

Sensitivity Analyses

To interrogate the temporal-stability assumption inherent in using 2017 Census values for `Percentage_S`, we conducted four complementary sensitivity analyses:

1. Perturbation of `Percentage_S` We generated 500 simulated datasets in which `Percentage_S` was randomly perturbed by $\pm 5\%$, reflecting plausible measurement error or temporal drift. For each simulation, $\text{Percentage_S}_i^* = \text{Percentage_S}_i \times (1 + \epsilon_i)$ with $\epsilon_i \sim N(0, 0.05)$. A negative-binomial model identical to the primary specification was refitted, and the empirical 2.5th, 50th and 97.5th percentiles of the RR estimates were summarised. This test assesses whether sampling variability in the Census covariate materially alters inference.

2. Trimming Extremes of `Percentage_S` We excluded villages in the lowest and highest deciles of `Percentage_S`, thereby focusing on the central 80% of the distribution. Repeating the negative-binomial regression on this trimmed sample evaluates sensitivity to outliers or extreme values that may be most susceptible to misestimation over time.

3. Mixed-Effects Negative-Binomial Model To allow for unmeasured heterogeneity at the district level, we fitted a mixed-effects negative-binomial model via `glmmTMB()`, including a random intercept ($1 \mid \text{District}$). This approach accounts for cluster-specific overdispersion and tests whether between-district variation confounds fixed-effects estimates.

4. Combined Trimming and Perturbation Finally, we applied the $\pm 5\%$ perturbation procedure to the trimmed sample (central 80% of `Percentage_S`) across 500 simulations. This joint analysis probes the most conservative scenario, where extreme covariate values are removed and residual measurement error remains.

Justification

Each sensitivity check targets a distinct threat to causal inference:

- *Perturbation* tests robustness to random error in `Percentage_S`, modelling drift since 2017.
- *Trimming* assesses influence of extremes, which may disproportionately affect estimates if misclassified.
- *Random-effects* controls for unobserved district-level confounding and correlation.
- *Combined trimming & perturbation* explores the joint impact of both biases.

Consistent rate ratios across these analyses strengthen confidence that our primary findings are not artefacts of Census-derived covariate instability, extreme observations, or cluster-level heterogeneity.

Sensitivity Analyses Results

To assess the robustness of our main negative-binomial findings to potential temporal instability in the 2017 Census–derived covariate `Percentage_S`, we conducted four supplementary analyses:

1. **Perturbation of `Percentage_S`:** 500 simulations in which `Percentage_S` was multiplied by $1 + \epsilon$, with $\epsilon \sim N(0, 0.05)$.
2. **Trimmed sample:** exclusion of observations below the 10th percentile and above the 90th percentile of `Percentage_S`.
3. **Mixed-effects NB model:** inclusion of a random intercept for `District`.
4. **Combined trimming perturbation:** trimming as above, then 500 perturbation simulations.

Table 1: Simulation intervals for perturbation of `Percentage_S` ($\pm 5\%$, 500 sims)

Variable	2.5%	Median	97.5%
(Intercept)	117.72	120.77	124.51
<code>CategorySanLimited_S</code>	1.29	1.32	1.35
<code>CategorySanSafely Managed_S</code>	0.29	0.32	0.36
<code>CategorySanUnimproved_U</code>	2.20	2.25	2.29
<code>CategorySanOpen Defecation</code>	≈ 0	≈ 0	≈ 0
<code>Percentage_S</code>	1.034	1.035	1.037

The perturbation analysis (Table 1) shows that the estimated rate ratios (RR) for all four service categories and for `Percentage_S` are highly stable under plausible $\pm 5\%$ fluctuations in the Census covariate. Notably, the median RR for `CategorySanLimited_S` was 1.32 (95 % simulation interval [SI]: 1.29–1.35), and for `Percentage_S` was 1.035 (95 % SI: 1.034–1.037).

Table 2: Estimates after excluding the lower and upper 10 % of `Percentage_S`

Variable	RR	2.5%	97.5%
(Intercept)	217.56	114.29	414.14
<code>CategorySanLimited_S</code>	0.82	0.59	1.14
<code>CategorySanSafely Managed_S</code>	0.53	0.33	0.83
<code>CategorySanUnimproved_U</code>	1.30	0.95	1.80
<code>CategorySanOpen Defecation</code>	≈ 0	≈ 0	∞
<code>Percentage_S</code>	1.01	1.00	1.02

When trimming the extremes of `Percentage_S` (Table 2), the magnitude of category effects was attenuated towards the null, and the confidence intervals widened. This suggests that extreme values of the Census covariate, if mismeasured due to temporal drift, could inflate the observed associations.

Table 3: Fixed-effects from mixed-effects NB model with random intercept for `District`

Variable	RR
(Intercept)	8.51
<code>CategorySanLimited_S</code>	1.32
<code>CategorySanSafely Managed_S</code>	0.32
<code>CategorySanUnimproved_U</code>	2.23
<code>CategorySanOpen Defecation</code>	≈ 0
<code>Percentage_S</code>	1.036

Table 4: Simulation intervals for combined trimming & perturbation (500 sims)

Variable	2.5%	Median	97.5%
(Intercept)	212.20	218.47	225.46
CategorySanLimited_S	0.79	0.82	0.84
CategorySanSafely Managed_S	0.50	0.53	0.57
CategorySanUnimproved_U	1.26	1.30	1.34
CategorySanOpen Defecation	≈ 0	≈ 0	≈ 0
Percentage_S	1.13	1.14	1.14

Allowing for between-district heterogeneity via a random intercept (Table 3) produced virtually identical rate ratios to the main model, indicating minimal confounding by district-level clustering.

Finally, the combined trimming and perturbation (Table 4) yields intervals for all key parameters that closely mirror those from the perturbation-only analysis, confirming the overall stability of our conclusions against plausible Census-related measurement error.

Conclusion. Across all four sensitivity checks, the central finding—that access rates differ markedly by service category and rise modestly with Percentage_S remains the same. This provides assurance that our inference, and results presented in the primary paper, are robust to both random fluctuation and potential temporal misalignment of the 2017 Census data.