

# Data Coverage — Findings

## Results

### Data Source Summary

Source	Countries	Year Range	Rows
Employment (ILOSTAT)	208	1969–2025	165,262
Wages (ILOSTAT)	134	1991–2025	126,987
Hours (ILOSTAT)	170	1991–2025	98,814
PWT	183	1950–2019	12,810

### Country Viability

Category	Count
In all four sources	121
CHIP-viable (emp + wages + PWT)	123
Missing wage data	74
Missing PWT data	32

**123 countries** have sufficient data to estimate CHIP. The binding constraint is wage data: ILOSTAT wages cover only 134 countries vs 208 for employment.

### Data Quality Tiers

Tier	Criteria	Countries	Recommendation
Excellent	15+ years	79	Use for all analyses
Good	8–14 years	10	Use for most analyses
Fair	3–7 years	28	Use with caution
Sparse	< 3 years	6	Consider excluding

### Recommended Analysis Range

**2000–2019.** The common year range across all sources is 1991–2019, but wage coverage before 2000 is sparse (few countries reporting). PWT ends at 2019, which caps the upper bound.

## Interpretation

### The Data Landscape

The CHIP estimation depends on the intersection of four data sources, each with different coverage characteristics:

- **Employment** is the broadest source (208 countries, back to 1969), providing a solid foundation for labor market analysis.
- **Wages** are the binding constraint. Only 134 countries report occupation-level wages to ILOSTAT, and many of those have gaps. The 74-country gap between employment and wage coverage represents economies where we know how many people work but not what they earn.
- **PWT** covers 183 countries but ends at 2019. This ceiling means CHIP estimation cannot extend to recent years without an alternative source for capital stock and GDP data.

- **Hours** data (170 countries) is more complete than wages but less than employment, reflecting that hours-worked surveys are more common than wage surveys.

## Quality Distribution

The 79 “excellent” countries (15+ years of employment data within the viable set) form a robust core for CHIP estimation. These include all major OECD economies plus many large developing countries (e.g., ARG, BRA, COL, EGY, IND, IDN, MEX, THA, ZAF).

The 6 “sparse” countries (< 3 years) contribute negligible weight in any aggregation scheme and can be safely excluded without affecting global estimates.

The 28 “fair” countries (3–7 years) are a judgment call: they add geographic breadth but may introduce noise from limited temporal coverage. These are best included in cross-sectional analyses but excluded from time series work.

## Original Study Exclusions

The original study excluded Nepal (NPL), Pakistan (PAK), Sri Lanka (LKA), Bangladesh (BGD), Egypt (EGY), Jordan (JOR), and Palestine (PSE) due to data quality concerns. Our coverage analysis shows:

- BGD, EGY, PAK, NPL, LKA all appear in the “excellent” tier (15+ years of employment data). However, the original exclusions were likely based on *wage* data quality, not employment coverage.
- JOR and PSE appear in the viable set but with potentially thinner wage coverage.

This suggests the original exclusions were conservative — some of these countries may have adequate data for inclusion, but wage quality (unit consistency, occupation coverage) requires case-by-case review beyond what this aggregate coverage analysis captures.

## Implications for Other Studies

1. **Baseline:** Uses all available countries (99 end up in the final estimate after pipeline filtering). The coverage analysis confirms this is appropriate — the viable set is large enough for a robust global average.
2. **Timeseries:** The year-by-year country count varies from 5 (early 2000s) to 64 (2014/2018). Coverage analysis explains why: wage reporting to ILOSTAT expanded dramatically around 2010. Time series results before 2005 are driven by a handful of countries.
3. **Weighting (planned):** With 123 viable countries, there is sufficient diversity to test GDP vs labor-force vs unweighted aggregation. The quality tier analysis suggests that restricting to “excellent” countries (79) for a sensitivity check is a natural robustness test.
4. **Stable panels:** The timeseries study found only 11 countries with 70%+ coverage across 2000–2019. Coverage analysis confirms this is consistent with the overall pattern: most countries have employment data but many lack consistent wage reporting.

## Limitations

1. **Employment-based quality tiers.** The quality tiers count years of *employment* data, not wage data. A country with 20 years of employment data but only 3 years of wage data would be rated “excellent” but would contribute minimally to CHIP estimation. A wage-specific quality tier analysis would be more informative.
2. **No occupation-level granularity.** This analysis counts countries and years but does not assess whether all 9 ISCO occupation groups are reported. A country that reports only 3 of 9 occupations is technically “viable” but produces less reliable wage ratios.

3. **No unit consistency check.** ILOSTAT wages are reported in various units (hourly, monthly, etc.) and currencies. This analysis does not verify that reported wages are in comparable units. The pipeline handles this via filtering to USD wages, but the coverage analysis doesn't reflect the post-filtering reduction.
4. **Static snapshot.** Coverage changes as ILOSTAT updates its database. Re-running this analysis periodically would track improvements in global wage data availability.