225 What does simulations of the Office for Students b3 regulations tell us about how fair and effectively it can identify areas below specified thresholds.

Alexander Bradley, Vladislav Areshka

The University of Portsmouth, Portsmouth, United Kingdom

Research Domains

Employability, enterprise and graduate careers (EE)

Abstract

Increasing participation has led to increasing debts for students and government in funding higher education which has led to increase focus on assessing the 'value for money' higher education. In England this culminated in the b3 regulation where 60% of full-time first degree undergraduate need to be in positive outcomes 15 months from their course end. This study uses a simulation methodology to explore how effectively the OfS can identify the true population value of positive outcomes, the precision of those estimates and accuracy with which it can classify courses/universities below the 60% threshold. We find that whilst the OfS can accurately estimate the true population value it cannot do it with sufficient precision to enable correct classification below the 60% threshold when sample sizes are small and/or the percentage of the population is small nor when the true population value approaches the 60% threshold.

Full paper

Around the world participation in higher education has increased leading to increased debts for students and governments culminating in a focus on the value for money of universities (Bondar et al. 2020; OECD 2017). In England the Office for Students met that call by introducing the b3 regulation and in particular the progression metric which requires 60% of full-time undergraduate to have a positive outcome by 15 months of their course end (Office for Students (OfS) 2022). This study sought to evaluate how effectively the proceed metric could identify the true population value, how precise its estimates were and crucially how well it could identify courses below the 60% threshold. The consequences to universities of falling below this metric are severe not just with reputation damage but also with financial fines or even the cessation of degree awarding powers.

This study uses a simulation methodology where synthetic data for a virtual population is repeatedly generated and sampled from under specific conditions which can be varied across specific factors (Morris, White, and Crowther 2019). We varied three factors, first the percentage of graduates in the population with a positive outcome which went from 20% to 95% by 5% increments. Second, the population size was manipulation from 40 students up to 1000 student increased by 10 students each iteration. Third, we varied the percentage sampled from 30%, minimum threshold set by OfS, to 90% each time increasing by 5%. The simulation procedure followed three phases: data generation, sampling of the data and calculating statistics from each of the sample. We focussed on four main outcome measures bias, coverage probability, precision and misclassification. Bias is the mismatch between population and sample estimates. Coverage probability refers to the proportion of true population values inside a confidence interval. Precision is the width of the confidence intervals and misclassification is the likelihood that a population above or below the 60% threshold is accurately identified as being above or below the threshold.

The results showed that generally the level of bias i.e. the sample estimates of positive outcomes roughly matched the true population estimate and as the sample size and percentage of the sample increased bias reduced as the variability in the bias estimate. Both points highlight the importance of a) maximising response rates to the graduate outcome survey from which the B3 regulations are calculated and b) ensuring comparisons are not done at too granular level like courses where result will vary widely. The coverage probability was good for both 95% and 90% confidence interval indicating that around 95% and 90% of the time the sample value did indeed fall within the confidence interval as would be expected. Third, the range of confidence intervals, precision of confidence intervals, could vary widely which would make it difficult for the OfS to correctly identify areas below the 60% threshold. For example, sample size of 50 or less students had confidence interval that had an average range of 26% for the 95%

confidence interval and 22% on average for the 90% confidence interval. Even for samples of 100 or less the 95% confidence interval varied by 14.93% on average and 12.55% for 90% confidence interval. Finally, as the population percentage in positive outcomes become closer to the 60% threshold it becomes increasingly difficult to accurately show it is below the 60% threshold. For example, in a sample size of 50 or below with a true population value of 55% in positive outcomes in 94.5% of samples the upper confidence limit of the 95% confidence interval was above 60% suggesting not enough certainly in that course being below the threshold. This highlights that judgements about what are above or below the threshold using 95% or 90% confidence interval becomes increasingly difficult with small samples or small percentage of the population being surveyed.

This simulation powerfully shows the very serious challenging of trying to regulating at a micro-level like course or an area where sample sizes are less than 50 or even less than a 100 as the sample estimates or the true population value will vary more, the confidence intervals will be far large and therefore discerning who is above or below threshold will be exceptionally challenging especially as the population value approaches the threshold. Another important but perhaps hidden assumption of using a 95% or 90% confidence interval is the in-built acknowledgement that means sometime areas above particular threshold will incorrectly be identified as below.

References

Bondar, Tamara Ivanivna, Nataliia Viktorivna Telychko, Hanna Vasylivna Tovkanets, Tetiana Dmytrivna Shcherban, and Vasyl Ivanivych Kobal. 2020. “Trends in Higher Education in EU Countries and Non-EU Countries: Comparative Analysis.” Revista Romaneasca Pentru Educatie Multidimensionala 12 (1Sup1): 77–92. doi:10.18662/rrem/12.1sup1/224

Morris, Tim P., Ian R. White, and Michael J. Crowther. 2019. “Using Simulation Studies to Evaluate Statistical Methods.” Statistics in Medicine 38 (11): 2074–2102. doi:10.1002/sim.8086

OECD. 2017. ‘Benchmarking Higher Education System Performance: Conceptual Framework and Data, Enhancing Higher Education System Performance.’

Office for Students. 2022. ‘Consultation on a New Approach to Regulating Student Outcomes