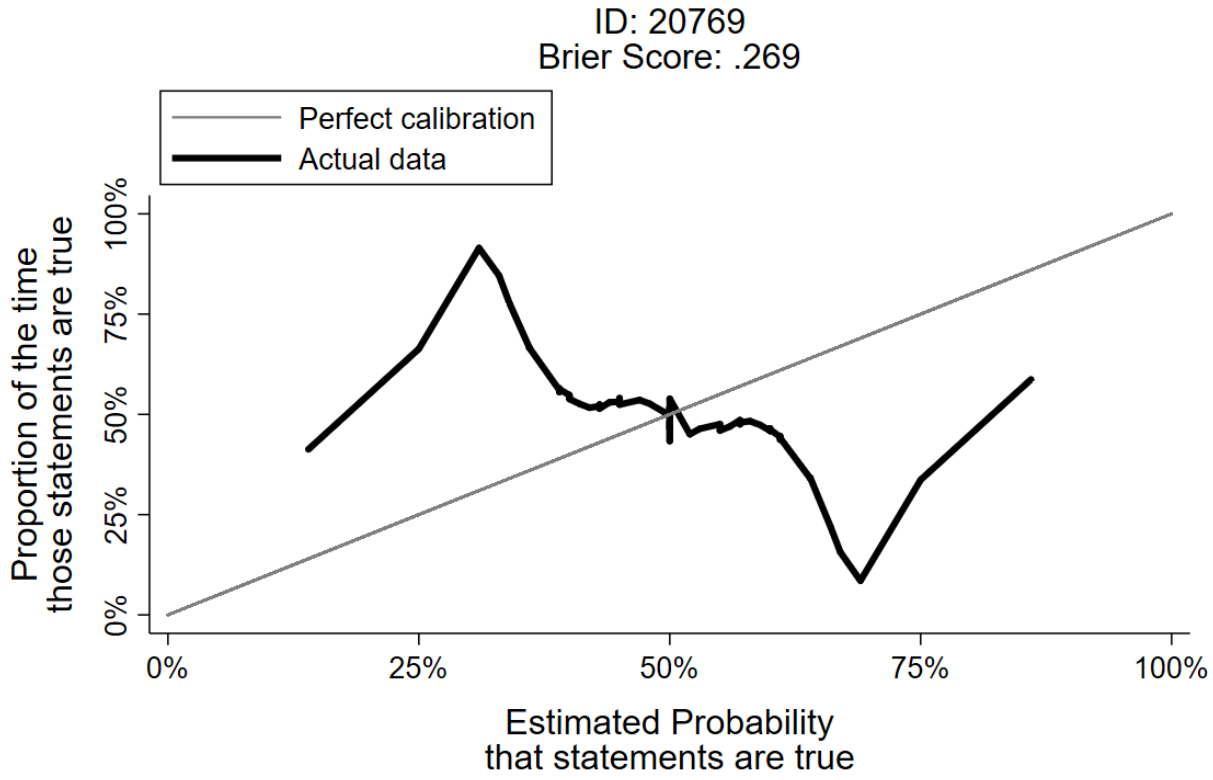

ID: 13456
Brier Score: .38
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

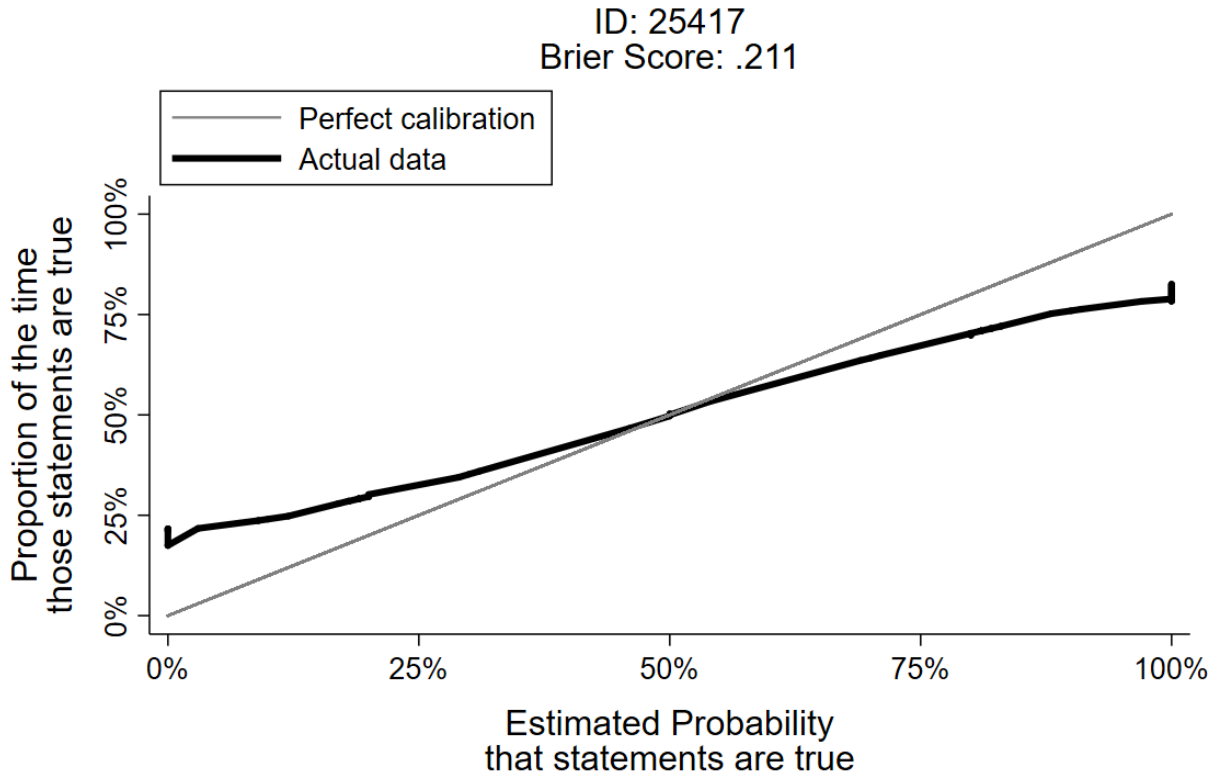

ID: 15484
Brier Score: .247
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

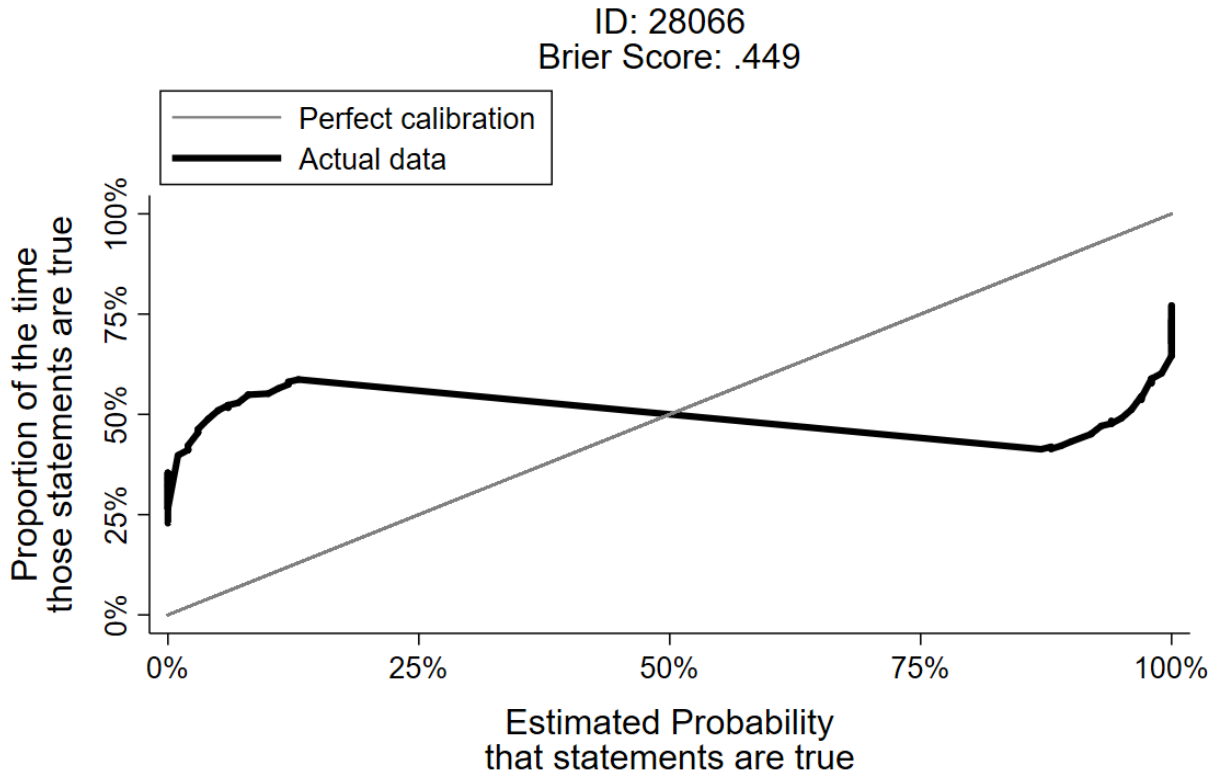

ID: 17106
Brier Score: .49
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

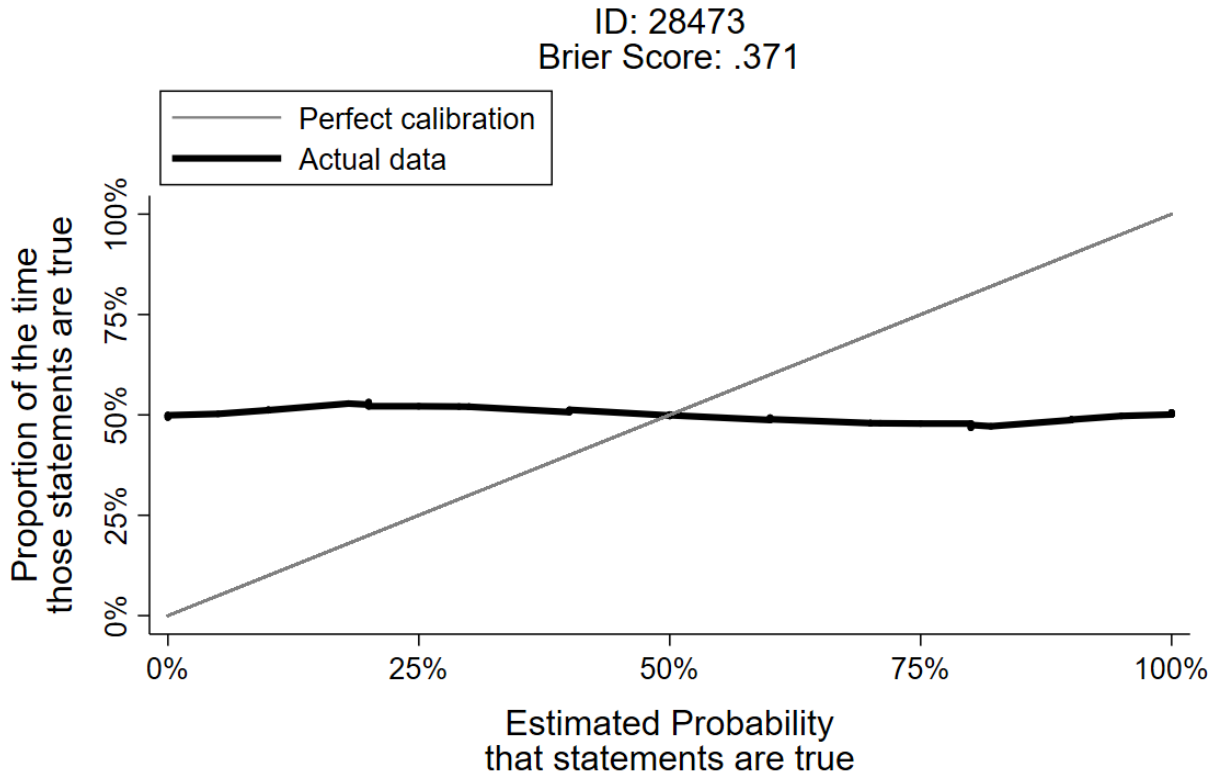

ID: 18275
Brier Score: .313
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

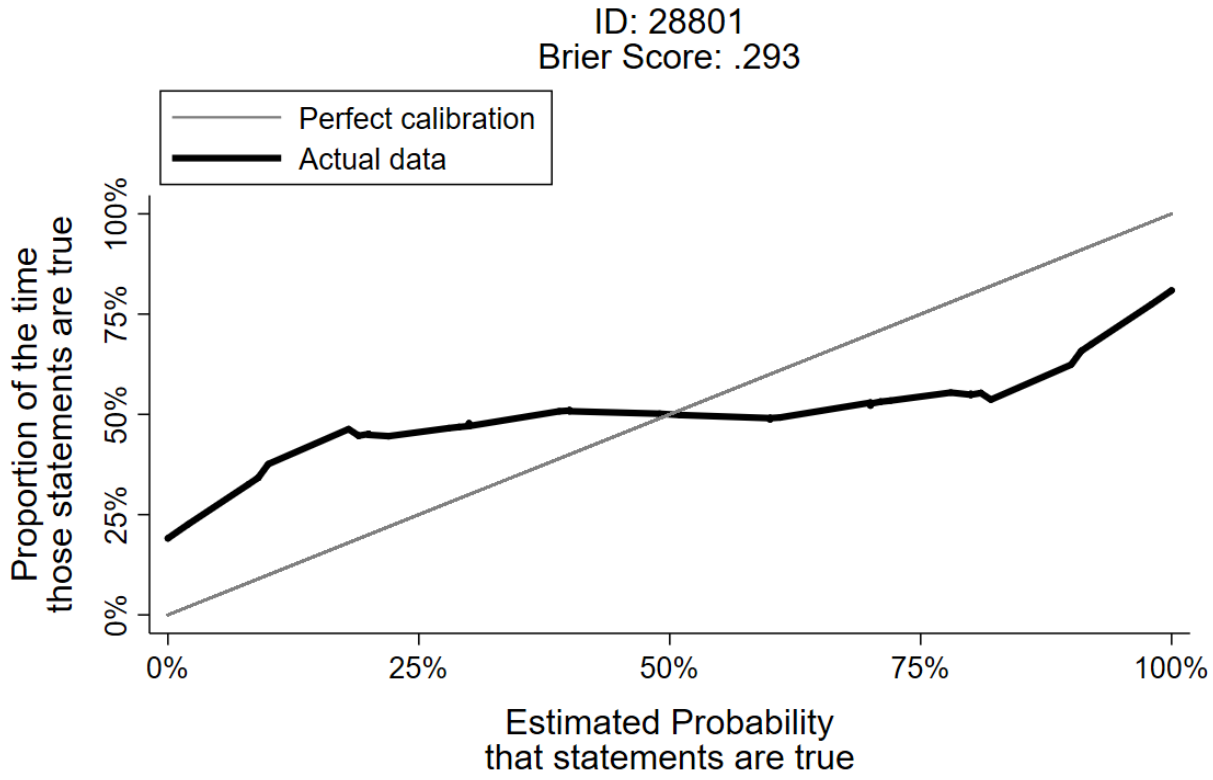

ID: 18657
Brier Score: .354
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

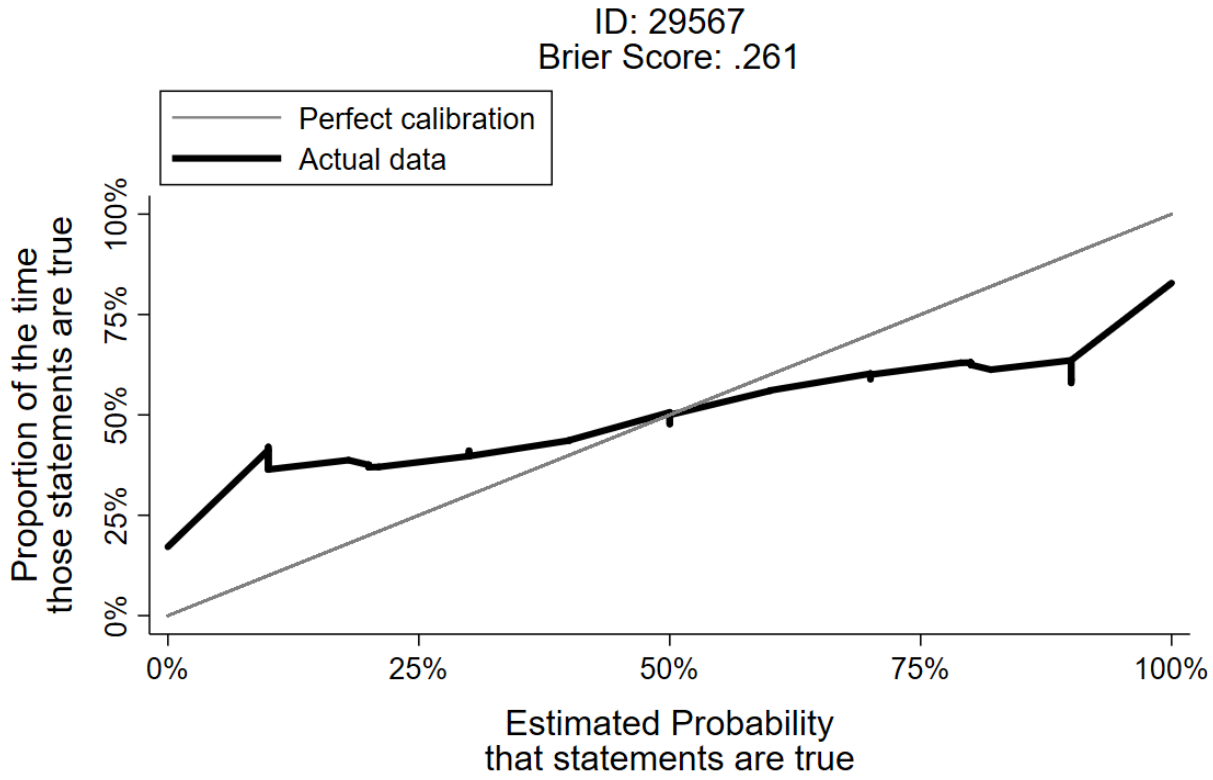

ID: 20769
Brier Score: .269
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

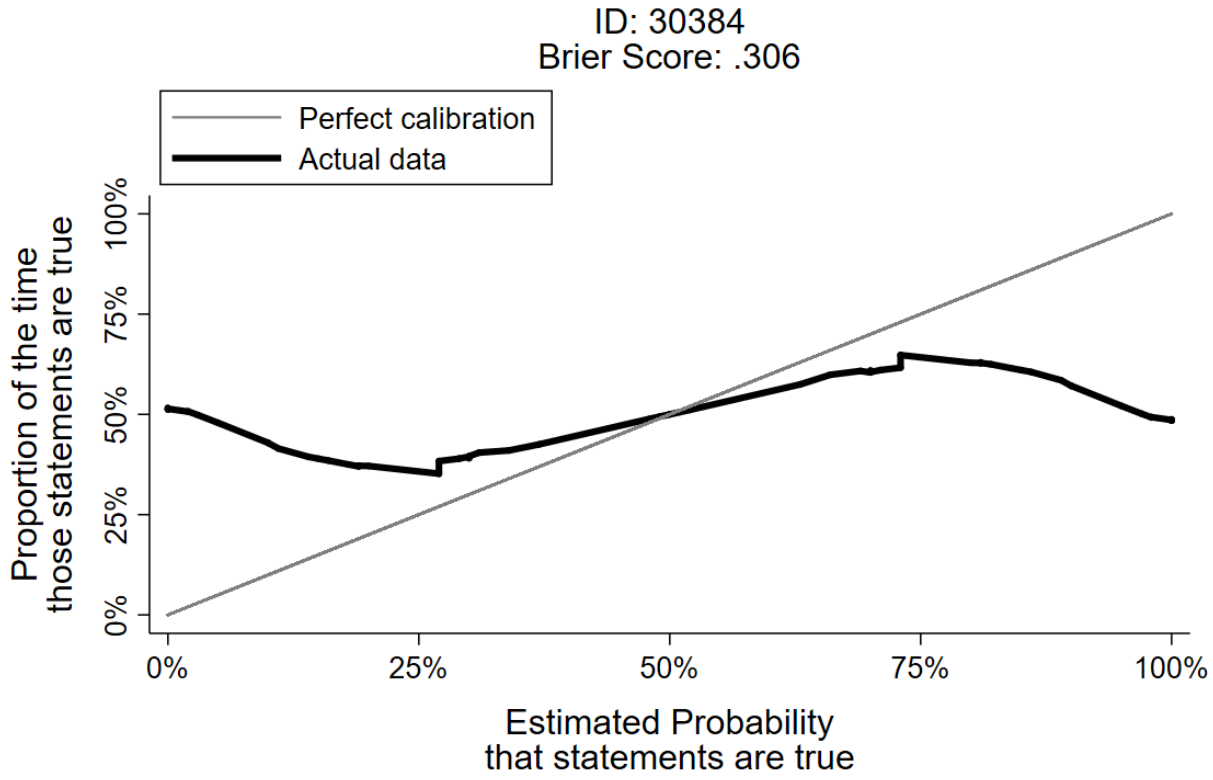

ID: 25417
Brier Score: .211
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

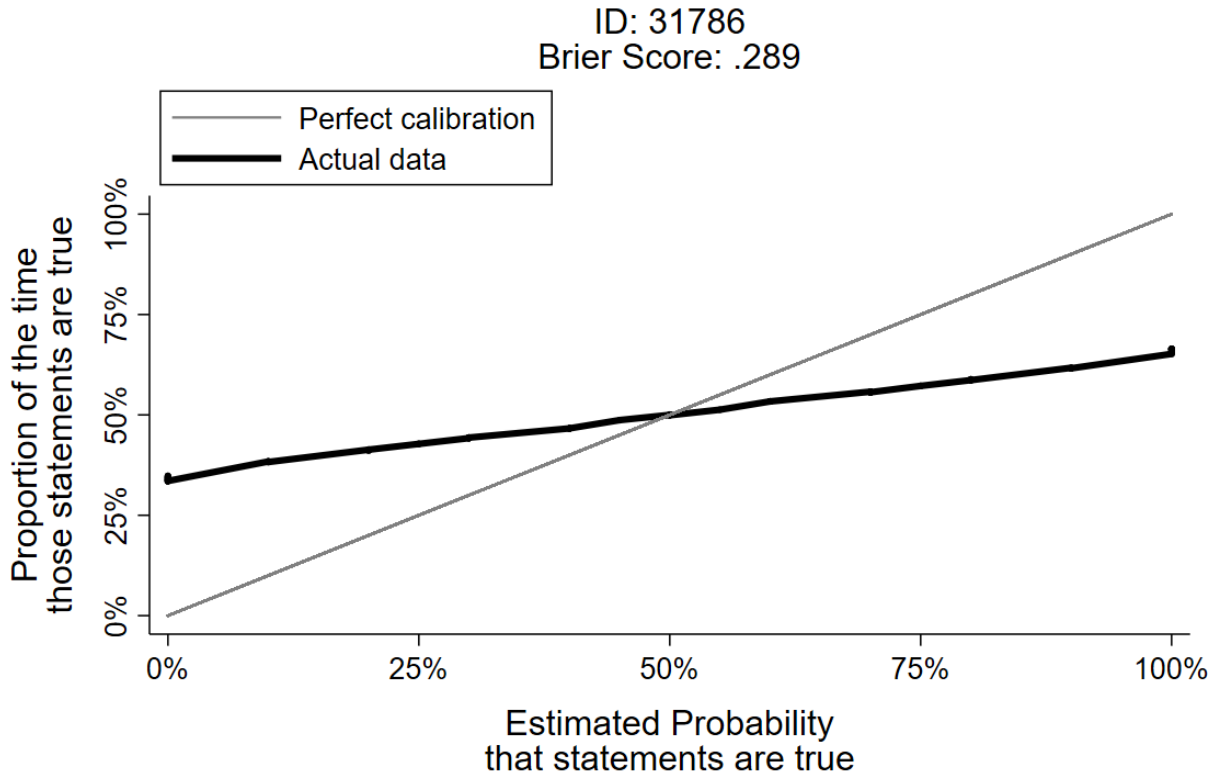

ID: 28066
Brier Score: .449
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

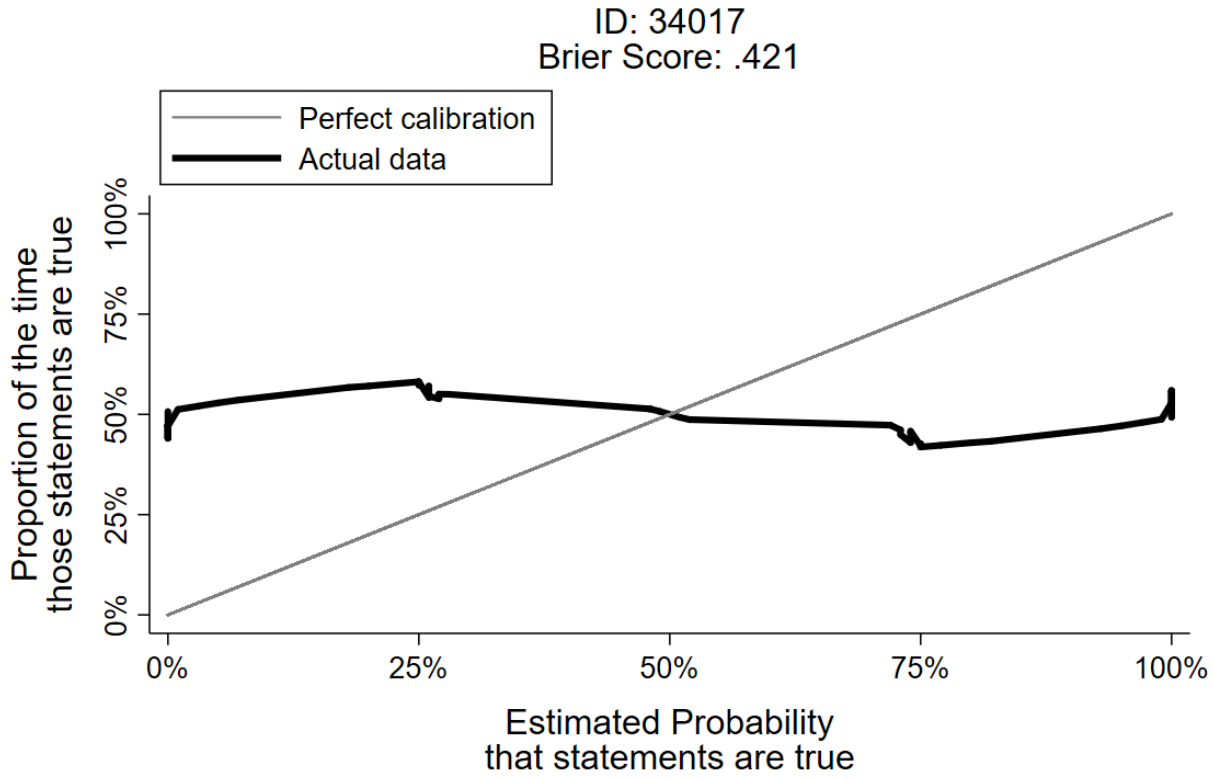

ID: 28473
Brier Score: .371
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

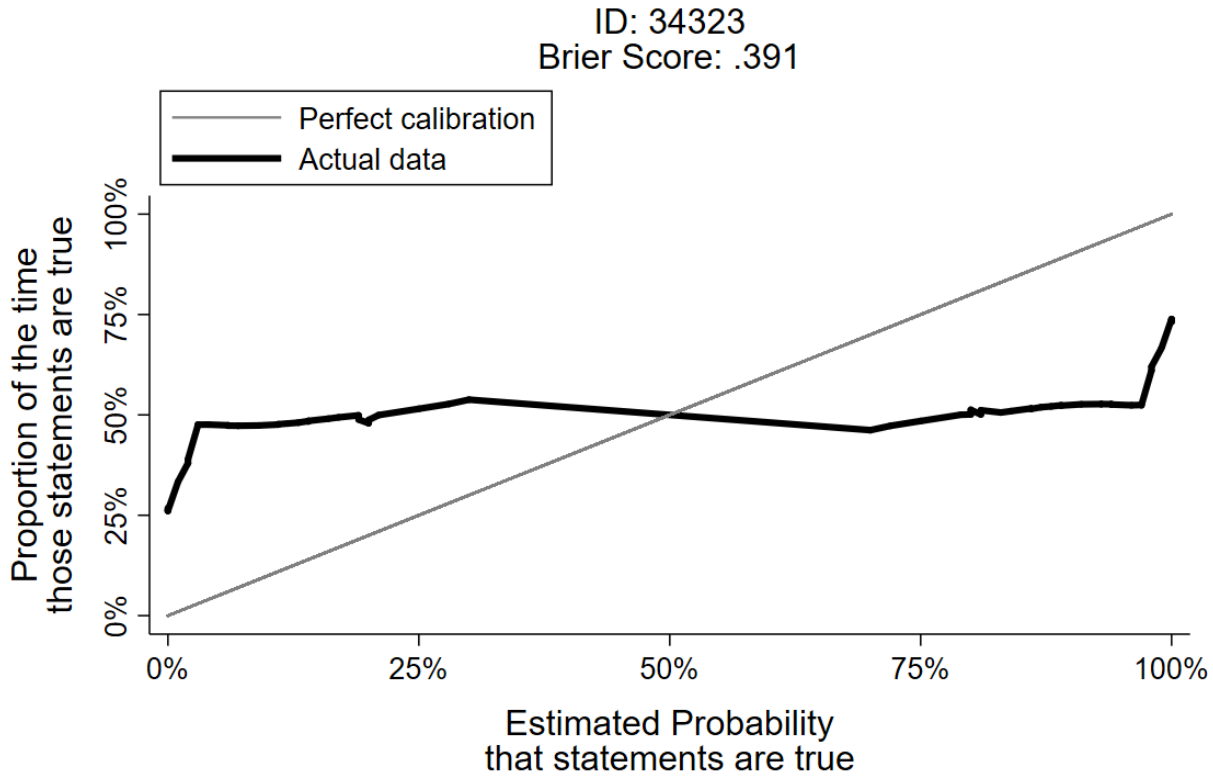

ID: 28801
Brier Score: .293
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

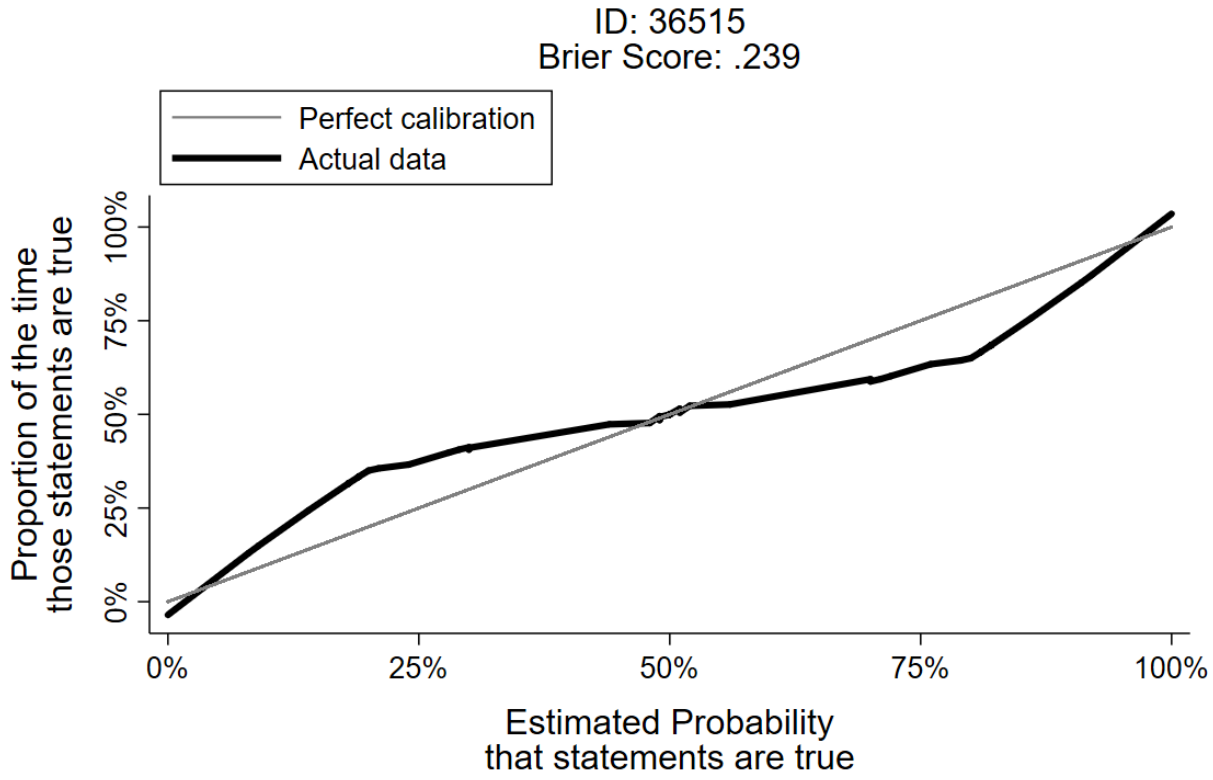

ID: 29567
Brier Score: .261
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

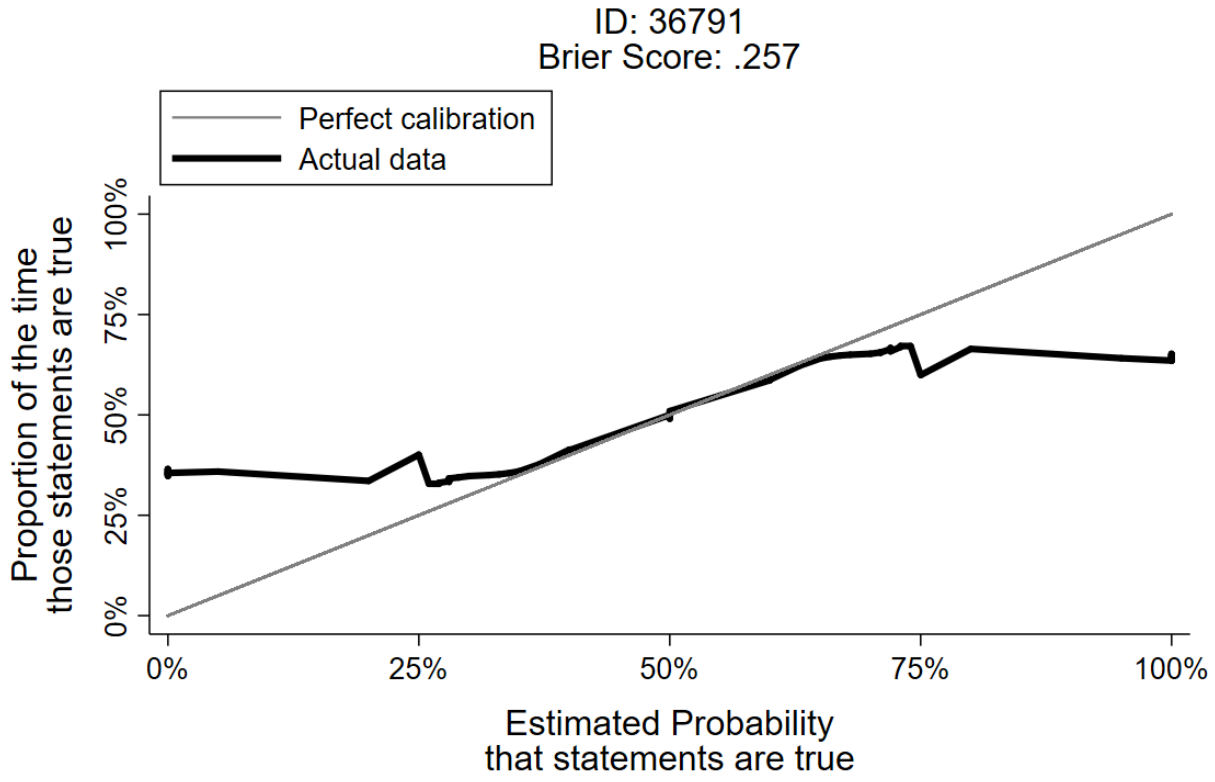

ID: 30384
Brier Score: .306
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

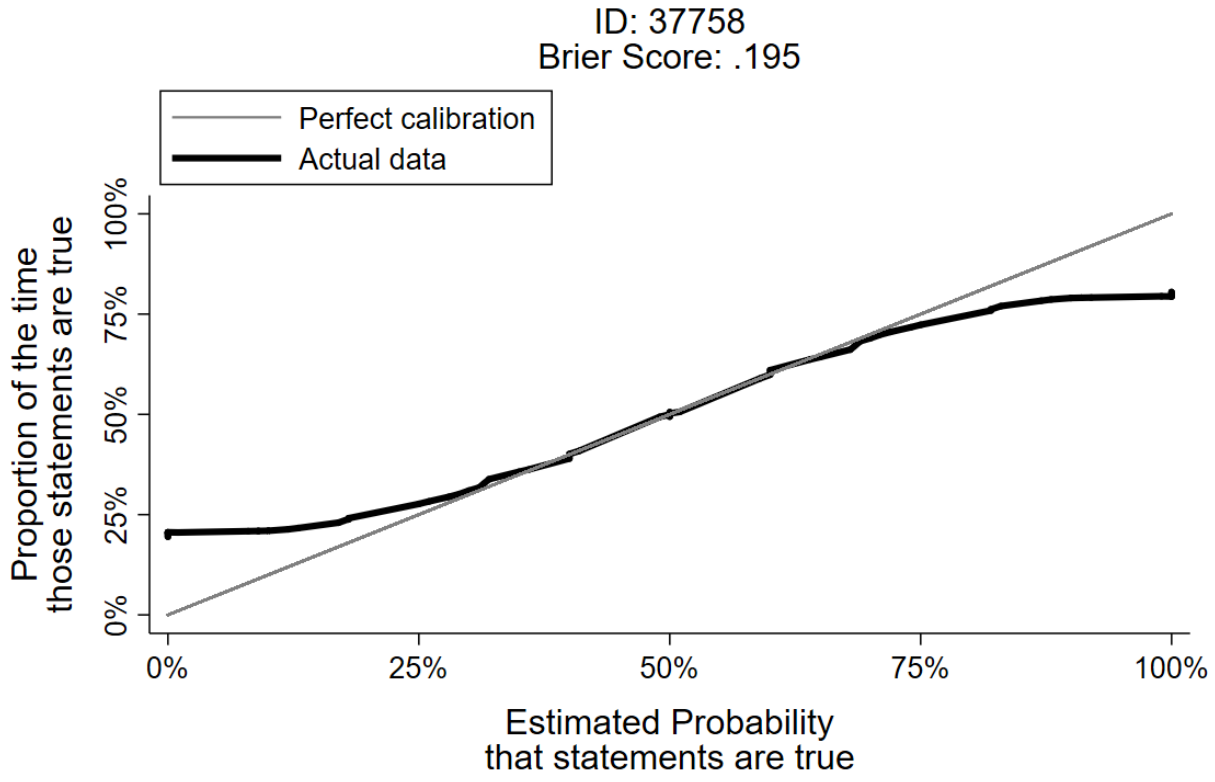

ID: 31786
Brier Score: .289
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

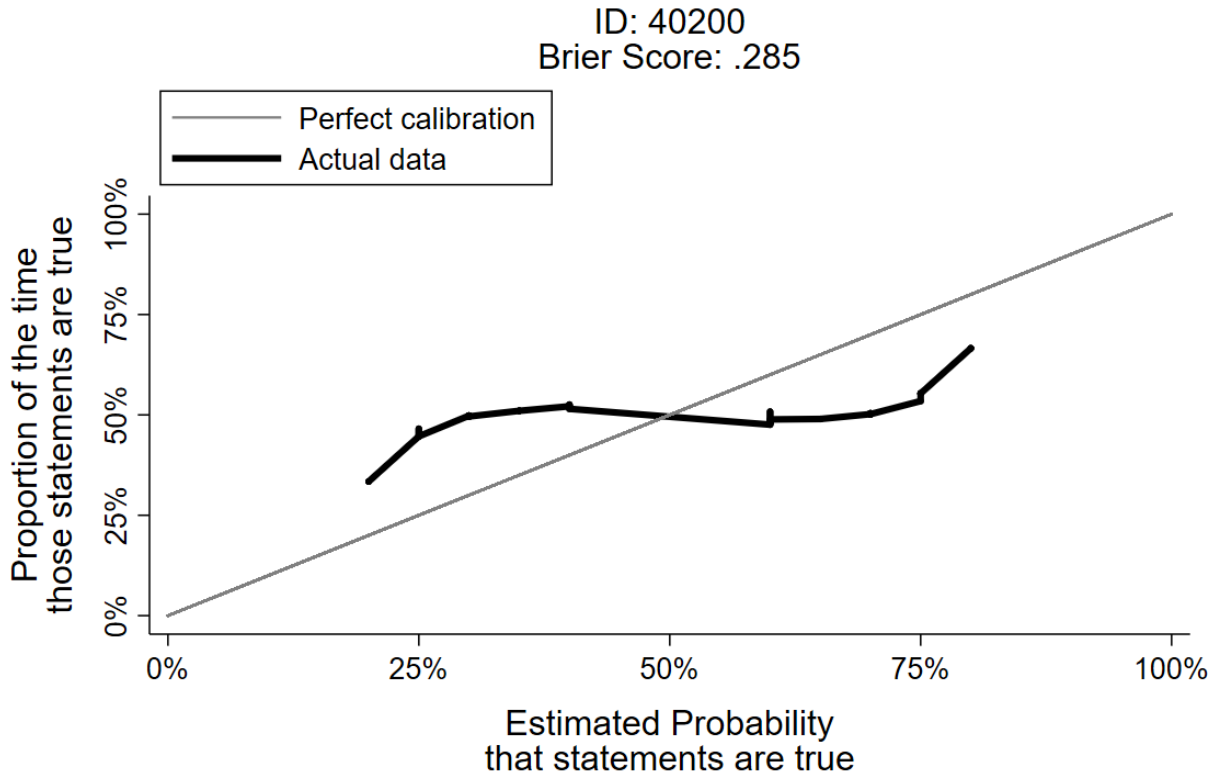

ID: 34017
Brier Score: .421
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

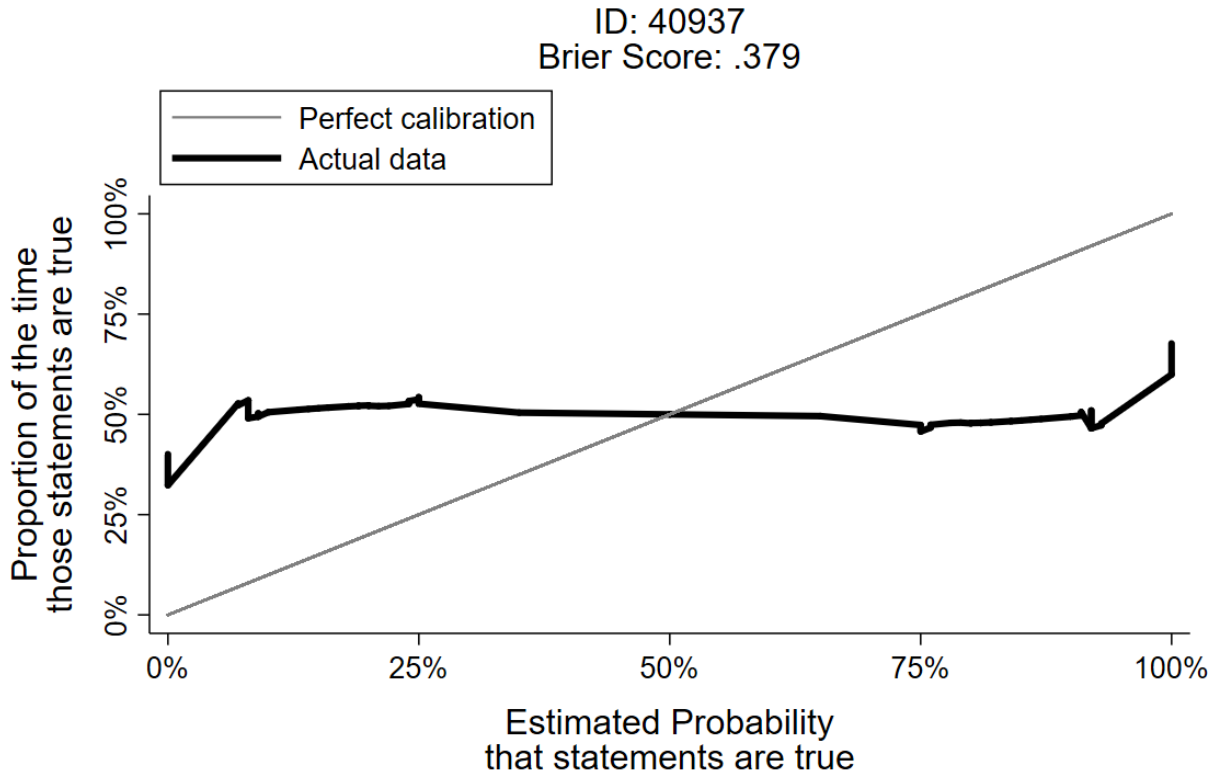

ID: 34323
Brier Score: .391
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

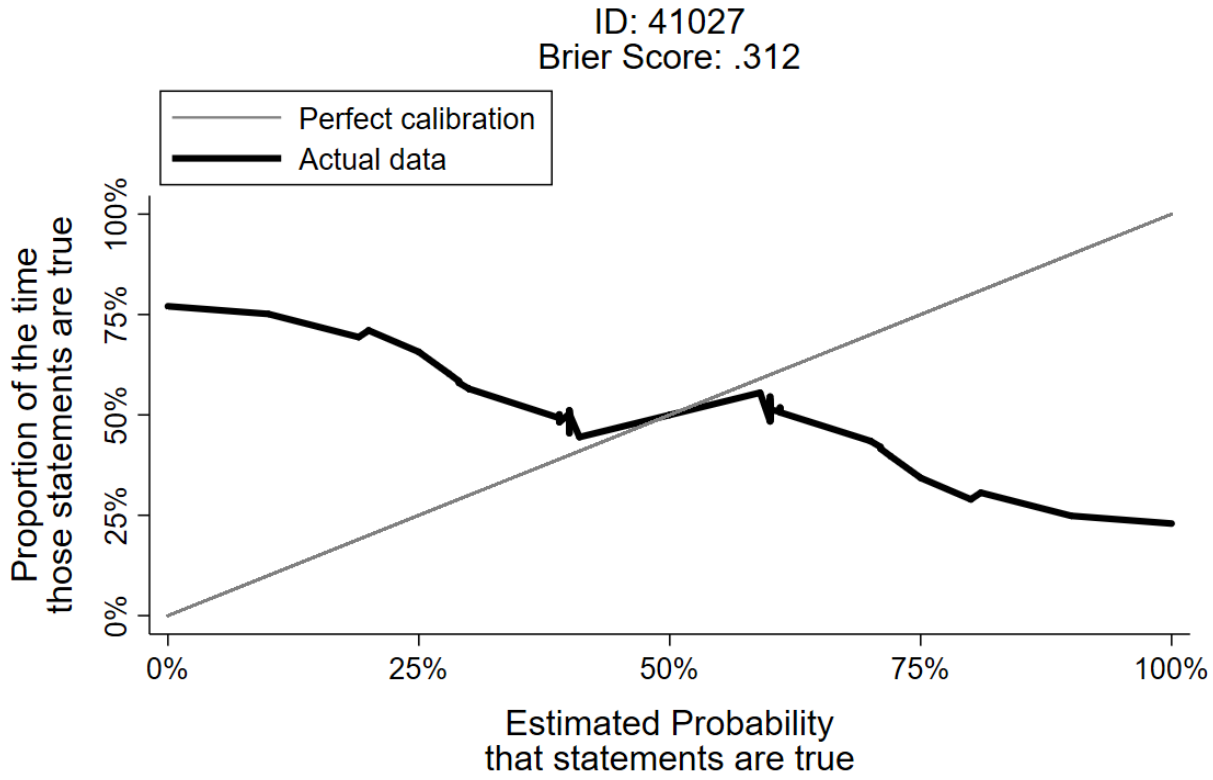

ID: 36515
Brier Score: .239
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

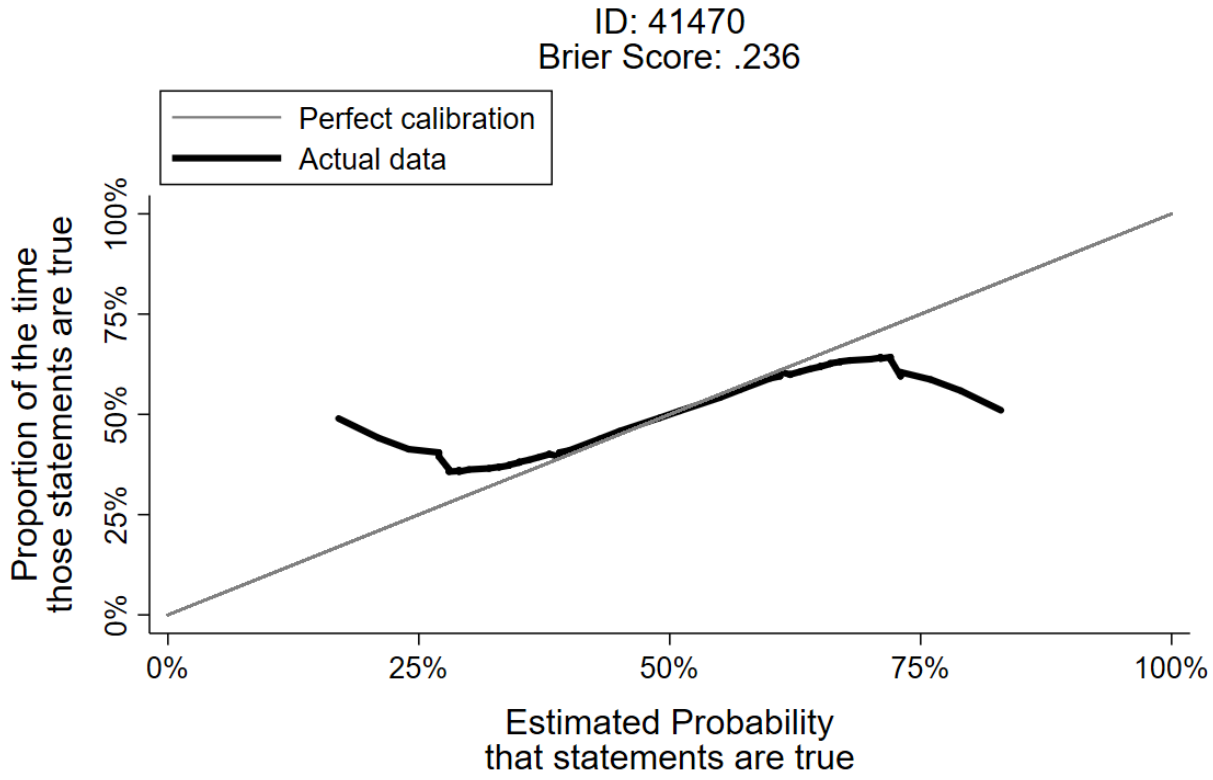

ID: 36791
Brier Score: .257
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

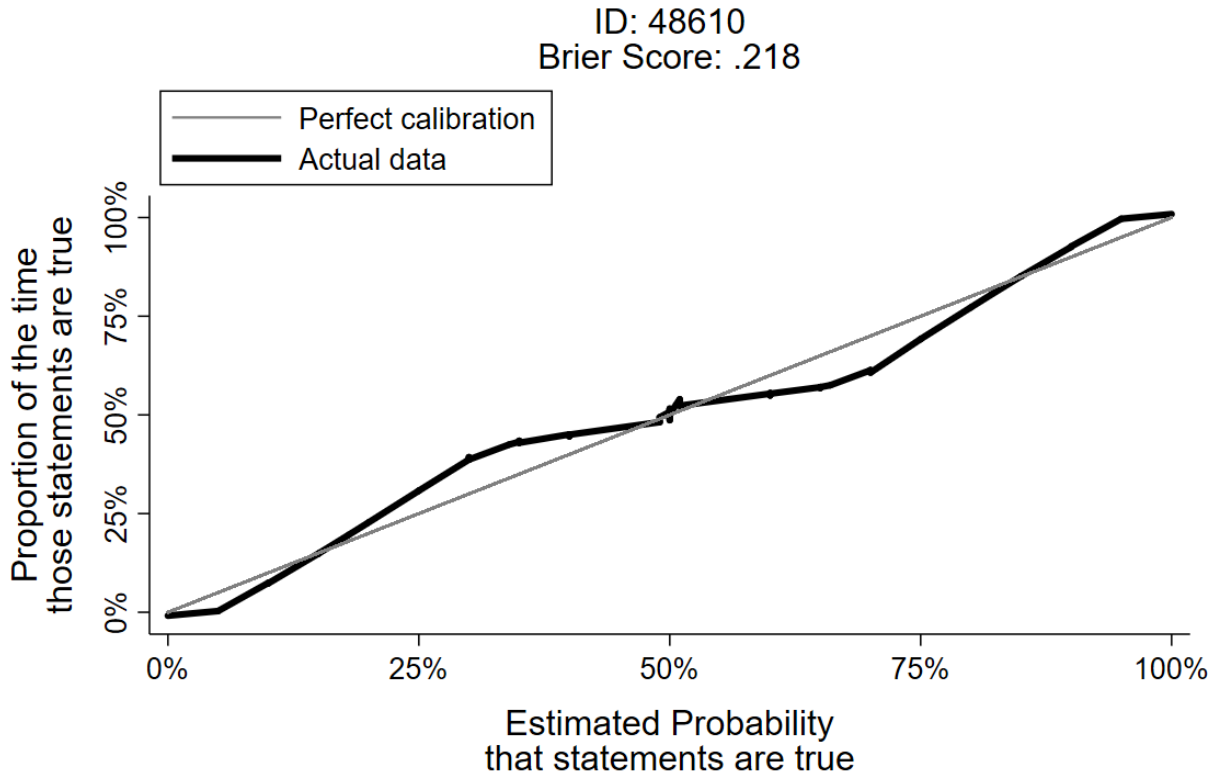

ID: 37758
Brier Score: .195
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

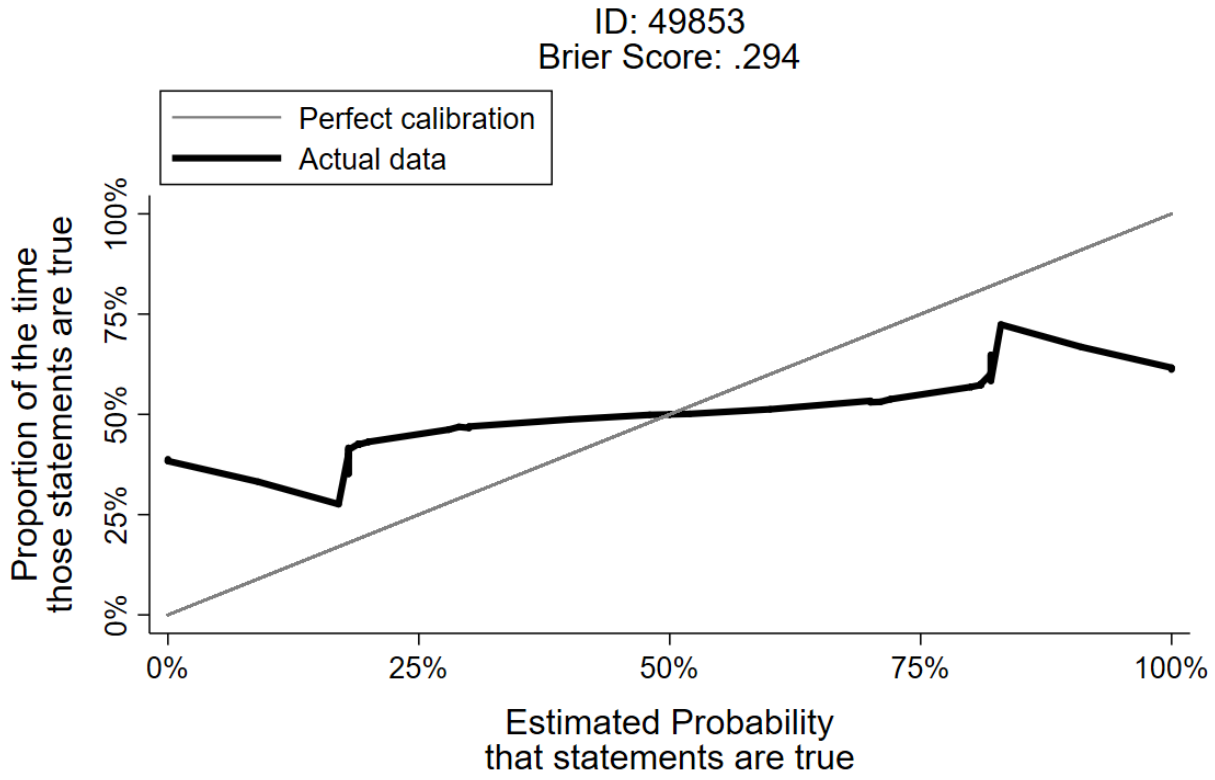

ID: 40200
Brier Score: .285
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

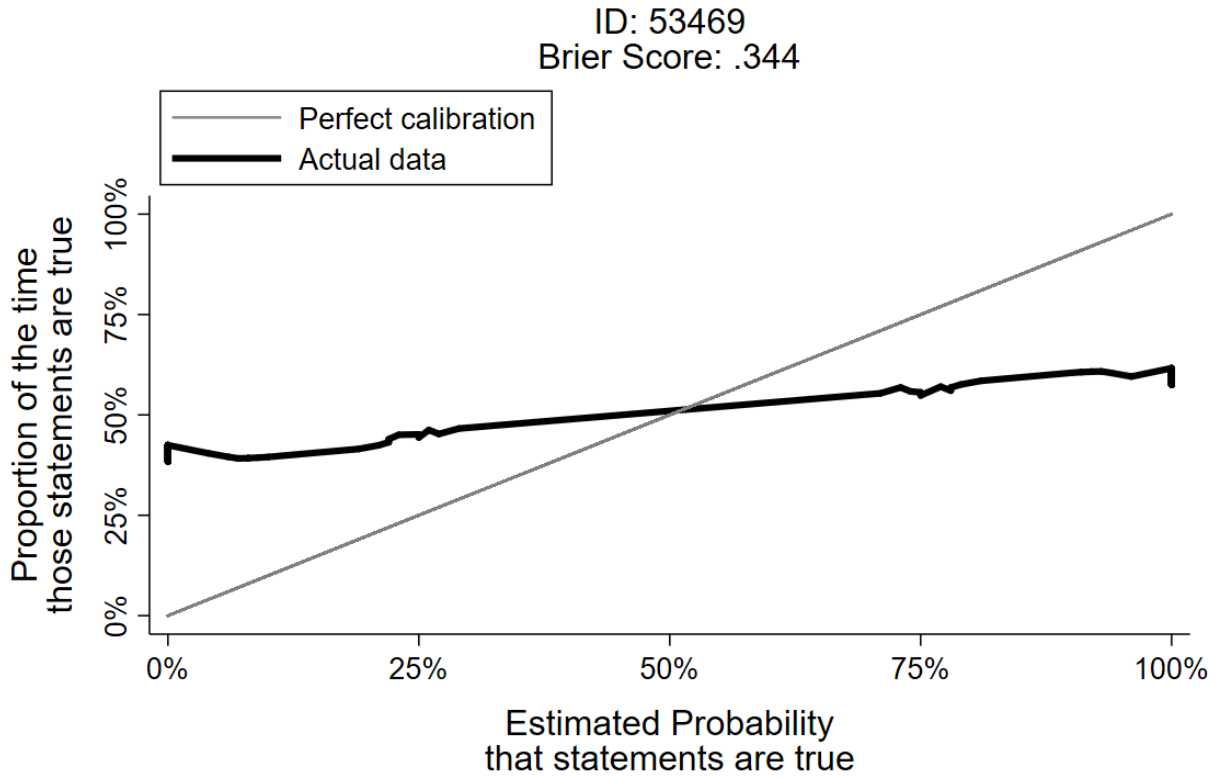

ID: 40937
Brier Score: .379
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

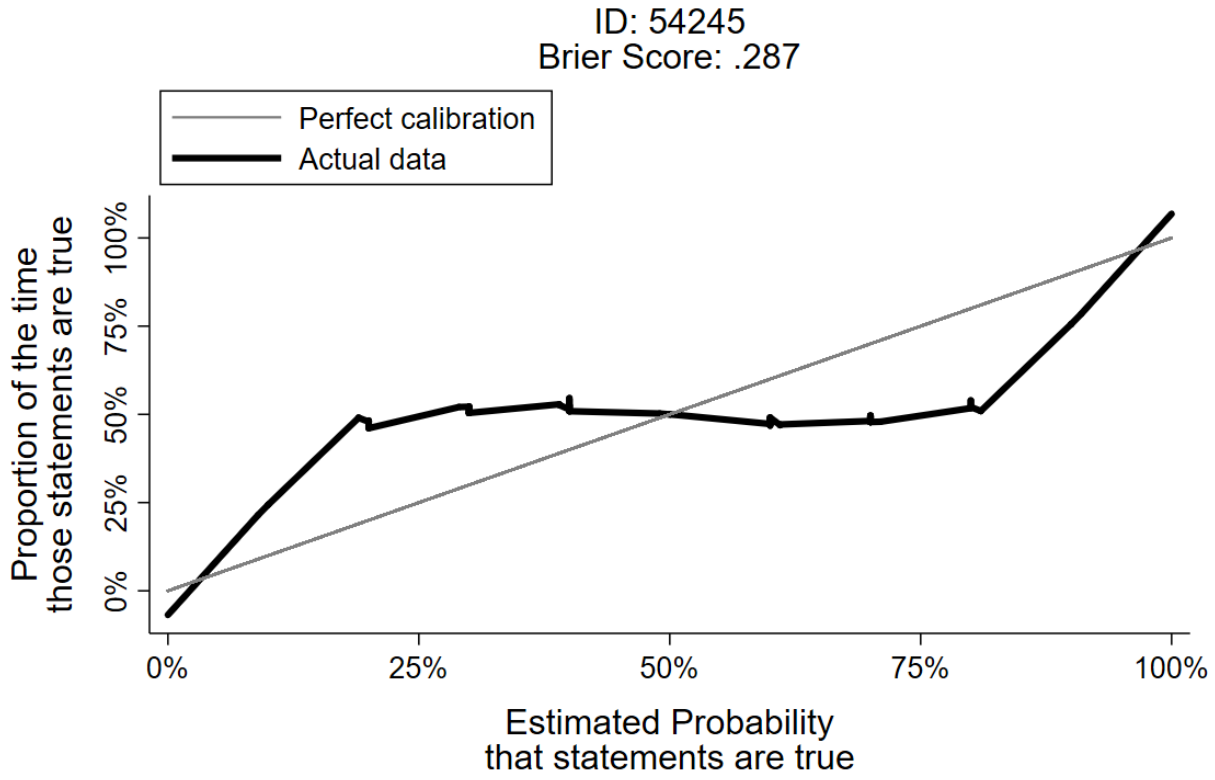

ID: 41027
Brier Score: .312
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

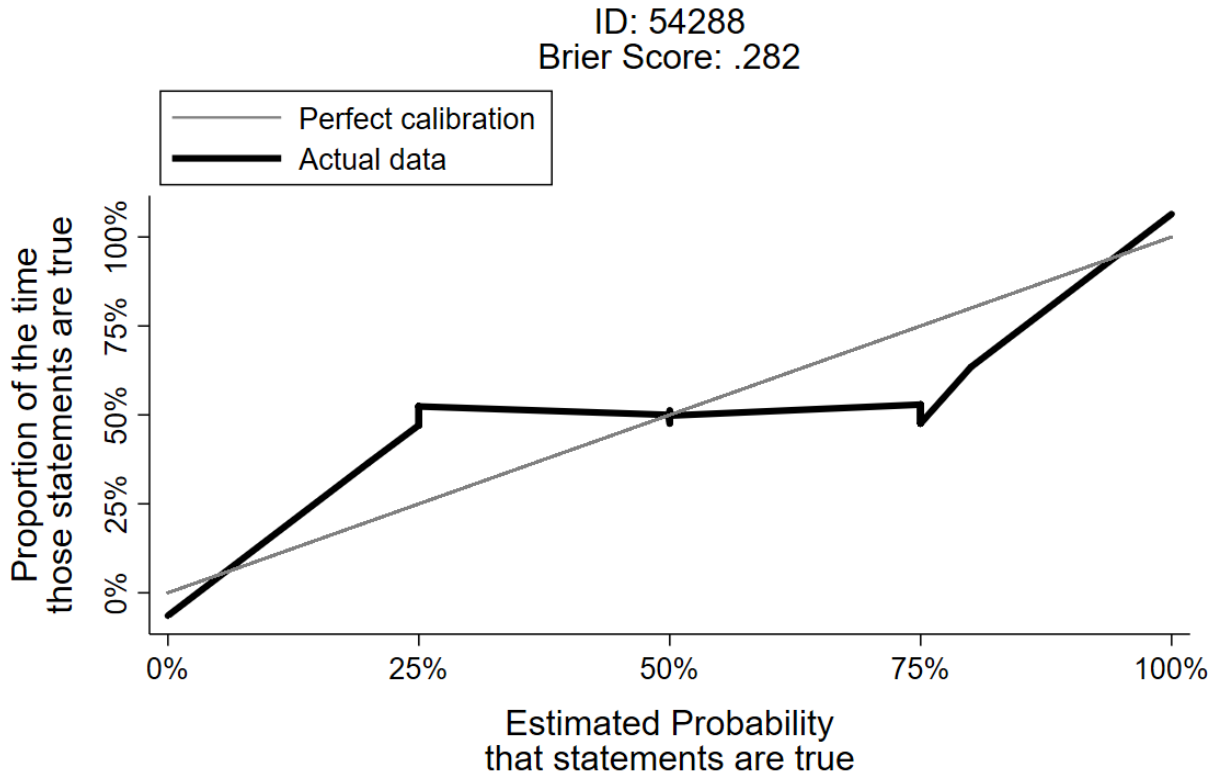

ID: 41470
Brier Score: .236
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

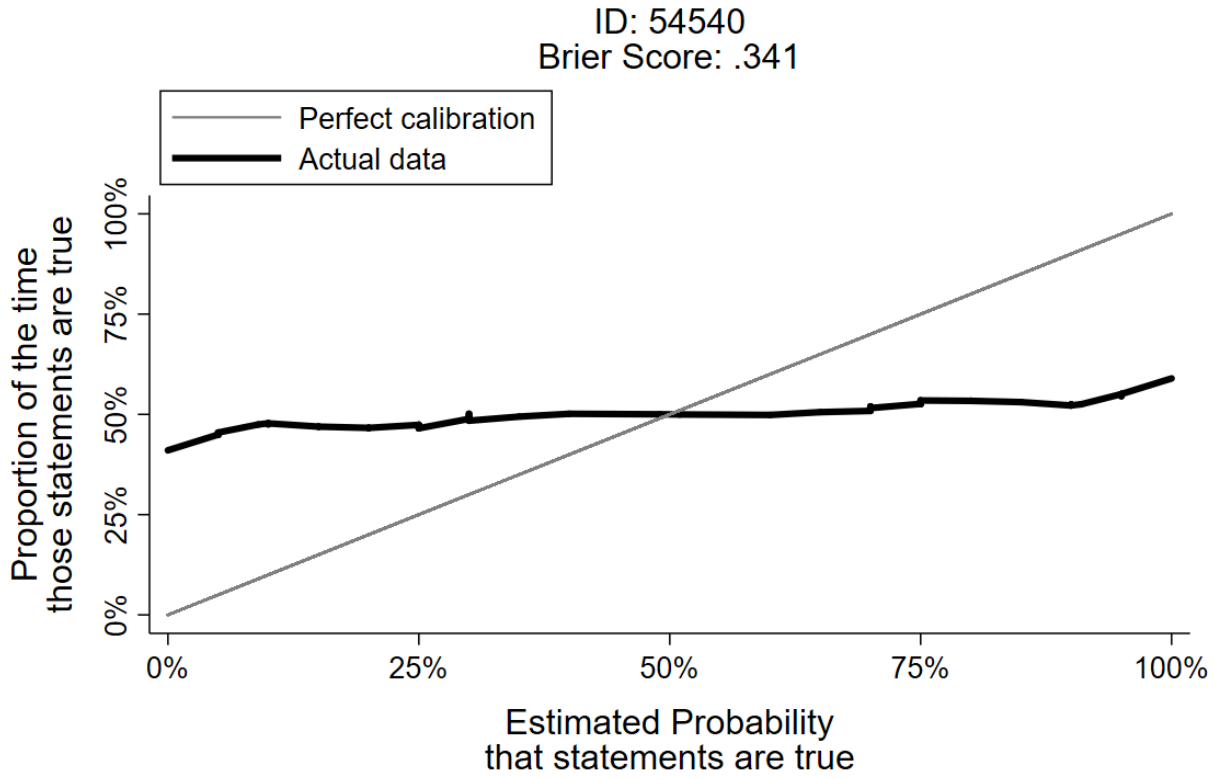

ID: 48610
Brier Score: .218
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

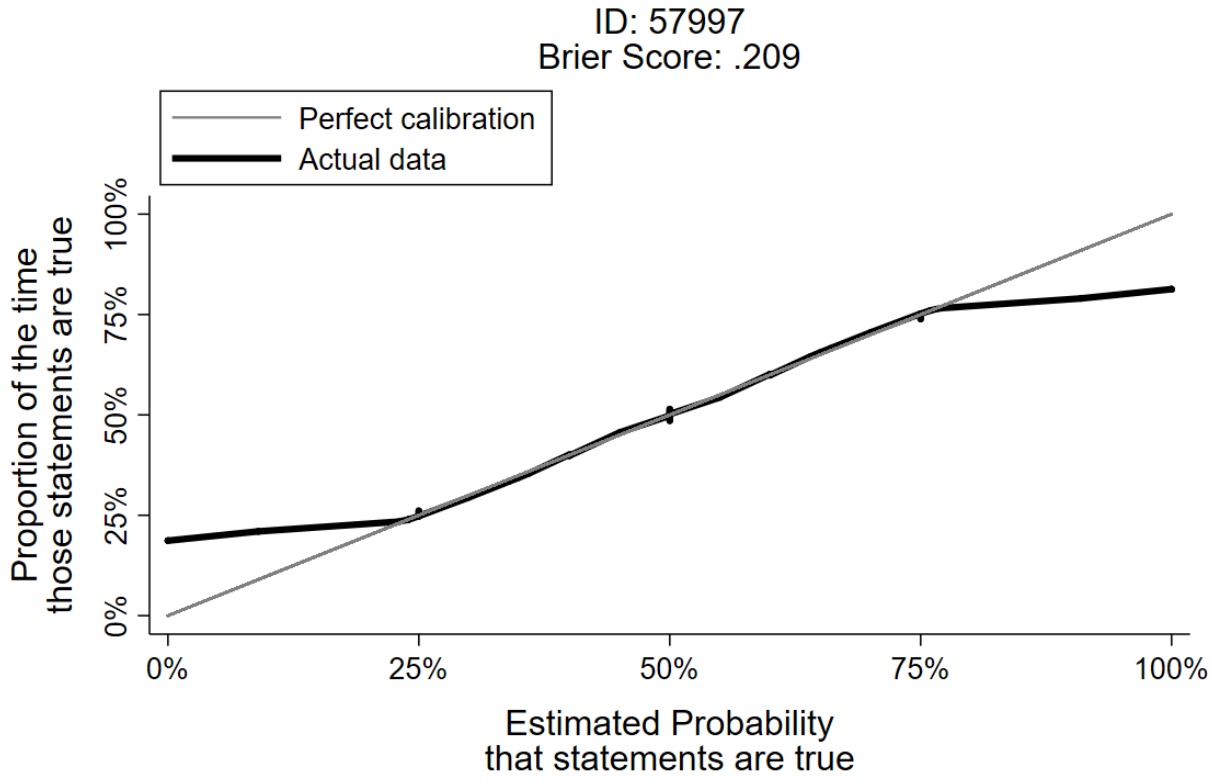

ID: 49853
Brier Score: .294
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

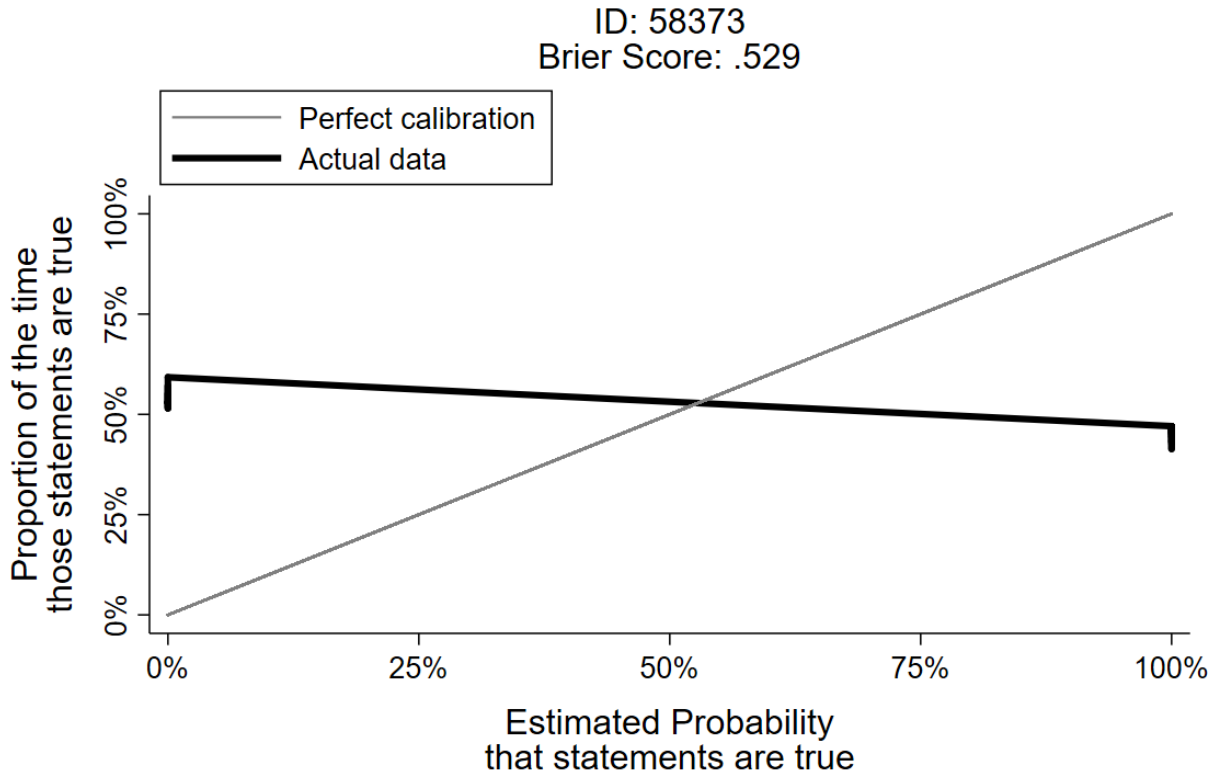

ID: 53469
Brier Score: .344
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

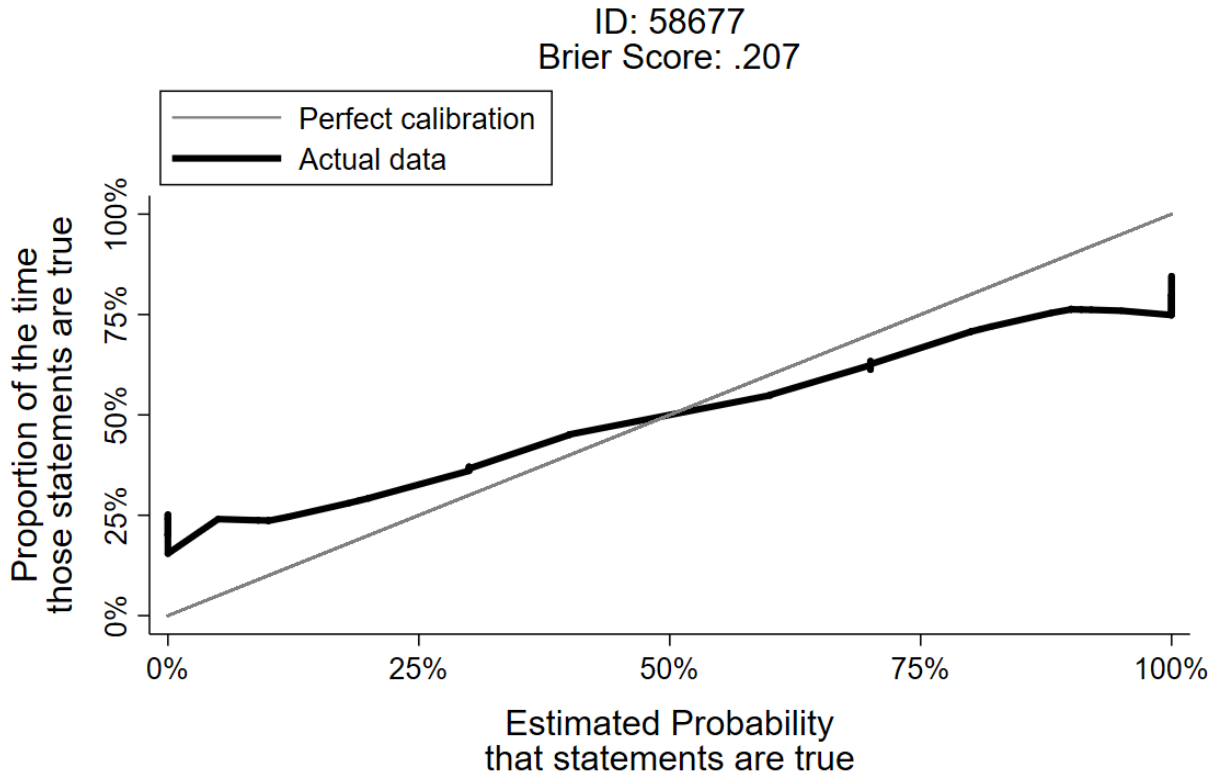

ID: 54245
Brier Score: .287
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

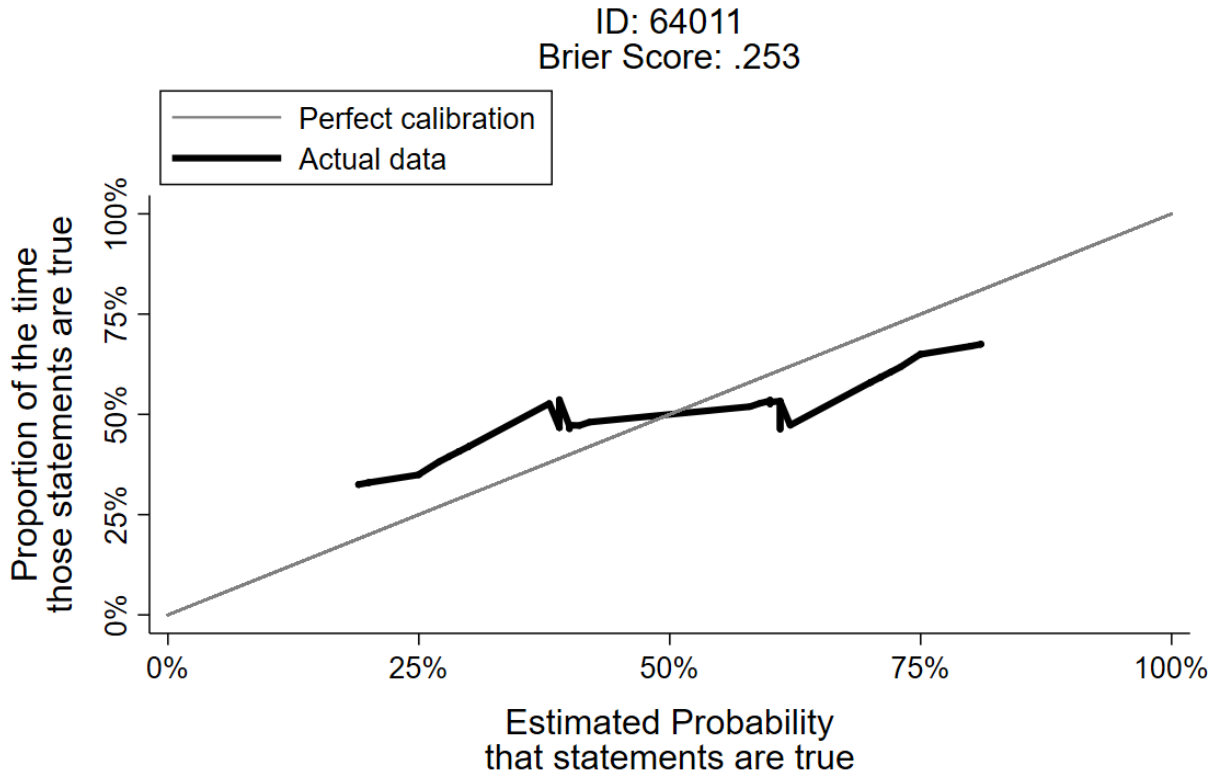

ID: 54288
Brier Score: .282
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

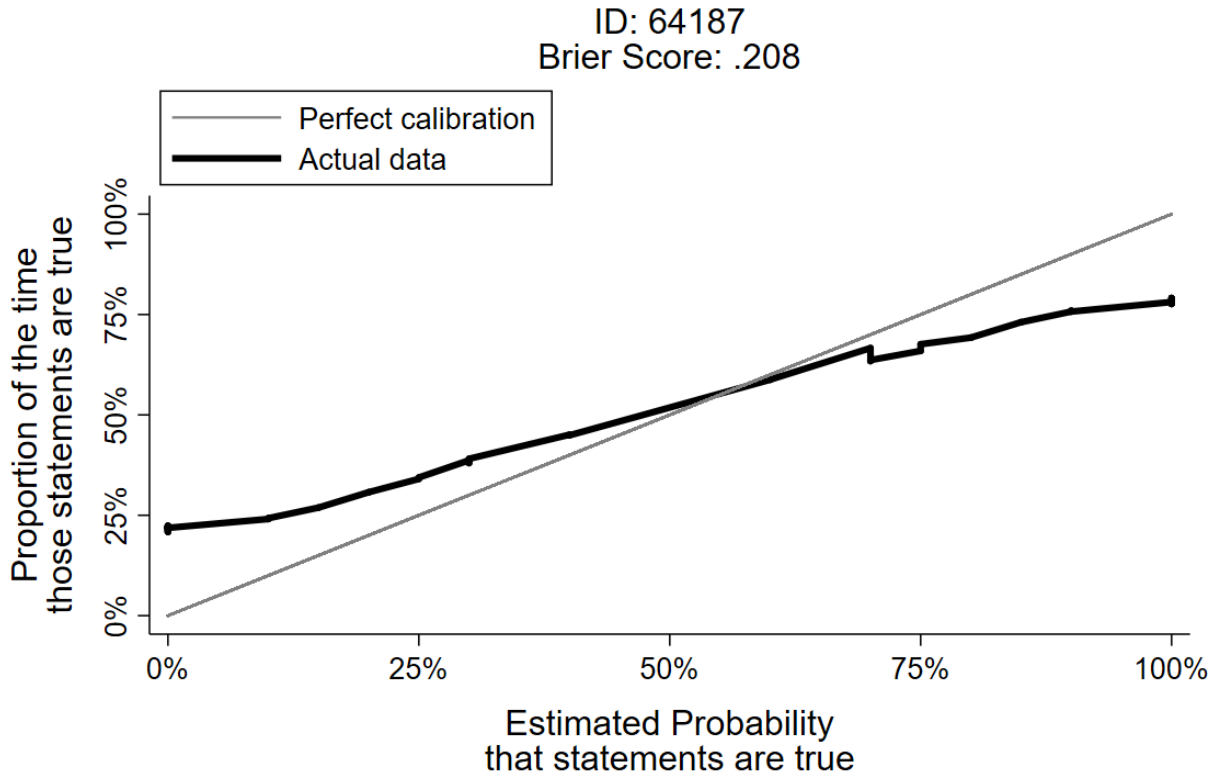

ID: 54540
Brier Score: .341
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

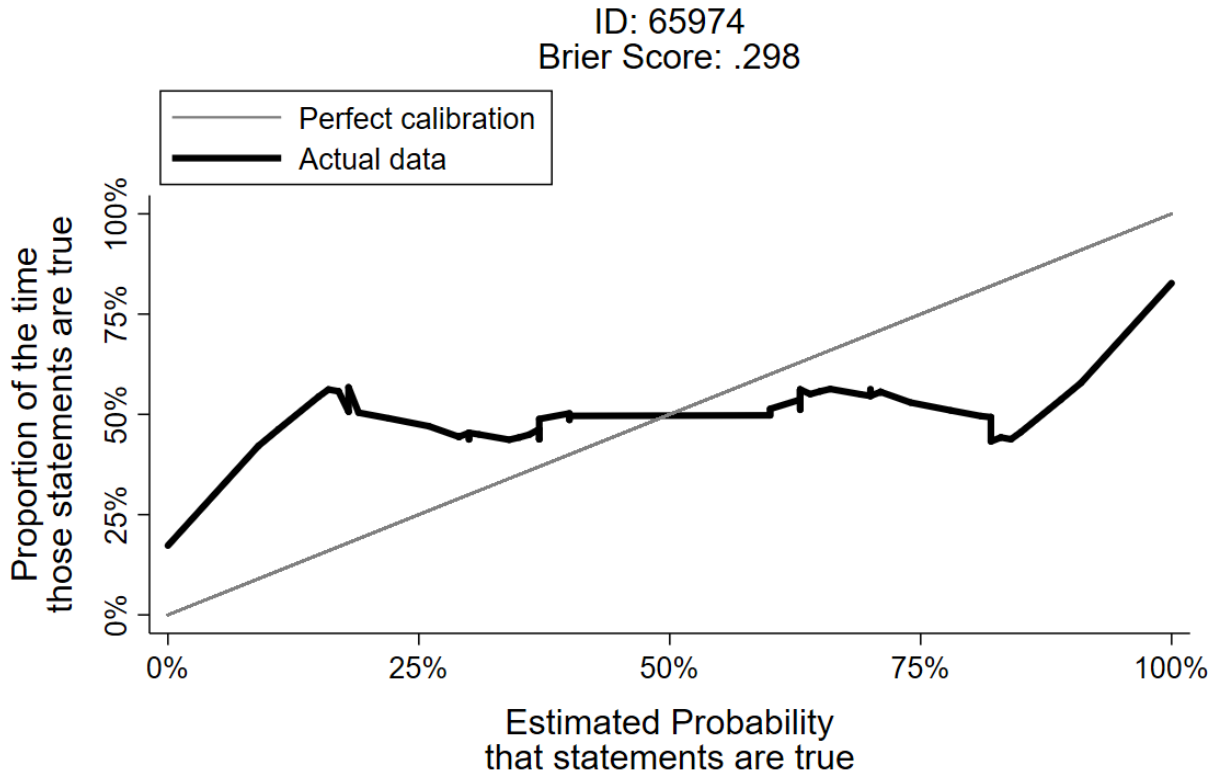

ID: 57997
Brier Score: .209
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

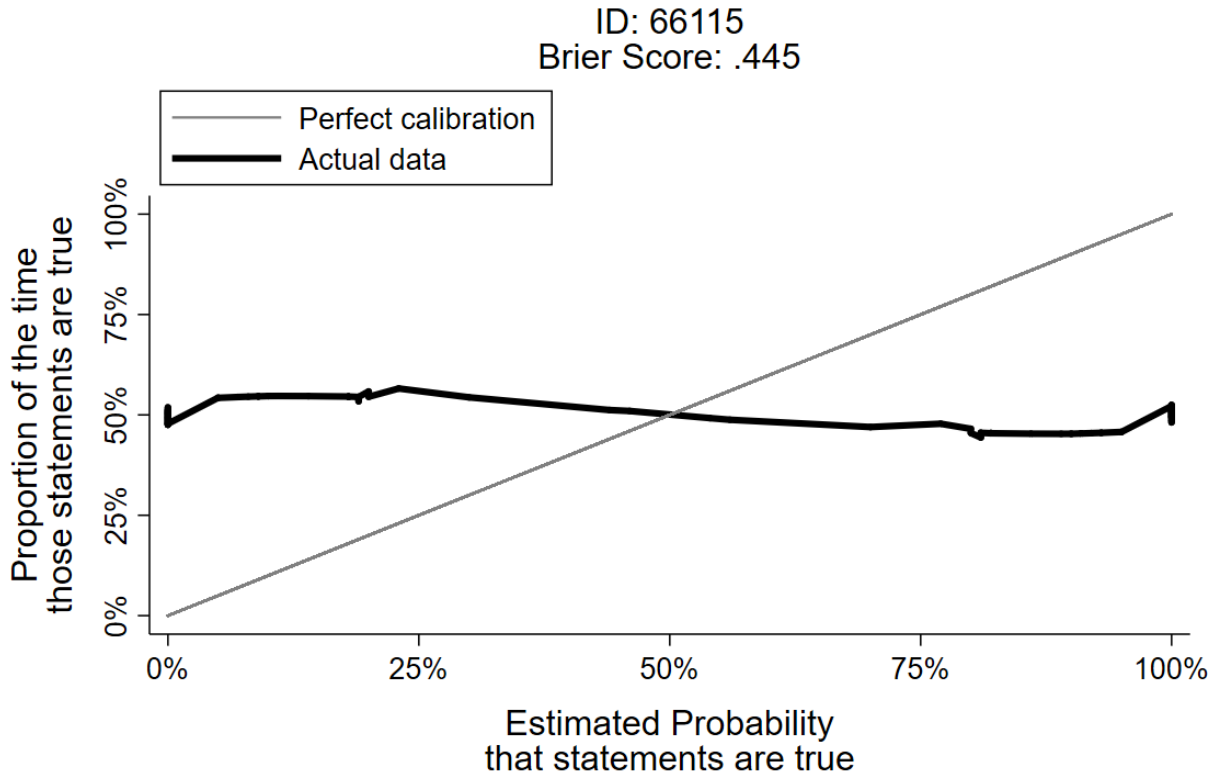

ID: 58373
Brier Score: .529
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

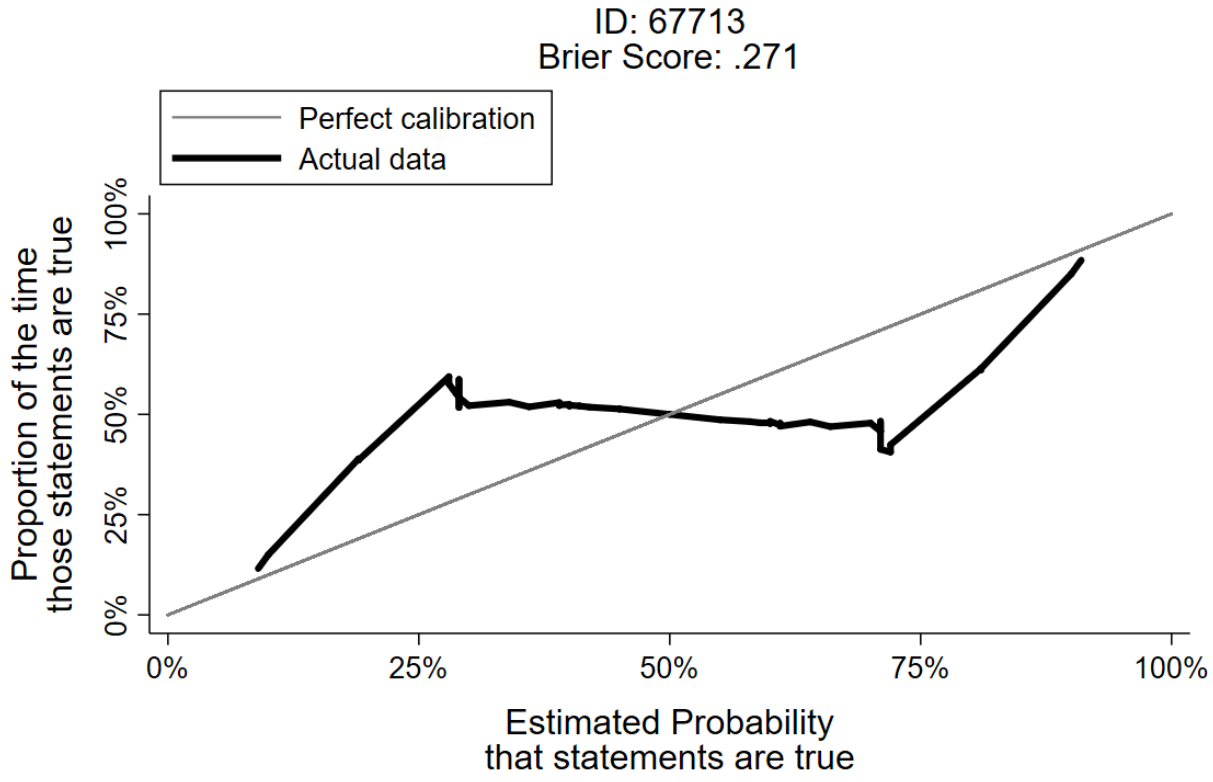

ID: 58677
Brier Score: .207
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

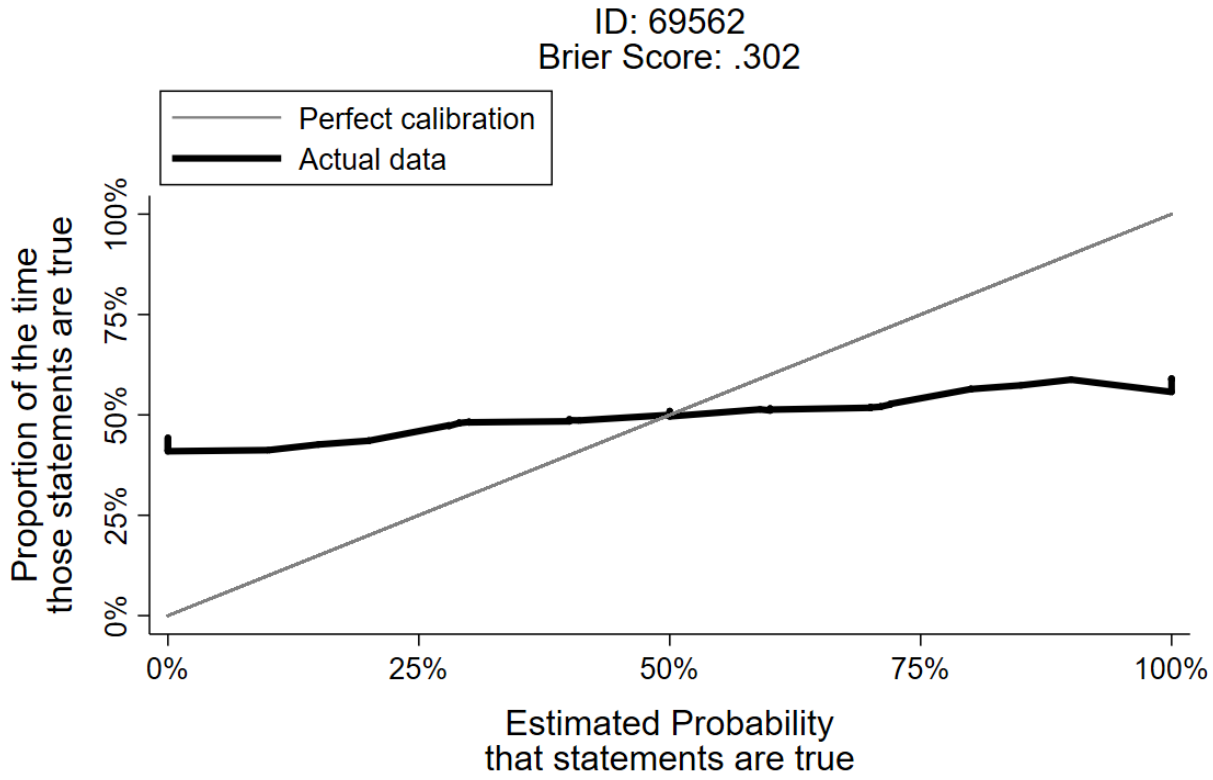

ID: 64011
Brier Score: .253
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

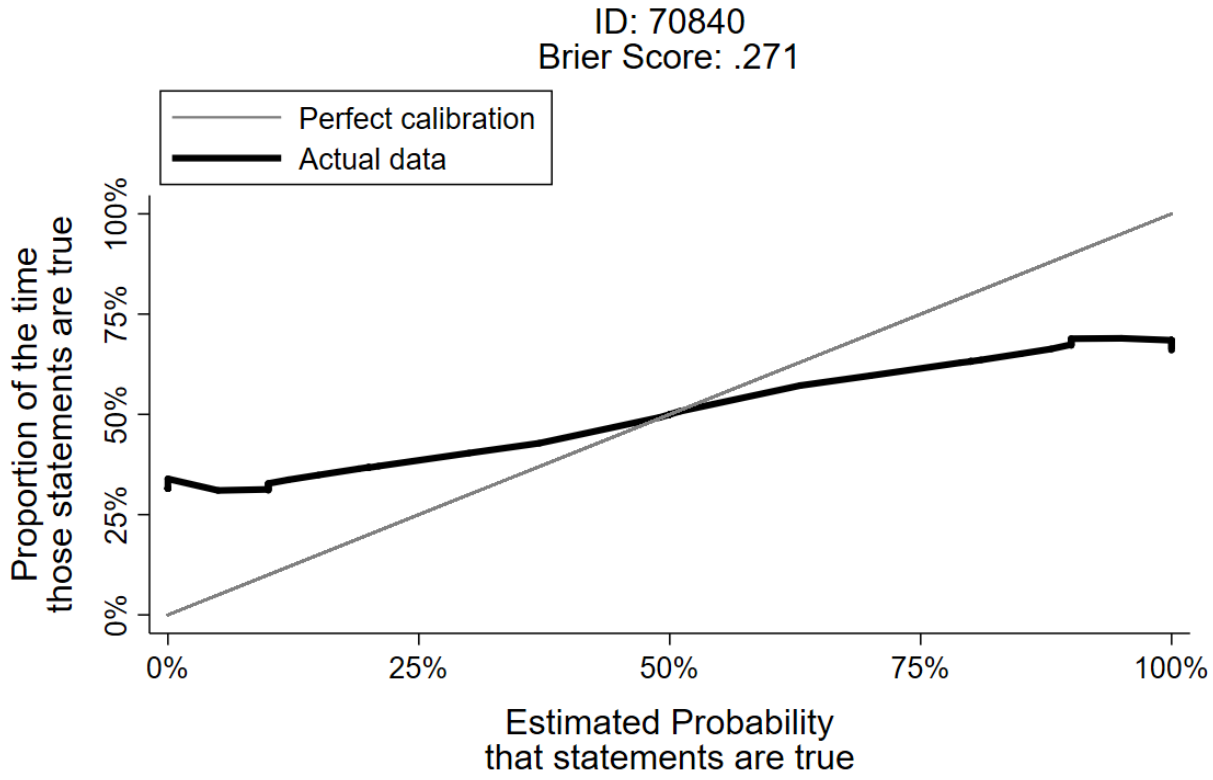

ID: 64187
Brier Score: .208
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

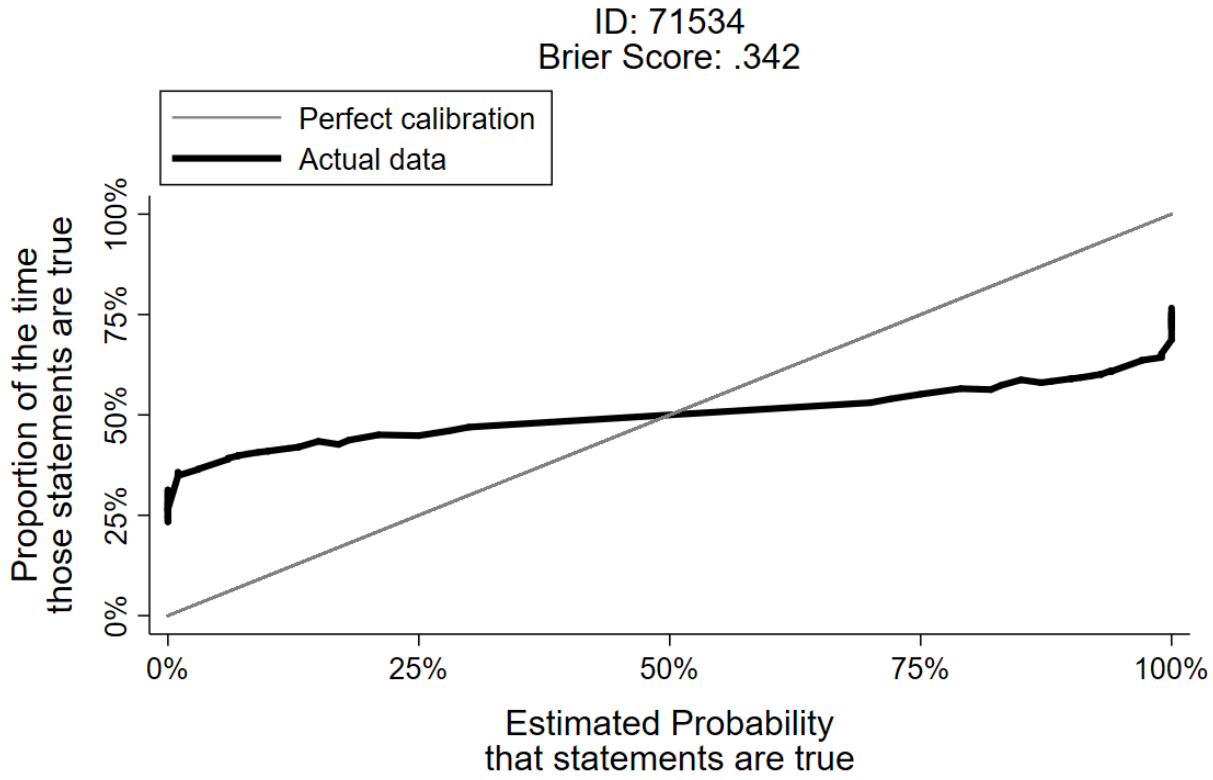

ID: 65974
Brier Score: .298
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

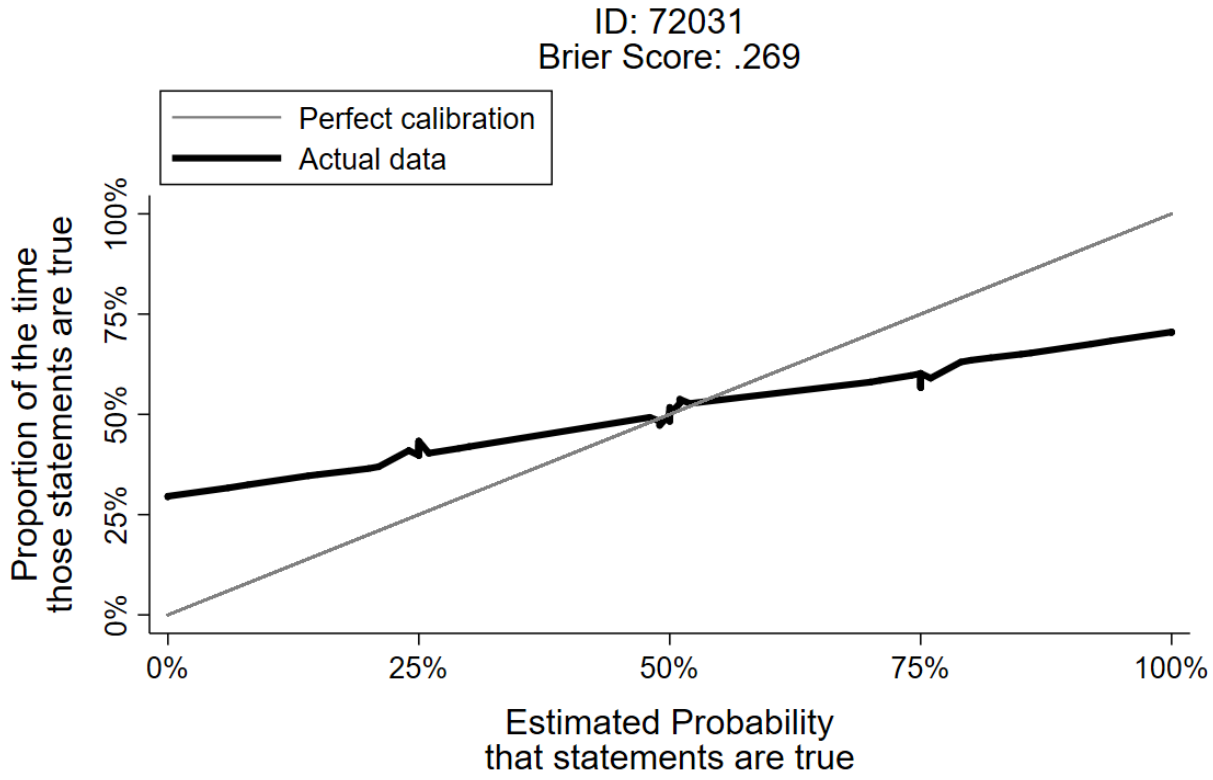

ID: 66115
Brier Score: .445
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

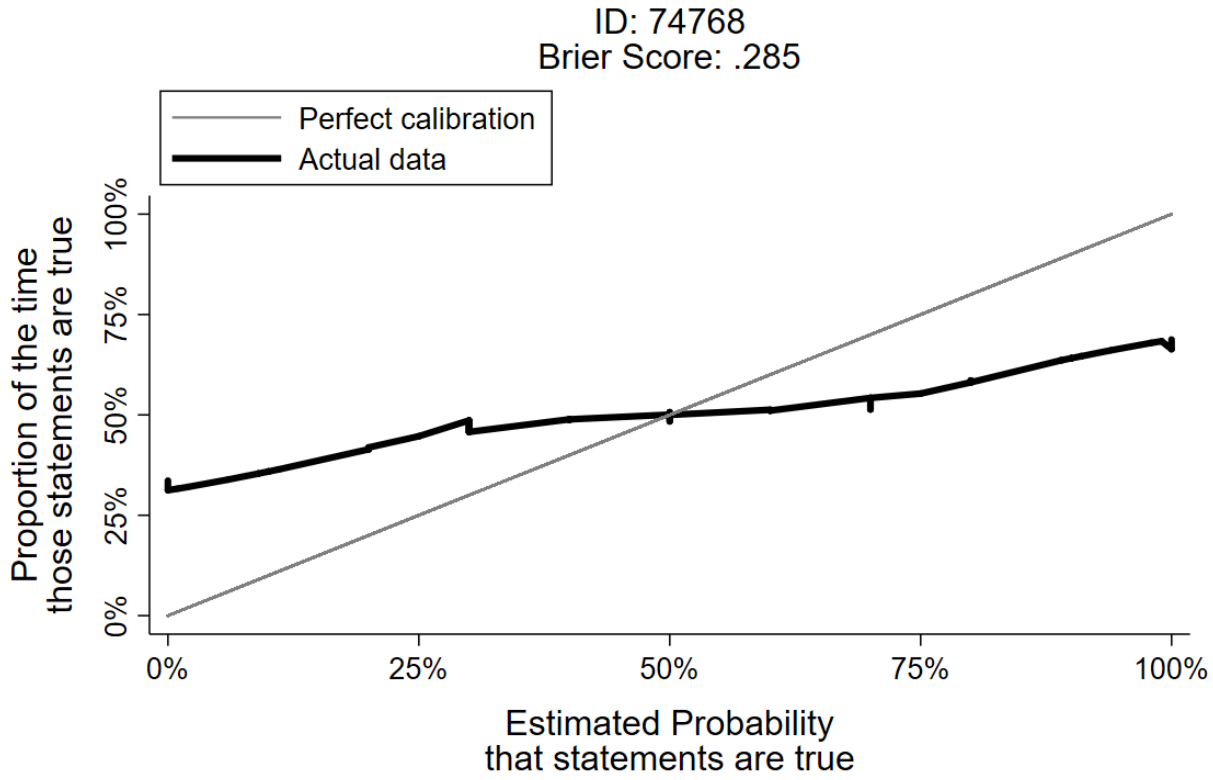

ID: 67713
Brier Score: .271
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

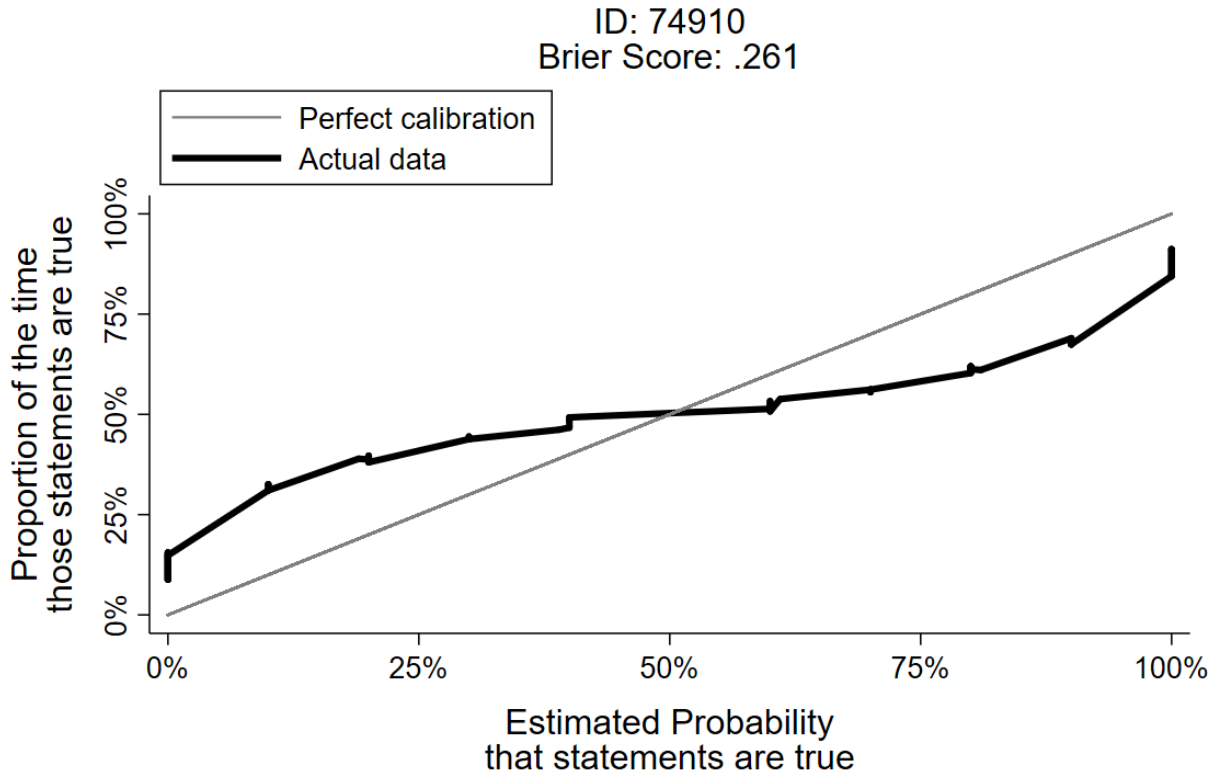

ID: 69562
Brier Score: .302
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

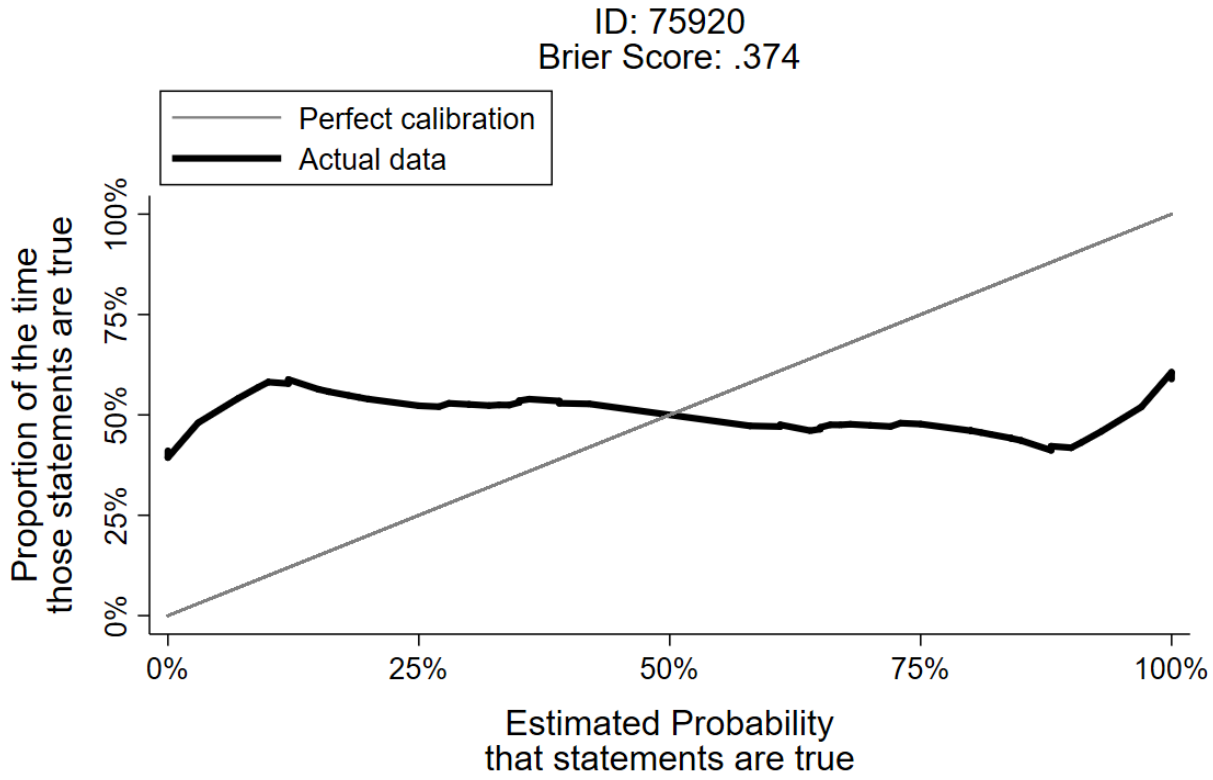

ID: 70840
Brier Score: .271
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

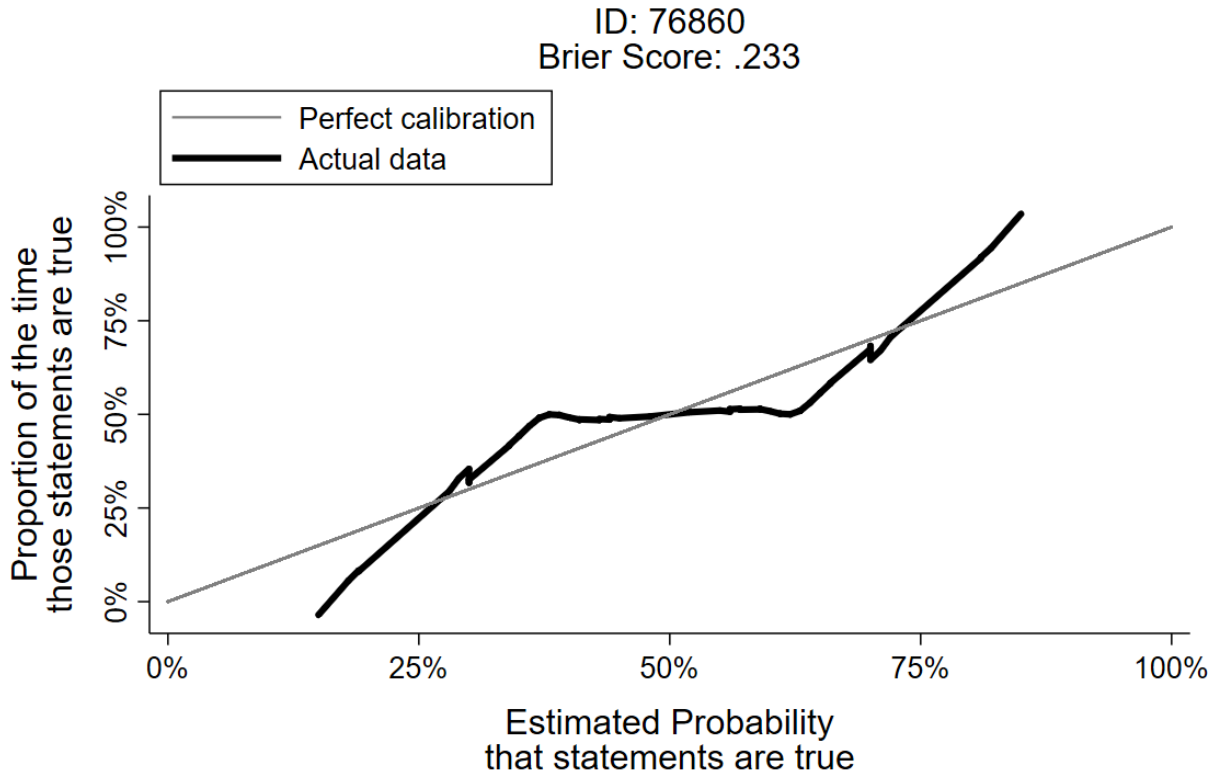

ID: 71534
Brier Score: .342
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

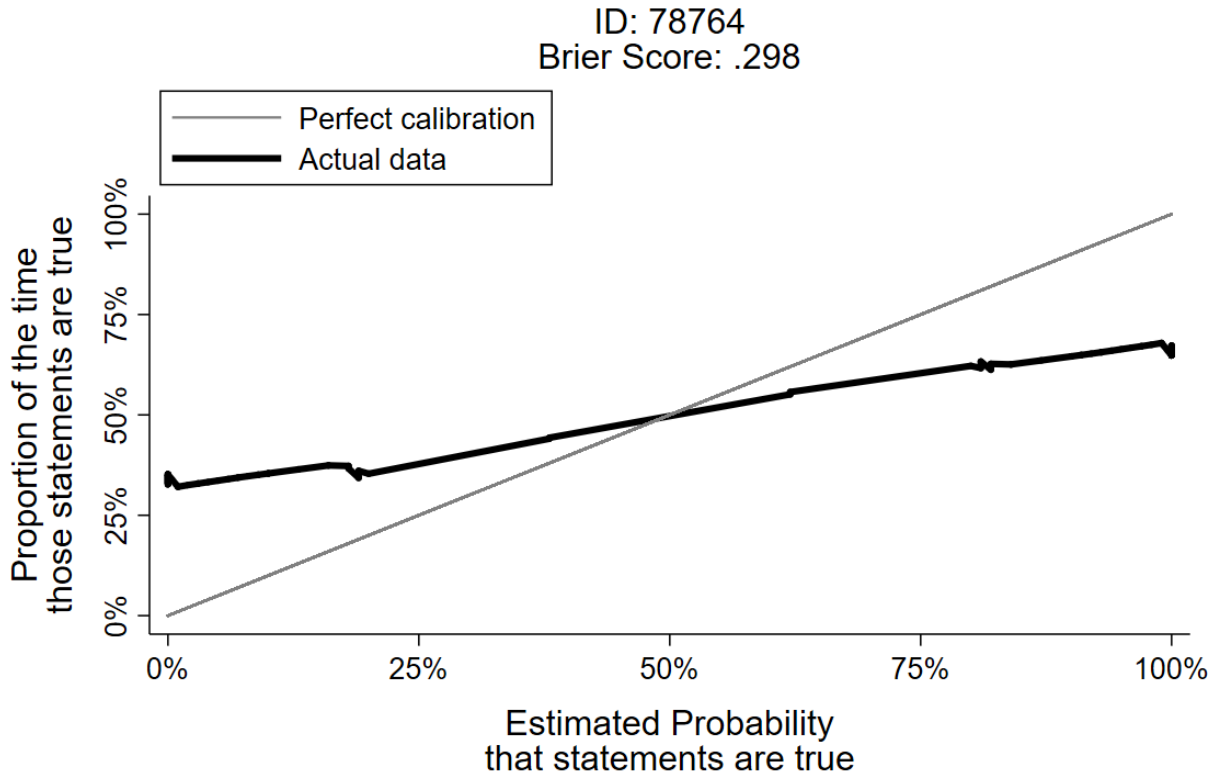

ID: 72031
Brier Score: .269
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

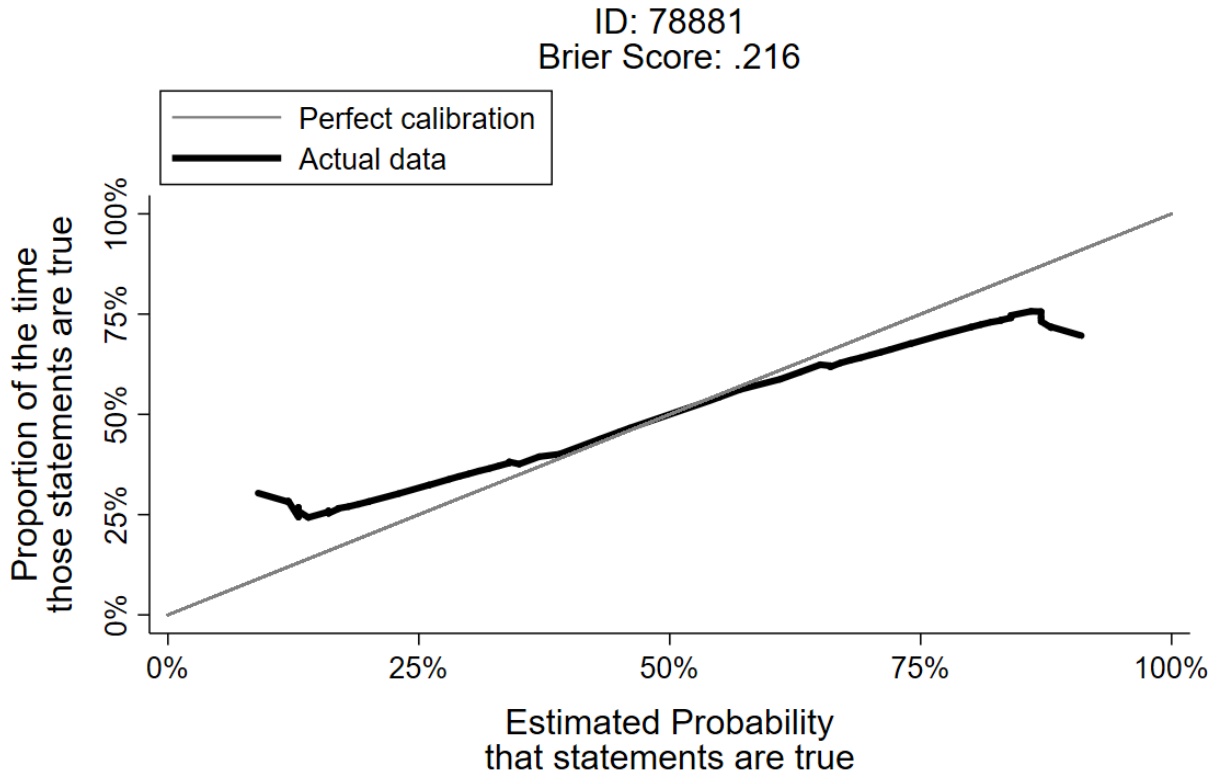

ID: 74768
Brier Score: .285
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

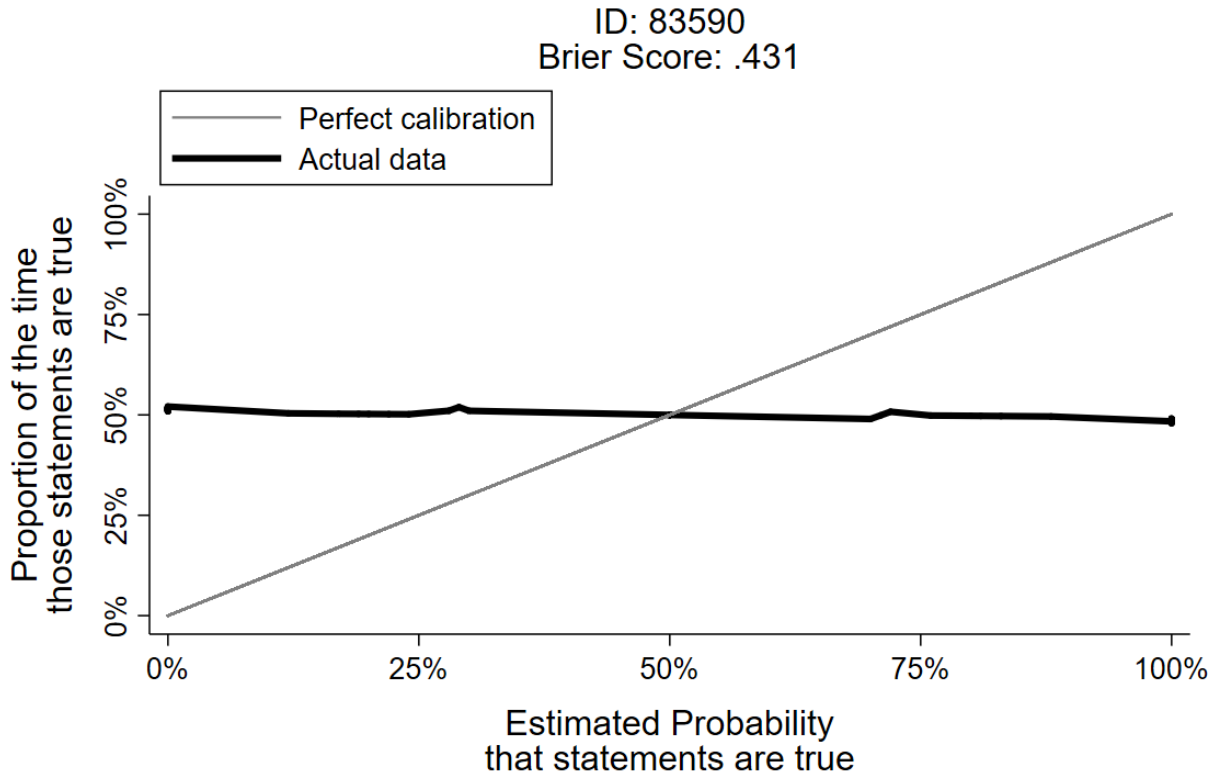

ID: 74910
Brier Score: .261
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

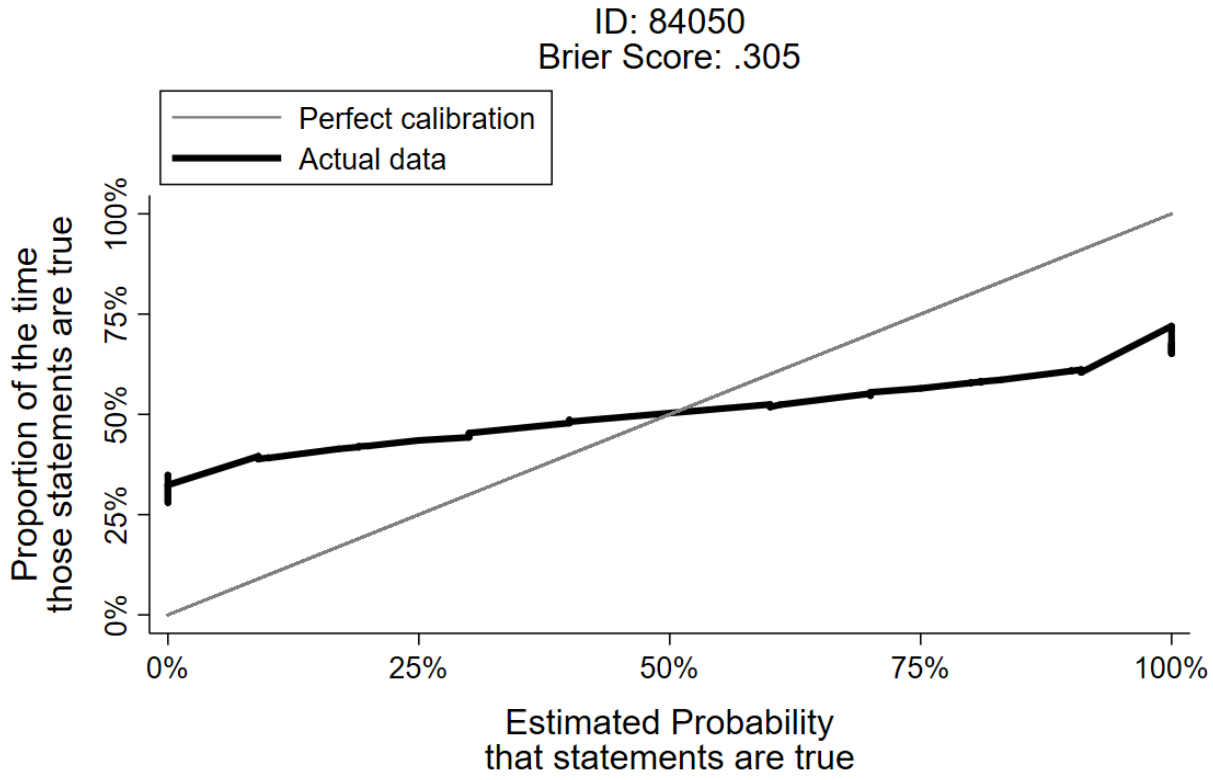

ID: 75920
Brier Score: .374
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

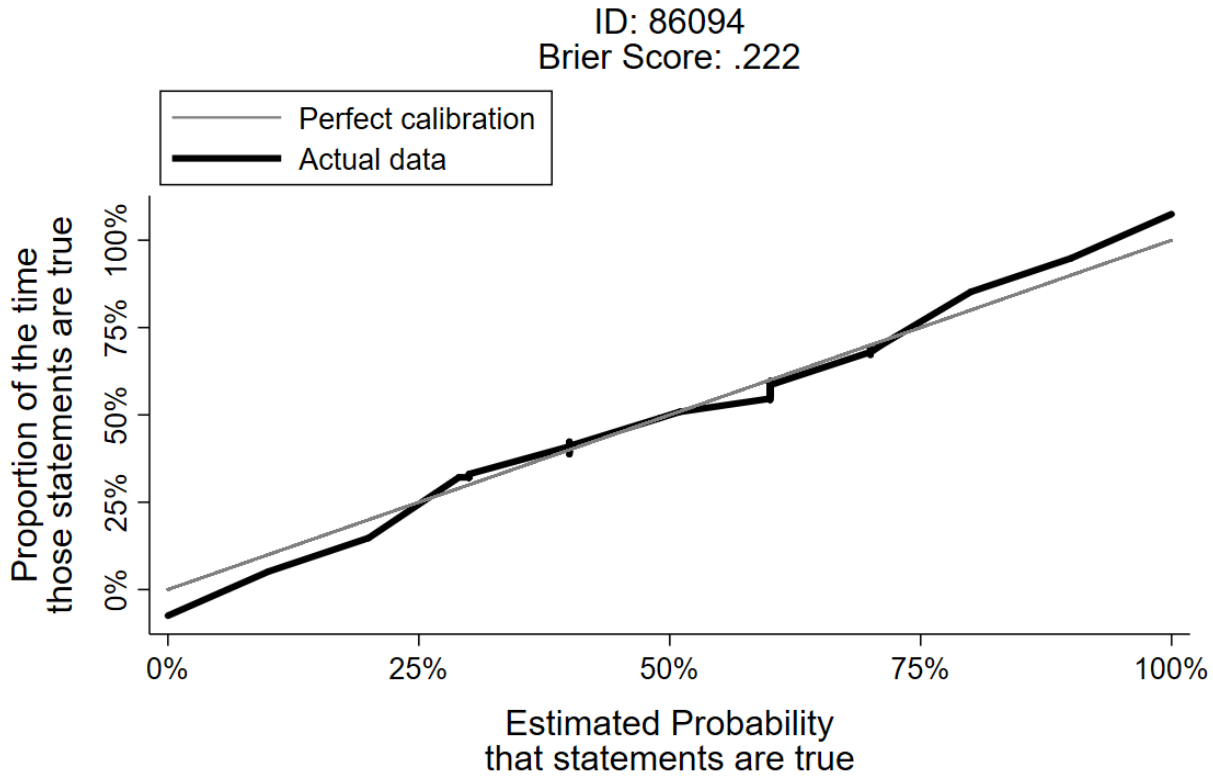

ID: 76860
Brier Score: .233
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

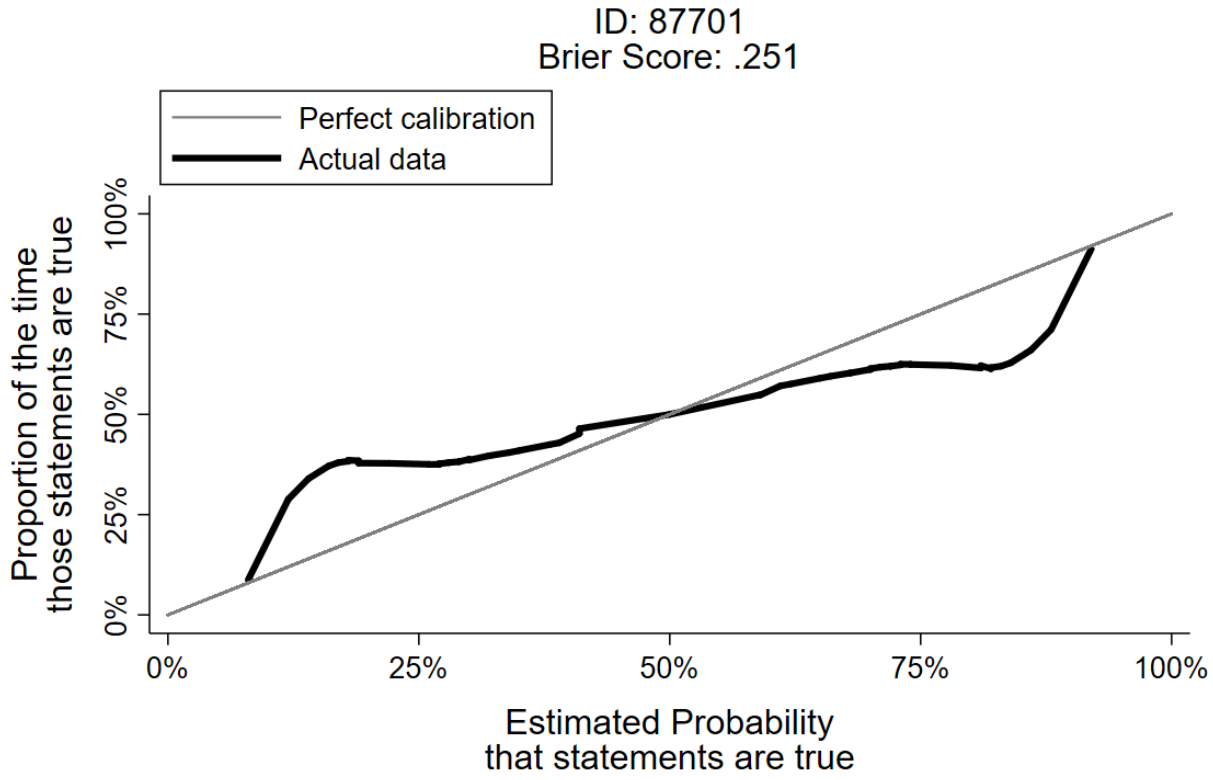

ID: 78764
Brier Score: .298
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

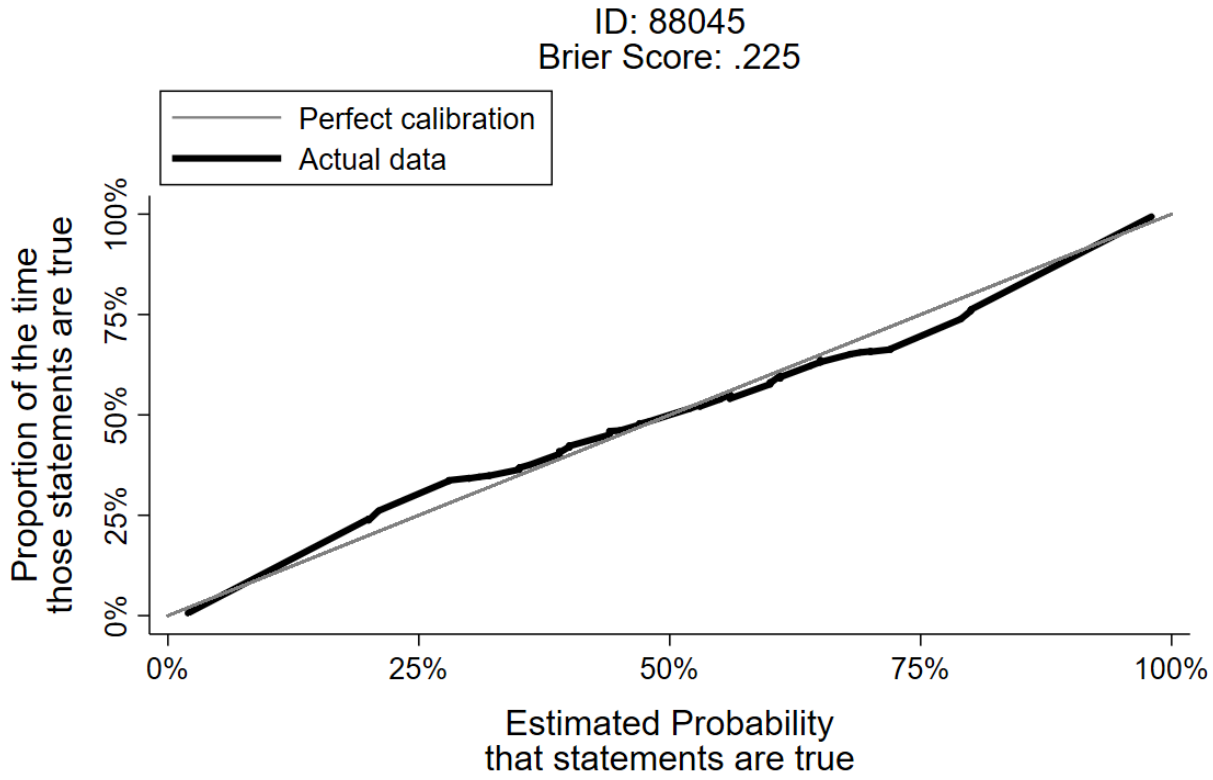

ID: 78881
Brier Score: .216
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

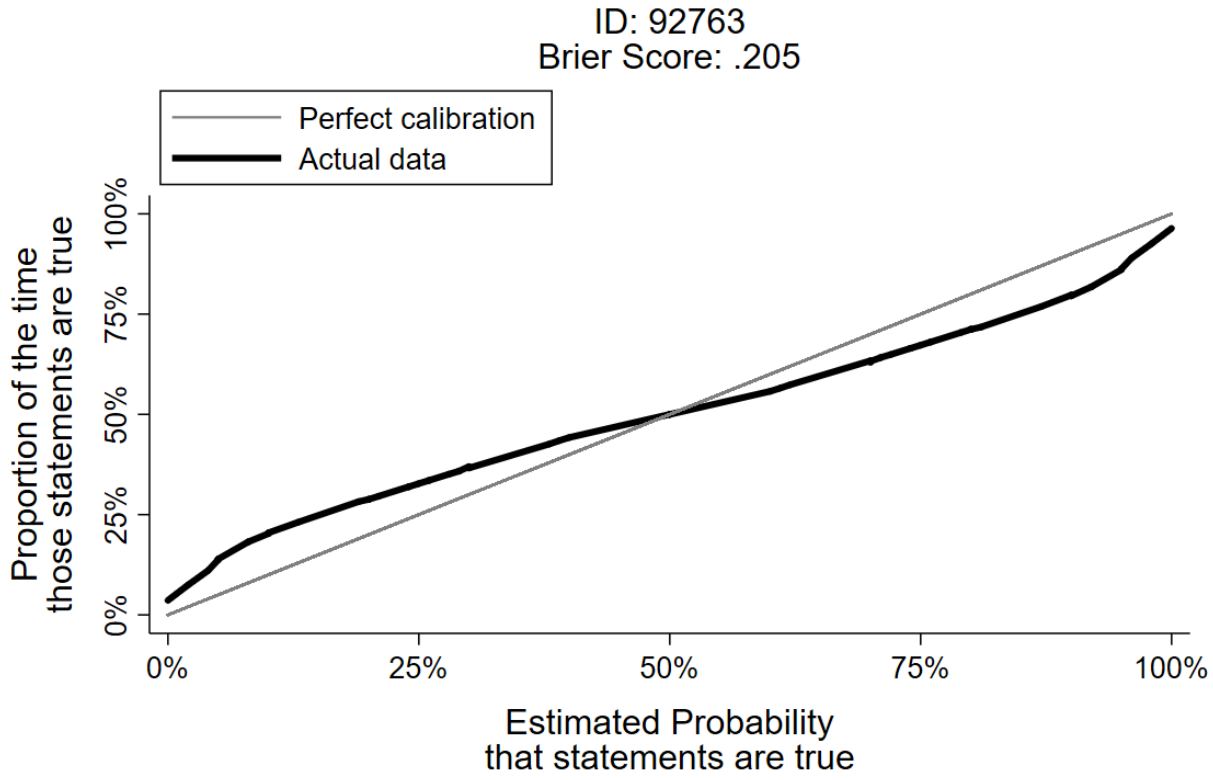

ID: 83590
Brier Score: .431
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

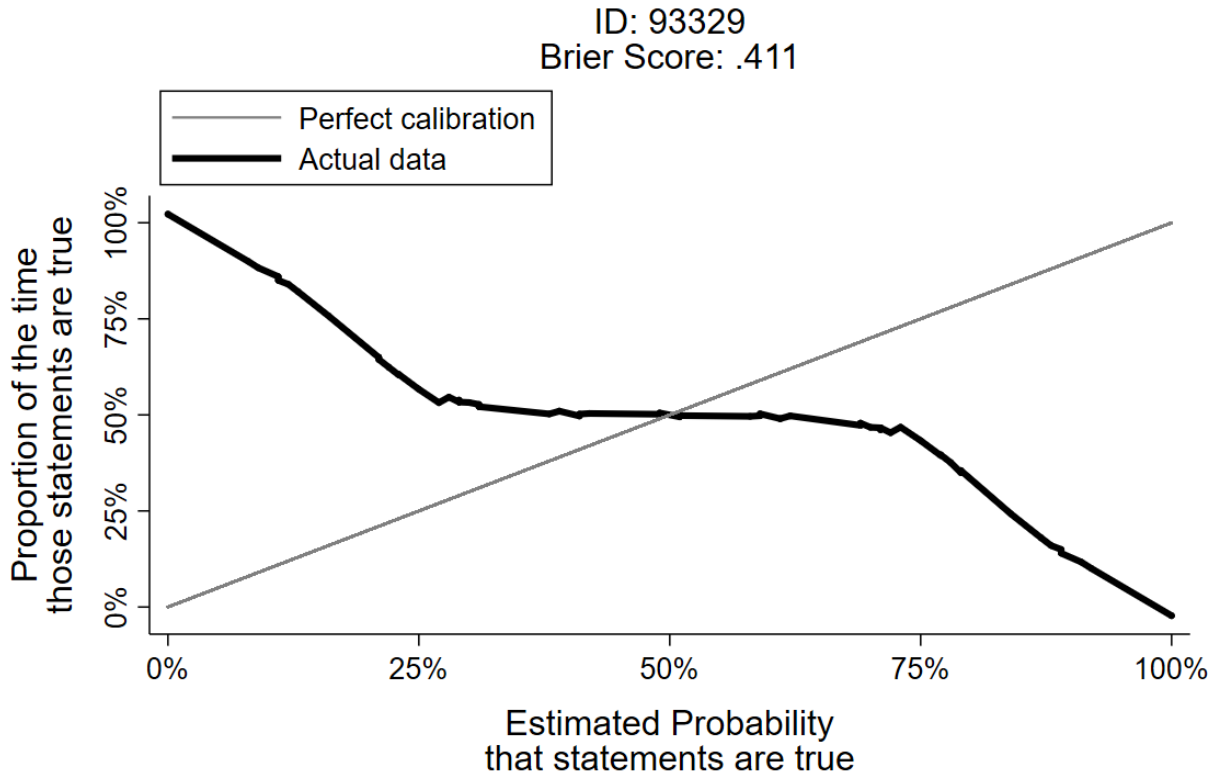

ID: 84050
Brier Score: .305
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

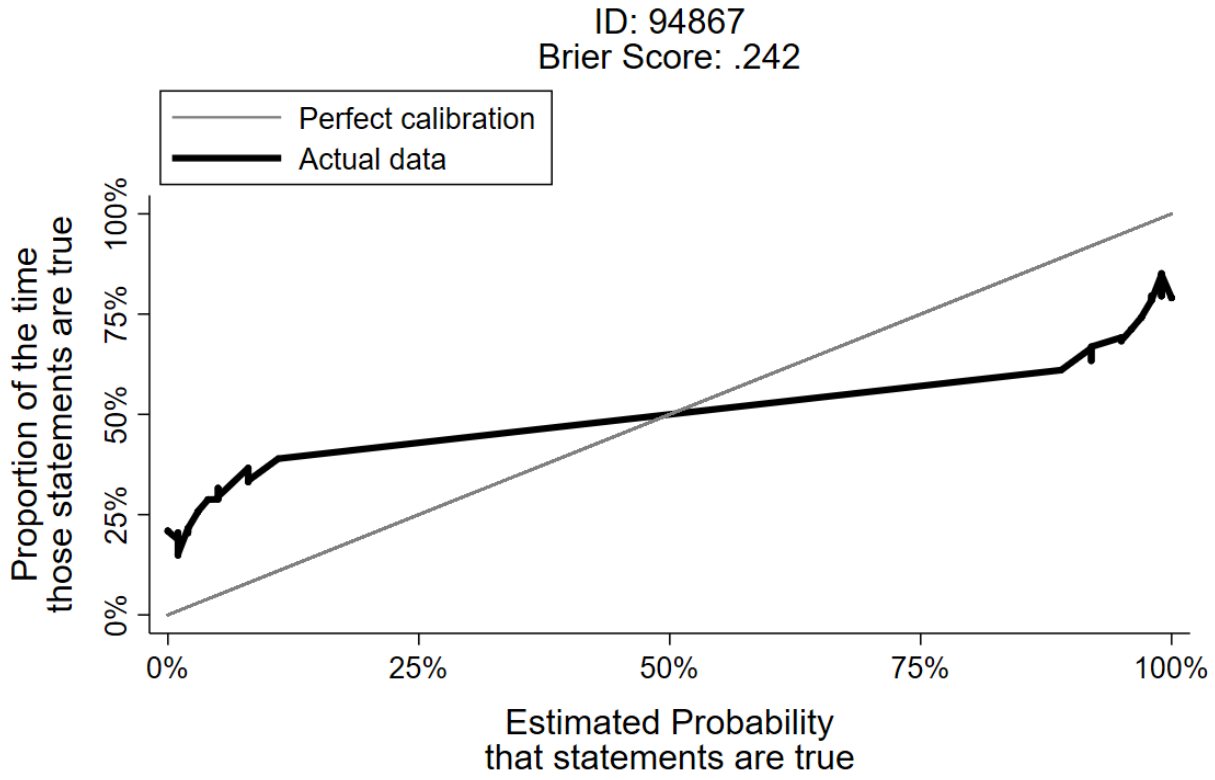

ID: 86094
Brier Score: .222
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

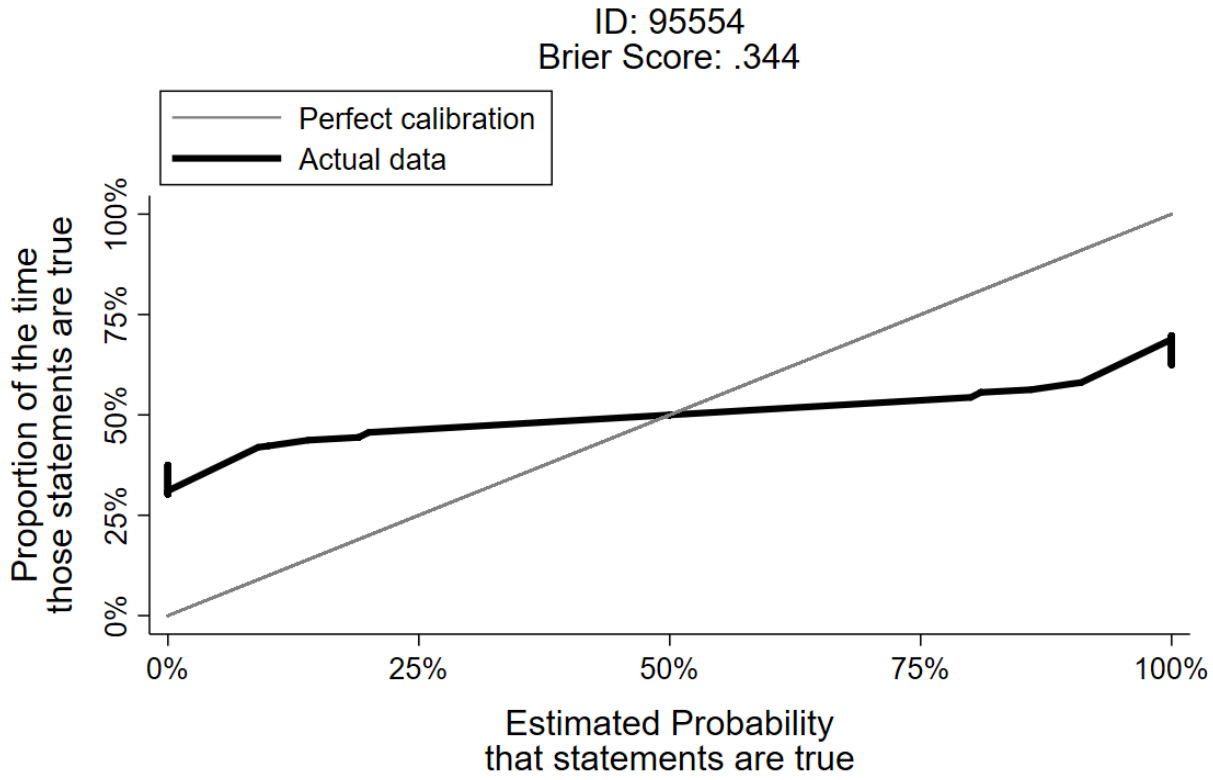

ID: 87701
Brier Score: .251
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

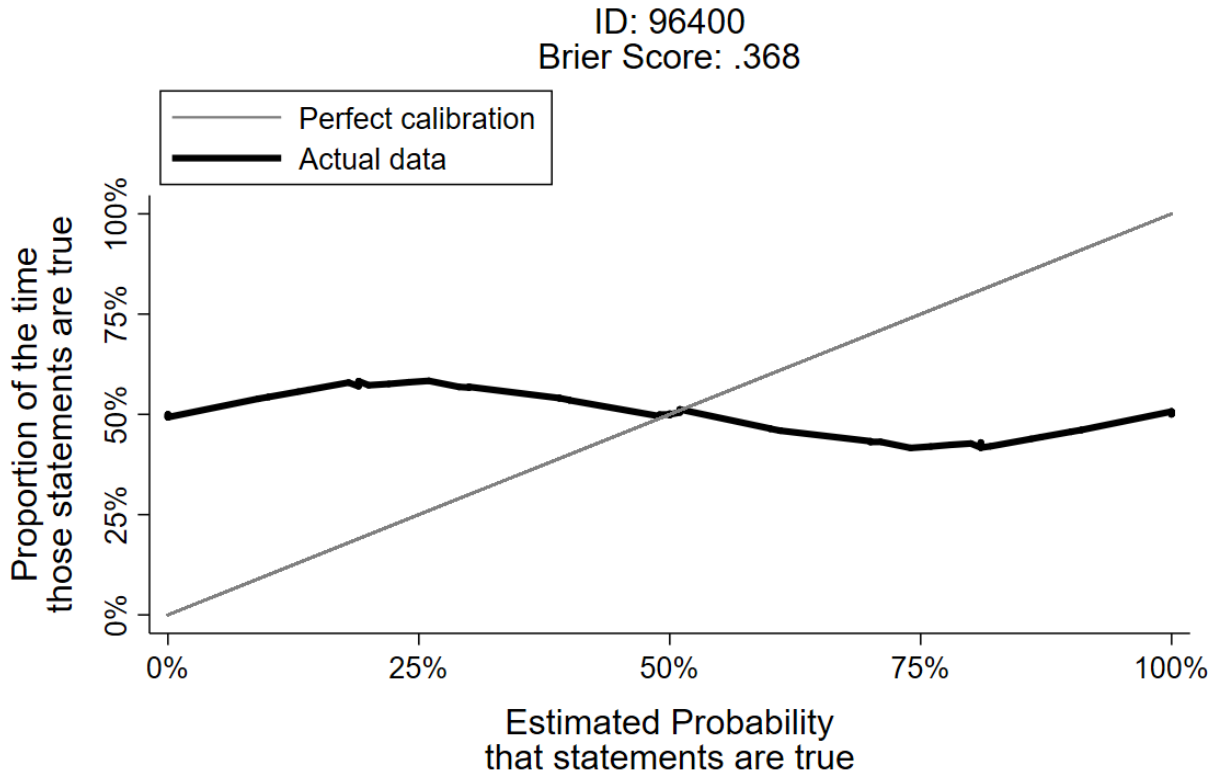

ID: 88045
Brier Score: .225
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

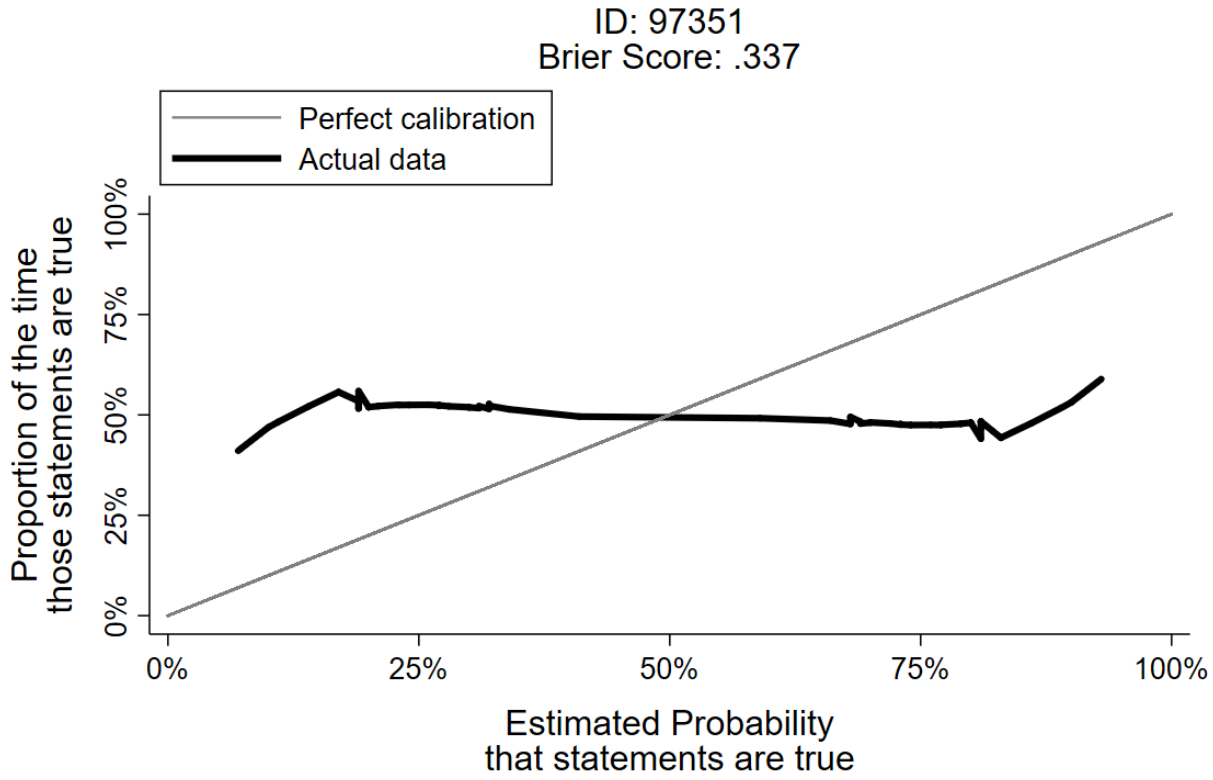

ID: 92763
Brier Score: .205
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

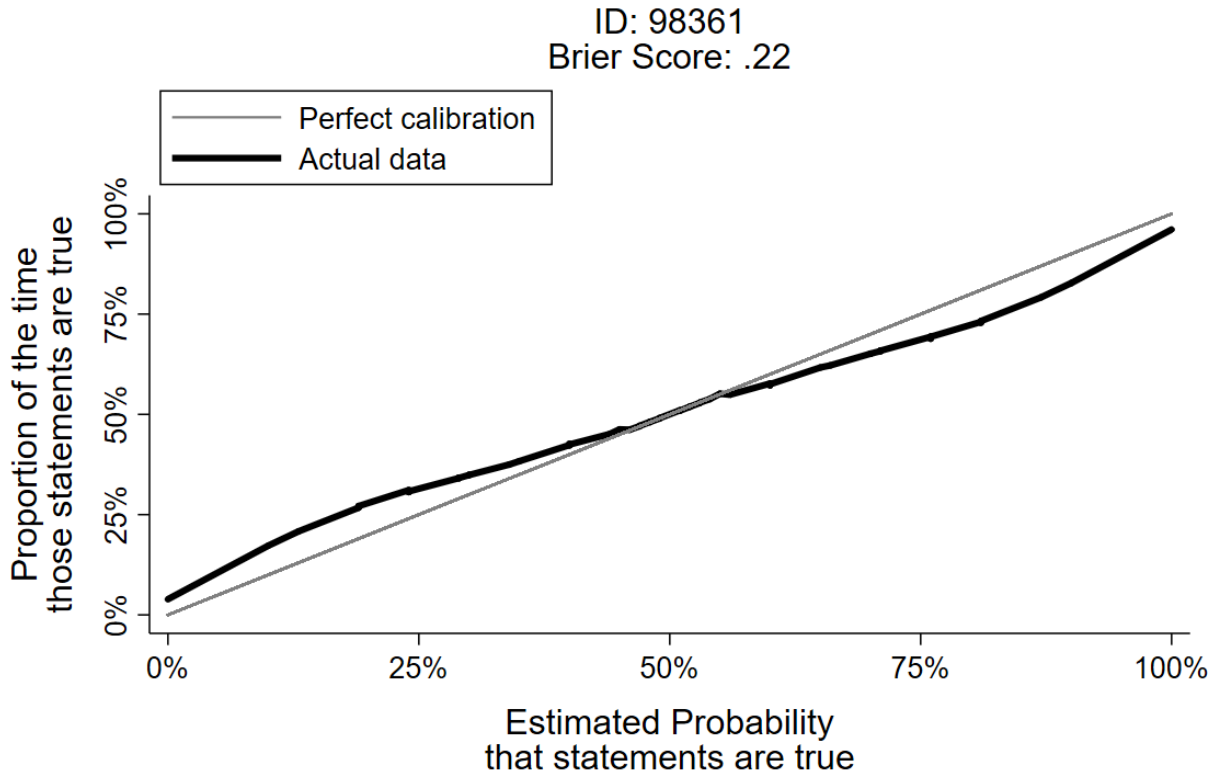

ID: 93329
Brier Score: .411
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true

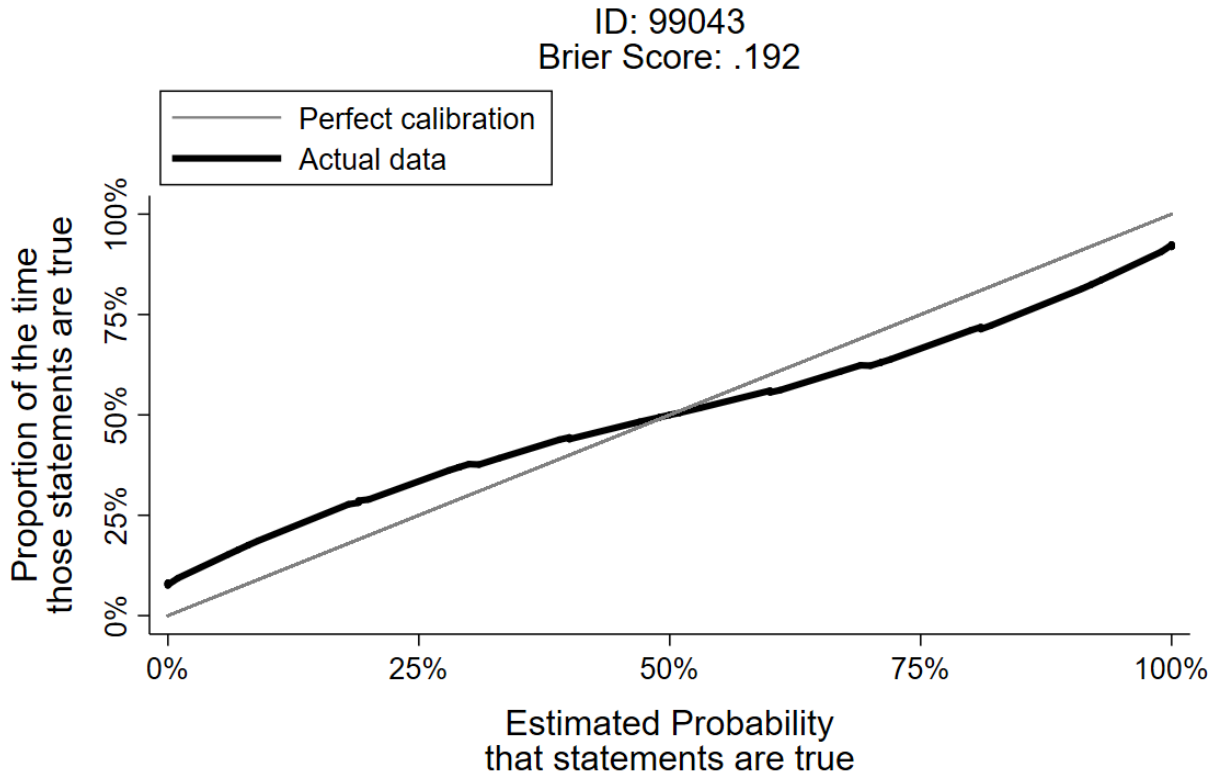

ID: 94867
Brier Score: .242
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true


ID: 95554
Brier Score: .344
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true


ID: 96400
Brier Score: .368
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true


ID: 97351
Brier Score: .337
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true


ID: 98361
Brier Score: .22
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true


ID: 99043
Brier Score: .192
Perfect calibration
Actual data
Proportion of the time those statements are true
0%
25%
50%
75%
100%
Estimated Probability that statements are true