Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 14 April 2026
Revised 20 May 2026
Accepted 22 June 2026

Natural Resources for Human Health 2026, Vol. 6 (3s): 27–31

KEYWORDS:
heterogeneous treatment effects; NAAC; NIRF; higher education; regression to the mean; India; small-sample causal inference.

Does Infrastructure Investment Pay Off Differently by Governance Type? A Small-Sample Heterogeneous-Effects Analysis of NAAC-to-NIRF Trajectories

Mathiazhagan K[1,*], Dr. Jagan Mohan R[2], Dr. Lokeshmaran A[3], Dr. Indira S[4]

[1]Statistician / PhD Scholar, Department of Community Medicine, Mahatma Gandhi Medical College and Research Institute, Sri Balaji Vidyapeeth.

[2]Deputy Director, C-DREaMs, Sri Balaji Vidyapeeth.

[3]Associate Professor / Statistician, Department of Community Medicine, Mahatma Gandhi Medical College and Research Institute, Sri Balaji Vidyapeeth.

[4]Assistant Professor / Statistician, Department of Community Medicine, Shri Sathiya Sai Medical College and Research Institute, Sri Balaji Vidyapeeth.

Corresponding Author: Mathiazhagan K, mathi11vicky11@gmail.com

ABSTRACT: A recurring policy question in Indian higher education is whether the same quality-assurance investment yields different competitive payoffs depending on an institution's governance type: whether it is centrally funded, State funded, or a self-financing Private/Deemed university. This paper examines one specific version of that question empirically: does a university's NAAC-assessed Infrastructure and Learning Resources score (Criterion 4) predict a subsequent gain in its NIRF overall ranking score, and does that relationship differ across governance types? Using 48 Indian universities for which both a NAAC criterion-level snapshot and at least two subsequent NIRF ranking cycles (2016-2025) were available, we split institutions into higher- and lower-infrastructure groups by a median split on Criterion 4 and estimated the average and type-moderated association with annualised NIRF score change, first with an unadjusted comparison of means, then with a covariate-adjusted linear model, and finally with an explicit treatment-by-governance-type interaction term. We found no statistically detectable overall association between infrastructure standing and subsequent NIRF gains (adjusted effect = 0.37 points/year, SE = 0.49, p = .45), and no statistically detectable heterogeneity by governance type (largest interaction term p = .66). The only robust predictor of annualised change was the institution's own baseline NIRF score, which was negatively associated with subsequent gains (b = -0.06, p = .047), consistent with a regression-to-the-mean or ceiling effect rather than any effect of infrastructure investment itself. Because the governance-type subgroups are small (as few as 6 Central universities with usable longitudinal data), we treat these null results as inconclusive rather than as evidence of no effect, and we set out the sample size a confirmatory replication would need. All analyses were conducted in Python..

1. INTRODUCTION

India's National Institutional Ranking Framework (NIRF) rewards institutions for outcomes placement rates, research output, graduate destinations, and a stakeholder perception survey that are widely assumed to depend, among other things, on the physical and digital infrastructure a university can offer its students and faculty. NAAC's accreditation process separately scores each institution's Infrastructure and Learning Resources under Criterion 4, on a scale that captures library holdings, laboratory equipment, campus connectivity, and similar inputs. It is a natural policy question whether investment reflected in a strong Criterion 4 score actually precedes measurable gains in NIRF standing, and whether that payoff is the same for a fully government funded Central university, a State university operating under tighter budget constraints, and a self financing Private or Deemed university that may be able to redeploy capital more flexibly.

This question has an obvious policy stake: if infrastructure investment pays off more for one governance type than another, funding bodies allocating limited capital works budgets most consequentially the state governments that fund State universities would want to know whether the return on that investment is comparable to what Central or private institutions realise from similar spending. Anecdotally, State universities are often assumed to under-invest in infrastructure relative to need and to see correspondingly little competitive payoff even when they do invest, but this has rarely been tested with matched before/after institutional data.

We test it here using the same matched NAAC-NIRF sample described in our companion paper on predictive fairness (Author, in preparation), restricted to the subset of institutions for which a genuine two timepoint outcome a change in NIRF score following the NAAC assessment could be constructed. We had originally intended a causal-forest-style analysis of heterogeneous treatment effects, following Wager and Athey (2018), but the usable longitudinal subsample (N = 48, with as few as 6 Central universities) is far too small to support a non-parametric forest reliably; we say so explicitly, and instead use a transparent, appropriately powered linear moderation design, which is the honest analytical choice at this sample size.

2. BACKGROUND

Heterogeneous treatment effect (HTE) estimation has moved from purely parametric interaction models toward flexible machine-learning estimators causal forests (Athey & Imbens, 2016; Wager & Athey, 2018), and meta-learner frameworks such as the T-learner and X-learner (Kunzel, Sekhon, Bickel, & Yu, 2019) precisely because real-world moderators of treatment effect are rarely known in advance and rarely linear. These methods, however, require substantially larger samples than classical regression to estimate subgroup-specific effects with acceptable precision, particularly when the moderating subgroups themselves are unevenly sized, as governance type is in Indian higher education (a small number of Central universities compared to a much larger population of State and private institutions). Applied guidance on treatment-effect heterogeneity (Abadie & Cattaneo, 2018) is consistent in recommending that researchers report power explicitly and avoid over-fitting subgroup effects when cell sizes are in the single digits, which is the situation we face for Central universities here.

3. Data and Methods

3.1 Constructing a two time point outcome

The companion cross-sectional dataset (N = 132) paired each NAAC-accredited university with a single NIRF cycle. For this analysis we returned to the full NIRF panel (2016-2025) and, for each of those 132 institutions, checked whether a further NIRF cycle existed one to four years after the cycle originally paired with its NAAC assessment. Forty-eight institutions had at least one such follow-up observation. For these institutions we computed the change in NIRF overall score between the baseline cycle and the latest available follow-up cycle within that window, and divided by the number of years elapsed (1 to 4 years, median 2) to obtain an annualised score change, since the follow-up horizon was not constant across institutions.

The resulting subsample of 48 comprises 6 Central, 17 State, and 25 Private/Deemed universities. We flag at the outset that the Central-university cell is very small, and any subgroup estimate for that group should be read as suggestive at most.

3.2 Treatment definition and estimation strategy

'Treatment' was operationalised as scoring above the sample median on NAAC Criterion 4 (Infrastructure and Learning Resources; median = 3.76 on the underlying criterion GPA scale), giving 23 higher-infrastructure and 25 lower-infrastructure institutions. This is a coarser design than the continuous dose-response or propensity-weighted designs used in larger causal-inference studies, chosen deliberately because more flexible estimators are not identifiable with confidence at N = 48 split three ways. We estimated the treatment effect in three steps, in order of increasing adjustment: (i) an unadjusted Welch's t-test comparing mean annualised change between the higher- and lower-infrastructure groups; (ii) a linear regression of annualised change on the treatment indicator, controlling for baseline NIRF score and NAAC CGPA; and (iii) a moderated regression adding governance-type indicators and their interactions with the treatment indicator, to test whether the treatment effect

Natural Resources for
Human Health
eISSN: 2583-1194

differs by type. All analyses were conducted in Python (pandas, NumPy, SciPy), with regression standard errors computed directly from the ordinary least squares residuals; given the small sample, we report exact coefficients and p-values throughout rather than relying on asymptotic approximations alone, and we discuss statistical power explicitly in Section 6.

4. RESULTS

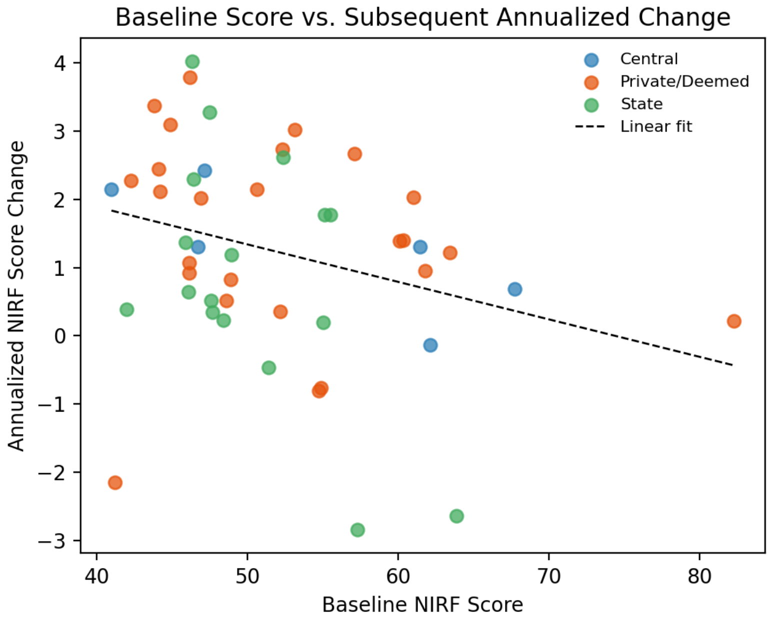

Figure 1. Baseline NIRF score plotted against subsequent annualised NIRF score change, by institution type. The downward-sloping fit line is consistent with regression to the mean rather than a treatment effect of infrastructure investment.

The unadjusted comparison showed almost no difference between groups: institutions above the infrastructure median gained an average of 1.24 NIRF points per year, versus 1.23 points per year for institutions below the median (difference = 0.02, Welch $t = 0.04$, $p = .97$). This null result held after covariate adjustment: the regression-adjusted treatment effect was 0.37 points per year ($SE = 0.49$, $p = .45$), controlling for baseline NIRF score and NAAC CGPA.

The only statistically reliable predictor of annualised change in this model was the institution's own baseline NIRF score ($b = -0.06$, $SE = 0.03$, $p = .047$): institutions that started from a higher NIRF score tended to gain less in subsequent years, and those starting lower tended to gain more. This pattern is the signature of regression to the mean (or a ceiling effect close to the top of the ranking), not evidence of any causal driver we measured, and we interpret it as such rather than as a substantive finding about infrastructure.

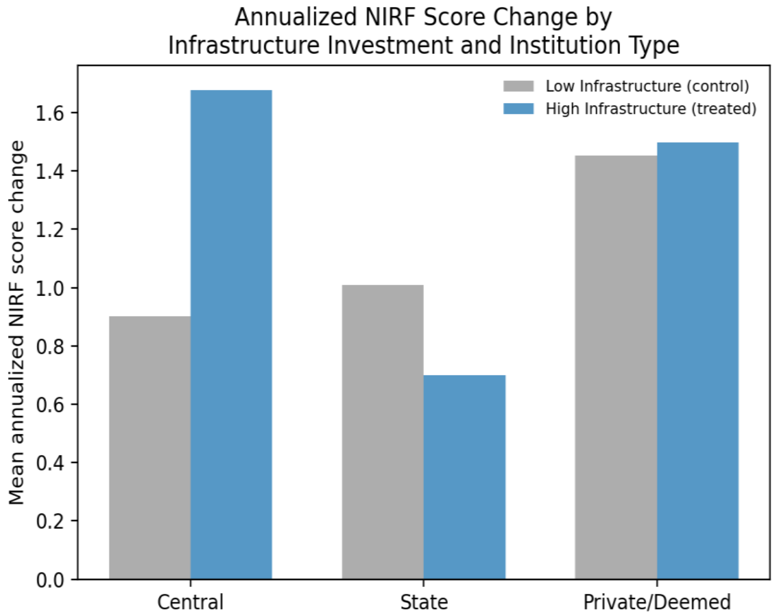

Figure 2. Mean annualised NIRF score change by infrastructure group and institution type. Error in these small cells (as few as n = 3 for Central universities) is not shown because confidence intervals at this cell size are not informative; see Table 2 for exact subgroup sizes.

Institution type	Low-infra n	High-infra n	Low-infra mean	High-infra mean	Diff.	p
Central	3	3	0.90	1.68	0.78	.38
State	9	8	1.01	0.70	-0.31	.74
Private/Deemed	13	12	1.45	1.50	0.05	.94

Table 1. Subgroup means for annualised NIRF score change, by institution type and infrastructure group (unadjusted t-tests).

No subgroup comparison in Table 1 approached significance, and the moderated regression's treatment-by-type interaction terms were similarly null (Private/Deemed x treatment: b = 0.54, p = .71; State x treatment: b = -0.64, p = .66). The point estimates move in different directions across groups positive for Central, negative for State, near-zero for Private/Deemed which is exactly the pattern one would expect to see by chance alone when splitting 48 observations three ways, and we do not read a substantive story into it.

5. DISCUSSION

Two things are worth taking from this analysis, and both are about what we did not find. First, we found no evidence that a stronger NAAC infrastructure score precedes larger subsequent gains in NIRF standing, either on average or for any particular governance type. This does not mean infrastructure investment has no payoff it may operate on a longer horizon than the one to four years available in our data, or its effect may be confounded with unmeasured investment in faculty, research support, or industry linkages that move together with infrastructure spending, diluting any isolated signal. Second, the one pattern that was statistically reliable higher-starting institutions gaining less, lower-starting institutions gaining more is a well-known statistical phenomenon (regression to the mean) rather than a substantive policy finding, and we flag this explicitly so it is not mistaken for evidence that under-performing institutions are 'catching up' for any particular reason.

We think it is more useful to readers to report this honestly as an underpowered null result than to force a heterogeneity story out of 48 observations split into cells as small as three. The pattern of point estimates in Table 1 is not implausible on its face (Central universities showing the largest apparent gain from infrastructure investment, State universities showing none), but with six Central universities in total, that estimate could easily reverse with a handful of different institutions in the sample, and we do not think it meets the bar for a policy claim.

6. LIMITATIONS AND STATISTICAL POWER

This paper's central limitation is sample size, and we address it directly rather than in passing. With 48 total observations split roughly evenly into treatment and control, and further split three ways by governance type, the smallest subgroup cells contain only three institutions per treatment arm. A conventional power calculation for detecting a medium-sized interaction effect (Cohen's $f^2 = .15$) at 80% power in a moderated regression with the covariates used here would require an estimated total sample in the range of 90-120 observations, roughly double what was available. We therefore present this analysis as an exploratory, appropriately transparent first pass rather than a confirmatory test, and we recommend that a future replication either extend the longitudinal window (more NIRF cycles per institution, once more become available) or draw on a larger matched sample using AISHE-code-based linkage rather than name matching, which would substantially increase N, particularly for the Central-university subgroup. We also note that our treatment definition (a median split on a continuous criterion score) discards information relative to a continuous dose-response or propensity-score design, a simplification made deliberately given the sample size, but one that a larger-sample replication would not need to make.

7. CONCLUSION

Within the constraints of a modest longitudinal subsample, we found no reliable evidence that NAAC-assessed infrastructure investment predicts subsequent NIRF score gains, overall or differentially by governance type. The clearest signal in the data was a regression-to-the-mean pattern in which institutions with lower starting NIRF scores gained more over time regardless of infrastructure standing. Given the small and uneven subgroup sizes involved, particularly for Central universities, we regard this as an inconclusive rather than a negative finding, and we set out the sample size that a properly powered confirmatory study would require.

REFERENCES

[1] Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. Annual Review of Economics, 10, 465-503.

[2] Athey, S., & Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences, 113(27), 7353-7360.

[3] Kunzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences, 116(10), 4156-4165.

[4] National Assessment and Accreditation Council. (2019). NAAC Manual for Universities (Revised edition). Bengaluru: NAAC.

[5] National Institutional Ranking Framework. (2025). NIRF India rankings, 2016-2025. Ministry of Education, Government of India. https://www.nirfindia.org

[6] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

[7] Sharma, R., & Devi, U. (2022). Convergence and divergence between NAAC and NIRF: An institutional perspective. Higher Education for the Future, 9(1), 45-61.

[8] Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242.