Strathclyde
Discussion Papers
in Economics

University of Strathclyde Business School

Health Insurance and Health-Seeking Behaviour: The Distributional Effects of Public Health Insurance in Nepal

Sabin Subedi

No. 26 – 5

June 2026

Department of Economics
University of Strathclyde, Glasgow

Health Insurance and Health-Seeking Behaviour: The Distributional Effects of Public Health Insurance in Nepal*

Sabin Subedi †

June 22, 2026

Public health insurance is one of the largest fiscal interventions in low- and middle-income countries, but who actually benefits varies sharply across settings. This paper provides the first causal evaluation of Nepal's National Health Insurance Programme (NHIP), a voluntary household-level scheme rolled out in stages across all 77 districts between 2016 and 2022. Combining a novel district-quarter panel of all health-facility visits with individual-level registration and claims records, I estimate that NHIP eligibility raises total health-facility visits by 14 percent, with similar gains across outpatient, emergency, and diagnostic margins. Effects emerge gradually and remain at their peak level once reached. The gains are unequally distributed: utilisation rises only in low-poverty districts, while high-poverty districts show no significant effect. Evidence from a 2019 facility-reassignment reform points to a supply-side explanation: in poor districts, the network of empanelled facilities is too sparse for coverage to translate into care.

Keywords: Health Insurance, NHIP, Poverty

JEL-Classification: H51, I12, I13, I14, J16, O15

*Correspondence: sabin.subedi@strath.ac.uk. All mistakes are my own.
I thank my supervisors Markus Gehrsitz and Stuart McIntyre for detailed feedback. I also thank Prakash Pathak, Mukesh Adhikari, Prashant Bhandari, Prajjwal Dhungana, Rory Allanson, Owen O'Donnell, Alexander Thompson, Hector Gutierrez Rufrancos, and Sam Grant for their helpful feedback. I am grateful to the Ministry of Health and Population for providing novel data, and to Suresh Tinkari from the Health Insurance Board for facilitating access to the health insurance claims data and for helpful discussions. I thank participants of the Health Economists' Study Group Winter Conference, the Strathclyde PGR Conference, the Economics PhD Reading Group at the University of Glasgow, and the Scottish Graduate Programme in Economics Conference for their discussions around the paper and formative feedback.
† : Affiliated to University of Strathclyde

1 — Introduction

Public health insurance has emerged in the past two decades as the principal alternative to tax-funded provision and out-of-pocket payment for healthcare in the developing world, with major schemes in Thailand, Mexico, China, India, and Indonesia each extending formal coverage to tens of millions of previously uninsured households (Kurowski et al. 2023). Together they form the institutional core of the global push toward Universal Health Coverage, the goal of ensuring that all individuals have access to needed services without facing financial hardship (WHO and Bank 2023). These schemes raise healthcare utilisation on average (Erlangga et al. 2019; Gruber, Hendren, and Townsend 2014; Gruber, Lin, and Yi 2023), but who captures the gains varies sharply across settings. Mexico's Seguro Popular closed 98 percent of the poor-rich infant-mortality gap, with mortality reductions concentrated in poor municipalities (Conti and Ginja 2023). A free-care randomised experiment in rural Ghana produced the opposite pattern: utilisation gains accrued more strongly to wealthier households even within a poor study population (Powell-Jackson et al. 2014). This paper asks where Nepal's National Health Insurance Programme (NHIP) sits on that spectrum, and whether the answer is driven by demand-side frictions such as premium affordability and information, or by supply-side constraints on the empanelled-facility network[1] through which insured care is delivered.

NHIP is a voluntary,[2] household-level contributory insurance scheme launched in 2016 and rolled out across all 77 districts of Nepal by 2022. Premiums are set at NPR 3,500 per year per household of five (USD 25; PPP USD 96), and coverage extends to a national network of empanelled public and private hospitals. The staggered district-by-district rollout creates variation in treatment timing that pre-treatment district characteristics (baseline utilisation, poverty, facility count, population) do not jointly predict (Section 4), consistent with the institutional record of mid-rollout HIB reorganisation and leadership turnover. I exploit this within a Callaway and Sant'Anna (2021) staggered difference-in-differences design that accommodates treatment-timing and cohort heterogeneity and avoids the negative-weighting concerns of static two-way fixed effects. The analysis uses two novel administrative sources: a district-quarter HMIS panel covering

1. The empanelled network is the set of facilities contracted to deliver care under the scheme.

2. Enrolment is legally mandated under the Health Insurance Act (2017), but the mandate is not enforced in practice; participation is therefore de facto voluntary.

all health-facility visits between 2014 and 2022, and individual-level registration and claims data from the Health Insurance Board for the mechanism analysis. A 2019 reform reassigned NHIP enrollees to their nearest empanelled public facility within an unchanged network. Because no facilities entered or exited at the reform date, the reduced-form effect on claims identifies a pure access shock; the reform is also a natural policy experiment that tests whether reshuffling enrollees across the existing facility network closes the access gap.

The intention-to-treat (ITT) effect of NHIP eligibility is a 14 percent increase in total health-facility visits on average. The gains are unequally distributed: the most affluent quarter of districts gains roughly 20 percent on total visits, while high-poverty districts show no significant effect on nearly all margins. The aggregate 14 percent figure is therefore driven by districts that were already better-off, and NHIP widens the poverty-utilisation gap. The programme's stated secondary objective of reducing poverty through improved healthcare access is therefore unmet on the utilisation margin. The dynamic response emerges gradually, with effects becoming statistically significant roughly five quarters after rollout and plateauing at twelve to fourteen quarters, with positive effects across outpatient, emergency, and diagnostic-service margins. Roughly half of the aggregate response runs through empanelled private facilities; restricting to public facilities alone attenuates the headline effect to about 7 percent (Section 7), without changing the equity pattern.

The mechanism evidence comes from the 2019 reform, which reassigned each enrollee's designated first-point-of-contact facility to the empanelled public facility nearest their home while leaving the set of facilities unchanged. In low-poverty districts, where the network is dense, reassignment brought enrollees substantially closer and claims rose. In high-poverty districts the network was already too sparse for reassignment to reduce distance much, and claims rose much less in response. The muted response in poor districts therefore reflects how little distance the policy could compress where the public network is sparse, rather than weaker demand for care. The policy implication generalises beyond Nepal. The distributional incidence of public health insurance is governed not by who is enrolled but by where care can actually be obtained: where the empanelled network is sparse, a nominal entitlement does not translate into utilisation. Expanding the supply of providers in underserved areas is therefore the binding lever for reaching the poor, and administrative rules that merely redistribute enrollees across a fixed network are no substitute for it.

This paper makes three contributions. First, it provides the first causal evaluation of Nepal's NHIP at the national level, using a novel district-quarter HMIS panel and individual-level registration and claims records from the Health Insurance Board, and contributes to the literature on the utilisation effects of public health insurance in low- and middle-income countries (LMICs) (Erlangga et al. 2019; Das and Do 2023; Dupas and Jain 2026; Gruber, Hendren, and Townsend 2014; Gruber, Lin, and Yi 2023). Second, it contributes to the equity literature (Powell-Jackson et al. 2014; Malani et al. 2024; Banerjee et al. 2021; Dupas and Jain 2024; Conti and Ginja 2023) by using individual-level claims to decompose the high-versus-low-poverty utilisation gap into enrolment, probability-of-claim, and intensity components via a three-factor Kitagawa decomposition. The poor-district shortfall is present on all three margins rather than a single bottleneck, which motivates the supply-side mechanism analysis. Third, it contributes to the supply-side mechanism literature (Dupas and Jain 2026; Boone et al. 2023; Dupas and Jain 2024) by using individual-level claims data from the Health Insurance Board to examine empanelment as a mechanism (Section 6).

1.1. Related Literature

There is a large literature evaluating public health insurance expansions in low- and middle-income countries.[3] Taken together, this literature suggests that large public insurance programmes can play a substantial role in raising healthcare utilisation and reducing out-of-pocket spending. The consensus is qualified in two important ways. First, evidence on health outcomes is surprisingly limited and rarely positive, with provider-side behavioural responses absorbing much of the policy intent (Das and Do 2023; Lagomarsino et al. 2012).[4] Second, the magnitude of utilisation effects varies widely across schemes: Erlangga et al. (2019) review 68 evaluations and report positive utilisation effects in 32 of 40, while Dupas and Jain (2026) review ten LMIC studies and find mixed and often statistically insignificant effects on utilisation.

Two large staggered rollouts provide the closest methodological precedents to this paper. Gruber, Hendren, and Townsend (2014) evaluate Thailand's 30 Baht universal-coverage scheme

3. For a comprehensive catalogue of LMIC insurance evaluations, see Appendix 1 of Gruber, Lin, and Yi (2023).

4. Out-of-pocket spending falls more robustly. Georgia's MIP cut 30-day OOP by 50 percent (Das and Do 2023). But of the seven studies Dupas and Jain (2026) examine that measured OOP effects, only Ghana and Mexico recorded meaningful declines. Provider-side moral hazard (induced demand, prices above mandated rates) remains the dominant unresolved channel.

and find a 12 percent rise in inpatient utilisation for the previously poor, alongside a collapse of the pre-programme provincial poverty–infant mortality correlation that they attribute to supply-side capitation reform.[5] Gruber, Lin, and Yi (2023) use a triple-difference design on the staggered county-level rollout of China's New Cooperative Medical Scheme and document a 12 percent reduction in age-adjusted mortality alongside up to a 77 percent rise in the probability of medical care use among 800 million rural Chinese. NHIP's headline 14–23 percent utilisation gain (Section 5) places it inside this range.

What the literature has not settled is who captures these gains. The evidence splits sharply. Mexico's Seguro Popular concentrated mortality reductions in pre-programme poor municipalities, closing 98 percent of the poor–rich infant-mortality gap with zero effects in rich municipalities (Conti and Ginja 2023). A free-care randomised experiment in rural Ghana produced the opposite pattern: utilisation gains accrued more strongly to wealthier households within the trial sample, even as out-of-pocket reductions ran more evenly across the wealth distribution (Powell-Jackson et al. 2014). India's Bhamashah Swasthya Bima Yojana raised male and female utilisation by 19 and 17 percent respectively after an empanelment expansion that cut distance to private hospitals by two-thirds, leaving the gender gap intact (Dupas and Jain 2024). On the demand side, Banerjee et al. (2021) randomise premium subsidies and registration assistance under Indonesia's JKN Mandiri scheme and find that even a full one-year premium subsidy combined with assisted online registration does not lead to full enrolment, with enrolment instead stalling at roughly 30 percent, ruling out demand-side instruments alone as the binding margin in voluntary contributory schemes.

The paper proceeds as follows. Section 2 describes Nepal's healthcare context and the design of NHIP. Section 3 introduces the data sources. Section 4 sets out the empirical strategy and identification assumptions. Section 5 presents the main treatment effects and the equity heterogeneity. Section 6 examines the mechanism. Section 7 discusses threats to identification, and Section 8 concludes.

5. Hospital capitation payments rose four-fold and the largest funding flows went to poor provinces, a supply-side architecture that Section 6 contrasts with NHIP.

2 — Background

2.1. Country Context

Nepal is a federal democratic republic in South Asia with a population of 29.2 million, classified by the World Bank as a lower-middle income country with a Gross National Income per capita of USD 1,380 (PPP USD 5,272) in 2022 (National Statistics Office 2023; World Bank 2025a). The country is divided into 7 provinces and 77 districts, with three tiers of government (federal, provincial, and local) established under the 2015 Constitution. Despite sustained economic growth over the past two decades, Nepal faces deep structural inequalities. Seventeen percent of the population remains multidimensionally poor, with poverty concentrated in the mountainous and hill districts of the mid- and far-western regions (GoN 2021). The rural-urban divide is pronounced, with over 80 percent of the population living in rural areas where access to services (health, education, financial) is constrained by difficult terrain and limited infrastructure (National Statistics Office 2023).

Nepal organises its healthcare system in three tiers. At the base, primary health care centres, health posts, and community health units serve as the first point of contact in rural areas. District hospitals form the second tier, providing inpatient, surgical, and diagnostic services for each of the 77 districts. Tertiary referral hospitals and specialised centres, concentrated in the Kathmandu Valley and a handful of regional cities, constitute the third tier. The private sector, comprising hospitals, nursing homes, clinics, and pharmacies, plays a substantial role, particularly in urban areas, and accounts for a growing share of service delivery: the number of private hospitals rose from 16 in 1990 to over 300 by 2014, and private facilities now hold over two-thirds of the country's hospital beds and employ 60 percent of its doctors (Adhikari et al. 2021). The government provides subsidised basic health services at public facilities, alongside free care for selected terminal diseases and free maternal/neonatal services under programmes such as the Safe Motherhood Programme. Nevertheless, out-of-pocket expenditure dominates health financing: nominal per capita OOP spending was USD 50 (PPP USD 191) in 2023, accounting for 59 percent of current health expenditure (World Health Organization 2025). This heavy reliance on direct payments leaves households, particularly poor ones, financially vulnerable to health shocks. Life expectancy at birth is 70 years, marginally above the lower-middle income country average of

69 years (World Bank 2025b), but this aggregate figure masks considerable geographic variation, with remote districts recording substantially worse outcomes.

2.2. The National Health Insurance Programme of Nepal

The Government of Nepal implemented the National Health Insurance Programme (NHIP) in 2016 with the primary goal of reducing household out-of-pocket (OOP) health expenditure and the secondary goal of reducing poverty. In 2017, the government established the Health Insurance Board (HIB) to implement health insurance nationwide with a mandate of covering every citizen. Pre-NHIP insurance coverage in Nepal was historically negligible: an early community-based scheme at Patan Hospital dates from 1976, and district-level pilots followed in the 2000s in collaboration with the BP Koirala Institute and the Government of Nepal, but social health insurance contributed essentially zero to current health expenditure as late as 2015 (Ayer et al. 2024). There were a few programmes run by the Ministry of Health and Population (MoHP) which provided free treatment for essential non-communicable diseases like cardiovascular diseases and malignancy alongside the Safe Motherhood Programme (SMP) that provided free maternal/neonatal care to pregnant women. SMP is detailed in Section 5.5.3.

Figure 1 shows the phase-wise implementation of health insurance across districts from 2016 to 2022. Although NHIP is mandated by law, limited enforcement capacity has rendered it de facto voluntary: households contribute premiums and must renew coverage annually (Ayer et al. 2024). The programme is primarily funded through government budget allocation and premiums collected from the enrolled. By the end of 2022, 6,045,192 individuals (approximately 23 percent of the total population) were insured under NHIP. A family of five members can register for health insurance at the annual premium of NPR 3,500 (USD 25; PPP USD 96)[6], with any additional member costing NPR 700 (USD 5; PPP USD 19). Once registered the members of NHIP are entitled to free care up to a maximum of NPR 100,000 (USD 721; PPP USD 2,755) per family per year. Any additional member increases the household coverage maximum by NPR 20,000 (USD 144; PPP USD 551). Alongside this, the government also provides subsidies to ultra-poor households, senior citizens, and people with disabilities or selected terminal diseases. Elderly members above the age of 70 or

6. USD 1 = NPR 138.7 at the time of writing. PPP-adjusted values use the World Bank 2022 ICP private-consumption PPP conversion factor for Nepal of NPR 36.3 per international dollar.

people with selected terminal diseases receive additional coverage of NPR 100,000 (USD 721; PPP USD 2,755) (HIB 2024). Relative to household incomes, the premium represents about 1 percent of annual income for a five-person household where all five members earn at Nepal's national poverty line of NPR 72,908 (USD 526; PPP USD 2,008) per person per year in 2022/23 (World Bank Group 2024), the threshold below which 20.3 percent of the population fall. The single-earner case is more realistic for many Nepali households: with one earner at the per-person poverty line, household income is NPR 72,908 (one-fifth of the household poverty threshold), and the premium share rises to about 5 percent.

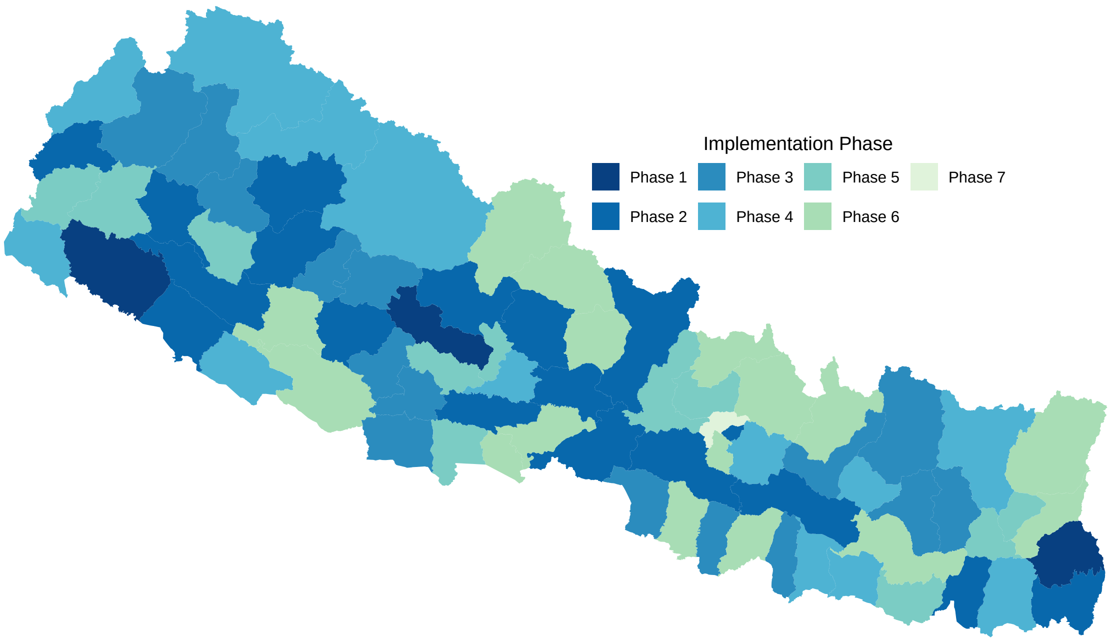
NHIP covers preventive, curative, inpatient, emergency, surgical, and rehabilitative care, alongside medicines, diagnostics, and health aid equipment, and covers up to NPR 2,000 (USD 14; PPP USD 55) of emergency ambulance service. It excludes services such as expensive medical aid equipment, plastic surgery, most dental services, and artificial insemination. At registration, households select a preferred first point of contact, which delivers primary services and refers patients upward when needed. By design, the system operates cashlessly: enrolled members face no payment at the point of service for covered care. The empanelled facility files a claim with HIB on the patient's behalf and is reimbursed directly; patients themselves never file claims.[7]

Take-up grew rapidly with the rollout: cumulative individual enrolment rose from fewer than 14,000 in fiscal year 2015/16 (three pilot districts) to approximately 1.1 million by 2017/18, 4.6 million by 2020/21, and 6.0 million by 2021/22, when the programme had reached all 77 districts (DoHS 2024). By the end of fiscal year 2022/23 (mid-2023), cumulative enrolment had grown to 7.2 million individuals, or 24.6 percent of the population (DoHS 2024). Supply-side capacity expanded alongside but unevenly: 451 facilities were empanelled with HIB by end-2022[8], yet only 343 of Nepal's 753 local government units host at least one empanelled provider, and HIB's 20-person claims-review team faced a daily inflow of 25,000 to 30,000 claims against a manual-review capacity of approximately 6,000 (Ayer et al. 2024).

7. In practice, insured members frequently pay out-of-pocket for prescribed medicines because empanelled-facility pharmacies do not stock the full benefit-package list, requiring purchase from outside retailers (Ayer et al. 2024).

8. The network has since grown to 501 facilities; the geospatial analyses in section 6 use this expanded snapshot.

Figure 1 — Coverage of Health Insurance in Nepal



Source: Author's own calculations based on the data collected from annual reports, press releases and notices.
Notes. This figure illustrates the spatial variation of the coverage of National Health Insurance Programme across different districts from 2016 to 2022

2.3. Rollout of NHIP

The government rolled out NHIP in multiple phases at the district level, Nepal's second-level administrative division, as shown in Figure 1. Figure 2 shows the implementation timeline from 2016 to 2022. The first phase began in April 2016 with three pilot districts (Kailali in the Terai plains, Baglung in the western hills, and Ilam in the eastern hills), selected to represent Nepal's major ecological zones rather than on the basis of health or economic characteristics. Ecological-zone selection is correlated with poverty and development gradients in Nepal; the balance check (Table A3) absorbs these dimensions, though the long-horizon dynamic ATTs remain most sensitive to the pilot cohort's trajectory. The programme then expanded gradually, reaching all 77 districts by the end of fiscal year 2078/79 (2021/22).

The institutional history of the programme is important for understanding the rollout process. The government initially established the Social Health Security Development Committee (SHSDC), a semi-autonomous body under the Ministry of Health, in February 2015 to design and pilot the programme. The SHSDC launched the first three pilot districts in 2016. Parliament enacted the Health Insurance Act in 2017, transforming the SHSDC into the Health Insurance Board (HIB) as a fully autonomous body. The permanent implementing institution thus did not exist in its final legal form until *after* the second phase of implementation had already begun. This institutional transition, combined with significant political instability across the rollout period 2015–2022, including the aftermath of the 2015 earthquake, a new constitution, the transition to federalism, and six prime ministers over this period, meant that the rollout proceeded through administrative improvisation rather than systematic planning.[9]

Several features of the rollout support the view that district selection was not driven by observable determinants of healthcare utilisation. First, the rollout sequence followed no formal selection criteria beyond administrative feasibility and the aim of maintaining balanced representation across Nepal's seven provinces.[10] The official implementation records show that within each fiscal year, registration opening dates varied by days or weeks across districts, with service commencement dates standardised to quarterly windows (Falgun 1 [mid-February], Baisakh 1 [mid-April], Shrawan 1 [mid-July], or Kartik 1 [mid-October]), reflecting administrative batching rather than deliberate sequencing.[11] Second, the composition of implementation cohorts defies any systematic ordering: the third cohort (FY 2073/74) includes Bhaktapur, an urban, relatively affluent district adjacent to Kathmandu, alongside Jumla, one of Nepal's most remote and poorest districts. Such pairings are inconsistent with selection based on infrastructure readiness, poverty,

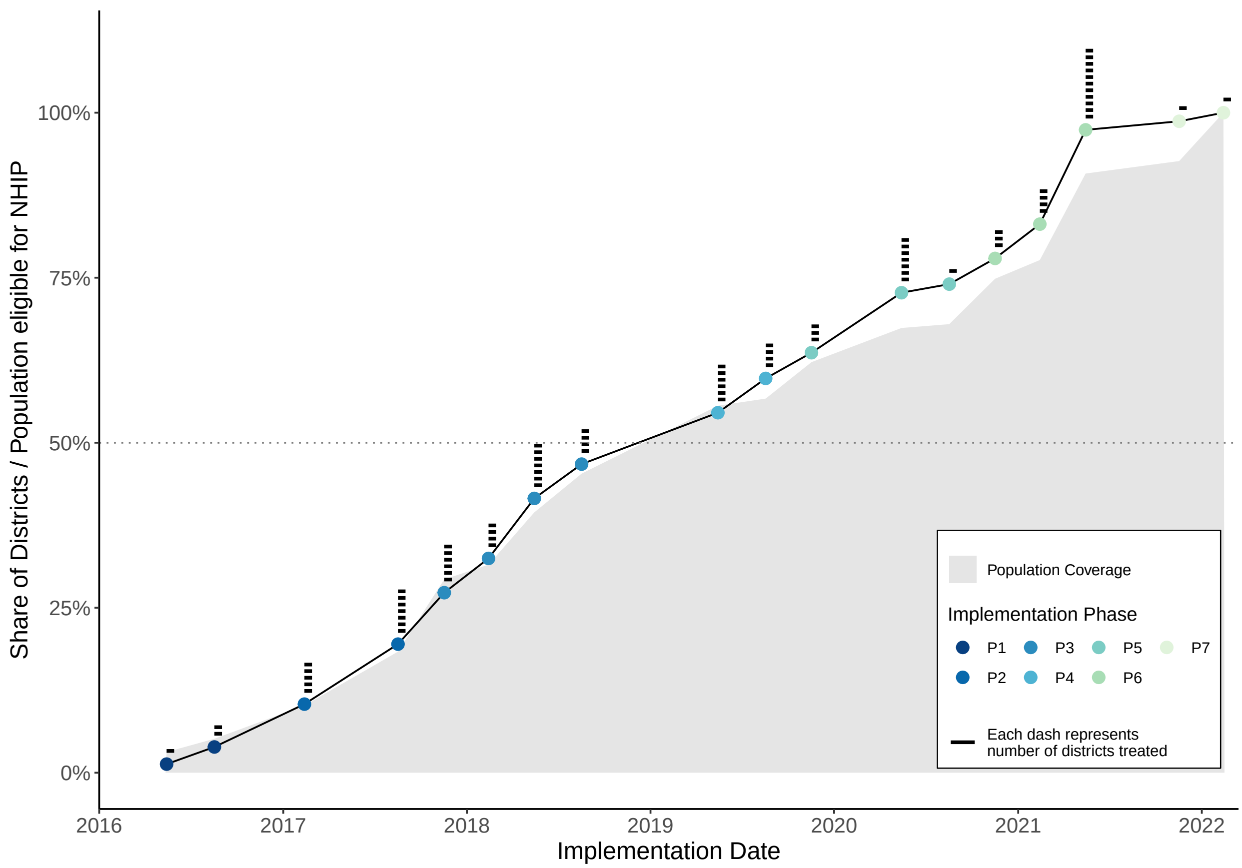
9. The World Bank had designed a formal impact evaluation of the pilot phase in 2014, combining a randomised enrolment encouragement design with a matched difference-in-differences fallback across 153 VDCs in six districts. A baseline survey was conducted in March–May 2014, but the evaluation was ultimately cancelled and no results were published (Ambel, Seff, and Belay 2014). The absence of a completed experimental evaluation shows the value of the quasi-experimental approach taken here.

10. This characterisation reflects interviews with HIB officials involved in the programme since 2016 (Bharati et al. 2025), who confirmed that no formal selection rule beyond administrative feasibility and provincial balance was applied. Neither HIB annual reports nor DoHS publications document any selection criteria, and a screening of contemporaneous news coverage similarly turned up no documentation of a selection process.

11. The full district-level implementation timeline, including exact registration and service commencement dates, is compiled from HIB annual reports, press releases, and public notices (Health Insurance Board 2017, 2019).

or administrative capacity. In addition, changes in HIB leadership during the rollout period further disrupted any consistent selection logic that might have been applied across phases.

Figure 2 — Rollout of the National Health Insurance Programme



Source: Author's own calculations based on the data collected from annual reports, press releases and notices.
Notes. This figure illustrates the implementation of the National Health Insurance Programme across different districts from 2016 to 2022. Dash at each point indicates the number of districts that were newly covered by health insurance at that time.

In Table 1, I present pre-treatment district characteristics across treatment timing cohorts, splitting districts into early- and late-treated groups at the median implementation quarter. None of the differences are statistically significant, with all standardised differences below 0.20 and most below 0.10, indicating a high degree of balance in observable characteristics and baseline health service utilisation. This suggests that the timing of the health insurance rollout was not strongly correlated with observable district characteristics.

Table 1 — Pre-Treatment Characteristics by Treatment Timing Cohort

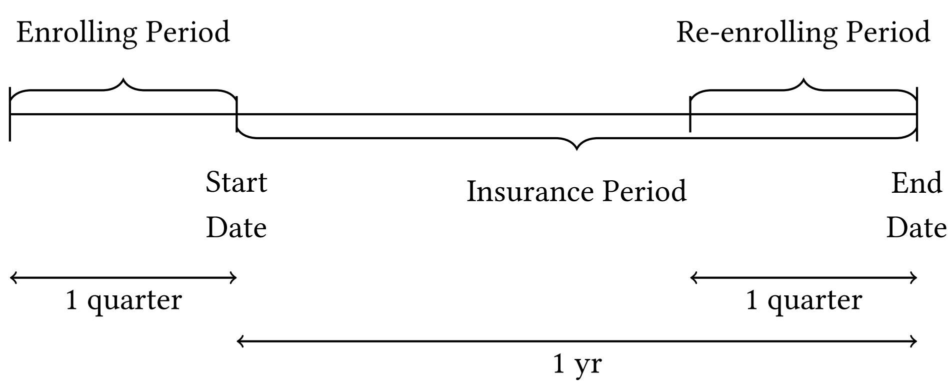
Variable	Mean (Early)	Mean (Late)	Diff.	Std. Diff.	P-value
Demographics					
Population (1000s)	388.43	408.51	20.09	0.06	0.80
Poverty incidence	0.29	0.27	-0.02	-0.16	0.49
No. of health facilities	111.11	114.92	3.81	0.08	0.72
Female population (1000s)	179	185.97	6.97	0.05	0.83
No. of households (1000s)	69.73	75.31	5.57	0.09	0.70
Female headed household share	0.07	0.07	0	0.02	0.92
Migrants share	0.27	0.25	-0.02	-0.15	0.53
Agricultural occupation share	0.84	0.83	-0.01	-0.09	0.70
Dependent Variables					
Total visits (1000s)	70.57	81.75	11.17	0.15	0.51
OPD visits (1000s)	67.95	65.92	-2.03	-0.04	0.86
Emergency visits (1000s)	4.57	6.25	1.68	0.17	0.45
Diagnostic services (1000s)	7.84	14.12	6.28	0.18	0.41
Total visits, ages 0–9 (1000s)	14.28	14.73	0.45	0.04	0.86
Total visits, ages 10–19 (1000s)	14.71	15.34	0.62	0.06	0.80
Total visits, ages 20–59 (1000s)	32.97	41.34	8.37	0.19	0.39
Total visits, ages 60+ (1000s)	8.61	10.34	1.74	0.15	0.50
Female share of visits	0.58	0.58	-0.01	-0.19	0.40
No of Districts	37	40			

Source: Demographic variables are the authors' own calculations based on the National Census 2011.
Notes: The table reports district-level pre-treatment averages by treatment timing cohort. Differences are defined as later-treated minus earlier-treated districts. Standardised differences are calculated as mean differences divided by the pooled standard deviation. P-values are from Welch two-sample t-tests. "No. of health facilities" refers to the full set of public health-service delivery points reported in DoHS records, comprising hospitals, primary health care centres (PHCCs), health posts, sub-health posts, and community health units; this is distinct from the 501 NHIP-empanelled facilities nationwide.

2.4. NHIP Procedures

The enrolment process follows a structured quarterly cycle, illustrated in Figure 3. There are four fixed service start dates each year: approximately mid-February, mid-April, mid-July, and mid-October (corresponding to Falgun 1, Baisakh 1, Shrawan 1, and Kartik 1 in the Nepali calendar). The quarter preceding each start date serves as the enrolment window, during which all new registrants are batched together for the same service commencement date. Renewal follows an analogous cycle: existing members can renew only during the quarter immediately before their coverage expires; missing this window results in a lapse of coverage until the next renewal cycle. Premiums are due in full at the time of registration or renewal, prior to the corresponding service start date; partial or installment payments are not accepted.

Figure 3 — NHIP Enrolment Timeline



At the local level, enrolment assistants based in each ward (the lowest administrative unit) conduct outreach to identify willing households and collect registration information, including household composition, demographics, and the preferred first point of contact (FPOC) facility. This information is verified and processed at the district-level HIB office, which formally enrols the household and issues insurance cards. The household head typically registers the family as a unit, paying the NPR 3,500 premium for up to five members, with additional members at NPR 700 each. Payment is collected locally, either through the enrolment assistant or at designated collection points. Ultra-poor households, identified through a government poverty listing, are exempt from premium payments and receive free enrolment.[12]

12. The ultra-poor classification follows the Ministry of Land Management, Cooperatives and Poverty Alleviation's

Public communication is conducted through both enrolment assistants and mass-media channels. In fiscal year 2020/21, the Board commissioned more than 22,000 local FM-radio insertions, over 3,000 television spots, audio jingles in Nepali, Bhojpuri, Maithili, Doteli, and Tharu, and distributed approximately 300,000 brochures, alongside a toll-free helpline (Health Insurance Board 2021a). Awareness nonetheless remains uneven: field interviews and provincial-level surveys identify low public awareness as a leading reason for the gap between Madhesh province (8 percent of provincial population enrolled) and Koshi (42 percent) (Ayer et al. 2024).

Upon registration, each member selects a first point of contact facility, typically the nearest public health centre, which serves as the gateway to all insured services. Primary care, diagnostics, and medicines are provided directly at the FPOC. If the patient requires specialist or inpatient care beyond the FPOC's capacity, the facility issues a referral to a secondary or tertiary provider, which can be either public or private, provided the facility is empanelled with HIB. The referral chain is important: services accessed without a valid referral from the FPOC are generally not covered, creating an incentive for patients to route through the designated facility. Emergency care is the exception: patients can access any empanelled facility directly in emergencies, with costs covered up to NPR 2,000 for ambulance services. At the end of 2022, approximately 33 percent of total households had at least one member enrolled in the NHIP programme, though individual-level coverage rates are considerably lower because not all household members within enrolled families register. By end-2022, approximately 57 percent of cumulative individual enrollees had renewed their coverage at least once (cumulative-ever renewal rate), against a one-period annual renewal rate of around 81 percent in FY 2021/22 (Ayer et al. 2024; DoHS 2024). The one-quarter gap between registration and benefit eligibility is the basis for the anticipation window used in the empirical strategy (Section 3).

2075 Guidelines. From January 2022, the treatment programme for ultra-poor citizens was formally integrated into HIB's system.

3 — Data

3.1. *NHIP Rollout*

I track NHIP implementation using DoHS annual health reports (2015–2023), HIB notices, and HIB press releases. For each district I record two dates: the date registration opened, and the date implementation began (when insurance benefits became claimable). I define the treatment date G_i as the implementation date. The registration window opens exactly one quarter before implementation in every district; individual registrations occur within that window, so individual-level gaps from registration to eligibility range from zero to one quarter. I set the anticipation parameter $\delta = 1$ quarter to bound the longest individual gap (Section 4).

3.2. *District Health Information System*

I use facility-reported administrative data from the District Health Information System (DHIS2). DHIS2 captures what facilities report monthly to the central system: counts of visits, service utilisation, referrals issued, family planning encounters, immunisations, and reportable diseases including malaria, leprosy, HIV/AIDS, and tuberculosis. The unit of observation is therefore the facility-reported aggregate, not the individual patient encounter. DHIS2 covers both public and private health services. Reporting to DHIS2 was mandated for all public health facilities beginning in 2014, two years before NHIP's launch, so the reporting infrastructure predates the programme. Reporting is not compulsory for the private sector, but DHIS2 data collection was a government priority before NHIP launched, and 422 of 451 empanelled facilities (93.5%) report through DHIS2 as of 2022.

Not all DHIS2 indicators are reported uniformly across districts: some variables are collected only in a subset of districts, while others are available throughout the panel. To avoid selection on data availability, I restrict the analysis to variables that were collected for all districts throughout the 2016–2022 study period; outcomes with incomplete district coverage are excluded from the main specifications.

The district-level aggregated data used in this study are compiled from a consistent set of facilities that reported throughout the 2016–2022 study period, effectively constituting a balanced facility panel. This construction ensures that the estimated utilisation changes reflect changes in patient visits at the same facilities over time, rather than the entry or exit of reporting units.

3.3. Health Insurance Board Claims Data

As a secondary data source, I use administrative claims records from the Health Insurance Board (HIB). The HIB records each insurance-financed visit at empanelled facilities, including the date, facility, patient demographics (age and gender), and district of residence. Unlike DHIS2, which records all visits a facility reports regardless of insurance status, the claims data record only visits for which the treating facility filed a reimbursement claim with HIB on the insured patient's behalf. I aggregate these records to the district–quarter level to match the structure of the main analysis data. The claims data are available for treated districts from the quarter of NHIP implementation onward, covering the period up to 2022. I use these data in the cross-source corroboration exercise (Table 5) to compare the magnitude of DiD-implied utilisation changes with the observed volume of insurance-financed visits, providing a consistency check on the plausibility of the main estimates. I also use the claims data in the mechanism analysis (Section 6). Because claims data record only insurance-financed visits, while the DiD captures the net change in total visits regardless of insurance status, the two quantities measure overlapping but distinct objects. Claims include inframarginal substitution (visits that would have occurred without NHIP but are now paid by insurance), while the DiD additionally captures any spillover increases among uninsured residents. The relative magnitude is therefore not directionally predetermined.

3.4. Poverty Incidence Data

I classify districts and Ilakas — pre-2017 sub-district administrative units, each grouping several Village Development Committees (VDCs) — by poverty incidence using small area estimates produced by the Central Bureau of Statistics (CBS) and the World Bank (CBS and World Bank 2013). The estimates apply the Elbers-Lanjouw-Lanjouw (Elbers, Lanjouw, and Lanjouw 2003) approach: a household consumption model is fit on the 2010/11 Nepal Living Standards Survey (NLSS-III) and applied to the 2011 Population and Housing Census to predict the poverty headcount ratio against the national poverty line of NPR 19,261 per person per year, with Monte Carlo standard errors reported at both the district and Ilaka levels. The median district poverty rate is 26 percent. I use this median as a binary split for the heterogeneity analysis in Section 5.2: districts above the median are classified as poor, districts below as non-poor. For the mechanism analysis in Section 6, I use the Ilaka-level poverty rate directly to exploit within-district variation in poverty exposure.

Two caveats are worth noting. First, the SAE estimates are a single cross-section corresponding to 2010/11. Since NHIP rollout began in 2016, I treat poverty status as a fixed pre-treatment baseline. Second, Nepal restructured its local administrative units in 2017, replacing the VDC and Ilaka system with new federal municipalities. The SAE estimates are reported on the pre-2017 geography while the HIB claims data record beneficiary residence on the post-2017 geography. I bridge the two using a deterministic concordance between the pre- and post-2017 administrative boundaries.

4 — Empirical Method

4.1. Empirical Strategy

My primary goal is to identify the causal effect of health insurance on healthcare utilisation and related outcomes. A naive comparison of districts that implement health insurance with those that do not is biased because adopters likely differ on unobservables that also affect outcomes.

To address this bias, I exploit the gradual rollout of health insurance and use a Difference-in-Differences (DiD) method for causal identification of the effect of the health insurance programme. This method exploits the variation in timing of implementation, with some districts implementing earlier than others. It compares the evolution of outcomes for the earlier-treated groups with groups not yet treated, controlling for district time-invariant confounders and time-specific effects that impact all districts.

4.2. Estimator

My target parameter is the group-time average treatment effect on the treated, $ATT(g,t) = \mathbb{E}[Y_t(g) - Y_t(0) \mid G_i = g]$, where $Y_t(g)$ and $Y_t(0)$ denote the potential outcomes at time t for districts first treated in quarter g versus never treated, and G_i is district i's first quarter of NHIP eligibility (with $G_i = \infty$ for not-yet-treated districts). $ATT(g,t)$ is the intention-to-treat effect of NHIP eligibility for districts in cohort g at quarter t, the average effect of becoming eligible regardless of individual enrolment. I estimate it using the Callaway and Sant'Anna (2021) (CS) estimator, which accommodates treatment-timing and group heterogeneity.

Prior literature shows that the two-way fixed effects (TWFE) estimator, which regresses the outcome on group and time fixed effects, can be biased in a staggered-timing setup because it imposes an assumption of homogeneous effects in treatment timing and homogeneous effects in groups. This homogeneity is unlikely: districts that implement health insurance earlier may become more efficient over time, with better enrolment strategies and service delivery. If there is any treatment timing heterogeneity, the static TWFE estimator can be biased due to a negative-weights problem (Borusyak, Jaravel, and Spiess 2024; Chaisemartin and D'Haultfœuille 2020; Goodman-Bacon 2021) (see Roth et al. 2023, for elaboration). The choice among these estimators depends on data structure and context.

I aggregate the data to the quarterly level, where Y_{it} is the outcome for district i at time t. The CS estimator uses "not-yet-treated" districts (those without NHIP by time t) as controls. Its identifying assumption is weaker than that of static two-way fixed effects: parallel trends needs to hold only between each treated cohort and its not-yet-treated comparisons over the cohort's post-treatment horizon, rather than across all districts and all time periods at once (Roth et al. 2023). The estimator is also more efficient when outcomes are serially correlated, as in this setting. The CS estimator also accommodates anticipation periods and covariate inclusion when parallel trends holds only conditional on covariates.

The set of distinct treatment cohorts is $\mathcal{G} = \{g : g \in \{9, \ldots, \tau\}\}$, where τ denotes the final observed quarter and the first cohort is treated in quarter 9. Using the potential outcomes framework, $Y_{it}(0)$ is the potential outcome of district i at time t if it remains untreated, and $Y_{it}(g)$ the potential outcome if first treated in quarter g. Define $\mathbb{1}\{G_i = g\}$ as the indicator for district i belonging to cohort g. The observed outcome can then be written as:

$$Y_{it} = Y_{it}(0) + \sum_{g \in \mathcal{G}} (Y_{it}(g) - Y_{it}(0)) \cdot \mathbb{1}\{G_i = g\} \tag{1}$$

The estimand introduced above can be expressed formally as:

$$ATT(g, t) = \mathbb{E}\left[Y_t(g) - Y_t(0) \mid G_i = g\right] \tag{2}$$

The CS estimator allows aggregation of group-time $ATT(g, t)$ estimates into summary measures. For the main analysis I use "group" aggregation as the measure of the overall ATT effect.

For each cohort g, define the cohort-specific average effect as:

$$\theta(g) = \frac{1}{\tau - g + 1} \sum_{t=g}^{\tau} ATT(g,t) \tag{3}$$

The overall group aggregation then averages across cohorts, weighted by cohort size:

$$\theta^O = \sum_{g \in \mathcal{G}} \theta(g) \cdot P(G_i = g \mid G_i \leq \tau) \tag{4}$$

$\theta(g)$ is the average treatment effect of NHIP eligibility among all districts in cohort g across all post-treatment periods. Thus θ^O first calculates the average ATT for each cohort, then averages across cohorts to produce the summary measure.

I also estimate dynamic (event-study) ATT effects that show how the treatment effect evolves with time since adoption. Defining $e = t - g$ as the number of quarters since treatment, the dynamic ATT at event time e is:

$$\theta_{es}(e) = \sum_{g \in \mathcal{G}} \mathbb{1}\{g + e \leq \tau\} \cdot P(G_i = g \mid G_i + e \leq \tau) \cdot ATT(g, g+e) \tag{5}$$

$\theta_{es}(e)$ is the average effect across all cohorts observed exactly e quarters after treatment, weighted by the probability of belonging to cohort g conditional on being observed at event time e.

The validity of this strategy rests on two key assumptions.

Assumption 1 (Irreversible treatment). *Once district i becomes eligible for NHIP at quarter G_i, eligibility remains in place: G_i is fixed for $t \geq G_i$ throughout the sample.*

Assumption 1 holds by design; a violation would appear as G_i changing across periods, which we do not observe.

Assumption 2 (Parallel trends with limited anticipation). *Let $\delta \geq 0$ denote the anticipation period. For each cohort $g \in \mathcal{G}$ and for all $t \geq 2$:*

$$\mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid G_i = g\right] = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid G_i > t + \delta\right] \tag{6}$$

Under Assumption 1 and Assumption 2, Equation 2 provides a causal interpretation. Assumption 2 states that, in the absence of treatment, the evolution of potential outcomes for cohort g districts would have been the same as for districts not yet treated by time $t + \delta$. The limited anticipation clause allows districts to adjust their behaviour up to δ periods before treatment, but not earlier. As described in Section 3, the registration window opens exactly one quarter before eligibility in every district, with individual registrations occurring within that window; I set $\delta = 1$ quarter to bound this administrative delay. Parallel trends is therefore evaluated relative to the period before registration rather than before eligibility onset. The assumption stated here is unconditional: no covariates enter the conditioning set, consistent with my use of the doubly-robust estimator with no covariates (Section 4.3). Subgroup splits in Section 5.2 invoke this assumption separately within each poverty group.

I justify Assumption 2 on empirical and institutional grounds. The empirical defence is direct: I regress rollout timing on pre-treatment district characteristics (baseline utilisation levels, population, number of health facilities, and poverty rates) and find that none of the observables individually or jointly predict treatment timing (F-test p-value = 0.35; Table A3). The balance table in Table 1 further shows that standardised differences between early- and late-treated districts are uniformly small (all below 0.20). The institutional context is consistent with this result: NHIP rollout timing was determined by administrative feasibility and provincial balance, independent of district-level health-utilisation trajectories (Section 2.3). The implementing institution itself underwent a major legal transformation mid-rollout, and multiple changes in government and HIB leadership disrupted any consistent selection logic. The official implementation records confirm that districts with vastly different characteristics, such as Bhaktapur (urban, low-poverty) and Jumla (remote, high-poverty), were placed in the same implementation cohort.

The pre-trend evidence corroborates this institutional justification rather than substituting for it (Roth et al. (2023) shows that binary pre-trends tests have limited power for staggered designs). Event-study plots in Figure 4 show pre-treatment coefficients close to zero with no systematic pattern across primary outcomes. The Rambachan and Roth (2023) honest-DiD sensitivity analysis (Figure A3) shows the headline total-visits result survives violations up to $\bar{M} = 0.3$, post-treatment deviations as large as 30% of the maximum pre-treatment deviation. Falsification tests show that NHIP has no effect on outcomes covered by a pre-existing free programme (the Safe Motherhood

Programme), which would not be expected to respond to insurance eligibility.

4.3. Inference

The CS estimator can be estimated using outcome regression, inverse probability weighting (IPW) and doubly robust methods. Although the identification remains the same with all the methods, the doubly robust method is more robust to misspecification than the other two (Sant'Anna and Zhao 2020). Therefore, in all my specifications I use the doubly robust method, with standard errors clustered at the district level using the multiplier bootstrap with 1,000 iterations. I report simultaneous confidence bands, which are wider than pointwise bands but account for dependence across group-time ATTs and asymptotically cover the entire path of effects (Callaway and Sant'Anna 2021). I use short gaps (varying base period) for the dynamic ATTs, which means that I estimate the treatment effect if the treatment had occurred in that period. However, choosing short gap over long differences (universal base period) does not affect my interpretation or the look of the event study plot, which is mainly because the post-adoption ATTs are the same regardless of the base period. As a robustness check, I re-estimate the main specifications using a universal base period and confirm that the results are unchanged (Table A9).

Since I test multiple utilisation outcomes simultaneously, I address the concern of multiple hypothesis testing by applying the Romano and Wolf (2005) stepdown procedure, which controls the familywise error rate while accounting for the dependence structure across test statistics. I implement this using the influence functions from the CS estimator to construct the joint covariance matrix across the seven count-based utilisation outcomes: total visits, OPD visits, emergency visits, their three new-visit counterparts (new total, new OPD, new emergency), and diagnostic services. The female share of visits is excluded from the set of outcomes because it is a level proportion rather than a log count, and conceptually captures composition rather than volume. All seven outcomes remain statistically significant at the 5% level after the Romano–Wolf correction (Table A4).

5 — Results

5.1. Effects of the health insurance programme on health visits

Table 2, Panel A reports the effects of health-insurance eligibility on hospital visits. Outcomes are aggregated to the district-quarter level, and I use the group-aggregated $ATT(g, t)$ from the CS estimator for the analysis of the overall treatment effect.

I find that eligibility for health insurance increases total visits, including all inpatient, outpatient, and emergency services, to health service providers by 14.2 percent.[13] The results for outpatient services follow a similar pattern. Total outpatient department (OPD) visits increase by 13.7 percent. The largest effects appear in emergency care, where total emergency visits increase by 19.6 percent. At pre-treatment district-quarter means of approximately 102,000 total visits, 78,000 OPD visits, and 9,300 emergency visits, these effects translate to roughly 14,500, 10,700, and 1,800 additional visits per quarter, respectively. This response is itself notable. In high-income settings, emergency utilisation is generally thought to be insensitive to insurance: households seek emergency care when a genuine emergency arises, whatever the cost. That emergency visits respond to NHIP at all suggests this assumption does not travel to a lower-middle-income setting, where the cost of care appears to have deterred some households from seeking even emergency treatment. The relatively large emergency effect may additionally reflect NHIP's policy design, which exempts emergency admissions from the referral requirement at the first point of contact, allowing patients to substitute toward emergency services to bypass the referral chain; the aggregate ATT does not separate these two channels. Appendix Table A1 reports these effects in per-capita terms (visits per 1,000 population) to facilitate comparison of magnitudes across districts of different sizes.

NHIP empanelment is not restricted to public providers. Although enrolment is anchored to a public first point of contact facility, the referral chain explicitly permits access to any empanelled provider, public or private, for specialist and inpatient care. Of the 501 facilities empanelled with HIB, 426 (85%) are public and 75 (15%) are private. The public-only robustness reported in Table A10 shows that restricting the sample to public facilities attenuates the total-visit response by roughly half, indicating that a substantial portion of the measured increase is delivered through

13. Throughout this paper, I convert log-linear coefficients to percentage effects using the exact transformation $(e^{\hat{\theta}} - 1) \times 100$, which avoids the approximation error of $\hat{\theta} \times 100$ that grows with the magnitude of $\hat{\theta}$.

empanelled private providers.

Table 2 — Effects of Health Insurance Eligibility on Healthcare Utilization

	Estimate (Is NHIP eligible = 1)	Std. Error	Pre-treatment Mean
Panel A: Overall Healthcare Visits			
Total Visits	0.1331***	(0.0284)	101,980
OPD Visits	0.1281***	(0.0353)	78,357
Emergency Visits	0.1790***	(0.0665)	9,275
Diagnostic Services	0.2089**	(0.1009)	32,108
Female Share of Visits	-0.0038*	(0.0020)	0.5806
Panel B: Number of OPD Visits by Disease Category			
OPD Communicable	0.1472***	(0.0308)	8,287
OPD Non-Communicable	0.1363***	(0.0439)	36,829
OPD Injuries	0.1302***	(0.0409)	13,322
OPD CVD	0.0072	(0.1251)	848
OPD Cancer	-0.0064	(0.0491)	938
Fit statistics			
Number of Cohorts	17		
Number of Time Periods	30		

Note: All outcomes are log-transformed except Female Share of Visits, which is a level proportion. Panel A reports effects on total visits, Outpatient Department (OPD) visits, emergency visits, female share of visits, and diagnostic services (laboratory tests and radiographic procedures combined). Panel B disaggregates OPD visits by disease category. Pre-treatment mean shows the average outcome value before NHIP implementation. Clustered (District) standard errors in parentheses, bootstrapped with 1000 iterations. Estimation Method: Doubly Robust. Control Group: Not Yet Treated. Anticipation Periods: 1. Signif. Codes: *** 99% ** 95% * 90% uniform confidence band does not include 0.

Panel A also reports the effect on diagnostic services, defined as the combined count of laboratory tests and radiographic procedures (X-ray, ultrasound, CT, MRI, ECG, and six other imaging modalities, all empanelled under NHIP). Insurance eligibility raises diagnostic-service use by 23.2 percent. At a pre-treatment mean of about 32,100 procedures per district-quarter, this implies roughly 7,500 additional procedures per district per quarter. The 23.2 percent diagnostic response is roughly $1.7\times$ the 13.7 percent OPD response, implying that each OPD visit under NHIP triggers more diagnostic services on average. Possible explanations include intensification of diagnostic workup per visit; a provider-side response, as physicians can now recommend diagnostic services without regard to a patient's ability to pay; a shift in patient composition toward conditions that require more imaging or laboratory testing; or a coding shift toward separately billable diagnostic procedures; the data do not distinguish these.

The final row of Panel A reports the effect on the female share of visits, which enters the regression in levels rather than logs so that the coefficient reads as a direct change in composition. NHIP eligibility reduces the female share by 0.38 percentage points off a pre-treatment mean of 58.1 percent. The effect is statistically distinguishable from zero at the 10 percent level and indicates that the utilisation gains accrued slightly more to male than to female visits. This shift is small. The gender pattern in Nepal differs from what Dupas and Jain (2024) document under India's BSBY scheme, where female visits constituted only 45 percent of utilisation and where distance and out-of-pocket costs disproportionately suppressed female care-seeking. Two features of the Nepali setting account for the contrast. First, the pre-treatment female share was already 58 percent. Women were already the majority of healthcare users in Nepal before NHIP, so the baseline gender gap motivating Jain and Dupas's analysis does not apply here. Second, NHIP enrolment is at the household level. A single annual premium covers a family of five, which removes any incentive for households to allocate insurance access disproportionately to male members.

Panel B of Table 2 presents the effects of health insurance eligibility on OPD visits by disease category. OPD visits rise by similar amounts across categories: 15.9 percent for communicable diseases, 14.6 percent for non-communicable diseases, and 13.9 percent for injuries and surgical cases. These estimates are close to the 13.7 percent increase in total OPD visits: insurance does not disproportionately affect any particular condition type. However, health insurance eligibility has no significant effect on OPD visits for cardiovascular diseases or malignancies. This finding is consistent with existing policy: the Ministry of Health and Population operates a separate programme for essential non-communicable diseases, including cardiovascular conditions and cancers, which provided free treatment prior to the introduction of health insurance. This also acts as a falsification test, showing that health insurance does not affect outcomes unrelated to its coverage.

Table 3 examines heterogeneity in the effects of health insurance eligibility across age groups. Health insurance increases health service visits across all age categories, with the largest effects observed among those aged 60 and above (19.9 percent). The effects for the 10–19 and 20–59 age groups are similar in magnitude, at 15.7 percent and 15.3 percent respectively. The smallest effect appears among children aged 0–9, at 9.6 percent. These findings are consistent with the positive relationship between age and healthcare demand. Two institutional features may sharpen this age pattern beyond demand alone. First, NHIP's benefit design is more generous at older ages: members above 70 receive their own coverage ceiling of NPR 100,000, separate from the shared household maximum (Section 2), so insurance relaxes the budget constraint by more for the oldest group than uniform coverage would imply. Second, the muted response among children aged 0–9 likely reflects, in part, that much of their care was already free before NHIP, through pre-existing public programmes such as the Safe Motherhood Programme, immunisation, and childhood-illness and nutrition services, leaving less scope for insurance to raise utilisation. The pattern across age groups therefore reflects not only underlying demand but also pre-existing entitlements and NHIP's own age-graduated generosity.

Table 3 — Effects of Health Insurance Eligibility on Health Visits by Age Group

	Estimate (Is NHIP eligible = 1)	Std. Error	Pre-treatment Mean
Age 0-9	0.0917***	(0.0320)	18,272
Age 10-19	0.1461***	(0.0368)	17,143
Age 20-59	0.1420***	(0.0282)	52,668
Age 60+	0.1812***	(0.0334)	13,896
Fit statistics			
Number of Cohorts	17		
Number of Time Periods	30		

Note: All the outcomes are log transformed. Clustered (District) standard errors in parentheses bootstrapped with 1000 iterations. Estimation Method: Doubly Robust. Control Group: Not Yet Treated, Anticipation Periods: 1. Signif. Codes: *** 99% ** 95% * 90% uniform confidence band does not include 0.

5.2. *Poverty Heterogeneity*

A secondary objective of NHIP is to reduce poverty, similar to other health insurance programmes in lower-middle-income countries. If the programme affects rich and poor districts disproportionately, it can widen the inequality gap. To test for this, I study the heterogeneous effect of health insurance on utilisation outcomes by poverty incidence. In Table 4, I split districts at the median pre-treatment 2011 World Bank poverty incidence rate into high-poverty (above median) and low-poverty (below median) groups. The effects on health visits are positive and statistically significant for low-poverty districts but indistinguishable from zero in high-poverty districts for total visits, OPD visits, emergency visits, non-communicable diseases, and injuries. The one partial exception is OPD communicable disease, where high-poverty districts show a 9.5% increase (coefficient 0.091, $p < 0.10$), meaningful but smaller than the 14% effect in low-poverty districts (0.130, $p < 0.01$). The pattern is suggestive of NHIP facilitating care for communicable conditions even in poor areas, although the difference across OPD categories (communicable at $p < 0.10$; injuries and non-communicable null) is within sampling noise and a clean acute-versus-discretionary split is not supported by the data: injuries are acute and symptomatic yet show no high-poverty response. The overall effects of health insurance are therefore driven by low-poverty districts, a finding that runs against the programme's poverty-reduction objective.

The corresponding dynamic event studies, reported in Figure A2 of the Appendix, refine this reading. Pre-treatment coefficients are statistically zero in both subgroups, indicating that parallel trends hold within each half of the poverty distribution and not only in the pooled sample. Post-treatment, the two trajectories diverge sharply. In low-poverty districts, effects grow steadily with event time and become statistically distinguishable from zero from $e = 5$ onwards. In high-poverty districts, by contrast, estimates remain close to zero throughout the post-treatment window and are never statistically distinguishable from zero under simultaneous bands.

The median split itself understates how concentrated the gains are. Table A2 in the Appendix reports a three-group breakdown (most affluent quarter Q1, middle half Q2–Q3, and highestpoverty quarter Q4) and shows that only Q1 drives the aggregate effect: Q1 shows large, significant effects across all outcomes ($\approx 20\%$ on each); Q2–Q3 shows no significant effects on any outcome; and Q4 shows an 18% increase in OPD communicable-disease visits (coefficient 0.169, $p < 0.01$) but null effects on every other outcome. The gains therefore accrue to the most affluent quarter rather than the less-poor half, which further exacerbates the equity concern. Communicable-disease care therefore reaches even the most poverty-stricken districts, but every other margin of utilisation remains unchanged there. Note that the average across subgroup estimates differs from the pooled ATT because each subgroup-specific estimand (the ATT for districts within that subgroup, conditional on within-subgroup parallel trends) is recovered using within-subgroup not-yet-treated controls. The Callaway–Sant'Anna estimator therefore constructs a different counterfactual within each split than in the full-sample estimation. Section 6 investigates the channels through which this gap arises.

Table 4 — Effects of Health Insurance Eligibility by District Poverty Level

	High Poverty (above median)		Low Poverty (below median)	
	Estimate	Std. Error	Estimate	Std. Error
Panel A: Overall Healthcare Visits				
Total Visits	0.0373	(0.0425)	0.1264***	(0.0356)
OPD Visits	0.0212	(0.0523)	0.1218***	(0.0384)
Emergency Visits	0.1203	(0.0967)	0.1686**	(0.0708)
Panel B: OPD Visits by Disease Category				
OPD Communicable	0.0909*	(0.0484)	0.1301***	(0.0395)
OPD Non-Communicable	0.0498	(0.0526)	0.1095**	(0.0490)
OPD Injuries	0.0566	(0.0447)	0.1242***	(0.0456)
Fit statistics				
Number of Cohorts	11		15	
Number of Time Periods	27		30	

Note: This table presents heterogeneous effects of health insurance eligibility by district poverty level. Districts are split at the median pre-treatment poverty incidence rate (2011 World Bank). All outcomes are log-transformed. Panel A shows effects on overall visits. Panel B disaggregates OPD visits by disease category. Clustered (District) standard errors in parentheses. Estimation Method: Doubly Robust. Control Group: Not Yet Treated. Anticipation Periods: 1. Signif. Codes: *** 95% uniform confidence band does not include 0. The high-poverty subgroup has fewer time periods (27 vs 30) than the low-poverty subgroup because earlier-treated cohorts exhaust the not-yet-treated control pool sooner in high-poverty districts; post-treatment horizons remain sufficient given the 5-quarter effect onset documented in the dynamic ATTs.

5.3. *Dynamic ATT*

The static estimates in the preceding sections summarise the overall effect of NHIP eligibility, but they mask how these effects evolve over time. Figure 4 plots dynamic ATTs from the CS estimator, tracing the treatment effect at each quarter relative to implementation. I use a varying base period specification: pre-treatment estimates compare each event-time to the period immediately before it (useful for inspecting parallel trends), while post-treatment estimates are level effects relative to the last pre-anticipation period.

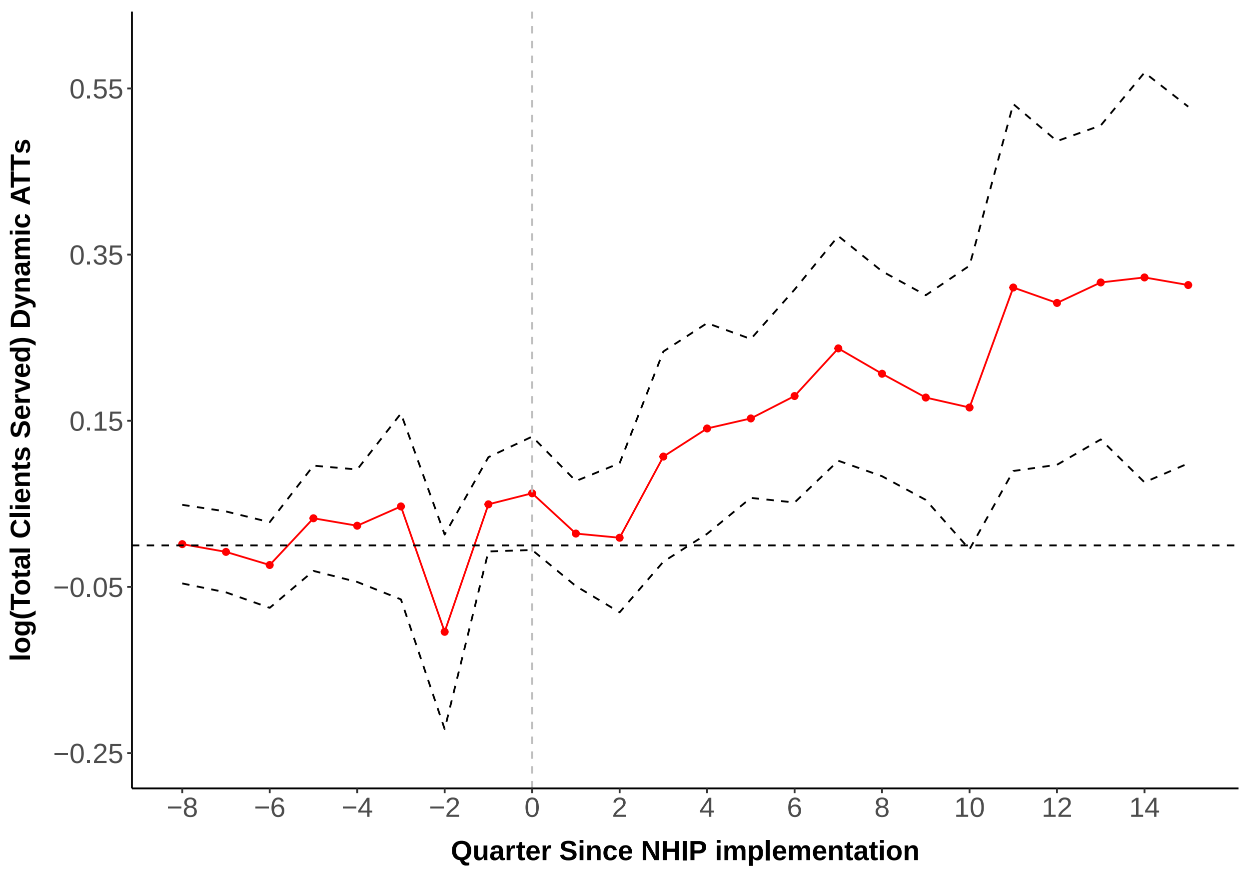
Pre-treatment coefficients corroborate the parallel trends assumption defended in Section 4; the leads are not themselves a test of identification. They hover around zero, rarely exceeding 0.05 in absolute value, and remain statistically indistinguishable from zero across the pre-treatment window. Healthcare utilisation in districts about to receive NHIP does not appear to have been diverging from utilisation in not-yet-treated districts ahead of implementation.

The post-treatment dynamics show a clear pattern. In the first few quarters after implementation, the treatment effects are small and not significantly different from zero. NHIP coverage takes time to build. Registration must be processed and insurance cards distributed before households learn of their benefits and utilisation responds. The effects then grow steadily, becoming statistically significant five quarters after implementation. By quarters 12–14, the point estimates reach approximately 0.35 log points, corresponding to a 42 percent increase in total visits, or roughly 43,000 additional visits per district-quarter against the pre-treatment mean of about 102,000.

The effects appear to plateau beyond this horizon rather than continuing to grow, suggesting that the programme reaches a coverage ceiling as the pool of willing enrollees in each district saturates. The widening confidence intervals at longer horizons reflect a mechanical feature of staggered adoption designs: only the earliest-treated cohorts contribute to long-horizon estimates, so the effective sample shrinks as event time increases. This does not indicate that the effects themselves are becoming less precise in any economic sense; it simply means that fewer districts inform these estimates.

Dynamic ATTs for other outcomes exhibit a similar pattern: near-zero pre-treatment leads followed by gradual post-treatment emergence that levels off at long horizons. This consistency across outcomes reinforces the interpretation that the measured effects reflect genuine changes in healthcare utilisation driven by insurance coverage, rather than artefacts of a particular outcome measure or specification choice.

Figure 4 — Total Clients Served Dynamic ATTs



Notes. This figure shows dynamic ATTs for total clients served by health service providers. The outcome is log-transformed total clients served. The x-axis shows quarters since NHIP implementation, with quarter 0 being the first implementation quarter. The shaded area represents a 95% simultaneous confidence band, constructed using the bootstrap procedure in Callaway and Sant'Anna (2021).

5.4. Summary

The headline behavioural response to NHIP is substantial but unequally distributed. Total visits rise by 14 percent, with similar magnitudes across disease categories and across age groups, peaking among those aged 60 and above. The aggregate effect, however, is driven entirely by low-poverty districts. Splitting districts at the median pre-treatment poverty rate, only low-poverty districts show statistically significant gains on total visits, OPD visits, emergency visits, non-communicable diseases, and injuries; the three-group split sharpens this further and shows that the gains accrue specifically to the most affluent quarter of districts. NHIP's secondary objective of reducing poverty through improved health-care access is therefore unmet on the utilisation margin: the gains concentrate in the districts that were already better off.

Utilisation gains do not translate directly into welfare gains. The estimated 14 percent could reflect previously unmet need that households would have valued even at the pre-NHIP out-of-pocket price, or moral hazard induced by the lower out-of-pocket price after insurance, or some combination. The data here do not separate these channels, so the welfare interpretation of the headline number remains open.

The aggregate utilisation response is in line with the LMIC literature on large public insurance expansions, which finds positive average utilisation effects in most evaluations (Erlangga et al. 2019; Gruber, Hendren, and Townsend 2014; Gruber, Lin, and Yi 2023; Dupas and Jain 2026). The equity result is what differentiates the Nepal experience. The closest comparison is Mexico's Seguro Popular: Conti and Ginja (2023) find that SP reduced infant mortality by 10 percent in poor municipalities and closed 98 percent of the poor–rich IMR gap, with zero effects in rich municipalities. NHIP produces the mirror image. The same-direction comparison comes from a free-care randomised experiment in rural Ghana, where Powell-Jackson et al. (2014) find utilisation gains accruing more strongly to wealthier households within a poor study population; this paper documents the same qualitative pattern at the district level under a nationwide staggered rollout, though Section 6 traces it to a different (supply-side) mechanism than the demand-side free-care lever studied there. The mechanism that produces the equity result, examined in Section 6, is supply-side rather than demand-side, paralleling supply and provider-network frictions documented for India's RSBY (Malani et al. 2024) and the empanelment constraint in

Dupas and Jain (2024).

Effects emerge gradually rather than at the moment of programme implementation. Post-treatment coefficients are small and insignificant in the first four quarters, become statistically significant from the fifth quarter onwards, and plateau at twelve to fourteen quarters at approximately 0.35 log points. The most natural reading is that coverage takes time to translate into care-seeking: registration must be processed, insurance cards must be distributed, households must learn that they are covered, and empanelled facilities must adjust capacity to the new claim flow. Conti and Ginja (2023) find a similar timing pattern for Mexico's Seguro Popular. Their summary specification places the entire infant-mortality effect in the bin of three or more years post-implementation rather than in the immediate zero-to-two-year bin, and they attribute the lag to the same combination of awareness and supply-side adjustment.

5.5. Robustness and Falsification

5.5.1 Robustness Checks

The headline result, a 14 percent increase in total visits, is supported by robustness checks. The checks address sample composition, estimator and specification choice, and inference. The full tables are in Appendix A.6.

Alternative samples. Two sample restrictions test whether the result is driven by particular districts. Restricting to public facilities only (Table A10) preserves significance at 1% for all three primary outcomes; emergency and OPD effects are largely unchanged, while total visits attenuate to roughly half (0.071 vs. 0.133), consistent with private facilities absorbing a substantial share of the total-visit response rather than with reporting artefacts among newly empanelled private providers. Excluding Kathmandu (the last district to receive NHIP and by far the largest urban district) leaves all three primary outcomes significant at 1%, with somewhat larger point estimates for total visits (0.151 vs. 0.133) and emergency visits (0.261 vs. 0.179) (Table A11), confirming that the result is not produced by an atypical urban outlier.

Estimator and specification. Four further checks address the chosen estimator and time aggregation. Replacing the varying base period with a universal base period (Table A9) leaves the post-treatment ATTs unchanged, consistent with the equivalence noted in Section 4. All outcomes are strictly positive at the district-quarter level and enter the main specifications under the pure $\log$ transformation. As a robustness check on the choice of monotonic transformation, I re-estimate all specifications using the inverse hyperbolic sine (Tables A14–A16); results are qualitatively and quantitatively unchanged. Alternative heterogeneity-robust staggered DiD estimators (the local-DID estimator of Chaisemartin and D'Haultfœuille (2020) and the imputation estimator of Borusyak, Jaravel, and Spiess (2024)) produce sign-consistent positive estimates for all three primary outcomes (Table A13). For total visits, the Borusyak estimate (0.126) tracks the Callaway–Sant'Anna headline (0.133) closely, while the de Chaisemartin estimate (0.082) is smaller; the gap reflects differences in weighting and target parameters across estimators. Finally, re-aggregating to monthly data and varying the anticipation window from $\delta = 0$ to $\delta = 3$ months (Table A12) shows that the effect strengthens monotonically with longer anticipation horizons. This shows that the choice of anticipation window matters for the results because of the administrative delay between registration and eligibility. A concerning reverse pattern would be visible in the pre-treatment event-study coefficients themselves, in the form of a downward pre-trend or selection effect that no choice of δ could absorb; Figure 4 shows no such pattern.

Inference. Two further procedures address inference concerns. The honest-confidence-interval procedure of Rambachan and Roth (2023) (Figure A3) shows that the total-visits effect remains bounded away from zero up to $\bar{M} = 0.3$ (post-treatment trend violations up to 30% of the largest pre-treatment deviation); the lower bound crosses zero at $\bar{M} = 0.4$. Calibrated against a pre-treatment deviation whose own confidence band contains zero, the $\bar{M} = 0.3$ threshold corresponds to a tolerated violation of roughly 24% of the headline coefficient. This is a moderate rather than maximal robustness margin; the economic interpretation is in Appendix A.5. To address multiple-hypothesis testing across the family of primary outcomes, I report family-wise-error-rate-adjusted p-values following Romano and Wolf (2005) (Table A4); the main-outcome significance survives the adjustment.

5.5.2 Cross-Source Validation

This subsection asks whether the district-level DiD estimates are corroborated by an entirely different administrative trail: the insurance claims that hospitals file with the Health Insurance Board (HIB) on behalf of insured patients. If the DiD effects reflect genuine insurance-driven utilisation, the implied increase in visits should appear as insured visits of comparable magnitude and composition in the claims records; if they instead reflect reporting artefacts in DHIS2, no such correspondence is expected. The HIB claims records are individual-level and independent of the DHIS2 facility reports used in the main analysis. Table 5 compares the magnitudes implied by the DiD coefficients with the average insurance-financed visits observed in the claims data for treated districts. Using the DiD coefficients and pre-treatment DHIS2 means, I compute the implied per-district–quarter increase via $(e^{\hat{\theta}} - 1) \times \bar{Y}_{pre}$ and place these in a 95% confidence interval alongside the claims-observed averages. The female-share row, which enters the regression in levels rather than logs, uses an additive construction: $\bar{Y}_{pre} + \hat{\theta} \pm 1.96 \cdot \mathrm{SE}$. To make the comparison more meaningful, I aggregate the claims data to the same district–quarter level.

Table 5 — Comparison of DiD-Implied Utilization Changes with Insurance Claims

	Claims Data (Observed)	DHIS2 Implied [95% CI]	Table Reference
Total Visits	10,876	[8,211; 21,187]	Table 2
Female Share of Visits	0.5597	[0.5729; 0.5807]	Table 2
Age 0–9	707	[537; 3,051]	Table 3
Age 10–19	724	[1,316; 4,181]	Table 3
Age 20–59	6,409	[4,772; 11,486]	Table 3
Age 60+	2,865	[1,705; 3,887]	Table 3
Total Visits (High Poverty)	6,780	[-2,649; 7,553]	Table 4
Total Visits (Low Poverty)	12,234	[8,142; 30,291]	Table 4

Note: I first average number of insurance-financed visits per district among enrolled individuals in treated districts during the post-treatment period to make the figures comparable. DHIS2 figures are derived by applying the estimated DiD coefficients to pre-treatment mean utilisation levels and represent implied changes in total facility visits. The two data sources measure different populations (all visitors vs. enrolled claimants), so the comparison is one of magnitude and composition rather than a strict equivalence.

The two sources broadly agree, with two exceptions. In aggregate, the DiD estimates imply an increase of roughly 14,500 visits per district–quarter (against a pre-treatment mean of about 102,000); claims data record about 10,900 insured visits, well within the implied confidence interval. The pattern persists across age groups, with both sources showing larger utilisation increases among working-age adults. The two exceptions are ages 10–19, where DiD-implied visits substantially exceed observed claims, and the female share of visits, where the claims-observed value (0.560) sits 1.3 percentage points below the DHIS2-implied 95% CI [0.573, 0.581]. The latter likely reflects that women are slightly under-represented among insurance claimants relative to their share of all facility visits, consistent with differential enrolment or claiming patterns by gender. The poverty rows are consistent with the heterogeneity reading. Claims-observed visits in high-poverty districts (6,780) and low-poverty districts (12,234) both fall within the implied DiD ranges ($[-2{,}649; 7{,}553]$ and $[8{,}142; 30{,}291]$), and insured care is markedly less prevalent in poor districts. The low-poverty agreement is informative because the implied range is bounded away from zero; the high-poverty agreement is largely mechanical because the implied range itself crosses zero and is wide enough to accommodate most plausible claim levels. The high-poverty claims figure reflects actual insurance use by the small number of enrollees in those districts rather than a precisely estimated causal effect.

This comparison is a cross-source consistency check rather than a strict independence test. Both DHIS2 and HIB claims share recording incentives created by NHIP empanelment, so neither source rules out improvements in reporting as part of the response, and the magnitudes should be read as order-of-magnitude rather than point comparisons (the CS estimator produces a cohort-weighted ATT, while the claims figure is a simple post-period average). What the comparison does establish is that the DiD-measured effect, both in aggregate and in the poverty split, translates into individual-level insurance use of similar magnitude and similar composition, a pattern that is hard to sustain under a pure DHIS2 reporting-artefact explanation. The mechanism analyses in Section 6 use these same HIB records to decompose where in the utilisation pipeline the high-poverty gap accumulates.

5.5.3 Falsification Exercise

The falsification exercise uses outcomes covered by the Safe Motherhood Programme (SMP), a free-care scheme that predates NHIP by seven years. Launched in 2009 and implemented nationwide at the community level, the SMP aims to reduce maternal and neonatal morbidity and mortality by providing institutional delivery free of charge to any Nepali mother and newborn at any public health facility with birthing services, incentivising antenatal care, and covering transportation costs. The SMP therefore eliminated, well before NHIP arrived, the financial barriers to maternal care that NHIP would otherwise address. The exclusion is also institutional: HIB regulations explicitly bar empanelled providers from claiming services delivered under MoHP vertical programmes through NHIP, naming the Safe Motherhood Programme and Family Planning among the programmes covered by the bar. Together, the two regulatory features imply that NHIP eligibility should produce no detectable effect on SMP outcomes; a non-null estimate on maternal mortality, stillbirths, or institutional delivery would instead indicate broad time-varying confounders. Table 6 reports the falsification results. Health insurance has no detectable effect on the four safe-motherhood outcomes. Across simple, group, and calendar aggregations, none of the estimates can be distinguished from zero. The null is therefore over-determined by both regulatory features; failing to detect an effect is consistent with NHIP being correctly identified but is also predicted by the absence of any payment or eligibility channel through which a confound could operate.

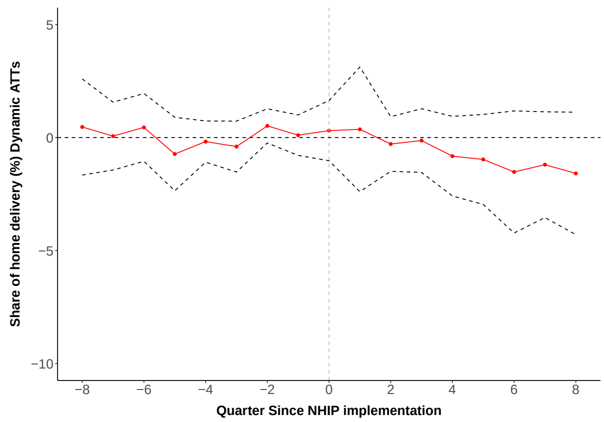
Table 6 — Health Insurance Eligibility on Safe Motherhood

	Share of home delivery	Share of delivery attended by SBA	Maternal Mortality Ratio	Ratio of still birth (per 1000 live births)
Model:	(1)	(2)	(3)	(4)
Simple Aggregation				
Is NHIP eligible = 1	-0.0103	0.0157	12.4370	0.6529
	(0.0071)	(0.0110)	(7.6634)	(1.6171)
Group Aggregation				
Is NHIP eligible = 1	-0.0034	0.0155	6.8191	1.6647
	(0.0081)	(0.0100)	(6.4903)	(1.7510)
Calendar Aggregation				
Is NHIP eligible = 1	-0.0051	0.0094	7.8861	0.4202
	(0.0063)	(0.0094)	(5.4218)	(1.4762)
Pre-Treatment Mean	0.07	0.89	12.27	17.11
Fit statistics				
Number of Cohorts	17	17	17	17
Number of Time Periods	30	30	30	30

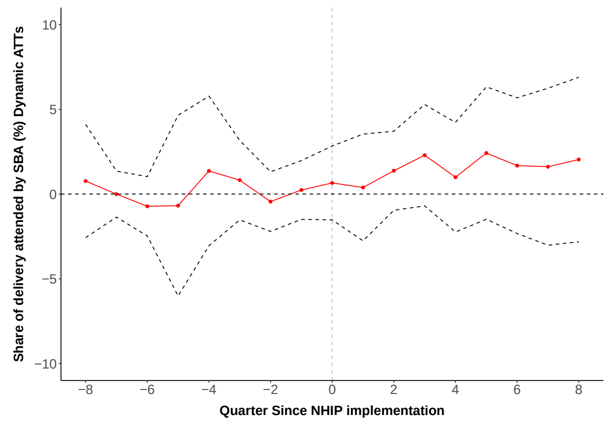
Note: Clustered (District) standard errors in parentheses bootstrapped with 1000 iterations. Estimation Method: Doubly Robust. Control Group: Not Yet Treated, Anticipation Periods: 1. Signif. Codes: *** 99% ** 95% * 90% uniform confidence band does not include 0.

I show the dynamic ATTs in Figure 5. The dynamic ATTs further show that the effect on these variables cannot be distinguished from zero throughout the study period.

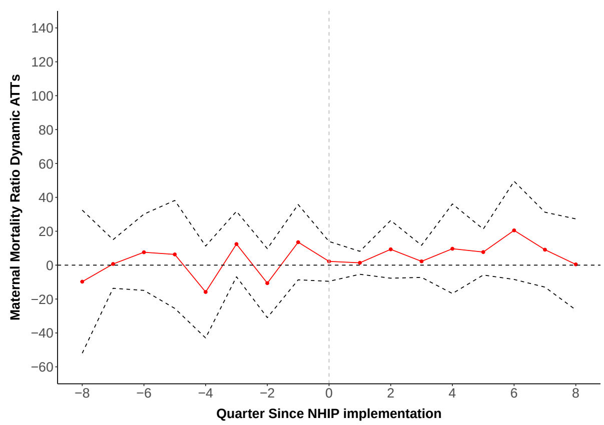
Figure 5 — Dynamic ATTs for Safe Motherhood outcomes



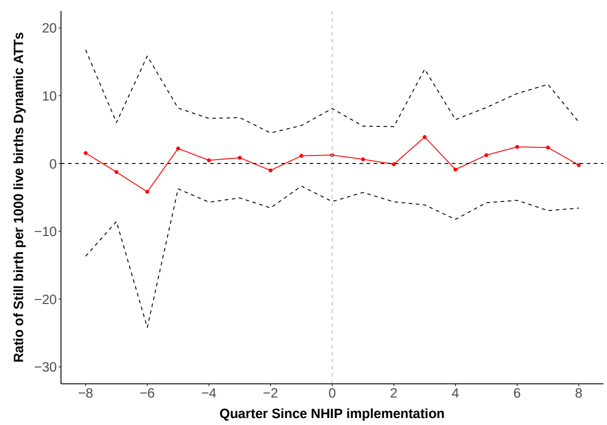
(a) Share of home delivery



(b) Share of delivery attended by SBA



(c) Maternal Mortality Ratio



(d) Ratio of Still Birth

Note: Panel (a) shows the dynamic ATTs for share of home delivery. Panel (b) shows dynamic ATTs for share of delivery attended by a skilled birth attendant. Panel (c) shows dynamic ATTs for maternal mortality ratio. Panel (d) shows dynamic ATTs for ratio of still births per 1000 live births. These dynamic ATTs are from CS model, where x-axis is quarter since NHIP implementation and y-axis is outcome. The black dotted lines represent a 95% uniform confidence interval.

The null on safe-motherhood outcomes is consistent with the main results not being driven by broad improvements in healthcare services or district-specific trends. The four SMP outcomes are all expressed as ratios (institutional-delivery share, SBA-attended share, maternal mortality per live birth, still-birth ratio), which are partially insulated against uniform reporting improvements that scale numerator and denominator equally; the mortality ratios are additionally underpowered given baseline rates. The binding check against reporting artefacts is therefore the public-facilities-only restriction (Table A10, discussed in Section 7), which removes private facilities whose reporting behaviour is most plausibly tied to NHIP empanelment incentives.

6 — Mechanism: Why Gains Concentrate in Low-Poverty Districts

Section 5.2 established that NHIP's utilisation gains concentrate in low-poverty districts; the claims data in Section 5.5.2 confirm that insured visits per capita run several times higher in low-poverty than in high-poverty districts. Why are utilisation gains concentrated in low-poverty districts? Several channels could explain this pattern. The NPR 3,500 premium may be unaffordable to poor households despite being modest in absolute terms; 17 percent of Nepal's population is multidimensionally poor (GoN 2021). Low enrolment, driven by awareness gaps, documentation barriers, or distrust, may keep the programme from reaching the poor (Banerjee et al. 2021; Watson, Yazbeck, and Hartel 2021). Even among those who do enrol, demand-side frictions such as distance, opportunity costs, and low health literacy may suppress claiming. Supply-side constraints, especially too few empanelled providers, may leave coverage formally in place but practically inert (Dupas and Jain 2024; Lagomarsino et al. 2012). This section uses registration and claims data from HIB (claims are filed by empanelled facilities on behalf of insured patients) to map where in the utilisation pipeline the high-poverty gap appears, and a 2019 policy reform to provide causal evidence on the supply-side channel.

6.1. Supply vs Demand: Descriptive Evidence

I begin by examining enrolment patterns and healthcare infrastructure across poverty groups in Table 7, using individual-level registration and claims data from HIB to compute district-level averages.

Table 7 — Mechanisms Behind Poverty Heterogeneity

	High Poverty	Low Poverty
Panel A: Demand-side		
Registration Rate (per 1,000 pop.)	22.6	28.9
Share Claimed (conditional on enrollment)	0.253	0.342
Panel B: Supply-side		
Number of Registered Health Facilities	96	123
Number of Empanelled Facilities	4.5	8.3

Note: Author's own calculation using HIB registration and claims data. All figures are district-level averages over the post-treatment period (2016–2022). Registration rate is the average quarterly number of individuals enrolled in NHIP per 1,000 district population. Any Claim measures the proportion of enrollees who made at least one insurance claim over the post-treatment period. Number of health facilities includes all public and private healthcare providers in the district. Number of empanelled facilities refers to healthcare providers contracted to deliver NHIP-covered services.

Panel A of Table 7 shows that enrolment patterns differ substantially across poverty groups. Individual-level registration rates are 22 percent lower in high-poverty districts compared to low-poverty districts (22.6 versus 28.9 per 1,000 population), suggesting that demand-side barriers, whether premium costs, documentation requirements, or limited programme awareness, constrain take-up in poorer areas. Even among those who do enrol, utilisation rates diverge sharply: only 25 percent of enrollees in high-poverty districts make at least one insurance claim, compared to 34 percent in low-poverty districts. This pattern indicates that enrolment alone does not translate into equivalent healthcare access, a finding similar to what Banerjee et al. (2021) document for Indonesia's national health insurance programme.

Panel B reveals a potential explanation for the utilisation gap conditional on enrolment. High-poverty districts have 22 percent fewer healthcare facilities on average (96 versus 123), but the disparity in NHIP-empanelled facilities is sharper: poor districts average 4.5 empanelled providers compared to 8.3 in richer districts, a 46 percent gap. However, empanelment itself is endogenous. One reason empanelment can be lower in poorer districts is the online system. NHIP requires facilities to submit claims and complete empanelment through a fully online portal, and facilities in poorer districts may lack the IT infrastructure (reliable internet, computer hardware, and trained staff) to participate. With so few contracted providers, enrollees in poor districts face limited options for accessing covered care. They may need to travel longer distances to reach an empanelled facility, or forgo insurance benefits entirely and seek care from non-empanelled providers at full cost.

The split between supply and demand is not as clean as Panel A versus Panel B suggests. Empanelled providers themselves serve as an information channel: individuals who visit a hospital without insurance can learn about NHIP from the facility and find out how to enrol. Districts with fewer empanelled providers therefore have weaker information diffusion in addition to fewer service-delivery points, which means part of what reads as a demand-side gap in Panel A is itself downstream of the supply-side gap in Panel B.

Taken together, the descriptive evidence in Table 7 points to both supply-side constraints and demand-side frictions as contributors to the reduced utilisation response in high-poverty districts: Panel A reveals demand-side barriers at the registration and claiming margins, and Panel B documents supply-side gaps in empanelled providers. The remainder of this section quantifies these channels and isolates specific mechanisms.

6.2. Decomposing the Gap: Enrolment vs Utilisation

The descriptive patterns in Table 7 motivate a formal accounting exercise. To document where in the utilisation pipeline the high-poverty gap accumulates, I decompose the total gap in insurance claims per capita between high-poverty (above-median) and low-poverty (below-median) districts, consistent with the median split used in the main heterogeneity analysis, using a three-factor extension of the Kitagawa (1955) rate decomposition. Claims per capita are written as enrolment (enrollees/population) times probability of use (claimants/enrollees) times intensity (claims/claimant), and the gap is attributed to each margin using symmetric (midpoint) weights, with an explicit residual interaction term that captures joint disadvantages across margins. Each group's margins are computed by pooling district-level counts and dividing aggregates, which enforces the multiplicative identity $A \times B \times C =$ claims/pop exactly within each group. Unlike the Das Gupta (1978, 1993) variant, which constructs weights that redistribute interaction effects into main-effect components, I retain the interaction explicitly to make joint disadvantages visible; Appendix A.8 reports the Das Gupta counterpart as a robustness check.

Table 8 — Decomposition of the Poverty Gap in Insurance Utilization

	Claims per 1,000 pop.	Share of gap (%)
Panel A: Group means		
High-poverty districts	598.9	
Low-poverty districts	1440.1	
Total gap	-841.2	100
Panel B: Decomposition		
Enrollment margin	-427.9	51
	(200.4)	(19.1)
Probability of use	-187.4	22
	(96)	(8.8)
Intensity (claims per user)	-221.2	26
	(77.9)	(17.9)
Interaction	-4.8	1
	(6.3)	(0.4)

Notes: Three-way decomposition of the gap in insurance claims per 1,000 population between high-poverty (above-median) and low-poverty (below-median) districts, using a median split of the pre-treatment 2011 district poverty rate (World Bank) to match the main heterogeneity analysis. Claims per capita are decomposed as: (enrollees/population) × (claimants/enrollees) × (claims/claimant). The enrollment margin accounts for differences in enrollment rates, holding utilisation behaviour constant at the midpoint. The probability-of-use margin accounts for differences in the share of enrollees who file at least one claim. The intensity margin accounts for differences in the number of claims per claimant, conditional on any use. The interaction captures the joint contribution of simultaneous disadvantages across margins. Group-level rates are computed as aggregate totals (e.g., total registrants / total population within the poverty group). Standard errors in parentheses are bootstrapped (500 iterations, resampling districts within poverty groups).

Table 8 presents the decomposition results. Low-poverty districts generate roughly 2.4 times more insurance claims per capita than high-poverty districts (1,440 vs 599 per 1,000 population), yielding a total gap of 841 claims per 1,000. The enrolment margin contributes 51 percent of the gap, the probability-of-use margin 22 percent, and the intensity margin 26 percent. A small interaction term (1 percent) indicates that the three margins contribute nearly additively.

The point estimates put enrolment first, but the precision does not support a clean ranking. The three margin shares have substantial standard errors and their confidence intervals overlap with each other, so enrolment, probability-of-use, and intensity are not statistically distinguishable. The accounting therefore describes where in the pipeline the shortfall is located, not which margin is dominant. The three margins are also jointly determined: any one share is conditional on the others, so the figures are not policy counterfactuals.

What the table does establish is that the high-poverty shortfall is broad-based across the utilisation pipeline. Enrollees in poor districts are scarcer, are less likely to claim once enrolled, and file fewer claims when they do. A pure demand-side reading would have to explain why willingness to engage is independently depressed at each of three sequential margins. A simpler reading is that supply itself is binding throughout. As noted in the descriptive evidence above, empanelled providers serve as an information channel as well as a service-delivery point; fewer empanelled providers therefore mean fewer opportunities to learn about and register for NHIP, fewer opportunities to claim once enrolled, and fewer follow-on care episodes per claimant. The next two subsections test this supply-side reading directly.

6.3. Within-District Poverty Differentials

The analysis so far has used a district-level poverty classification, treating each district as a single unit. To separate demand-side barriers from district-level supply constraints, I now exploit within-district variation in poverty at the municipality level. Using Small Area Estimation poverty rates at the ilaka level from the 2011 Census, mapped to current municipality boundaries via a federal restructuring concordance (748 of 753 municipalities matched: 4 LGUs have no recorded NHIP registrations and 1 (Sunkoshi, Udayapur) is unmatched by the federal-restructuring concordance; the analysis therefore covers 100% of registered enrollees by construction. The 4 zero-registrant LGUs are themselves consistent with the equity reading: in places with no NHIP uptake, there is no within-municipality variation to identify), I construct a municipality$\times$quarter panel and progressively add controls to decompose the poverty differential.

District$\times$quarter fixed effects absorb all district-level time-varying confounders, including the number and type of empanelled providers, NHIP rollout timing, and district-level health infrastructure. The identifying variation is thus across municipalities within the same district-quarter. I then progressively add geography controls (elevation, terrain ruggedness, road access, land area), demand-side proxies (rural status, education), and a supply-side control (log number of unique assigned First Point of Contact facilities per municipality) to decompose the sources of the poverty differential.

Figure 6 shows the unconditional relationship between municipality poverty and insurance outcomes. The two panels tell different stories. Registrations per 1,000 population show a negative gradient with municipality poverty, but the slope is noisy and not visibly clean. Claims per registrant, in contrast, slope clearly downward across the poverty range: poorer municipalities file substantially fewer claims per registered person. The contrast is informative on its own. If the only difference between rich and poor municipalities were that fewer poor people register, claims per registrant should be flat: registered users would behave similarly across municipalities. Instead, registered users in poor municipalities claim less, which means something is constraining utilisation conditional on enrolment. The most natural candidates are the count of providers serving these municipalities and the quality of the services those providers can deliver.

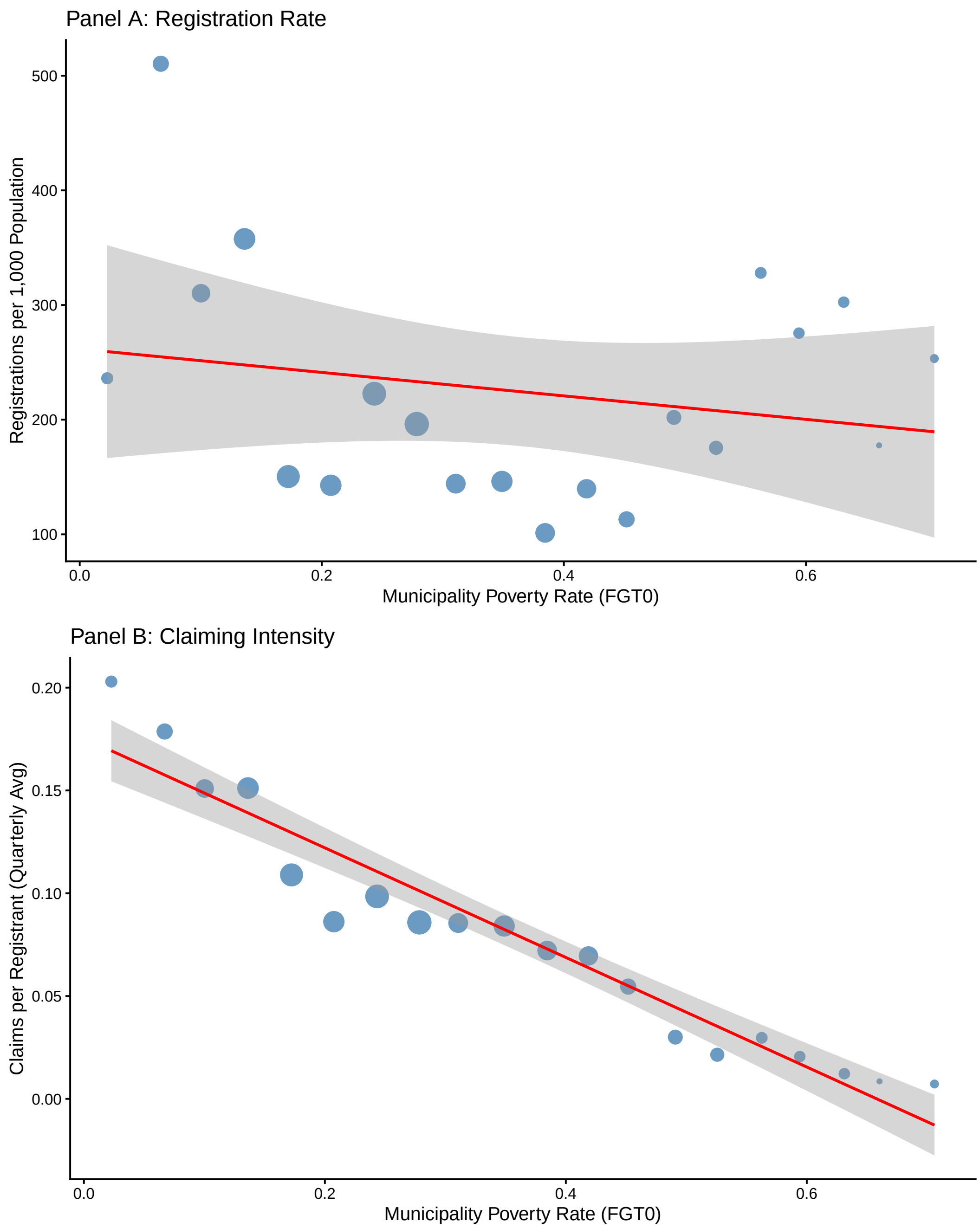


Figure 6 — Municipality Poverty and Insurance Outcomes

Table 9 formalises this pattern with a sequential decomposition that progressively adds controls.

Table 9 — Within-District Poverty Gradient: Registration and Claims

	Reg. per 1,000 pop				Claims per Registrant			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High Poverty VDC	-80.2^{***}	-60.1^{***}	-32.2	-8.725	-0.032^{***}	-0.027^{***}	-0.020^{***}	-0.015^{**}
	(22.21)	(21.81)	(20.88)	(18.91)	(0.007)	(0.007)	(0.006)	(0.006)
District $\times$ Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Urbanization + Education	No	No	Yes	Yes	No	No	Yes	Yes
Supply control	No	No	No	Yes	No	No	No	Yes
Observations	13,122	13,122	13,122	13,122	13,122	13,122	13,122	13,122
Unique municipalities	748	748	748	748	748	748	748	748
R^2	0.752	0.757	0.774	0.785	0.503	0.510	0.521	0.525

Notes: The unit of observation is a VDC$\times$quarter over the post-treatment period (2016–2022), restricted to VDC-quarters with positive registrations. High Poverty VDC is a binary indicator equal to one if the VDC's poverty headcount rate (FGT0) exceeds the national median (0.264). VDC-level poverty is the population-weighted ilaka-level headcount rate from Small Area Estimation (CBS/World Bank, 2011 Census), mapped to current VDC boundaries via a federal restructuring concordance. Geography controls: log elevation, log ruggedness, log land area, road access. Urbanization + Education: rural indicator, share with intermediate education or above. Supply control: log number of unique assigned First Point of Contact (FPOC) facilities per VDC—an exogenous measure reflecting designated provider availability, not utilization choices. Clustered (VDC) standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Column (1) shows the raw within-district poverty differential: municipalities with above-median poverty rates have 80 fewer registrations per 1,000 population and 0.032 fewer claims per registrant compared to low-poverty municipalities within the same district-quarter. Adding geography controls (column 2) attenuates the registration coefficient by 25%, suggesting remoteness plays a meaningful role. Adding urbanisation and education (column 3) produces further attenuation, with the registration coefficient falling to -32.2: rural municipalities register less, and education is positively associated with both outcomes, consistent with demand-side barriers related to information and health literacy. Columns (3) and (4) are not individually statistically significant for registration (p $\approx$ 0.12 and p $\approx$ 0.63 respectively), meaning the demand, supply, and residual shares should be interpreted as indicative of the pattern of attenuation rather than precisely estimated.

The critical test is column (4), which adds the number of assigned healthcare facilities (unique FPOCs) as a within-district supply control. This measure is predetermined: facilities are designated at registration, not chosen by enrollees. Adding supply attenuates the poverty coefficient further but it remains negative (-8.725). Sequentially absorbing geography, demand-side proxies, and the supply control reduces the registration differential by roughly 89%, with a residual 11% remaining after all three blocks. The block-by-block attenuations (25%, 35%, and 29% of the original differential) are sensitive to the order of inclusion; they describe how the conditional correlation moves as each block enters, not a causal decomposition of the differential into channels. The claims outcome maintains statistical significance through all columns (columns 5–8), providing a more precise decomposition for that margin. The analysis is restricted to municipality-quarters with positive registrations, excluding municipalities where no one enrolled; this likely attenuates the estimated poverty differential, as high-poverty municipalities with zero enrolment are omitted by construction.

The two patterns answer different parts of the question. The within-district registration shortfall in poor municipalities co-moves with structural correlates of poverty: once geography, urbanisation, and the count of FPOCs enter the specification, the registration coefficient is statistically indistinguishable from zero, with point estimate -8.725 and a wide confidence interval. This is consistent with the shortfall being driven by remoteness, rural status, and a thinner local supply network rather than poverty itself, though the imprecision of the residual coefficient leaves room for other channels; the four zero-registration municipalities are also excluded from the panel by construction. The claims-per-registrant differential, by contrast, persists through all four blocks of controls. Accounting for providers and adjusting for geography and demographic composition does not close this gap. Something below the level of FPOC count and basic geography is therefore constraining claiming behaviour among registered enrollees in poor municipalities. The most plausible candidates are the quality and depth of the facilities serving those municipalities, such as service breadth, hours of operation, specialist availability, and effective referral capacity, which the count of unique FPOCs does not measure. The next subsection examines distance to the assigned FPOC directly using a policy reform, which provides a sharper test of the supply-side channel.

6.4. *Access Barriers: Evidence from FPOC Reassignment*

To provide causal evidence on the role of access barriers, I exploit the 2019 policy change that restructured First Point of Contact (FPOC) designation. Prior to this reform, enrollees could choose any empanelled facility, public or private, as their FPOC; the 2019 reform reassigned individuals to a new FPOC. Crucially, the reform was an administrative reshuffling of existing assignments rather than a supply change: no facilities entered or exited the programme at the reform date, and the underlying network of empanelled facilities was held fixed. This exercise is similar in spirit to Dupas and Jain (2024), who exploit the 2018 expansion of hospital empanelment under India's BSBY programme to identify a distance channel in care utilisation. The mechanism differs in a substantive way. In Dupas and Jain (2024), distance falls because supply grows — new private hospitals enter the programme near previously underserved locations. In my setting, the supply of empanelled care is held constant; distance changes only because individuals are administratively reassigned to a different facility within the same fixed network. The reduced-form effect on claims therefore reflects the response to a pure access shock, holding supply fixed.

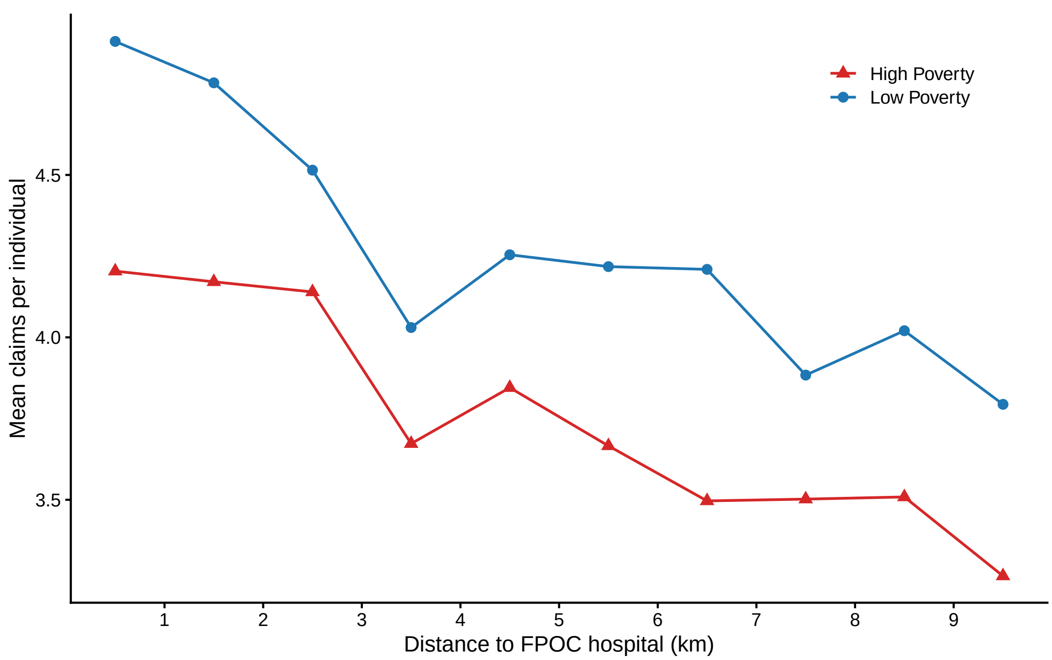
To isolate the distance channel from confounds that vary across provider ownership types (service breadth, staffing levels, payment procedures), I restrict the treated group to individuals whose pre- and post-2019 FPOCs are both public facilities. The control group is all individuals whose FPOC did not change, which includes both public-FPOC and private-FPOC stayers. Because controls did not switch FPOCs, no ownership-change channel operates on them. The identifying assumption is that the pre/post change in claims among non-switchers within the same ward provides the counterfactual trend for switchers, which requires public-FPOC and private-FPOC stayers to share parallel claiming trends within ward. Identification is local to the subpopulation of compliers who moved between public facilities, a LATE on the same provider type, which sidesteps monotonicity concerns arising from heterogeneous transition types. The restriction substantially shrinks the treated group but leaves the control group unchanged.

Distance from each individual's ward centroid to their FPOC facility is computed as a road distance: 97% of origin–destination pairs are routed using the Google Maps Distance Matrix API, with OSRM as a fallback (3%) for pairs Google failed to return. The location of each of the 501 empanelled facilities in the network was geocoded manually.

Identification compares reassigned individuals to non-reassigned individuals in the same ward, before and after the 2019 reform. The sample is restricted to individuals observed claiming in both the pre- and post-2019 periods, so the results describe the intensive margin of active claimants; the extensive margin is by construction outside the panel.[14]

Before turning to causal estimates, Figure 7 shows the unconditional relationship between distance to FPOC and claims, separately for high- and low-poverty districts. Both groups show a downward gradient: low-poverty enrollees decline from 4.91 claims per individual in the 0–1 km bin to 3.79 at 9–10 km, and high-poverty enrollees decline from 4.20 to 3.27 over the same range. Two patterns stand out — low-poverty enrollees claim more at every distance, and their gradient with distance is steeper. The FPOC reassignment exercise then asks whether moving an individual closer causally raises their claim count, and whether the response differs across poverty groups.

Figure 7 — Distance to FPOC and Claims: Unconditional Relationship



Notes. This figure plots mean claims per individual against distance from ward centroid to FPOC facility, separately for high- and low-poverty districts (median split of district poverty rate). The sample pools all individual-time observations across all enrollees, prior to the Pub→Pub restriction used in the regression analysis. Each point is a 1-km bin from 0 to 10 km.

14. If policy-induced attrition were differential across treatment status, the intensive-margin estimate would be biased. Appendix Table A22 shows that among pre-2019 claimants with a recorded FPOC, those with private FPOCs (more likely to be reassigned by the policy) had a lower attrition rate than those with public FPOCs, inconsistent with policy-induced selective attrition into the panel.

Table 10 reports the effect of FPOC reassignment on distance to care. The full-sample estimate is -2.18 km ($p < 0.01$): reassigned individuals end up 2.18 km closer to their FPOC on average. These standard errors are clustered at the ward level, the level at which FPOC reassignment varies; under the more conservative district-level clustering the full-sample first stage is no longer significant at conventional levels (district-clustered standard errors in brackets, Table 10), a caveat I carry to the instrumental-variables interpretation in Appendix A.7. The effect is highly heterogeneous across poverty groups. In low-poverty districts the policy delivers a -2.25 km reduction ($p < 0.01$); in high-poverty districts the effect is -0.76 km and statistically indistinguishable from zero. The interaction test rejects equality across poverty groups ($p = 0.042$). This pattern is consistent with binding supply-side constraints in high-poverty districts: where public facilities are sparse, the nearest public facility was likely already in use, so the reassignment policy has little distance to compress.

Table 10 — Effect of 2019 FPOC Policy on Distance to Care

	Full Sample	High Poverty	Low Poverty
	(1)	(2)	(3)
Treated × Post	-2.181***	-0.761	-2.250***
	(0.387)	(0.608)	(0.405)
	[1.475]	[0.540]	[1.573]
Fixed effects			
Individual FE	Yes	Yes	Yes
Ward × Post	Yes	Yes	Yes
Summary statistics			
Mean distance (km)	6.4	6.7	6.3
Unique FPOC hospitals	250	101	153
Observations	506,242	122,626	383,616
N (individuals)	262,842	64,036	198,806
N (wards)	1,587	739	848

Note: This table reports the effect of the 2019 FPOC policy change on road distance (km) to the designated First Point of Contact hospital. Treated × Post captures the differential distance change for individuals whose FPOC was reassigned following the restriction to public hospitals. The treated group is restricted to individuals whose pre- and post-2019 FPOCs are both public hospitals, holding provider ownership constant so the change is purely in distance; controls are all individuals whose FPOC did not change. All specifications include individual fixed effects and ward × post interactions to absorb ward-level time shocks. Ward-clustered standard errors in parentheses; district-clustered standard errors in brackets. Poverty groups defined by median split of district poverty rate. Distance outliers above 20km excluded. Distances are road distances from Google Maps (97.1%) and OSRM routing (2.9%); no straight-line fallback was required. Signif. Codes: *** 1% ** 5% * 10%.

Table 11 reports the reduced-form effect of FPOC reassignment on insurance claims. Reassigned individuals make 1.13 more claims per individual on average ($p < 0.01$), a 27% increase relative to the sample mean of 4.18 claims. The poverty heterogeneity mirrors Stage 1: low-poverty districts see a 1.15-claim increase, high-poverty districts a 0.59-claim increase, and the difference is highly significant ($p < 0.01$).

Table 11 — Effect of FPOC Policy on Insurance Claims

	Full Sample	High Poverty	Low Poverty
	(1)	(2)	(3)
Treated × Post	1.126***	0.592***	1.152***
	(0.091)	(0.169)	(0.095)
	[0.292]	[0.181]	[0.313]
Fixed effects			
Individual FE	Yes	Yes	Yes
Ward × Post	Yes	Yes	Yes
Summary statistics			
Mean claims	4.18	3.73	4.32
Unique FPOC hospitals	250	101	153
Observations	506,242	122,626	383,616
N (individuals)	262,842	64,036	198,806
N (wards)	1,587	739	848

Note: This table reports the reduced-form effect of the 2019 FPOC policy change on the number of insurance claims per individual. Treated × Post captures the differential change in claiming for individuals whose FPOC was reassigned following the restriction to public hospitals. The treated group is restricted to individuals whose pre- and post-2019 FPOCs are both public hospitals, holding provider ownership constant so the change is purely in distance; the control group is all individuals whose FPOC did not change. All specifications include individual fixed effects and ward × post interactions to absorb ward-level time shocks. Ward-clustered standard errors in parentheses; district-clustered standard errors in brackets. Poverty groups defined by median split of district poverty rate. Distance outliers above 20km excluded. Distances are road distances from Google Maps (97.1%) and OSRM routing (2.9%); no straight-line fallback was required. Signif. Codes: *** 1% ** 5% * 10%.

The combination of Stage 1 and reduced-form heterogeneity yields a clean supply-side interpretation. In low-poverty districts the policy delivers both a meaningful distance reduction and a substantial claims response. In high-poverty districts the policy delivers little distance reduction, and consequently little claims response. In low-poverty districts the implied per-km claims response (the Wald ratio of reduced-form claims to first-stage distance) is 0.51 claims per km; the corresponding high-poverty ratio divides by a first-stage distance effect that is statistically indistinguishable from zero (Table 10, column 2) and is therefore uninformative, as the weak-instrument diagnostics in Appendix A.7 confirm. The smaller total claims response in high-poverty districts is thus explained by the supply-side constraint on distance reduction itself—with fewer public facilities to which beneficiaries can be reassigned closer, the policy compresses less distance to begin with—rather than by any measured difference in per-km demand, which the data cannot identify in those districts.

Appendix subsection A.7 reports instrumental variable estimates as a sensitivity check (Table A17). The full-sample IV is -0.52 claims per kilometer, weak-IV-robust under Anderson–Rubin inference (95% CI $[-0.74, -0.41]$). The high-poverty IV is uninformative because the first-stage instrument is weak in that subsample (Kleibergen–Paap $F = 1.6$; AR confidence interval unbounded). I do not lean on the IV as a headline causal estimate because the exclusion restriction, that FPOC reassignment affects claims only through distance, is hard to defend; reassignment plausibly also disrupts continuity of care between patient and provider and alters administrative procedures for claiming. The average reduced-form effect of reassignment on claims does not require this exclusion. Interpreting the poverty heterogeneity as evidence of a distance mechanism does require that the non-distance channels of reassignment (continuity-of-care disruption, administrative friction, within-public-facility quality differences) do not vary systematically with district poverty. Appendix Section A.7 also documents robustness of the main result to alternative geographic time fixed effects (Table A18), distance bandwidths from 10 to 40 km (Table A20), progressive inclusion of COVID-exposed cohorts (Table A19), and leave-one-district-out exclusions for the high-poverty subsample (Figure A4). A within-public-sector hospital quality check (Table A21) rules out the alternative that reassignment moved beneficiaries to systematically higher-quality hospitals.

6.5. Summary

The mechanism analysis points to supply-side constraints as the leading explanation for the high-poverty utilisation gap. The gap accumulates across the entire utilisation pipeline: poor districts have fewer enrollees, a smaller fraction of enrollees who claim, and fewer claims per claimant. The three margins are not statistically rankable from the decomposition alone. Within-district analysis shows that the registration shortfall co-moves with structural correlates of poverty: rural status, urbanisation, and the count of empanelled providers. The claims-per-registrant differential, however, persists through these same controls, pointing to characteristics of the available facilities that the count of providers does not capture.

The 2019 FPOC reassignment exercise sharpens the supply-side reading with a quasi-experimental design. Reassignment lowers distance to the assigned facility, holding the supply of empanelled facilities fixed. In low-poverty districts the policy delivers a meaningful distance reduction and a substantial increase in claims; in high-poverty districts both effects are much smaller. The smaller total response in poor districts reflects the smaller distance reduction that the policy can deliver there; the per-km demand response cannot be separately identified in high-poverty districts, where the first-stage distance change is statistically indistinguishable from zero (Table 10). The leading explanation is therefore a supply-side constraint: in high-poverty districts, fewer public facilities are available to which beneficiaries can be reassigned closer.

The supply-side conclusion contrasts with much of the LMIC insurance literature, which emphasises demand-side instruments such as premium subsidies, registration assistance, and awareness campaigns. Banerjee et al. (2021) show that even a full one-year premium subsidy combined with assisted online enrolment leaves take-up of Indonesia's voluntary scheme below 30 percent, and Malani et al. (2024) document supply-side and learning frictions that choke off RSBY utilisation in India despite free coverage. The findings here align with that supply-side reading and complement Dupas and Jain (2024), who isolate a distance channel under India's BSBY through an empanelment expansion. The setting differs because the supply of empanelled facilities is held fixed at the reform date; the causal mechanism is therefore distance per se, not the bundle of distance plus new facility entry. The policy implication is direct: in poverty-affected areas, expanding and upgrading empanelled public facilities is likely a more effective lever for closing the utilisation gap than rule changes that reassign individuals across an unchanged network.

7 — Threats to Identification

The staggered rollout of NHIP between 2016 and 2022 overlaps three potentially bundled events that could confound the estimated treatment effects if their impact on healthcare utilisation varied systematically across districts in a manner correlated with NHIP timing.

2015 Gorkha Earthquake. In April 2015, a magnitude-7.8 earthquake struck central Nepal, causing widespread destruction of health infrastructure in 14 of the 77 districts.[15] The earthquake occurred approximately one year before the first NHIP pilot districts began implementation in 2016. If earthquake-affected districts were systematically earlier or later NHIP adopters, post-earthquake reconstruction and health system recovery could generate differential utilisation trends that the estimator would incorrectly attribute to insurance.

Two features of the setting mitigate this concern. The CS estimator compares each treated cohort to not-yet-treated districts at each point in time, with district fixed effects absorbing level differences; what would threaten identification is differential post-2015 *trends* correlated with NHIP rollout timing, not differences in baseline levels. The 14 earthquake-affected districts do differ from the rest in absolute terms (total visits are nearly $2\times$ higher and emergency visits roughly $3\times$ higher; the visit differences are not statistically significant given the small affected sample ($p > 0.25$), and earthquake-affected districts are also somewhat less poor on average (0.21 vs 0.29, $p = 0.01$; Table A5)). This reflects the urban composition of the NPC-14 list, which includes the Kathmandu Valley and other densely populated central-hills districts, not a feature that biases the staggered-DiD design. Most importantly, earthquake status does not predict NHIP rollout timing in a regression that conditions on baseline utilisation, facility count, and poverty rate (Table A6; bivariate $p = 0.15$, conditional $p = 0.45$). With only 14 affected districts against 63, the test has limited statistical power, but the variation absorbed by the controls does not enlarge the timing gap, which is what a recovery-confound mechanism would predict.

2017 Federalisation Reform. Nepal's 2015 Constitution established a three-tier federal system; provinces elected their first governments in 2017–2018. The federal restructure devolved health-facility staffing, drug procurement, and local infrastructure investment to provincial and local governments. Federalisation threatens identification only if its disruptive effects varied across districts in a manner correlated with NHIP rollout timing.

Two arguments rule out this concern. NHIP rollout was administered entirely by the central Health Insurance Board (HIB) under §15(c) and §41 of the Health Insurance Act 2074 (2017), which

15. The 14 severely affected districts are identified based on the National Planning Commission's Post-Disaster Needs Assessment (National Planning Commission 2015).

vest sole authority over service-provider listing and programme expansion in the Board. The first district (Kailali) entered the programme in April 2016, before provincial governments were elected, and the staggered district-by-district rollout continued under HIB authority through 2021. Provincial and local-level NHIP coordination committees were formalised only in 2078 BS (2021–22), at the very end of the sample, and the procedures issued under §41 of the Act (Health Insurance Board 2021c, 2021b) limit lower-tier roles to promotional cooperation and operational support, not eligibility, benefits, or rollout sequencing. Treatment timing is therefore institutionally exogenous to federalisation milestones.

Second, even if federalisation disruption did vary across districts, it would predict NHIP rollout timing mostly through observable district characteristics. The factors that determine susceptibility to federalisation disruption (population, poverty incidence, baseline health-facility infrastructure, and similar dimensions of pre-existing local-government capacity) are precisely the characteristics tested in the rollout balance regression (Table A3), where none individually or jointly predict treatment timing (F-test p-value = 0.35). The same balance evidence that supports parallel trends for the main analysis therefore also rules out a federalisation-driven confound operating through observable district characteristics. This argument relies on the assumption that the relevant dimensions of federalisation disruption are captured by the observables tested for balance; unobserved components of pre-existing local-government capacity, if independently correlated with NHIP rollout timing, would not be detected by the balance test.

COVID-19 Pandemic. The COVID-19 pandemic, beginning in March 2020, represents the most significant threat to identification. Nationwide lockdowns sharply reduced healthcare utilisation across all districts, and the pandemic's impact likely varied across districts depending on infection rates, lockdown severity, and health system capacity. Since NHIP continued to roll out to new districts during 2020–2022, later treatment cohorts experienced the pandemic during both their pre-treatment and post-treatment periods, while earlier cohorts experienced it only in later post-treatment periods. If COVID-19's impact on utilisation was correlated with NHIP timing, the parallel trends assumption could be violated.

The cohort-specific estimates in Table A7 disaggregate this threat. The pre-COVID cohorts in Panel A are positive in nine of ten cases and significant in seven; the one negative estimate comes from the single-district pilot cohort, where the ATT is identified off a single comparison and the sign likely reflects small-sample noise (Table A7, note †). The COVID-era cohorts in Panel B do not move in a consistent direction. The first two cohorts (23 and 25) deliver large positive ATTs (0.245^{***} and 0.290^{***}); the next two (26 and 27), implemented in the inter-lockdown window, turn negative at the 5% and 10% levels (-0.029 and -0.197); cohort 28, treated during the second lockdown, is strongly negative (-0.324^{***}); and cohort 29, treated just after the second lockdown ended, swings back to a positive ATT (0.109^{***}). The signs flip across the COVID era rather than line up. A pandemic-driven confound inflating the headline would push the COVID-era cohorts in the same direction; instead the pattern tracks lockdown timing, not NHIP rollout.

As a complementary check, I restrict aggregation to the ten pre-COVID cohorts (first treatment quarter on or before Q4 2019), keeping their full post-treatment horizons including pandemic-era calendar time and holding the control pool fixed at the headline specification (Table A8). The estimate is 0.172^{***}, larger than the headline 0.133. The comparison is not a clean COVID-bias diagnostic, however, since the pre-COVID cohorts have longer post-treatment horizons by construction and the dynamic ATT in Figure 4 grows over event time; the 0.172/0.133 ratio is consistent with longer accumulating dynamics among earlier cohorts as well as with a downward COVID bias on the headline. The check is best read as showing that COVID-era cohorts do not pull the headline up, rather than as pinning down a directional bias.

DHIS2 Reporting Composition. A potential concern with administrative data from the District Health Information System is that NHIP empanelment could change the composition or completeness of facility reporting, generating measured utilisation increases that reflect improved data capture rather than genuine changes in health-seeking behaviour. If newly empanelled private facilities begin reporting to DHIS2 upon joining the insurance network, the district-level aggregates would mechanically increase even absent any change in actual utilisation. The data construction in Section 3 addresses this directly: the district-level aggregates are compiled from a balanced panel of facilities that reported throughout 2016–2022, so the estimated effects capture changes in visits at the same facilities over time, not the entry of new reporting units. Any private facility that began DHIS2 reporting only after NHIP reached its district (a plausible response if empanelment generated the reporting incentive) is excluded by this restriction, so the supply-side expansion induced by NHIP is not counted in the estimated effect. The reported coefficients therefore understate the true utilisation response if NHIP induced new private capacity. The falsification outcomes provide an independent check: NHIP eligibility shows no detectable effect on the four Safe Motherhood Programme outcomes already covered by an earlier free-care scheme (Table 6), and no effect on the non-communicable disease categories (cardiovascular, malignancy) covered by separate ministry programmes. These outcomes are recorded through the same DHIS2 stream as the headline outcomes; a reporting-driven artefact would inflate them too, yet the null effects rule this out. The HIB-DHIS2 cross-check (Table 5) is consistent with the two administrative systems tracking the same response, but because both streams share NHIP empanelment incentives, it does not by itself rule out a common recording-improvement; the balanced-panel construction is the binding defence.

As a complementary robustness check, I re-estimate the main specifications restricting the sample to public facilities only, which have the most consistent reporting history (Table A10). All three primary outcomes remain positive and significant at 1%, but the magnitudes split: emergency visits are essentially unchanged (from 0.179 to 0.178 log points); OPD visits attenuate modestly (from 0.128 to 0.112, about 13% smaller); total visits attenuate by roughly half (from 0.133 to 0.071). The split localises the response: emergency and OPD effects sit in the public sector, while a substantial share of the total-visit response operates through empanelled private facilities.

Cross-District Spillovers (SUTVA). The validity of the staggered DiD design also requires that NHIP implementation in one district does not affect healthcare utilisation in other districts. The institutional design of NHIP effectively rules out the most salient spillover channel. Registration is strictly residence-based: individuals can only enrol in the district where they reside. Upon enrolment, each member is assigned a first point of contact facility within their district for primary care. Even when a patient is referred to a facility in another district, the referral is not covered by the insurance programme unless the patient is enrolled in that district. This rules out the more concerning direction of spillover, in which control-district residents cross into treated districts and inflate treated-area counts. A subtler reverse channel remains: NHIP-induced referrals from treated districts to tertiary facilities in not-yet-treated districts (predominantly Kathmandu) appear in the control district's DHIS2 data regardless of payer, biasing the DiD toward zero. This attenuates rather than overstates the estimates, so the main estimates are conservative under this channel. Excluding Kathmandu yields larger point estimates (Table A11), as the attenuation channel predicts. General-equilibrium effects on provider supply or information diffusion across district borders are not directly tested; primary healthcare delivery in Nepal is largely local, which bounds the magnitude of either channel.

8 — Conclusion

This paper provides the first causal evaluation of Nepal's National Health Insurance Programme. Using a district-quarter HMIS panel from 2014 to 2022 and a staggered difference-in-differences design that exploits the 2016–2022 rollout across all 77 districts, I estimate the intention-to-treat (ITT) effect of NHIP eligibility on healthcare utilisation. The ITT estimate is a 14 percent increase in total health-facility visits on average, with similar magnitudes across outpatient, emergency, and diagnostic-service margins. Effects emerge gradually rather than at the moment of implementation, becoming statistically significant five quarters after rollout and plateauing at twelve to fourteen quarters, consistent with registration and supply-side adjustment taking time to translate eligibility into care-seeking.

The aggregate response is unequally distributed. Utilisation gains concentrate entirely in low-poverty districts; in high-poverty districts, the headline effects are statistically indistinguishable from zero on nearly all margins. A three-group breakdown sharpens the result, showing that only the most affluent quarter of districts experiences large utilisation gains. The mechanism analysis identifies supply-side constraints as the leading explanation, with access shaping the registration shortfall and facility quality shaping the residual claims-per-registrant gap. The high-poverty shortfall is broad-based across enrolment, claiming, and intensity margins, with the three margins not statistically rankable from the decomposition alone. Within-district analysis shows that the registration shortfall co-moves with structural correlates of poverty: rural status, urbanisation, and the count of empanelled providers. The claims-per-registrant differential, by contrast, persists through these same controls, pointing to facility characteristics that the count of providers does not measure. The 2019 FPOC reassignment exercise sharpens this supply-side reading: where the policy compresses distance, claims rise; where it does not, claims do not. The smaller total response in high-poverty districts reflects how little distance the policy could compress there, where the public network is sparse; the data cannot identify whether per-kilometre demand itself differs by poverty.

The paper contributes to three literatures. It provides the first causal evaluation of Nepal's NHIP and adds to the LMIC literature on utilisation effects of public health insurance. It extends the equity literature by combining district-level ITT with individual claims, decomposing the equity gap across enrolment, claiming, and intensity margins, and showing that the programme's gains accrue to the most affluent district quartile and widen the poverty-utilisation gap. It complements the supply-side mechanism literature with a causal estimate that holds the supply of empanelled facilities fixed, isolating distance from the bundle of distance and new-facility entry studied by Dupas and Jain (2024). The equity result inverts the pattern Conti and Ginja (2023) document for Mexico's Seguro Popular, where mortality reductions accrued to poor municipalities; it aligns with the wealth gradient Powell-Jackson et al. (2014) document for Ghana, where free care benefited the better-off more even within a poor sample.

These findings carry a specific policy implication. In poverty-affected areas, expanding the supply of empanelled public facilities is likely a more effective lever for closing the utilisation gap than demand-side instruments or rule changes that reassign individuals across an unchanged network. Three caveats limit the interpretation. First, utilisation is not welfare; the headline 14 percent could in principle reflect either previously unmet need or moral hazard, and the estimates do not separate the two. Second, the HMIS data are provider-reported. Third, the within-district and FPOC analyses identify supply-side constraints as the leading factor; they do not pinpoint which dimension of supply, whether the count of empanelled facilities, the distance to them, or qualitative aspects such as service breadth, hours of operation, specialist availability, or referral capacity, matters most. Future work on these dimensions, and on long-run health outcomes once the relevant administrative data become available, will be necessary to translate the supply-side reading into a sharper set of policy levers. Expanding and upgrading empanelled public facilities in high-poverty districts, however, is the most direct lever the present findings identify.

References

Adhikari, Ramesh Prasad, Manisha Laxmi Shrestha, Emily N. Satinsky, and Nawaraj Upadhaya. 2021. "Trends in and determinants of visiting private health facilities for maternal and child health care in Nepal: comparison of three Nepal demographic health surveys, 2006, 2011, and 2016." *BMC Pregnancy and Childbirth* 21 (1): 1. https://doi.org/10.1186/s12884-020-03485-8. (Cited on page 6).

Ambel, Alemayehu A., Ilana Seff, and Tekabe Ayalew Belay. 2014. *Impact Evaluation of Nepal District Health Insurance Pilot 2014, Baseline Survey.* Strategic Impact Evaluation Fund (SIEF). Baseline collected March–May 2014; evaluation subsequently cancelled. World Bank. https://microdata.worldbank.org/index.php/catalog/3209. (Cited on page 10).

Anderson, T. W., and Herman Rubin. 1949. "Estimation of the parameters of a single equation in a complete system of stochastic equations." *Annals of Mathematical Statistics* 20 (1): 46–63. (Cited on pages 85, 86).

Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11:727–753. https://doi.org/10.1146/annurev-economics-080218-025643. (Cited on pages 85, 86).

Ayer, Rakesh, Suresh Tiwari, Shambhu Prasad Jnawali, and Rudi Van Dael. 2024. *A Study on Nepal's National Health Insurance Program* [in en]. Technical report TCS240257-2. Accepted: 2024-05-09 Last Modified: 2024-06-05T13:39+08:00. Asian Development Bank, May. Accessed November 28, 2025. https://www.adb.org/publications/study-nepal-national-health-insurance-program. (Cited on pages 7, 8, 14).

Banerjee, Abhijit, Amy Finkelstein, Rema Hanna, Benjamin A. Olken, Arianna Ornaghi, and Sudarno Sumarto. 2021. "The Challenges of Universal Health Insurance in Developing Countries: Experimental Evidence from Indonesia's National Health Insurance" [in en]. *American Economic Review* 111, no. 9 (September): 3035–3063. ISSN: 0002-8282, accessed April 23, 2024. https://doi.org/10.1257/aer.20200523. (Cited on pages 4, 5, 38, 39, 53).

Bharati, Rajani, Diana Romero, Alexis Pozen, James Sherry, Bhuwan Paudel, Mukesh Adhikari, and Prakash Acharya. 2025. "Experiences and Effect of Implementing Social Health Insurance (SHI) Program in Nepal-A Mixed Method Study." *PLOS Global Public Health* 5 (4): e0003492. https://doi.org/10.1371/journal.pgph.0003492. (Cited on page 10).

Boone, Claire E., Paul J. Gertler, Grace Makana Barasa, Joshua Gruber, and Ada Kwan. 2023. "Can a private sector engagement intervention that prioritizes pro-poor strategies improve healthcare access and quality? A randomized field experiment in Kenya." *Health Policy and Planning* 38 (9): 1006–1016. https://doi.org/10.1093/heapol/czad076. (Cited on page 4).

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies* 91, no. 6 (November): 3253–3285. ISSN: 0034-6527, accessed May 5, 2024. https://doi.org/10.1093/restud/rdae007. https://doi.org/10.1093/restud/rdae007. (Cited on pages 18, 32, 81).

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics,* Themed Issue: Treatment Effect 1, 225, no. 2 (December): 200–230. ISSN: 0304-4076, accessed May 5, 2024. https://doi.org/10.1016/j.jeconom.2020.12.001. https://www.sciencedirect.com/science/article/pii/S0304407620303948. (Cited on pages 2, 17, 21, 29, 72, 77, 81).

CBS, and World Bank. 2013. *Small Area Estimation of Poverty in Nepal.* Technical report. Kathmandu: Central Bureau of Statistics, National Planning Commission Secretariat, Government of Nepal. (Cited on page 16).

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects" [in en]. *American Economic Review* 110, no. 9 (September): 2964–2996. ISSN: 0002-8282, accessed May 5, 2024. https://doi.org/10.1257/aer.20181169. https://www.aeaweb.org/articles?id=10.1257/aer.20181169. (Cited on pages 18, 32, 81).

Conti, Gabriella, and Rita Ginja. 2023. "Who Benefits from Free Health Insurance? Evidence from Mexico." Previously circulated as IFS Working Paper 23/07, *Journal of Human Resources* 58 (1): 146–182. https://doi.org/10.3368/jhr.58.3.1117-9157R2. (Cited on pages 2, 4, 5, 30, 31, 59).

Das, Jishnu, and Quy-Toan Do. 2023. "The Prices in the Crises: What We Are Learning from 20 Years of Health Insurance in Low- and Middle-Income Countries" [in en]. *Journal of Economic Perspectives* 37, no. 2 (May): 123–152. ISSN: 0895-3309, accessed May 16, 2024. https://doi.org/10.1257/jep.37.2.123. https://www.aeaweb.org/articles?id=10.1257/jep.37.2.123. (Cited on page 4).

Das Gupta, Prithwis. 1978. "A General Method of Decomposing a Difference Between Two Rates into Several Components." *Demography* 15 (1): 99–112. https://doi.org/10.2307/2060493. (Cited on pages 40, 97, 98).

———. 1993. *Standardization and Decomposition of Rates: A User's Manual.* Current Population Reports P23-186. U.S. Bureau of the Census. (Cited on pages 40, 97, 98).

DoHS. 2024. *Annual Health Report 2079/80* [in en]. Technical report. Kathmandu, Nepal: Department of Health Services: Ministry of Health and Population. Accessed May 5, 2024. https://hmis.gov.np/wp-content/uploads/2024/03/Annual-Health-Report-2079-80.pdf. (Cited on pages 8, 14).

Dupas, Pascaline, and Radhika Jain. 2024. "Women Left Behind: Gender Disparities in Utilization of Government Health Insurance in India." *American Economic Review* 114, no. 10 (October): 3345–85. https://doi.org/10.1257/aer.20230521. https://www.aeaweb.org/articles?id=10.1257/aer.20230521. (Cited on pages 4, 5, 24, 31, 38, 47, 53, 59).

———. 2026. "Public Health Insurance in Low- and Middle-Income Countries." In *The Handbook of Social Protection: Evidence and New Directions for Low- and Middle-Income Countries,* edited by Rema Hanna and Benjamin A. Olken. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/14927.001.0001. https://direct.mit.edu/books/oa-edited-volume/6103/The-Handbook-of-Social-ProtectionEvidence-and-New. (Cited on pages 4, 30).

Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355–364. https://doi.org/10.1111/1468-0262.00399. (Cited on page 16).

Erlangga, Darius, Marc Suhrcke, Shehzad Ali, and Karen Bloor. 2019. "The impact of public health insurance on health care utilisation, financial protection and health status in low- and middle-income countries: A systematic review." *PLOS ONE* 14 (8): e0219731. https://doi.org/10.1371/journal.pone.0219731. (Cited on pages 2, 4, 30).

GoN. 2021. *Nepal Multidimensional Poverty Index 2021* [in en]. Discussion Papers. National Planning Commission, Government of Nepal. Accessed November 28, 2025. https://www.undp.org/nepal/publications/nepal-multidimensional-poverty-index-2021. (Cited on pages 6, 38).

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics,* Themed Issue: Treatment Effect 1, 225, no. 2 (December): 254–277. ISSN: 0304-4076, accessed May 5, 2024. https://doi.org/10.1016/j.jeconom.2021.03.014. https://www.sciencedirect.com/science/article/pii/S0304407621001445. (Cited on page 18).

Gruber, Jonathan, Nathaniel Hendren, and Robert M. Townsend. 2014. "The Great Equalizer: Health Care Access and Infant Mortality in Thailand." *American Economic Journal: Applied Economics* 6 (1): 91–107. https://doi.org/10.1257/app.6.1.91. (Cited on pages 2, 4, 30).

Gruber, Jonathan, Mengyun Lin, and Junjian Yi. 2023. "The largest insurance program in history: Saving one million lives per year in China." *Journal of Public Economics* 226:104999. https://doi.org/10.1016/j.jpubeco.2023.104999. (Cited on pages 2, 4, 5, 30).

Health Insurance Board. 2017. *Annual Report FY 2072/73.* Technical report. Kathmandu, Nepal: Health Insurance Board. (Cited on page 10).

———. 2019. *Annual Report FY 2074/75.* Technical report. Kathmandu, Nepal: Health Insurance Board. (Cited on page 10).

———. 2021a. *Annual Report FY 2077/78* [in ne]. Technical report. Kathmandu, Nepal: Health Insurance Board. (Cited on page 14).

Health Insurance Board. 2021b. *Local-Level Health Insurance Coordination Committee Formation/Operation Sample Procedure, 2078 [B.S.]* Sample Procedure. Issued under Section 41 of the Health Insurance Act, 2074 (2017). Kathmandu, Nepal: Government of Nepal, Ministry of Health and Population. https://hib.gov.np/content/8/local-level-health-insurance-committee-guidlines/. (Cited on page 55).

———. 2021c. *Provincial-Level Health Insurance Coordination Committee Formation/Operation Sample Procedure, 2078 [B.S.]* Sample Procedure. Issued under Section 41 of the Health Insurance Act, 2074 (2017). Kathmandu, Nepal: Government of Nepal, Ministry of Health and Population. https://hib.gov.np/content/7/province-level-insurance-committee-guidlines/. (Cited on page 55).

HIB. 2024. *Health Insurance Board Strategy Roadmap 2081/82* [in en]. Technical report. Kathmandu, Nepal: Department of Health Services: Health Insurance Board. Accessed March 28, 2025. https://hib.gov.np/public/uploads/shares/HIB_Strategy_Roadmap_2081_82.pdf. (Cited on page 8).

Kitagawa, Evelyn M. 1955. "Components of a Difference Between Two Rates." *Journal of the American Statistical Association* 50 (272): 1168–1194. https://doi.org/10.1080/01621459.1955.10501299. (Cited on page 40).

Kurowski, Christoph, Anurag Kumar, Julio Cesar Mieses Ramirez, Martin Schmidt, and Denise Valerie Silfverberg. 2023. *Health Financing in a Time of Global Shocks: Strong Advance, Early Retreat.* Publication Title: Health Financing in a Time of Global Shocks: Strong Advance, Early Retreat. World Bank, May. https://doi.org/10.1596/39864. (Cited on page 2).

Lagomarsino, Gina, Alice Garabrant, Atikah Adyas, Richard Muga, and Nathaniel Otoo. 2012. "Moving towards universal health coverage: health insurance reforms in nine developing countries in Africa and Asia." *The Lancet* 380, no. 9845 (September): 933–943. ISSN: 0140-6736, accessed April 23, 2024. https://doi.org/10.1016/S0140-6736(12)61147-7. https://www.sciencedirect.com/science/article/pii/S0140673612611477. (Cited on pages 4, 38).

Malani, Anup, Cynthia Kinnan, Gabriella Conti, Kosuke Imai, Morgen Miller, Shailender Swaminathan, Alessandra Voena, and Bartosz Woda. 2024. *Evaluating and Pricing Health Insurance in Lower-Income Countries: A Field Experiment in India.* Working Paper 32239. National Bureau of Economic Research. https://doi.org/10.3386/w32239. (Cited on pages 4, 30, 53).

National Planning Commission. 2015. *Nepal Earthquake 2015: Post Disaster Needs Assessment.* Kathmandu: National Planning Commission, Government of Nepal. https://www.npc.gov.np/images/category/PDNA_volume_BFinalReport.pdf. (Cited on pages 54, 75).

National Statistics Office. 2023. *National Population and Housing Census 2021 (National Report)* [in en]. Report. National Statistics Office, Government of Nepal. Accessed November 28, 2025. https://censusnepal.cbs.gov.np/results/downloads/national?type=report. (Cited on page 6).

Powell-Jackson, Timothy, Kara Hanson, Christopher J. M. Whitty, and Evelyn K. Ansah. 2014. "Who Benefits from Free Healthcare? Evidence from a Randomized Experiment in Ghana." *Journal of Development Economics* 107:305–319. https://doi.org/10.1016/j.jdeveco.2013.11.010. (Cited on pages 2, 4, 5, 30, 59).

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies* 90 (5): 2555–2591. https://doi.org/10.1093/restud/rdad018. (Cited on pages 20, 32, 73).

Romano, Joseph P., and Michael Wolf. 2005. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73 (4): 1237–1282. https://doi.org/10.1111/j.1468-0262.2005.00615.x. (Cited on pages 21, 32).

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics* 235, no. 2 (August): 2218–2244. ISSN: 0304-4076, accessed May 5, 2024. https://doi.org/10.1016/j.jeconom.2023.03.008. https://www.sciencedirect.com/science/article/pii/S0304407623001318. (Cited on pages 18, 20).

Sant'Anna, Pedro H.C., and Jun Zhao. 2020. "Doubly robust difference-in-differences estimators." *Journal of Econometrics* 219 (1): 101–122. ISSN: 0304-4076. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.06.003. https://www.sciencedirect.com/science/article/pii/S0304407620301901. (Cited on page 21).

Watson, Julia, Abdo S. Yazbeck, and Lauren Hartel. 2021. "Making Health Insurance Pro-poor: Lessons from 20 Developing Countries." Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23288604. *Health Systems & Reform* 7, no. 2 (July): e1917092. ISSN: 2328-8604, accessed May 8, 2024. https://doi.org/10.1080/23288604.2021.1917092. https://doi.org/10.1080/23288604.2021.1917092. (Cited on page 38).

WHO, and World Bank. 2023. *Tracking Universal Health Coverage: 2023 Global Monitoring Report.* _Eprint: https://elibrary.worldbank.org/doi/pdf/10.1596/40348. World Bank. https://doi.org/10.1596/40348. https://elibrary.worldbank.org/doi/abs/10.1596/40348. (Cited on page 2).

World Bank. 2025a. *World Bank Development Indicators* [in en]. Accessed November 28, 2025. https://data.worldbank.org. (Cited on page 6).

———. 2025b. *World Bank Development Indicators* [in en]. Accessed November 28, 2025. https://data.worldbank.org. (Cited on page 7).

World Bank Group. 2024. *Poverty & Equity Brief: Nepal.* Technical report. Washington, DC: World Bank Group, October. (Cited on page 8).

World Health Organization. 2025. *Global Health Expenditure Database.* Accessed via World Bank Open Data API; database last updated 12 December 2025. Accessed December 12, 2025. https://apps.who.int/nha/database. (Cited on page 6).

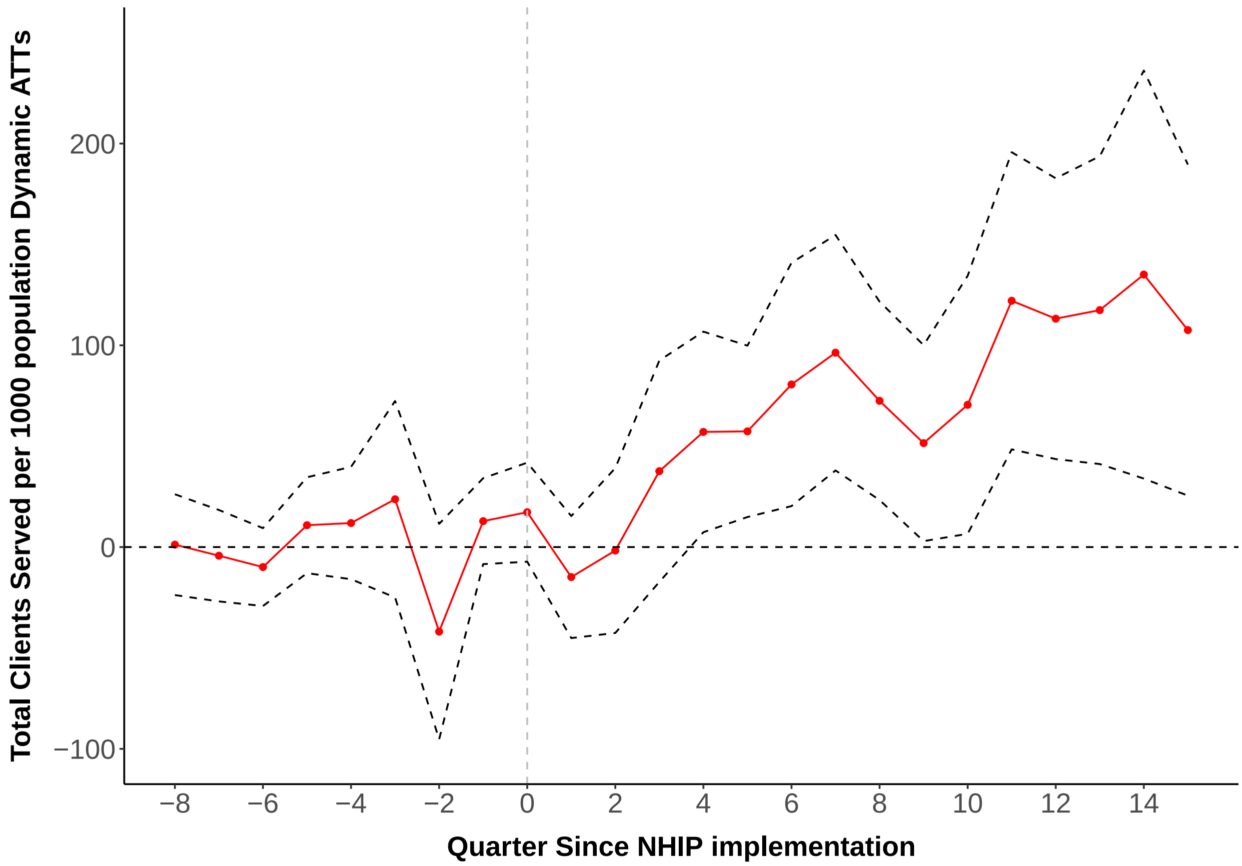
A — Appendix

Table A1 — Health Insurance Eligibility on Health Visits (per 1000 population)

	Total Clients Served	Total New Clients Served	Total OPD Visits	Total New OPD Visits
Model:	(1)	(2)	(3)	(4)
Group Aggregation				
Is NHIP eligible = 1	39.7522***	17.5118	38.0590***	28.6267***
	(7.3389)	(9.2275)	(7.2238)	(5.6907)
Pre-Treatment Mean	267.71	222.04	239.28	199.67
Fit statistics				
Number of Cohorts	17	17	17	17
Number of Time Periods	30	30	30	30

Note: All the outcomes are per 1000 population. Clustered (District) standard errors in parentheses bootstrapped with 1000 iterations. Estimation Method: Doubly Robust. Control Group: Not Yet Treated, Anticipation Periods: 1. Signif. Codes: *** 99% ** 95% * 90% uniform confidence band does not include 0.

Figure A1 — Total Clients Served per 1000 population Dynamic ATTs



Notes. This figure shows dynamic ATTs for total clients served per 1000 population by health service providers. The outcome is total clients served per 1000 population, x-axis shows quarters since NHIP implementation, quarter 0 being the first implementation quarter.

A.1. Poverty Heterogeneity: Three-Group Breakdown

Table A2 — Effects of Health Insurance Eligibility by District Poverty Level

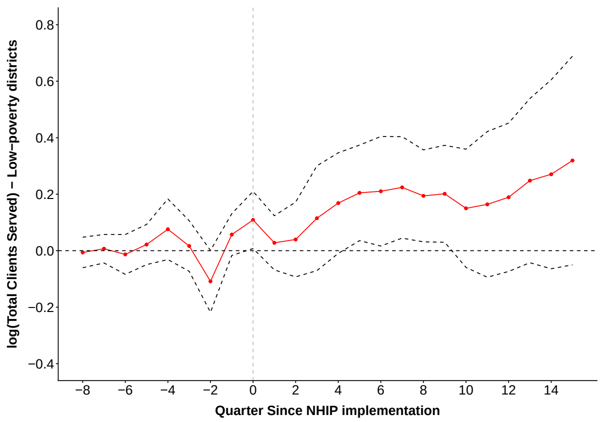
	High poverty (Q4)		Middle poverty (Q2–Q3)		Low poverty (Q1)	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Panel A: Overall Healthcare Visits						
Total Visits	0.0256	(0.0696)	0.0557	(0.0412)	0.2039***	(0.0622)
OPD Visits	0.0621	(0.0701)	0.0008	(0.0593)	0.2018***	(0.0621)
Emergency Visits	0.0082	(0.1309)	0.1390	(0.1078)	0.2435***	(0.0903)
Panel B: OPD Visits by Disease Category						
OPD Communicable	0.1690**	(0.0660)	0.0407	(0.0663)	0.1946***	(0.0511)
OPD Non-Communicable	0.0472	(0.0677)	-0.0058	(0.0694)	0.1994***	(0.0684)
OPD Injuries	0.0401	(0.0717)	0.0071	(0.0537)	0.2214***	(0.0735)
Fit statistics						
Number of Cohorts	9		14		12	
Number of Time Periods	27		27		30	

Note: This table presents heterogeneous effects by district poverty level using a three-group classification: low-poverty (bottom quartile, Q1), middle-poverty (interquartile range, Q2–Q3), and high-poverty (top quartile, Q4). All outcomes are log-transformed. Panel A shows effects on overall visits. Panel B disaggregates OPD visits by disease category. The results show that the overall treatment effect is driven entirely by the lowest-poverty districts (Q1), which exhibit large and statistically significant effects across all outcomes. Middle-poverty (Q2–Q3) districts show no statistically significant effects on any outcome. High-poverty (Q4) districts show no significant effects on overall visits, non-communicable disease, or injuries, with the single exception of OPD communicable-disease visits (0.169, $p < 0.05$). This is consistent with the median-split results in the main text and provides a sharper characterisation: the effect is concentrated specifically in the bottom quartile of the poverty distribution. Note that subgroup estimates cannot be arithmetically reconciled to the overall ATT in the main tables because the Callaway–Sant'Anna estimator constructs different counterfactuals in each subsample: not-yet-treated districts within the same poverty group serve as controls, rather than the full not-yet-treated pool. Clustered (District) standard errors in parentheses. Estimation Method: Doubly Robust. Control Group: Not Yet Treated. Anticipation Periods: 1. Signif. Codes: *** 1% ** 5% * 10%.

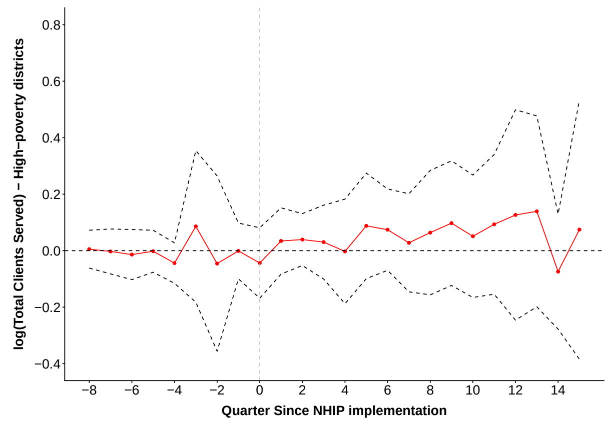
A.2. Dynamic Effects by Poverty Subgroup

Figure A2 reports event-study estimates for the median split by poverty incidence. Each panel applies the same Callaway–Sant'Anna estimator as the main event study (varying base period, anticipation = 1, doubly robust, not-yet-treated controls) to the corresponding subsample.

Figure A2 — Dynamic ATTs on log Total Clients Served, by poverty subgroup



(a) Low-poverty districts (below median, $N = 40$)



(b) High-poverty districts (above median, $N = 37$)

Note: Dynamic ATTs estimated separately on each half of the poverty distribution using the Callaway–Sant'Anna estimator with varying base period, anticipation = 1, doubly-robust adjustment, and not-yet-treated controls. Districts are split at the median of the 2011 World Bank district poverty incidence rate (median = 0.260). The x-axis is quarters since NHIP implementation; the y-axis is the dynamic ATT on log Total Clients Served. The black dotted lines represent 99% uniform confidence intervals.

A.3. Rollout Balance

Table A3 — Rollout Timing and Pre-Treatment Observables

	Coefficient	Std. Error
Avg. Total Visits (1000s)	0.0566	(0.0383)
Avg. OPD Visits (1000s)	−0.0161	(0.0343)
Avg. Emergency Visits (1000s)	−0.1556	(0.1921)
Population (1000s)	−0.0051	(0.0062)
No. of health facilities	−0.0150	(0.0326)
Poverty Rate	−3.3813	(5.8969)
Fit statistics		
Observations	77	
R^2	0.0882	
Joint F-statistic	1.129	
F-test p-value	0.3545	

Note: This table reports OLS estimates from regressing the treatment timing variable (first quarter of NHIP implementation) on pre-treatment district characteristics. Pre-treatment means are computed over all quarters before treatment. The joint F-test evaluates whether observables jointly predict rollout timing. Failure to reject supports the assumption that rollout timing is as-good-as-random with respect to district characteristics.

A.4. Multiple Hypothesis Testing

Table A4 — Romano-Wolf Multiple Hypothesis Testing Correction

	ATT	Std. Error	Unadjusted p	Romano-Wolf p
Total Visits	0.1331***	(0.0284)	0.0000	0.0000
OPD Visits	0.1281***	(0.0353)	0.0003	0.0011
Emergency Visits	0.1790**	(0.0665)	0.0071	0.0157
New Total Visits	0.1052***	(0.0283)	0.0002	0.0004
New OPD Visits	0.1169***	(0.0387)	0.0025	0.0079
New Emergency Visits	0.1729**	(0.0660)	0.0088	0.0167
Diagnostic Services	0.2089**	(0.1009)	0.0383	0.0389
Fit statistics				
Number of Districts	77			
Hypotheses Tested	7			
Simulations	10,000			

Note: This table reports Romano–Wolf stepdown adjusted p-values that control the familywise error rate across all utilisation outcomes tested simultaneously. The procedure uses the joint covariance structure of the test statistics, estimated via influence functions from the Callaway and Sant'Anna (2021) estimator, with 10,000 simulations. Significance stars reflect Romano–Wolf adjusted p-values. Signif. Codes: *** 1% ** 5% * 10%.

A.5. Sensitivity to Parallel Trends Violations

Figure A3 presents sensitivity analysis following Rambachan and Roth (2023). These plots show honest confidence intervals for the impact-period treatment effect ($e = 0$) under increasingly permissive violations of the parallel trends assumption using the relative magnitudes restriction (Δ^{RM}). The parameter $\bar{M}$ on the horizontal axis bounds the magnitude of post-treatment trend violations relative to the largest pre-treatment deviation. At $\bar{M} = 0$, parallel trends hold exactly; at $\bar{M} = 1$, post-treatment violations can be as large as the maximum observed pre-treatment deviation.

For total visits, Figure A3 shows that the honest confidence interval remains bounded away from zero up to $\bar{M} = 0.3$, meaning the result is robust to post-treatment trend violations up to 30% of the largest observed pre-treatment deviation. At $\bar{M} = 0.4$, the lower bound crosses zero. The result therefore rests on pre-treatment trends being approximately—but not necessarily exactly—parallel. The complementary evidence supporting this assumption is substantial: rollout timing is uncorrelated with pre-treatment observables (F-test $p = 0.35$), the event study plots show near-zero pre-treatment coefficients with no systematic pattern, earthquake-affected status is orthogonal to rollout timing (Table A6), and restricting the analysis to the pre-COVID cohorts yields a *larger* headline estimate rather than a smaller one (0.172 versus 0.133; Table A8), indicating that pandemic-era cohorts do not inflate the result.

The economic content of the $\bar{M} = 0.3$ breakdown is as follows. The largest pre-treatment $\widehat{ATT}(e)$ coefficient in absolute value is 0.104 log points, with a simultaneous 95% confidence band of $[-0.214, 0.006]$ that contains zero—this coefficient is statistically indistinguishable from zero under the simultaneous-band inference used in the Rambachan–Roth sensitivity setup. Calibrating against this point estimate, $\bar{M} = 0.3$ permits post-treatment trend violations of up to 0.031 log points per period—roughly 24% of the headline effect (0.133)—with the result continuing to reject the null. The robustness threshold therefore tolerates an unmodelled differential trend equal to nearly a quarter of the estimated impact, calibrated against a pre-treatment deviation that is itself a noisy estimate of zero. This is a moderate rather than maximal robustness, but it operates as a third filter on top of two stronger ones: balance on pre-treatment observables (Table A3, F-test $p = 0.35$) and the pre-COVID cohort restriction, which yields a larger rather than a smaller estimate (Table A8). For a violation at this $\bar{M} = 0.3$ relative-magnitude scale to threaten the result, an unobserved shock would have to clear all three filters simultaneously—which the statutory structure of NHIP rollout and the observed balance on district characteristics make difficult to construct.

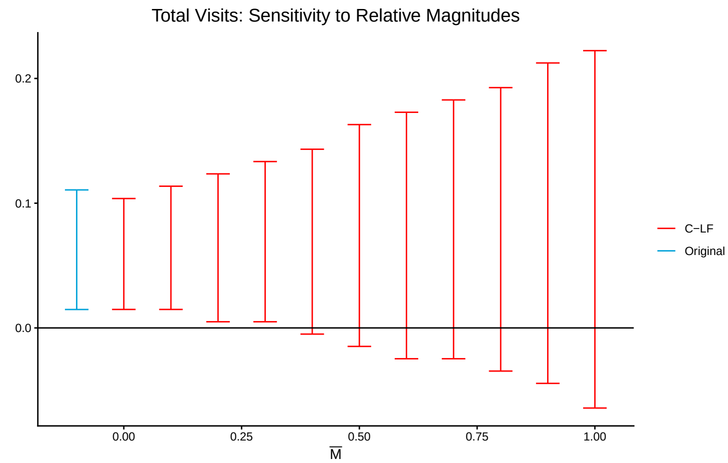


Figure A3 — Sensitivity of Total Visits Treatment Effect to Parallel Trends Violations

Table A5 — Balance: Earthquake-Affected vs Unaffected Districts

	Earthquake (N=14)	Non-Earthquake (N=63)	Difference	p-value
Total Visits	158,380	81,126	77,254	0.268
OPD Visits	95,286	72,373	22,913	0.405
Emergency Visits	18,945	5,850	13,096	0.259
No. of Hospitals	111.9	109.8	2.1	0.892
Poverty Rate	0.212	0.289	-0.077	0.013
Treatment Timing (Quarter)	22.9	20.2	2.7	0.195

Note: Pre-treatment district-level means for the 14 districts severely affected by the 2015 Gorkha earthquake versus the remaining 63 districts. Total visits, OPD visits, and emergency visits are quarterly averages computed over all pre-treatment periods for each district. Treatment timing is the quarter in which NHIP was first implemented. p-values are from two-sample t-tests. The 14 earthquake-affected districts are identified based on the National Planning Commission Post-Disaster Needs Assessment (National Planning Commission 2015).

Table A6 — Earthquake Status and NHIP Rollout Timing

	(1) Bivariate	(2) With Controls
Earthquake-Affected	2.683	1.477
	(1.834)	(1.956)
Avg. Total Visits		0.000014
		(0.000008)
No. of Hospitals		-0.025
		(0.021)
Poverty Rate		-2.083
		(5.911)
Observations	77	77
R^2	0.0277	0.0714

Note: OLS regression of NHIP treatment timing (quarter of first implementation) on earthquake-affected status. Column (1) is bivariate; column (2) adds pre-treatment average total visits, number of hospitals, and district poverty rate as controls. Standard errors in parentheses. Earthquake-affected status does not predict rollout timing in either specification.

Table A7 — Cohort-Specific Treatment Effects on Total Visits

Cohort	Calendar Period	Districts	ATT	Std. Error
Panel A: Pre-COVID cohorts				
9	Jul–Sep 2016	1	$-0.0654^{***\dagger}$	(0.0199)
10	Oct–Dec 2016	2	0.1828^{***}	(0.0664)
12	Apr–Jun 2017	5	0.0680	(0.0636)
14	Oct–Dec 2017	7	0.1890^{***}	(0.0563)
15	Jan–Mar 2018	6	0.2430^{***}	(0.0574)
16	Apr–Jun 2018	4	0.0511	(0.0641)
17	Jul–Sep 2018	7	0.2345^{***}	(0.0656)
18	Oct–Dec 2018	4	0.1878^{*}	(0.1003)
21	Jul–Sep 2019	6	0.1494^{**}	(0.0585)
22	Oct–Dec 2019	4	0.2543^{**}	(0.1060)
Panel B: COVID-era cohorts				
23	Jan–Mar 2020	3	0.2450^{***}	(0.0383)
25	Jul–Sep 2020	7	0.2901^{***}	(0.0493)
26	Oct–Dec 2020	1	-0.0293^{**}	(0.0135)
27	Jan–Mar 2021	3	-0.1969^{*}	(0.1187)
(Second lockdown: 29 Apr–1 Sep 2021)				
28	Apr–Jun 2021	4	-0.3244^{***}	(0.0641)
29	Jul–Sep 2021	11	0.1092^{***}	(0.0372)

Note: Group-level ATT estimates from the Callaway and Sant'Anna (2021) estimator. Each row reports the average treatment effect for districts first treated in the indicated quarter. Panel A includes cohorts that began implementation before the first COVID-19 lockdown (24 March 2020). Panel B includes cohorts that began implementation during or after the pandemic. Nepal experienced two nationwide lockdowns: 24 March–21 July 2020 and 29 April–1 September 2021. Cohorts 26–28, which implemented during the inter-lockdown and second lockdown periods, show negative effects consistent with programme disruption rather than identification failure. Clustered (district) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%. †Single pilot district; negative estimate likely reflects small-sample noise.

Table A8 — Robustness: Pre-COVID Cohorts (first_period $\leq$ 22), Full Post-Treatment Horizon

	Estimate (Is NHIP eligible = 1)	Std. Error	Pre-treatment Mean
Total Visits	0.1724***	(0.0269)	101,980
OPD Visits	0.1665***	(0.0379)	78,357
Emergency Visits	0.1843**	(0.0807)	9,275
Fit statistics			
Number of Districts	77		
Number of Cohorts	10		
Number of Time Periods	30		

Note: This table re-estimates the main specifications using only pre-COVID treatment cohorts (`first_period` $\leq$ 22, i.e. districts treated by Q4 2019), but keeps their full post-treatment horizons including pandemic-era calendar time. The control pool is unchanged from the headline: districts in cohorts 23–28 continue to serve as not-yet-treated controls in periods preceding their own treatment. The aggregation is restricted to pre-COVID cohorts using the influence-function machinery underlying the package's group aggregation, so cross-cohort covariance through shared controls is preserved. Compare with the headline (Table 2). All outcomes are log-transformed. Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

Table A9 — Robustness: Universal Base Period

	Estimate (Is NHIP eligible = 1)	Std. Error	Pre-treatment Mean
Total Visits	0.1331***	(0.0234)	101,980
OPD Visits	0.1281***	(0.0307)	78,357
Emergency Visits	0.1790***	(0.0646)	9,275
Fit statistics			
Number of Districts	77		
Number of Cohorts	17		
Number of Time Periods	30		

Note: This table re-estimates the main specifications using a universal base period instead of the varying (short gap) base period used in the main analysis. All outcomes are log-transformed. Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

Table A10 — Robustness: Public Facilities Only

	Estimate (Is NHIP eligible = 1)	Std. Error	Pre-treatment Mean
Total Visits (Public)	0.0712***	(0.0196)	72,972
OPD Visits (Public)	0.1122***	(0.0275)	58,849
Emergency (Public)	0.1776***	(0.0670)	4,899
Fit statistics			
Number of Districts	77		
Number of Cohorts	17		
Number of Time Periods	30		

Note: This table re-estimates the main specifications restricting the sample to public facilities only. All outcomes are log-transformed. Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

Table A11 — Robustness: Excluding Kathmandu

	Estimate (Is NHIP eligible = 1)	Std. Error	Pre-treatment Mean
Total Visits	0.1512***	(0.0213)	84,479
OPD Visits	0.1242***	(0.0333)	72,059
Emergency Visits	0.2606***	(0.0585)	6,328
Fit statistics			
Number of Districts	76		
Number of Cohorts	16		
Number of Time Periods	29		

Note: This table re-estimates the main specifications after excluding Kathmandu, which is the last district to receive NHIP (Apr–Jun 2022) and by far the largest in terms of healthcare utilisation. As a late-treated district, Kathmandu sits in the control group for nearly all other cohorts. Its atypical size and COVID recovery trajectory could disproportionately influence the estimates. All three outcomes remain significant at the 1% level, with total visits and emergency visits yielding larger point estimates than the full sample. Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

Table A12 — Robustness: Monthly Data with Varying Anticipation

	Estimate (Is NHIP eligible = 1)	Std. Error	Pre-treatment Mean
Panel A: Anticipation $\delta = 0$ months			
Total Visits	-0.0148	(0.0186)	35,998
OPD Visits	0.0144	(0.0254)	27,685
Emergency Visits	0.1437	(0.1106)	3,309
Panel B: Anticipation $\delta = 1$ months			
Total Visits	0.0398**	(0.0190)	35,998
OPD Visits	0.0338	(0.0248)	27,685
Emergency Visits	0.0899	(0.1196)	3,309
Panel C: Anticipation $\delta = 2$ months			
Total Visits	0.0956***	(0.0254)	35,998
OPD Visits	0.0860**	(0.0338)	27,685
Emergency Visits	0.1367**	(0.0570)	3,309
Panel D: Anticipation $\delta = 3$ months			
Total Visits	0.1639***	(0.0318)	35,998
OPD Visits	0.1452***	(0.0344)	27,685
Emergency Visits	0.1638***	(0.0580)	3,309
Fit statistics			
Number of Districts	77		
Number of Time Periods	91		

Note: This table re-estimates the main specifications using monthly rather than quarterly data, with varying anticipation periods ($\delta = 0, 1, 2, 3$ months). The main analysis aggregates to quarters with $\delta = 1$ quarter; this table verifies that results are robust to finer temporal resolution and different anticipation assumptions. All outcomes are log-transformed. Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

Table A13 — Robustness: Alternative Staggered DiD Estimators

	Total Visits	OPD Visits	Emergency Visits
Callaway & Sant'Anna (2021)	0.1331***	0.1281***	0.1790***
	(0.0284)	(0.0353)	(0.0665)
de Chaisemartin & D'Haultfœuille (2020)	0.0819***	0.0748***	0.0782*
	(0.0180)	(0.0262)	(0.0447)
Borusyak et al. (2024)	0.1259***	0.1101***	0.2790*
	(0.0229)	(0.0243)	(0.1478)

Note: This table compares the main Callaway and Sant'Anna (2021) estimator with two alternative heterogeneity-robust staggered DiD estimators: the local-DID estimator of Chaisemartin and D'Haultfœuille (2020) and the imputation estimator of Borusyak, Jaravel, and Spiess (2024). All three estimators are designed to produce unbiased estimates under heterogeneous treatment effects in staggered adoption designs. All outcomes are log-transformed. Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

A.6.1 Inverse Hyperbolic Sine Transformation

All count outcomes are strictly positive at the district-quarter level and enter the main specifications under the pure $\log$ transformation. As a robustness check on the choice of monotonic transformation, I re-estimate the main specifications replacing $\log$ with the inverse hyperbolic sine (IHS), $\text{asinh}(x) = \log\left(x + \sqrt{x^2 + 1}\right)$. For large values of x, IHS and $\log$ are nearly identical, so the IHS estimates serve as a check that the headline pattern is not driven by the specific log transformation. Tables A14–A16 show that all results are qualitatively and quantitatively unchanged.

Table A14 — Effects of Health Insurance Eligibility on Healthcare Utilization (IHS)

	Estimate (Is NHIP eligible = 1)	Std. Error
Panel A: Overall Healthcare Visits		
Total Visits	0.1331***	(0.0284)
OPD Visits	0.1281***	(0.0353)
Emergency Visits	0.1837***	(0.0709)
Diagnostic Services	0.2098**	(0.1011)
Female Share of Visits	-0.0038*	(0.0020)
Panel B: Number of OPD Visits by Disease Category		
OPD Communicable	0.1473***	(0.0307)
OPD Non-Communicable	0.1363***	(0.0439)
OPD Injuries	0.1302***	(0.0409)
OPD CVD	0.0144	(0.1314)
OPD Cancer	-0.0072	(0.0493)
Fit statistics		
Number of Cohorts	17	
Number of Time Periods	30	

Note: Re-estimation of Table 2 using the inverse hyperbolic sine (IHS) transformation instead of log. Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

Table A15 — Effects of Health Insurance Eligibility on Health Visits by Age Group (IHS)

	Estimate (Is NHIP eligible = 1)	Std. Error
Age 0–9	0.0917***	(0.0320)
Age 10–19	0.1462***	(0.0368)
Age 20–59	0.1420***	(0.0282)
Age 60+	0.1812***	(0.0334)
Fit statistics		
Number of Cohorts	17	
Number of Time Periods	30	

Note: Re-estimation of Table 3 using the IHS transformation. Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

Table A16 — Effects of Health Insurance Eligibility by District Poverty Level (IHS)

	High Poverty (above median)		Low Poverty (below median)	
	Estimate	Std. Error	Estimate	Std. Error
Panel A: Overall Healthcare Visits				
Total Visits	0.0373	(0.0425)	0.1264***	(0.0356)
OPD Visits	0.0212	(0.0523)	0.1218***	(0.0384)
Emergency Visits	0.1287	(0.1087)	0.1687**	(0.0708)
Panel B: OPD Visits by Disease Category				
OPD Communicable	0.0909*	(0.0483)	0.1301***	(0.0395)
OPD Non-Communicable	0.0498	(0.0526)	0.1095**	(0.0490)
OPD Injuries	0.0565	(0.0447)	0.1242***	(0.0456)
Fit statistics				
Number of Cohorts	11		15	
Number of Time Periods	27		30	

Note: Re-estimation of Table 4 using the IHS transformation. Subsamples defined by median split on district-level poverty incidence (2011 World Bank). Clustered (District) standard errors in parentheses. Signif. Codes: *** 1% ** 5% * 10%.

A.7. Distance Mechanism: Instrumental Variable Estimates

This appendix reports instrumental variable estimates as a sensitivity check on the main reduced-form analysis. The FPOC reassignment indicator (Treated $\times$ Post) instruments for distance to FPOC; the first stage corresponds to Table 10 in the main text, and the second stage is reported in Table A17 below under the preferred ward$\times$post fixed-effects specification.

The IV results corroborate the reduced-form interpretation in the main text. In the full sample, distance has a statistically significant negative effect on claims (-0.52 per km, $p < 0.01$), and the Anderson–Rubin 95% confidence interval $[-0.74, -0.41]$ confirms this conclusion under weak-instrument-robust inference (Anderson and Rubin 1949; Andrews, Stock, and Sun 2019). The same holds for the low-poverty subsample, where the IV coefficient is -0.51 with AR CI $[-0.73, -0.40]$. In the high-poverty subsample, however, the first-stage instrument is weak (Kleibergen–Paap rk Wald $F = 1.6$), reflecting that the FPOC policy delivered little distance change in those districts (see Table 10); the Anderson–Rubin confidence interval is unbounded, indicating that the data are uninformative about the per-kilometer distance effect in high-poverty districts.

I do not lean on the IV as the headline causal estimate. The exclusion restriction — that FPOC reassignment affects claims only through distance — is difficult to defend on institutional grounds: reassignment also plausibly disrupts continuity of care between patient and provider, alters administrative procedures for claiming, and shifts within-public-sector facility characteristics such as size and case-mix breadth. Any of these channels would violate exclusion. The IV is reported here for transparency and as evidence that the reduced-form result is not an artifact of the distance-utilisation correlation operating in an unexpected direction.

Table A17 — IV Estimates: Effect of Distance on Insurance Claims

	Full Sample		High Poverty		Low Poverty	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance (km)	-0.350***	-0.516***	-0.617	-0.778	-0.341***	-0.512***
	(0.055)	(0.076)	(0.383)	(0.637)	(0.055)	(0.077)
	[0.171]	[0.263]	[0.505]	[0.471]	[0.173]	[0.269]
Fixed effects						
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Ward × Post	No	Yes	No	Yes	No	Yes
Instrument: Treated × Post						
Kleibergen-Paap rk Wald F	31.8	31.8	3.7	1.6	30.2	30.9
Anderson-Rubin 95% CI	$[-0.735, -0.405]$		$(-\infty, +\infty)$		$[-0.730, -0.400]$	
Summary statistics						
Mean claims	4.18		3.73		4.32	
Mean distance (km)	6.4		6.7		6.3	
Observations	506,242		122,626		383,616	
N (individuals)	262,842		64,036		198,806	

Note: This table reports two-stage least squares estimates of the effect of distance (km) to FPOC hospital on the number of insurance claims per individual. Distance is instrumented by Treated × Post, the interaction of FPOC reassignment with the post-2019 indicator. The treated group is restricted to individuals whose pre- and post-2019 FPOCs are both public hospitals, holding provider ownership constant so the LATE identifies the distance channel; controls are all individuals whose FPOC did not change. Odd columns include individual fixed effects; even columns add ward × post interactions. Poverty groups defined by median split of district poverty rate. First-stage strength is reported as the Kleibergen-Paap rk Wald F, which is cluster-robust under ward clustering. The conventional rule of thumb is $F > 10$; recent literature suggests substantially higher thresholds when conventional standard errors are not robust to weak identification, motivating the Anderson-Rubin confidence interval reported below. The Anderson-Rubin 95% confidence interval is weak-IV-robust (Anderson and Rubin 1949; Andrews, Stock, and Sun 2019) and is the recommended inference procedure for just-identified IV; it is valid regardless of first-stage strength, and is unbounded when the instrument has no identifying power. AR CIs are reported for the preferred specification (Individual FE + Ward×Post) and span both columns within each sample. Ward-clustered standard errors in parentheses; district-clustered in brackets. Distance outliers above 20km excluded. Distances are road distances from Google Maps (97.1%) and OSRM routing (2.9%); no straight-line fallback was required. Signif. Codes: *** 1% ** 5% * 10%.

A.7.1 Distance Robustness

Table A18 reports the first-stage and reduced-form estimates under alternative geographic time fixed effects, varying the time-interaction unit from ward (the preferred specification) to VDC to district. The coefficients are stable in sign and magnitude across all three specifications, and statistical significance is preserved for the full-sample and low-poverty estimates throughout. The high-poverty first-stage coefficient is statistically insignificant under ward$\times$post and gains significance under coarser fixed effects, consistent with limited distance variation in those districts being absorbed differently as the geographic time-interaction unit widens; the high-poverty reduced form is robust at all three FE levels.

Table A18 — **Distance Analysis: Robustness to Geographic Time Interactions**

	Full Sample	High Poverty	Low Poverty
Panel A: First Stage (Distance $\sim$ Treated$\times$Post)			
Individual FE + Ward$\times$Post	-2.181***	-0.761	-2.250***
	(0.387)	(0.608)	(0.405)
Individual FE + VDC$\times$Post	-2.332***	-1.072*	-2.397***
	(0.403)	(0.648)	(0.422)
Individual FE + District$\times$Post	-2.393***	-1.391**	-2.449***
	(0.397)	(0.687)	(0.418)
Panel B: Reduced Form (Claims $\sim$ Treated$\times$Post)			
Individual FE + Ward$\times$Post	1.126***	0.592***	1.152***
	(0.091)	(0.169)	(0.095)
Individual FE + VDC$\times$Post	1.176***	0.602***	1.206***
	(0.091)	(0.153)	(0.096)
Individual FE + District$\times$Post	1.115***	0.548***	1.146***
	(0.094)	(0.149)	(0.099)
Observations	515,963	125,349	390,614
N (individuals)	262,842	64,036	198,806

Note: This table reports robustness of the first-stage and reduced-form estimates to alternative geographic time interactions. All specifications include individual fixed effects; rows vary the geographic level of the time interaction from ward (1,810 units) to VDC (324 units) to district (53 units). The preferred specification (Ward$\times$Post) corresponds to the main-text tables. Ward-clustered standard errors in parentheses. Poverty groups defined by median split of district poverty rate. Distance outliers above 20 km excluded. Signif. Codes: *** 1% ** 5% * 10%.

Table A19 tests whether the results are driven by COVID-19 contamination. The post-period spells overlap substantially with the pandemic, so I progressively restrict the sample to earlier enrolment cohorts with less COVID exposure. The reduced-form effect is stable across cohort restrictions, indicating that COVID contamination does not drive the main results; ward×post fixed effects further absorb ward-level pandemic shocks.

Table A19 — Distance Analysis: Robustness to COVID Exposure

Enrollment Window	COVID%	Full Sample	High Poverty	Low Poverty	N (ind.)
Panel A: First Stage (Distance ∼ Treated×Post)					
May 2019 only	∼14	-2.635***	0.107	-2.900***	35,845
		(0.532)	(1.147)	(0.569)	
May–Aug 2019	∼27	-2.318***	-0.040	-2.481***	85,224
		(0.455)	(0.896)	(0.483)	
May–Nov 2019	∼40	-2.177***	-0.493	-2.275***	123,870
		(0.445)	(0.690)	(0.469)	
May 2019–Feb 2020	∼52	-2.042***	-0.330	-2.137***	161,745
		(0.407)	(0.589)	(0.428)	
All cohorts	∼70	-2.181***	-0.761	-2.250***	258,531
		(0.387)	(0.608)	(0.405)	
Panel B: Reduced Form (Claims ∼ Treated×Post)					
May 2019 only	∼14	1.102***	0.500	1.160***	35,845
		(0.191)	(0.402)	(0.205)	
May–Aug 2019	∼27	1.223***	0.505	1.274***	85,224
		(0.153)	(0.401)	(0.160)	
May–Nov 2019	∼40	1.249***	0.533	1.291***	123,870
		(0.135)	(0.348)	(0.141)	
May 2019–Feb 2020	∼52	1.197***	0.712***	1.224***	161,745
		(0.131)	(0.259)	(0.137)	
All cohorts	∼70	1.126***	0.592***	1.152***	258,531
		(0.091)	(0.169)	(0.095)	

Note: This table reports robustness of the distance analysis to progressively wider enrollment cohort windows. Each row adds later-enrolling cohorts whose post-period spells have greater overlap with the COVID-19 pandemic (March 2020 onward). COVID% is the approximate share of post-period days falling after March 24, 2020. All specifications use individual FE + ward×post FE with ward-clustered standard errors. Distance outliers above 20 km excluded. Signif. Codes: *** 1% ** 5% * 10%.

Table A20 varies the maximum distance threshold from 10 to 40 km. The main specification uses 20 km. The reduced-form effect is stable across bandwidths, supporting 20 km as a defensible policy-relevant commuting-distance cutoff. The first-stage distance effect is similarly robust.

Table A20 — Distance Analysis: Robustness to Distance Bandwidth

Bandwidth	Full Sample	High Poverty	Low Poverty	N (ind.)
Panel A: First Stage (Distance $\sim$ Treated$\times$Post)				
0–10 km	-0.185	-0.497	-0.174	206,255
	(0.178)	(0.342)	(0.184)	
0–15 km	-1.633***	-0.984	-1.661***	241,170
	(0.346)	(0.745)	(0.360)	
0–20 km	-2.181***	-0.761	-2.250***	262,842
	(0.387)	(0.608)	(0.405)	
0–25 km	-2.595***	-2.337*	-2.613***	273,608
	(0.385)	(1.375)	(0.401)	
0–30 km	-2.714***	-1.388	-2.821***	280,145
	(0.371)	(1.260)	(0.389)	
0–40 km	-2.672***	0.055	-2.931***	286,773
	(0.415)	(1.101)	(0.444)	
Panel B: Reduced Form (Claims $\sim$ Treated$\times$Post)				
0–10 km	0.867***	0.311	0.886***	206,255
	(0.127)	(0.198)	(0.131)	
0–15 km	1.059***	0.655***	1.076***	241,170
	(0.108)	(0.196)	(0.112)	
0–20 km	1.126***	0.592***	1.152***	262,842
	(0.091)	(0.169)	(0.095)	
0–25 km	1.133***	0.571***	1.172***	273,608
	(0.081)	(0.151)	(0.086)	
0–30 km	1.102***	0.623***	1.140***	280,145
	(0.077)	(0.135)	(0.083)	
0–40 km	1.089***	0.631***	1.132***	286,773
	(0.072)	(0.118)	(0.078)	

Note: This table reports robustness of the distance analysis to varying maximum distance thresholds. Each row

Figure A4 reports a leave-one-district-out sensitivity check on the high-poverty reduced-form coefficient. Looping over the 26 high-poverty districts in the analysis sample, the reduced-form coefficient remains bounded between approximately 0.35 and 0.70, confirming that the high-poverty result is not driven by any single district. The 26-district FPOC sample differs from the 37 high-poverty districts in the main heterogeneity analysis (Table 4): the FPOC analysis requires Pub→Pub-reassigned individuals within the 20 km bandwidth, which drops 14 districts (mean 2.8 empanelled facilities vs 5.2 in the retained districts), including Nepal's most remote and least-supplied areas such as Humla, Darchula, Doti, Mugu, and Dolpa. This itself is consistent with the supply-side reading: the most supply-constrained districts cannot support the FPOC design. The LOO bounds therefore characterise robustness within the better-supplied half of high-poverty districts.

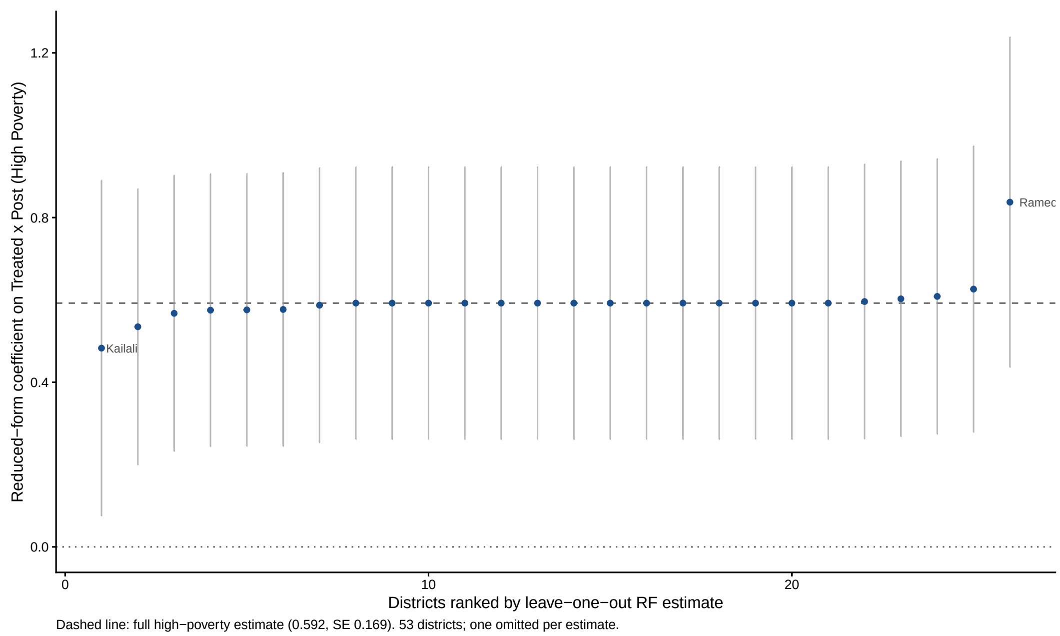


Figure A4 — Leave-One-District-Out Sensitivity: High-Poverty Reduced-Form Coefficient

A.7.2 Within-Public-Sector Hospital Quality

A concern with the distance interpretation is that FPOC reassignment may simultaneously change hospital quality: if reassigned public facilities are systematically better-equipped than original public facilities, the claims response could reflect quality rather than distance. To test this, Table A21 compares the diagnostic diversity (number of unique ICD codes observed pre-2019) of pre-reassignment and post-reassignment FPOCs for each Pub→Pub treated individual. Reassigned FPOCs have systematically *fewer* unique ICD codes on average (-204.5 codes per individual, $p < 0.001$), reflecting that the policy moved beneficiaries to smaller, closer public facilities. Combined with the increase in claims documented in Table 11, this rules out a "higher-quality hospital" channel: claims rose despite reassignment to less diagnostically broad facilities. The unique-ICD measure is mechanically correlated with hospital claim volume, so the result conflates "quality" and "volume"; this rests on the assumption that smaller-ICD-count hospitals are not systematically of higher quality on unobserved dimensions. The assumption is plausible but cannot be tested directly with the available data. A per-encounter measure (ICD codes per claim rather than per hospital) would partially address the volume confound and is left for future work.

Table A21 — Within-Public-Sector Hospital Quality Comparison (Pub→Pub Treated)

	Unique ICD codes (mean)
Pre-reassignment FPOC	617.6
Post-reassignment FPOC	413.2
Within-individual change (post − pre)	-204.5
Standard deviation	481.2
Paired t-test p-value	0.000
N (treated with both pre & post ICD data)	12,564

Note: For each Pub→Pub treated individual, this table compares the diagnostic diversity of their pre-reassignment FPOC against their post-reassignment FPOC. Quality is proxied by the number of unique ICD codes claimed at each hospital in the pre-2019 period (i.e., the measure is independent of post-policy claiming). The within-individual paired comparison absorbs individual heterogeneity. Sample: 15,323 Pub→Pub treated individuals; 12,564 have pre-2019 ICD data for both their pre- and post-reassignment FPOC. Reassigned FPOCs have systematically fewer unique ICD codes than original FPOCs (reflecting that reassignment moved beneficiaries to smaller, closer public facilities). Combined with the increase in claims documented in Table 11, this rules out a "higher-quality hospital" channel: claims rose despite reassignment to less diagnostically broad facilities, supporting distance reduction as the operative mechanism. CAVEAT: unique-ICD count is mechanically correlated with hospital claim volume, so a busier hospital sees more distinct diagnoses by chance; this proxy therefore conflates "quality" and "volume." However, the directional argument here only requires that smaller-ICD-count hospitals are not systematically of higher quality on other dimensions, which is plausible.

A.7.3 Attrition and Sample Selection

Since the balanced panel requires observation in both the pre- and post-2019 periods, individuals who stopped claiming after the policy change are excluded. If attrition were differential—higher among those more likely to be treated—the estimates would be biased. To test this, I classify all pre-2019 claimants with a recorded FPOC by their pre-period hospital type (public or private) and compare attrition rates. Table A22 shows that private-FPOC individuals, who were more likely to be reassigned by the policy, had a *lower* attrition rate (31.8%) than public-FPOC individuals (32.7%). A logistic regression confirms the difference is statistically significant after controlling for pre-period claiming intensity. This rules out differential attrition as a source of bias.

Table A22 — Attrition Test: Pre-Period Claimant Retention by FPOC Type

Pre-2019 FPOC Type	Pre-Period Claimants	Attrited (N)	Attrition Rate (%)	Mean Pre-Period Claims
Public	440,730	144,257	32.73	1.21
Private	165,024	52,491	31.81	1.18
Total	605,754	196,748	32.48	1.20

Logistic regression: Attrited $\sim$ Private FPOC + Pre-Claims

Private FPOC	-0.0672*** (SE: 0.0063)
Pre-period claims	-1.0626*** (SE: 0.0082)

Note: This table tests for differential attrition from the balanced claims panel used in the distance analysis. The sample consists of pre-2019 claimants with a recorded FPOC hospital—the same population from which the balanced panel was drawn (individuals without a recorded FPOC were never candidates for the distance analysis). Pre-2019 claimants are classified by their **last pre-cutoff spell** FPOC hospital, matching the definition used to build the balanced panel. "Attrited" indicates no post-2019 claims observed. **Caveat:** actual treatment status (FPOC reassignment) cannot be observed for attritors because they have no post-2019 claims from which to read post-period FPOC. We therefore proxy treatment exposure using pre-period FPOC type: individuals with private pre-FPOCs were mechanically reassigned by the 2019 policy (which restricts FPOC to public hospitals), while public pre-FPOCs may or may not have been reassigned (e.g., to a closer public facility). This is an imperfect proxy; the test is therefore conservative in detecting attrition driven by the policy. If the policy caused differentially higher attrition among those more likely to be treated, we would expect higher attrition among private-FPOC claimants. The logistic regression controls for pre-period claiming intensity. Hospital ownership classified from facility names; NGO classified as private.

A.8. Das Gupta Decomposition: Robustness

As a robustness check on the three-way decomposition reported in Table 8, I re-estimate the decomposition using the Das Gupta (1978, 1993) weighting scheme, which constructs main-effect weights that absorb the interaction term by construction and thus leaves no residual. The Das Gupta weights for three factors $Y = A \times B \times C$ are symmetric averages across orderings: the contribution of factor A is $(A_1 - A_2) \times [(B_1C_1 + B_2C_2)/3 + (B_1C_2 + B_2C_1)/6]$, with analogous expressions for B and C. Because the three main effects are constructed to sum exactly to $A_1B_1C_1 - A_2B_2C_2$, the method does not produce a separate interaction term—any joint disadvantage across margins is redistributed into the three main-effect components.

Table A23 — Das Gupta (1978, 1993) Decomposition: Robustness Check

	Claims per 1,000 pop.	Share of gap (%)
Panel A: Group means		
High-poverty districts	598.9	
Low-poverty districts	1440.1	
Total gap	-841.2	100
Panel B: Das Gupta decomposition		
Enrollment margin	-429.5	51.1
	(202.2)	(19.1)
Probability of use	-189	22.5
	(97.8)	(8.9)
Intensity (claims per user)	-222.8	26.5
	(79.7)	(17.9)

Notes: Three-factor decomposition of the gap in insurance claims per 1,000 population between high-poverty (above-median) and low-poverty (below-median) districts, using a median split of the pre-treatment 2011 district poverty rate (World Bank), and applying the Das Gupta (1978, 1993) interaction-free weights. For $Y = A \times B \times C$, the main-effect contribution of factor A is $(A_1 - A_2) \times [(B_1C_1 + B_2C_2)/3 + (B_1C_2 + B_2C_1)/6]$, with symmetric formulas for B and C; the three main effects sum exactly to the total gap by construction, so no residual interaction term appears. This differs from the naive symmetric weighting used in the main text (Table 8), which retains the interaction explicitly. Because the interaction in the main-text specification is only 1% of the total gap, the two decompositions yield nearly identical shares; each main effect gains approximately one-third of the redistributed interaction. Standard errors in parentheses are bootstrapped (500 iterations, resampling districts within poverty groups).

Table A23 reports the results. The Das Gupta decomposition assigns 51 percent of the gap to the enrolment margin, 23 percent to the probability-of-use margin, and 27 percent to the intensity margin—each within half a percentage point of the corresponding naive-symmetric share in Table 8. The near-identity is mechanical: because the interaction in the main-text decomposition is only 1 percent of the total gap, redistributing it across the three main effects shifts each share by only fractions of a percentage point. The substantive conclusion is therefore unchanged under either weighting convention: the high-poverty gap accumulates in roughly comparable proportions across enrolment, claiming conditional on enrolment, and intensity conditional on any use. The main-text specification is preferred because it makes the (small) joint disadvantage across margins visible rather than redistributing it silently.