

Original research

Geographic Concentration of RCI-Listed Rehabilitation Training Institutions in India: A Reproducible National Register Analysis

Koti Reddy Saripalli

Bharath Healthcare Laboratories Private Limited (Pinnacle Blooms Network), Plot No. 50/A, Raghavendra Nagar Colony, Jeedimetla, Suchitra, Hyderabad 500055, Telangana, India

Correspondence: Koti Reddy Saripalli; kotii@bhclpl.org; ORCID: 0009-0007-7866-0737

Manuscript BHCL-PUB-2026-09-02 | Version 2.0 | 15 September 2026

Abstract 251 words | Main text 2618 words

Revised manuscript. Not peer reviewed.

Abstract

Background: Reliable information on training institutions can support rehabilitation workforce planning, but regulatory listings do not directly measure active capacity or professional supply. This study describes the geographic distribution and concentration of institutional entries in a fixed Rehabilitation Council of India (RCI) register snapshot.

Methods: A descriptive census reconstructed every institution code and printed serial number from the 113-page RCI list dated 4 February 2026. Source-heading state assignments were checked against code prefixes. Every final institution label was reconciled with source text; extraction exceptions were reviewed with AI assistance. All 36 states and union territories were retained, including those absent from the source. Jurisdictional counts, cumulative shares, the Herfindahl-Hirschman index (HHI) and Gini coefficient were calculated. Course counts, seats, current approval status and population-adjusted rates were outside scope.

Results: The reconstruction recovered 1,068 unique codes and the complete serial sequence 1–1,068. Entries were present in 32 jurisdictions: 27 states and five union territories. Fourteen institutional display labels were corrected or restored without changing state totals. Uttar Pradesh and Rajasthan accounted for 482 entries (45.13%); the five largest counts totalled 663 (62.08%) and the ten largest 847 (79.31%). Four jurisdictions were not represented in the source. Gini across all 36 jurisdictions was 0.6914 and HHI was 0.1216. Excluding unrepresented jurisdictions reduced Gini to 0.6528.

Conclusions: RCI-listed institutional entries are geographically concentrated. The source-linked dataset offers a reproducible organisational frame for further workforce research. These counts do not establish active training capacity, workforce shortage or population-adjusted inequity; those questions require additional programme-period, student-flow, employment and need data.

Keywords: rehabilitation; health workforce planning; training institutions; India; geographic concentration; administrative data; reproducibility

Background

Rehabilitation workforce planning requires reliable information about how and where professionals are trained. The World Health Organization's Rehabilitation 2030 initiative identifies a multidisciplinary workforce and stronger national and subnational planning as priorities.[1] World Health Assembly resolution WHA76.6 places rehabilitation within health-system strengthening and universal health coverage.[2] These commitments create a practical information requirement: countries need data that distinguish training institutions, educational opportunities, professional qualifications and the workforce available to deliver services. An institution list can support the first of these tasks, but its analytical unit must remain clear.

WHO's Guide for Rehabilitation Workforce Evaluation takes a labour-market and competency perspective, connecting workforce description with the problems and opportunities experienced by workers and the systems that employ them.[3] Educational geography is one component of that broader assessment. The location of an institution can identify an organisational contact point for planning, surveys and collaboration. It does not indicate how many students enter or complete training, whether graduates remain in the same jurisdiction, or whether residents can obtain rehabilitation services. Administrative records become more useful when these distinctions are maintained during extraction and analysis.

In India, the Rehabilitation Council of India (RCI) regulates rehabilitation and special-education training and recognises institutions offering relevant qualifications.[4] Its list dated 4 February 2026 contains institution codes, names and addresses alongside programme, approval-period, extension and intake information.[5] RCI describes the public list in terms of institutions and their academic-year status, with intake per batch.[6] Because a single institutional entry may contain several courses and historical periods, translating the document into a current capacity estimate requires additional course-level adjudication. The

institutional identifiers and state headings support a narrower, directly reproducible geographical analysis.

This study describes the geographic distribution and concentration of distinct RCI-listed

30 institution codes in that single register snapshot. It also provides a source-linked institution dataset and reproducible validation process. The questions are: how many institutional entries are represented in each state or union territory; how much of the national listing is concentrated in a small number of jurisdictions; and what additional information is required before this administrative geography can inform workforce allocation decisions? The study
35 makes no priority claim to be the first analysis of rehabilitation training in India.

Methods

Study design and analytical unit

This was a descriptive census of a fixed public administrative document. The analytical unit was one distinct RCI institution code printed in the 4 February 2026 list. All numbered

40 institutional entries were eligible, including entries whose programme information was historical, conditional, incomplete or otherwise insufficient to establish current activity. An entry was not removed solely because its course details were unavailable to the institution-level extraction. A code identifies a register entry; it was not assumed to equal a unique university, corporate owner, physical campus or currently operating training provider.

45 Different codes were retained even when institutional names were similar.

The geographic frame included India's 28 states and eight union territories.[7] A jurisdiction with no source entry was retained with a count of zero and labelled 'not present in this source'. This is a statement about the selected register, not a comprehensive assessment of educational provision in that territory. The study did not combine other regulatory lists,

50 distance-mode lists, clinical service directories or provider-owned data with the source census.

Source provenance and extraction

The official source was a 113-page PDF headed 'List of Approved Institutions upto 04-02-2026'.^[5] The file was retrieved from the government publication host linked to RCI. The

55 source URL, snapshot date, audit date and SHA-256 checksum are supplied in the provenance manifest. The checksum fixes the exact document analysed, so a later replacement at the same URL cannot silently change the study population. Extraction and analysis for this revision were completed on 15 September 2026 using Python and pdfplumber 0.11.9.

60 Institution codes were extracted directly from their printed column using text coordinates. Printed serial numbers were extracted independently from the adjacent serial-number column. State-heading events were located on each page and carried forward in reading order until the next heading. Institution names and address excerpts were obtained from detected table cells where available; coordinate-bounded extraction from the name-and-address column supplied

65 the remaining entries. These two name-extraction routes were required because some institutional entries began at the bottom of a page or were not returned as complete table cells. Page numbers and vertical coordinates were retained as source locators.

Geographic and label reconciliation

The source heading was the primary basis for state assignment. Historical or variant headings

70 were normalised using an explicit crosswalk, including Orissa to Odisha, Pondicherry to Puducherry, Uttranchal to Uttarakhand, Chattisgarh to Chhattisgarh and Tamilnadu to Tamil Nadu. Jammu and Kashmir was represented as a union territory; Delhi denotes the National Capital Territory of Delhi. Each heading-derived assignment was cross-checked against a separately specified institution-code prefix map. Prefixes were treated as a consistency check

75 rather than an automatic substitute for the source heading; in particular, JK and JKH were distinguished.

Source-derived codes and geographic assignments were reconciled against the institution table accompanying the earlier draft. Display labels from that table were retained only when they matched the source name-and-address text after case and punctuation normalisation.

80 Non-matching or empty labels were reviewed against the source and corrected in a versioned crosswalk. The released label is a source-supported identifier for readers, not a claim to establish the institution's complete current legal name. Source terminology within institutional names was preserved. Contact metadata were removed from the released name-and-address excerpts; the original public document remains the reference for complete
85 entries.

Validation and reproducibility

Validation covered every included record. Checks required 1,068 unique codes, a complete sequence of printed serial numbers from 1 to 1,068, no unmatched codes between the reconstructed and earlier institution tables, agreement between heading-derived and prefix-
90 derived state assignments, and a source-text match for every final institution label. One serial-number extraction rule was broadened to allow the printed number 714 without a terminal full stop. The record-level validation file identifies the source page and name-extraction route for each code.

These were computational checks with AI-assisted review of extraction exceptions and
95 targeted source inspection. They were not independent double extraction by two human reviewers, and no such procedure is claimed. Full-record checks establish consistency with this source document; they cannot establish that the regulator's underlying entries were complete, current or error-free. A portable script rebuilds the released institution table, validation file, state counts, figure coordinates and concentration statistics from the
100 checksum-verified PDF and the supplied label crosswalk. It does not depend on a course-intake dataset.

Descriptive analysis

For each jurisdiction, the number of distinct institution codes and percentage of the national total were calculated. Counts were sorted in descending order, with alphabetical ordering for ties. Cumulative shares were reported for the two, five and ten largest counts. The median jurisdictional count was calculated both across all 36 jurisdictions and across the 32 represented jurisdictions. States and union territories remained visible individually rather than being combined into broad regions whose boundaries could obscure the distribution.

Concentration was summarised by two complementary measures. The Herfindahl-Hirschman index (HHI) was calculated as the sum of squared jurisdictional shares, $HHI = \sum (n_j/N)^2$, where n_j is a jurisdiction's count and N is the national total. Its reciprocal is the number of equal-share jurisdictions that would produce the same concentration; it is an algebraic summary, not a target number of training locations. The Gini coefficient was calculated as $\sum_j \sum_k |n_j - n_k| / (2KN)$, where K is the number of jurisdictions. A Lorenz curve plotted

cumulative institution share against cumulative jurisdiction share after sorting counts from smallest to largest. A sensitivity description recalculated Gini among represented jurisdictions only.

All concentration measures weight administrative jurisdictions equally. They do not adjust for population, land area, educational demand, disability prevalence or service need, and an equal-count reference is not an equity standard. No population-adjusted rates were calculated. No hypothesis tests or sampling confidence intervals were used because the analysis enumerated the selected document rather than sampling its entries; uncertainty about register coverage remains a separate limitation. Calculations retained full precision, with percentages generally displayed to two decimal places.

125 **Variables outside the analysis**

The analysis excluded course counts, intake or seat totals, annual training capacity, current programme approval status, ownership, professional supply and patient outcomes.

Programme strings and intake tokens were not used to select or weight institutions. A source entry can contain a superseded qualification, a later qualification, differing intake figures and

130 conditions applying to particular periods. Establishing an active capacity estimate would therefore require a separately validated institution-by-programme-by-period dataset. Keeping that task outside this study preserves a consistent institution-level denominator.

Results

Source coverage and record validation

135 The source reconstruction recovered 1,068 unique institution codes and the complete printed serial sequence from 1 to 1,068. All codes matched the earlier institution table, and all heading-derived state assignments agreed with both the code-prefix check and the earlier table after jurisdiction-name normalisation. Table-cell extraction supplied name-and-address text for 980 records; coordinate-bounded extraction supplied 88. Fourteen display labels

140 required correction or restoration, principally because address tokens had been placed within names or an institutional name was missing. All final labels matched their source text. These label changes did not alter any state count.

Distribution across states and union territories

Entries were present in 32 of 36 jurisdictions: 27 states and five union territories. The listed

145 state entries totalled 1,022 and union-territory entries 46. Uttar Pradesh had the largest count, 254 (23.78%), followed by Rajasthan, 228 (21.35%), Haryana, 75 (7.02%), Maharashtra, 59 (5.52%), and Madhya Pradesh, 47 (4.40%). The complete distribution is shown in Table 1 and Figure 1. Table rankings describe the number of entries in this source, without adjustment for jurisdiction size.

150 Andaman and Nicobar Islands, Dadra and Nagar Haveli and Daman and Diu, Lakshadweep
and Nagaland were not represented in the source. Among represented jurisdictions, Ladakh
had one entry and Tripura two. Thirteen jurisdictions each had between one and five entries;
together they accounted for 46 entries (4.31%). The median count was 11 across all 36
jurisdictions and 16.5 across represented jurisdictions. These measures describe a distribution
155 with several large counts and many small counts.

Concentration of institutional entries

Uttar Pradesh and Rajasthan together accounted for 482 entries (45.13%). The five largest
counts totalled 663 (62.08%), and the ten largest totalled 847 (79.31%). The other 26
jurisdictions collectively accounted for 221 entries (20.69%). At the opposite end of the
160 distribution, the 18 jurisdictions with the smallest counts, including the four unrepresented
jurisdictions, contained 53 entries (4.96%). Table 2 summarises these concentrations without
assigning normative thresholds.

The HHI was 0.1216, equivalent to 1,215.8 on a 0–10,000 scale. Its reciprocal was 8.23
equal-share jurisdictions. The Gini coefficient across all 36 jurisdictions was 0.6914.

165 Excluding the four jurisdictions with no source entry reduced Gini to 0.6528; concentration
therefore remained apparent within the represented jurisdictions. Figure 2 shows the
corresponding Lorenz curve. The diagonal is an equal-count reference across administrative
units and does not describe an expected or desirable population distribution.

Discussion

170 Principal findings and meaning of concentration

The RCI snapshot provides a geographically extensive but concentrated set of institutional
entries. Most states and several union territories are represented, yet two states account for
nearly half of the listing and ten for almost four-fifths. The high concentration relative to

equal jurisdictional counts is visible without assigning each institution an assumed number of

175 seats or professionals. Complete recovery of the printed code and serial sequences supports the internal completeness of the reconstructed institution census for this specific document.

Concentration has several possible explanations that this design cannot distinguish.

Jurisdictions differ in population, higher-education systems, historical institutional development, demand for particular qualifications, regulatory participation and patterns of

180 student migration. Some training institutions may serve a national catchment, while others primarily recruit locally. Consequently, a high count is not evidence of oversupply and a small count is not a quantified workforce deficit. The present results locate questions for a subsequent needs-informed assessment; they do not identify an optimal distribution of institutions.

185 **How the dataset can inform workforce planning**

The immediate value of this resource is a defined, source-linked organisational frame.

Regulators, educational researchers and workforce planners can use institution codes to organise follow-up information on current activity, course availability, annual admissions, completion, professional registration and graduate destinations. Each additional variable

190 should have its own reference period and validation rule. Linking these stages would help separate training opportunities from realised workforce supply, consistent with the labour-market perspective of GROWE.[3] Such linkage would require attention to missing records, changes in names or ownership, and qualifications that follow different regulatory pathways.

For jurisdictions with few or no entries in the selected register, a useful next step is to verify

195 what forms of training are available through other lists or arrangements and where prospective students obtain training. This may reveal local provision missing from the frame, reliance on other jurisdictions, or constraints requiring further study. Decisions about expanding institutions should then consider faculty, clinical placements, demand, financing

and employment opportunities. The present census cannot rank those options or show that
 200 any particular technology, service model or provider can resolve a workforce shortage.

Regulatory documents as reusable research infrastructure

The extraction process illustrates why identifiable institutional units should be separated from
 more complex programme histories. Codes and headings supported full-record reconciliation,
 while page-spanning entries required a different name-extraction route. The source itself is an
 205 administrative publication, not a database designed around this research question. Preserving
 source locators and reporting the checks performed allows another analyst to inspect a
 disputed entry rather than infer how a total was constructed. The approach is applicable to
 other countries where public regulatory information remains distributed in document tables.

A regularly versioned machine-readable register could reduce this reconstruction burden. A
 210 useful structure would keep institution identifiers, locations, programme identifiers, approval
 periods, status conditions and intake limits in linked tables, with a stated update date and
 change history. This is a data-management recommendation arising from the study's
 extraction experience, not an assessment of institutional compliance. Stable identifiers,
 provenance and reusable processing code are also consistent with the FAIR principles for
 215 scientific data stewardship.[8] The supplied package implements those elements at the
 institution level while leaving unverified capacity fields outside the release.

Strengths and limitations

Strengths include a fixed official source, reconstruction of every printed institutional code,
 explicit retention of unrepresented jurisdictions, a complete serial-number check, two
 220 geographic consistency checks and source-text reconciliation of every final label. The
 analysis is reproducible without private provider data. Its measures have transparent formulas
 and a single denominator, making the descriptive findings straightforward to update when

another comparable snapshot becomes available. No claim is made that an institution-level census replaces programme-level or workforce-level evaluation.

225 Several limitations constrain interpretation. First, source completeness is not real-world completeness: the list may contain legacy entries or omit relevant institutions outside its scope. Second, this study did not determine whether each listed institution was active or had an approval valid in a particular academic year. Third, one code need not correspond to one physical teaching site, and similar names may belong to separate entries; the unit is therefore
230 explicitly the RCI code. Fourth, the geographic analysis stops at the state/union-territory level and cannot describe within-state travel distances, rural access or local clustering.

Fifth, institutional counts give equal weight to providers of very different size and educational scope. Sixth, concentration statistics are sensitive to the choice and number of administrative units and cannot be interpreted as population-adjusted inequity. Seventh, the
235 document contains no information about students' origins, graduates' destinations, professional participation or service outcomes. Finally, source checks in this revision were computational and AI-assisted, without independent human double extraction. Public release of the source locators, crosswalk and code enables further scrutiny, but does not remove the possibility of a shared extraction or interpretation error.

240 **Conclusions**

The 4 February 2026 RCI register contains 1,068 distinct institutional entries across 32 states and union territories. Their geographic distribution is concentrated: two states contain 45.13% and ten jurisdictions 79.31% of the entries. A complete source-linked institution dataset can support subsequent verification and workforce-planning research. Moving from
245 this administrative geography to decisions about training capacity or service access requires additional, period-specific information on programmes, student flows, professional registration, employment and population need.

Declarations

Ethics approval and consent to participate

250 The study analysed publicly available institutional records. It enrolled no participants and used no individual-level clinical or student data. No ethics-committee approval or waiver is claimed.

Consent for publication

Not applicable: no participant-level information, photographs or case descriptions are
255 included.

Availability of data and materials

The audited institution table, state counts, figure coordinates, record-validation file, label crosswalk, change log, source manifest and portable reproduction script accompany this manuscript as supplementary materials. The official source is available at reference 5. The
260 source manifest identifies the exact PDF by its SHA-256 checksum. No repository deposition or dataset DOI is claimed.

Competing interests

Koti Reddy Saripalli is affiliated with Bharath Healthcare Laboratories Private Limited, which operates Pinnacle Blooms Network and has an organisational interest in rehabilitation
265 workforce development. The study uses a public regulatory document and contains no analysis of Pinnacle treatment outcomes, product performance or proprietary patient data.

Funding

No external funding was reported for the original study. Bharath Healthcare Laboratories Private Limited provided organisational support. No funding was received from RCI or WHO
270 for this analysis.

Authors' contributions

Koti Reddy Saripalli is the sole named author and corresponding author, responsible for the study question, interpretation and manuscript accountability. Computational extraction, formal analysis, source reconciliation and draft revision used the AI-assisted workflow

275 described in Methods. Final scientific review and publication approval are responsibilities of the named author; no second human extraction or independent human replication is represented as completed.

Acknowledgements

The Rehabilitation Council of India made the source document publicly available. Use of that

280 document does not imply RCI endorsement of the analysis or its interpretation. No additional contributors are named.

Use of artificial intelligence

OpenAI Codex assisted with source-document extraction, code development, computational checks, review of label exceptions and manuscript drafting. All published numbers were

285 calculated by executable Python code from the reconstructed institutional records. The workflow checked every code, serial number, heading-based state assignment and final label against the source. AI assistance is not equivalent to independent human verification; the validation record and limitations describe the checks actually performed.

References

- 290 1. World Health Organization. Rehabilitation 2030. Geneva: WHO. Accessed 15 September 2026.
<https://www.who.int/initiatives/rehabilitation-2030>
2. World Health Assembly. Strengthening rehabilitation in health systems. Resolution WHA76.6. Geneva: World Health Organization; 2023.
https://apps.who.int/gb/ebwha/pdf_files/WHA76/A76_R6-en.pdf
- 295 3. World Health Organization. Guide for rehabilitation workforce evaluation. Geneva: WHO; 2023. ISBN 978-92-4-007167-4. <https://www.who.int/publications/i/item/9789240071674>

4. Rehabilitation Council of India. Objectives. New Delhi: RCI. Accessed 15 September 2026.
<https://rehabcouncil.nic.in/objectives/>
5. Rehabilitation Council of India. List of Approved Institutions upto 04-02-2026. New Delhi: RCI;
300 2026. 113 pages. Source snapshot audited 15 September 2026. [Available online](#)
6. Rehabilitation Council of India. Regular Mode: Recognized Institutions. New Delhi: RCI.
Accessed 15 September 2026. <https://rehabcouncil.nic.in/regular-mode-inst/>
7. Election Commission of India. About ECI. New Delhi: ECI. Accessed 15 September 2026.
<https://www.eci.gov.in/about-eci/>
- 305 8. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific
data management and stewardship. Scientific Data. 2016;3:160018. doi:10.1038/sdata.2016.18.
<https://doi.org/10.1038/sdata.2016.18>

Table 1. RCI-listed institutional entries by state or union territory, 4 February 2026

State or union territory	Type	Listed institutions	Share (%)
Uttar Pradesh	State	254	23.78
Rajasthan	State	228	21.35
Haryana	State	75	7.02
Maharashtra	State	59	5.52
Madhya Pradesh	State	47	4.40
Tamil Nadu	State	43	4.03
Karnataka	State	39	3.65
Odisha	State	36	3.37
Delhi	UT	33	3.09
Kerala	State	33	3.09
West Bengal	State	32	3.00
Andhra Pradesh	State	28	2.62
Punjab	State	21	1.97
Telangana	State	20	1.87
Gujarat	State	19	1.78
Jharkhand	State	18	1.69
Bihar	State	15	1.40
Chhattisgarh	State	15	1.40
Himachal Pradesh	State	7	0.66
Assam	State	5	0.47
Chandigarh	UT	5	0.47
Manipur	State	5	0.47
Arunachal Pradesh	State	4	0.37
Puducherry	UT	4	0.37
Sikkim	State	4	0.37
Uttarakhand	State	4	0.37
Goa	State	3	0.28
Jammu and Kashmir	UT	3	0.28
Meghalaya	State	3	0.28
Mizoram	State	3	0.28

State or union territory	Type	Listed institutions	Share (%)
Tripura	State	2	0.19
Ladakh	UT	1	0.09
Andaman and Nicobar Islands	UT	0	0.00
Dadra and Nagar Haveli and Daman and Diu	UT	0	0.00
Lakshadweep	UT	0	0.00
Nagaland	State	0	0.00
India	28 states + 8 UTs	1,068	100.00

Each institution code is counted once. Zero means no code from that jurisdiction was found in this source snapshot. Counts do not establish current institutional activity, training capacity or population-adjusted access. Delhi denotes the National Capital Territory of Delhi. Ties are ordered alphabetically; displayed percentages may not sum exactly because of rounding.

Table 2. Concentration measures for listed institution counts

Measure	Result	Interpretation
Two largest jurisdictional counts	482 / 1,068 (45.13%)	Uttar Pradesh and Rajasthan
Five largest jurisdictional counts	663 / 1,068 (62.08%)	Adds Haryana, Maharashtra and Madhya Pradesh
Ten largest jurisdictional counts	847 / 1,068 (79.31%)	Cumulative count share
Smallest-count half of jurisdictions	53 / 1,068 (4.96%)	18 of 36 jurisdictions, including four zeros
Median jurisdictional count	11	All 36 jurisdictions
Gini coefficient	0.6914	Equal weighting of 36 jurisdiction counts
Gini among represented jurisdictions	0.6528	32 jurisdictions with at least one entry
HHI, unit scale	0.1216	Sum of squared jurisdictional shares
Reciprocal HHI	8.23	Equivalent number of equal-share jurisdictions

These are descriptive concentration measures for institution-code counts. They do not measure competition, population-adjusted inequity, service availability or workforce shortage. No external classification thresholds are applied.

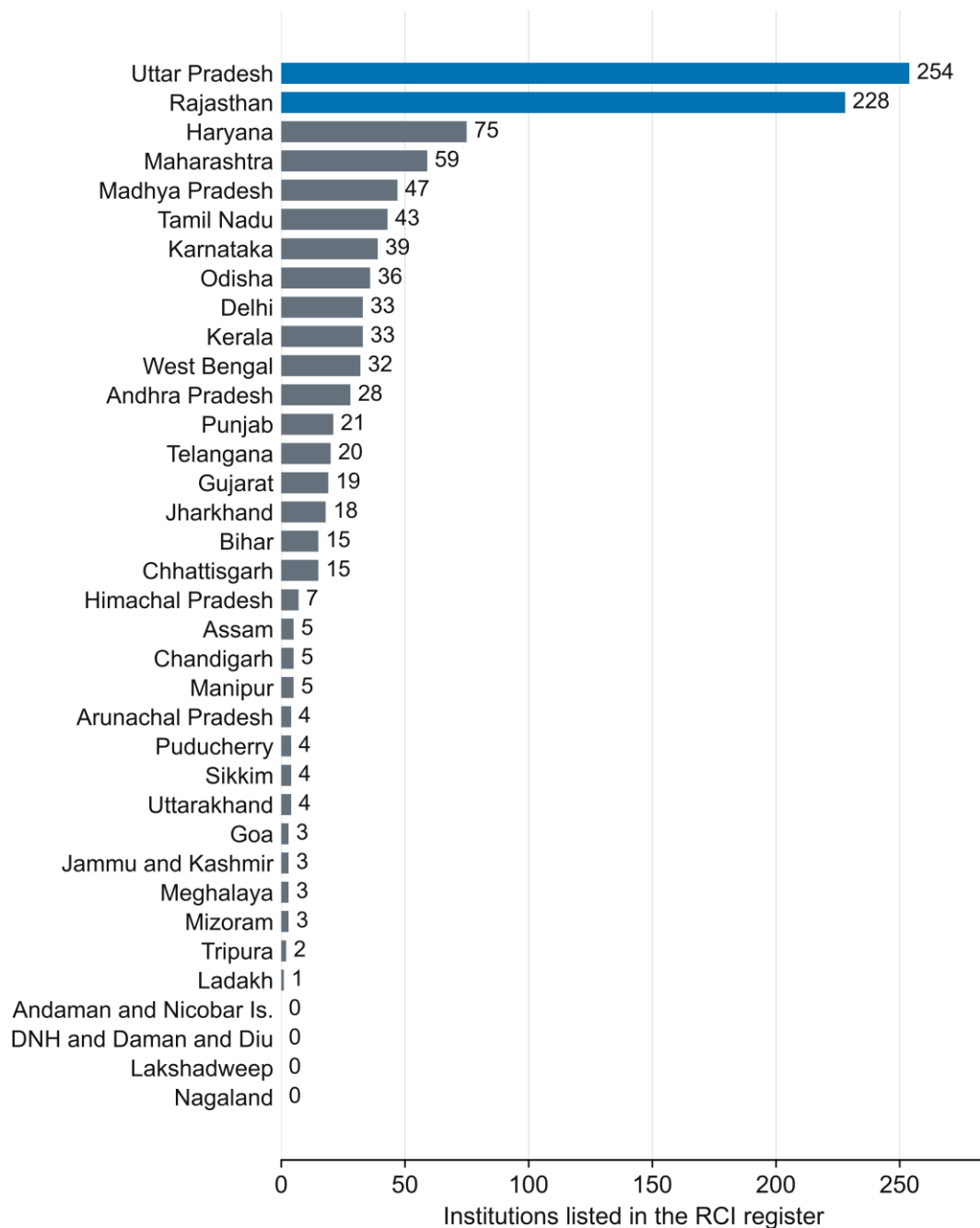


Figure 1. Geographic distribution of listed institutions. Number of distinct RCI institution codes by state or union territory in the 4 February 2026 source. All 36 jurisdictions are shown, including four with no entry. Bars represent register entries, not current training seats, active professionals or rates per population. Source: workforce_state_counts.csv.

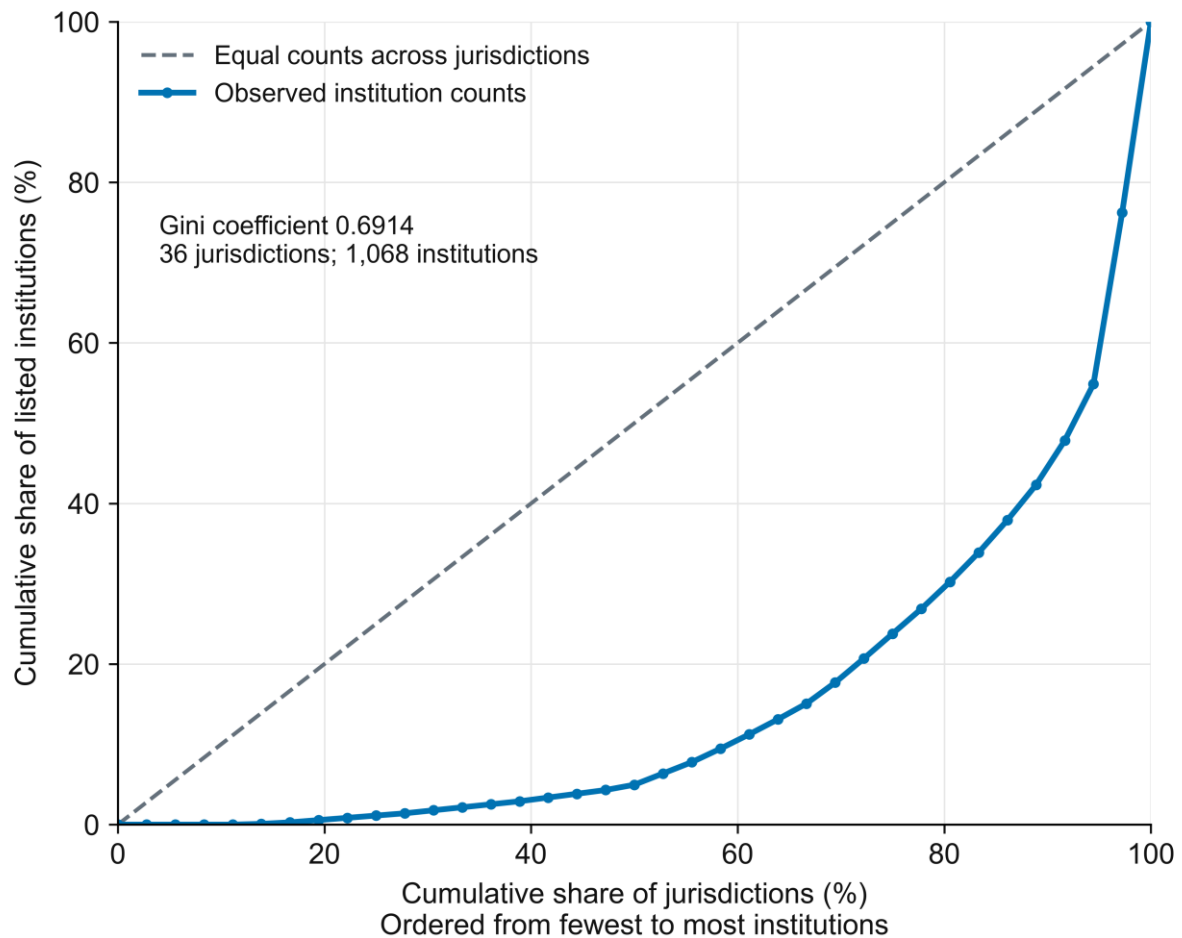


Figure 2. Concentration of listed institutions across jurisdictions. Lorenz curve for institution-code counts across all 36 states and union territories, ordered from the smallest to the largest count. The horizontal axis shows the cumulative share of jurisdictions and the vertical axis the cumulative share of listed institutions. The diagonal represents equal numbers of entries in each jurisdiction, not equal access per person or an equity target. Gini = 0.6914. Source: workforce_lorenz.csv.

Supplementary material

Companion material to BHCL-PUB-2026-09-02 | 2.0

Supplementary data and reproducibility package

The package contains `workforce_institutions_audited.csv` (1,068 rows), `workforce_state_counts.csv` (36 rows), `workforce_lorenz.csv` (37 curve coordinates), `workforce_record_validation.csv`, `workforce_source_headings.csv`, `workforce_label_crosswalk.csv`, `workforce_change_log.csv`, `workforce_metrics.json`, `workforce_source_manifest.json`, `reproduce_workforce.py` and a README with software requirements and field definitions.

The institution-level `source_name_address_excerpt` is a contact-stripped source excerpt, not a guaranteed complete mailing address. The institution code is the primary identifier. Labels preserve source naming, including historical institutional terminology.

The earlier course-approval CSV and its capacity-derived figures are excluded from this version. Reproducing this study does not require them.

Version 2.0 scope statement

This revision replaces the earlier course-and-capacity analysis with a reconstructed institution-geography census. It removes unvalidated programme counts, seat totals, ownership categories, field flags and population-adjusted rates. All geographic counts were regenerated from the original source, all labels were source-reconciled, and new concentration analyses were calculated. These are changes to analytical scope and data reconstruction; they do not assert changes in the real-world institutions between versions.