03-06-2025

# Beyond Numbers:
## Integrating Geospatial Tools and Small Area Estimates for Effective Program and Coverage Evaluation

Research and Development Solutions

www.glocalevalweek.org

# Objectives

By the end of today, you'll understand:

1. **Why** traditional population estimates are failing your programs
2. **How** building-level geospatial analysis can supplement your M&E
3. **How** Small Area Estimates can give you precision at a fraction of the cost
4. **A real-world model** that's reaching 450,000 people with GPS-verified accuracy

# Quick Poll

# Agenda

We'll move from problem to solution to real-world application - from the "denominator crisis" plaguing Pakistani programs to a digital PHC model that's changing how we think about coverage and cost-effectiveness.

- Part 1 - The Challenge - Pakistan's Populations & The Denominator Crisis
- Part 2 - The Solution - Geospatial Population
- Part 3 - The Innovation - Small Area Estimates for Precision Targeting
- Part 4 - Real Application - Real Application

# Part 1: The Challenge Beyond Official Population Counts

# Pakistan's Data Desert

**District**

Data exists, but not readily available

Also, what if I don't want to implement at district level

**Sub- District**

Estimates unavailable

**Union Council**

"Just a Dream"

**Household**

...

# The Coverage Illusion

You're implementing a health intervention. Which of these programs is a  success?

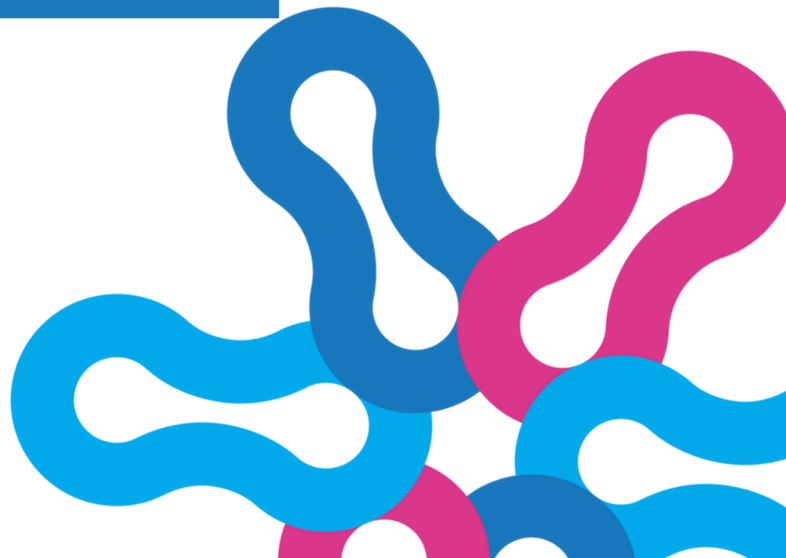
Scenario A: Your program reaches 1,500 people.

Scenario B: Your program reaches 2000 people.

# The Truth Behind the Numbers

| | People Reached | Target Population | Coverage Rate | Conclusion |
|---|---|---|---|---|
| Scenario A | 1,500 | 2500 | 60% | Excellent success! |
| Scenario B | 2000 | 10,000 | 20% | Program failure |

**The Fundamental Question:**
Which scenario represents your current programs?



# The Resource Allocation Trap

**Option 1**
**Spread resources thinly across large areas**
**Result: Minimal impact**

**Option 2**
**Concentrate resources in small areas**
**Result: Uncertain if reaching right people**

# Why Impact Measurement Fails

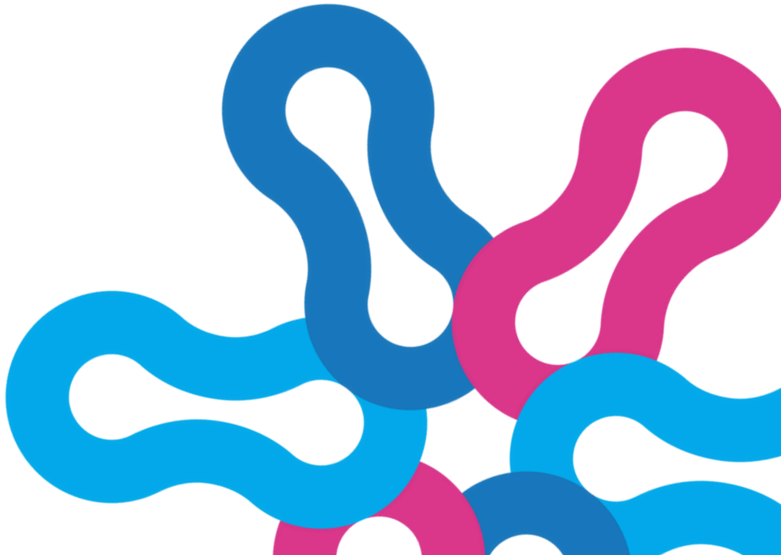
When we don't know our true denominator:

- "High coverage" programs may be reaching tiny populations
- "Low impact" areas may have massive unreached populations
- Cost-per-beneficiary calculations are misleading
- Scaling decisions are based on false premises

## The Intervention Paradox

How many "large-scale" interventions are actually "boutique" programs that aren't "moving the needle" because we never knew where the needle was?





# Real World Example: Orangi Town

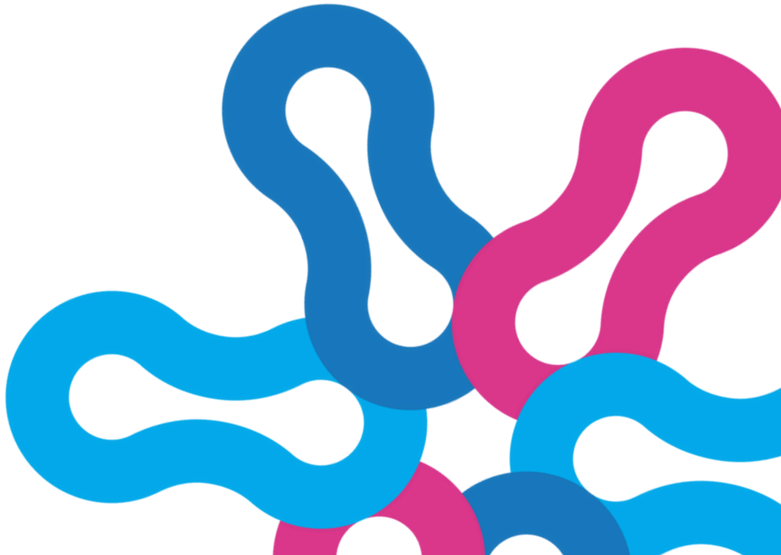
## One of Asia's largest slums

- Official census data: Not available at required resolution
- Population density distribution: Complete mystery
- Program implementation: Must happen here anyway
- Impact measurement: Based on guesswork

Yet this is where programs must be implemented,

where impact must be measured,

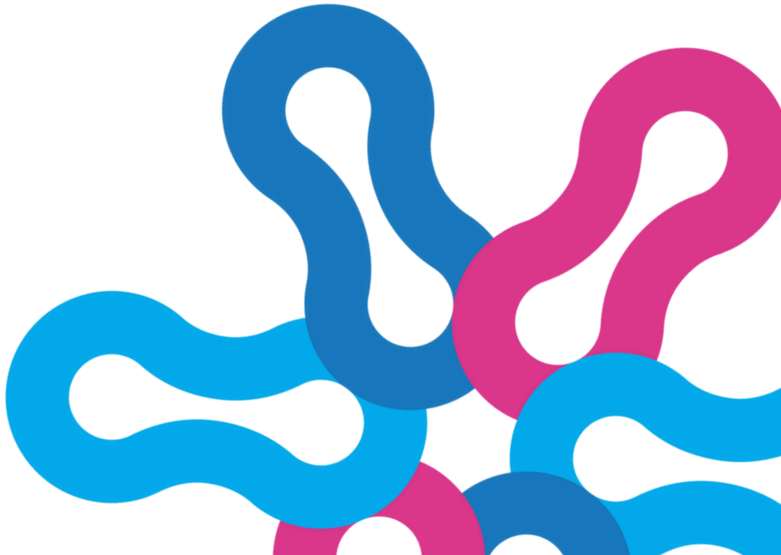
where lives must be changed.



# Building-Level Population Innovation

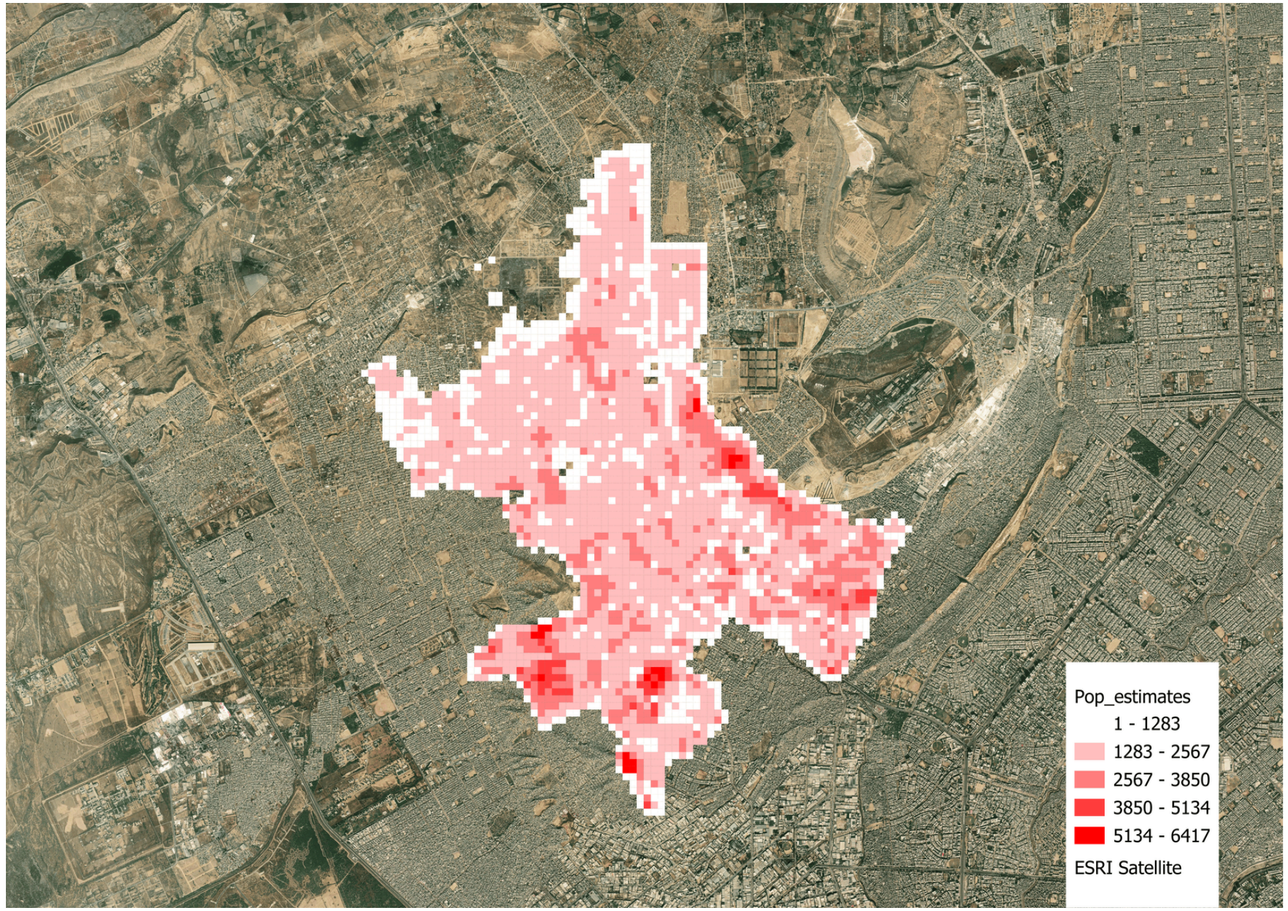
## The RADS Approach: From Rooftops to Reality

1. Buildings First: Using open datasets (Google, Microsoft, OpenStreetMap)
2. Households Second: Height analysis + commercial vs residential classification
3. Population Third: Census ratios applied to actual household counts



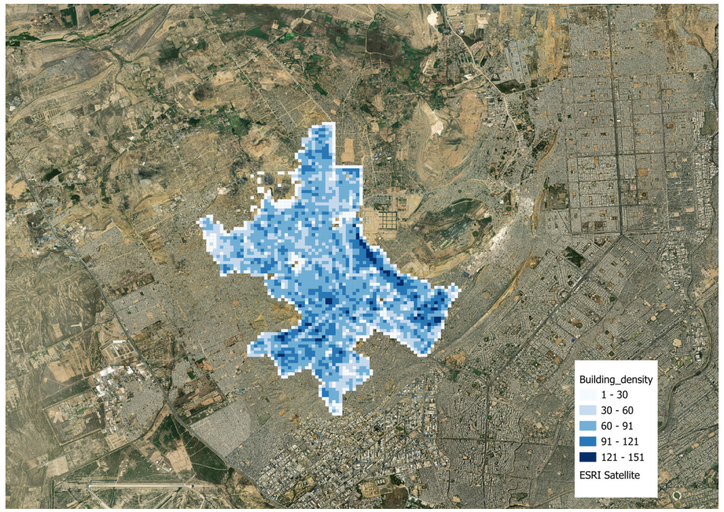
# Building-Level Population Innovation

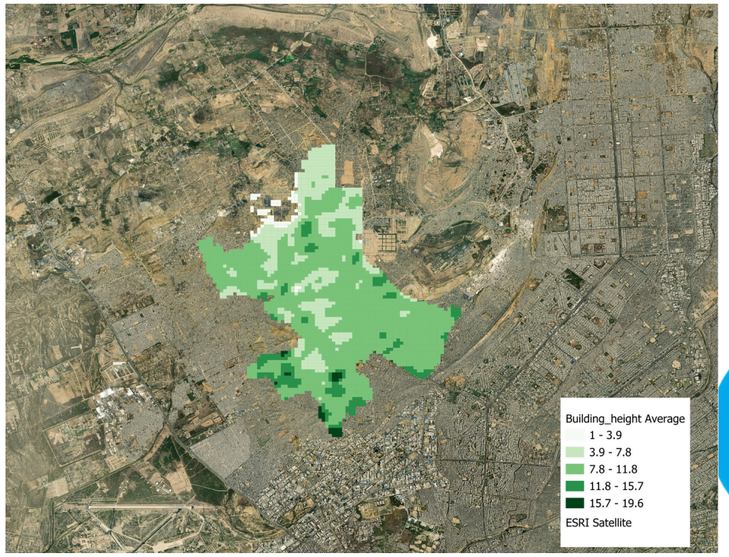
## The RADS Approach: From Rooftops to Reality



# Building-Level Population Innovation

## The RADS Approach: From Rooftops to Reality



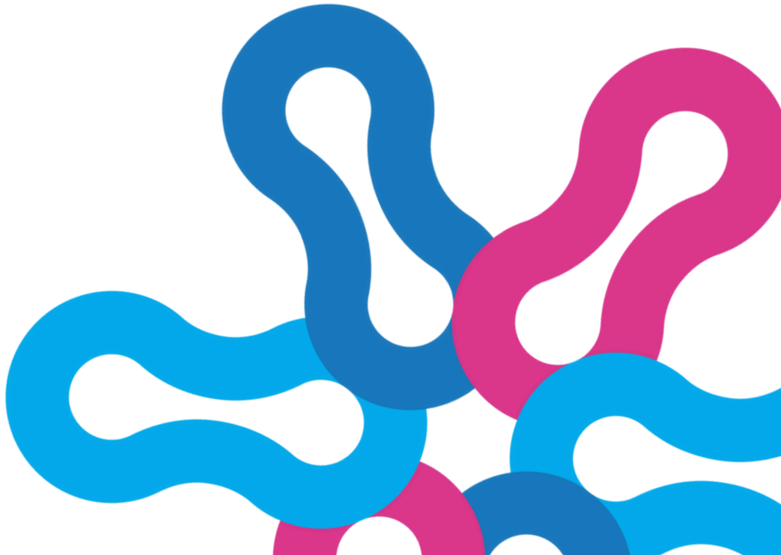



# Part 2: Geospatial Population Estimation Methodology

# The Building-Based Technical Approach

## The RADS Approach: From Rooftops to Reality

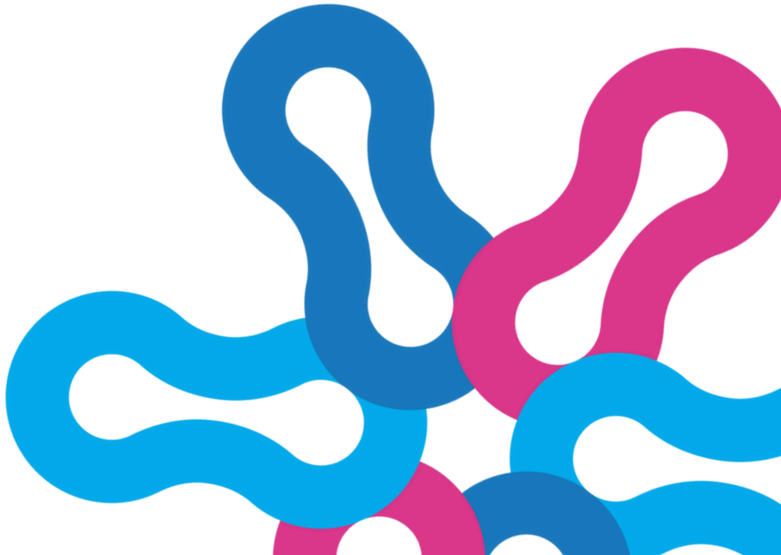
- Step 1: Open Building Dataset
  - Quality Considerations:
    - Data freshness varies by region
    - Confidence thresholds for filtering
    - Completeness assessment methods
- Step 2: Point of Interest (POI) Analysis - Commercial vs Residential Classification:
    - Healthcare facilities, schools, markets, businesses
    - Religious buildings, community centers
    - Transportation hubs, industrial areas
    - Residential density indicators



# The Building-Based Technical Approach

## The RADS Approach: From Rooftops to Reality

- Step 3: Height-Based Classification - Building Height Analysis:
    - Single-story residential patterns
    - Multi-story residential complexes
    - Commercial building identification
    - Mixed-use building handling
- Step 4: Household Density Calculations - Census Ratio Application:
    - Persons per household ratios from latest census
    - Regional variations in household size
    - Urban vs rural adjustment factors
    - Informal settlement considerations



# Why This Matters for M&E:
# True Coverage vs Absent Coverage Rates

| Method | People Reached | Population Estimate | Coverage Rate | Decision |
|---|---|---|---|---|
| **Before Geospatial** | 1,000 | 1,500 | 67% | Scale program |
| **After Geospatial** | 1,000 | 8,000 | 12.5% | Complete redesign needed |

## Real-World Implications

- Budget allocation based on accurate need assessment
- Program design matched to actual population distribution
- Success metrics that reflect true impact

# Integration with Traditional Tools

## Multi-Source Data Fusion:

- Mobile Data Integration:
  - Call detail records for population movement patterns
  - Mobile phone penetration rates for demographic weighting
  - Real-time population fluctuation tracking
- Satellite Imagery Enhancement:
  - High-resolution imagery for building verification
  - Temporal analysis for settlement growth tracking
  - Land use classification for context

# Part 3: Small Area Estimates as Game Changers

# Understanding Small Area Estimation

## What Are Small Area Estimates?

Small Area Estimation (SAE) is a statistical technique that allows us to make reliable estimates for geographic areas or population subgroups where we have little or no survey data by leveraging auxiliary information and statistical modeling.

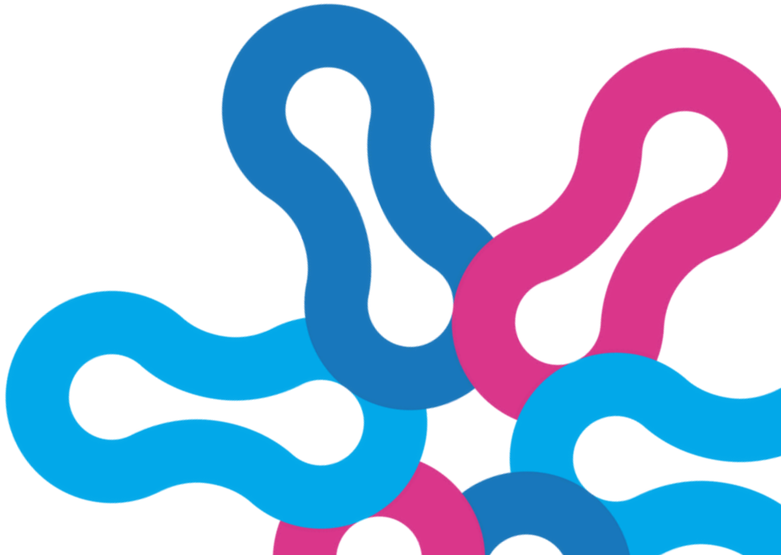
## Think of SAE like "smart guessing" with math backing it up

### The Problem

You want to know malnutrition rates in a specific neighborhood, but only surveyed 5 houses. Not enough to be confident!

### The SAE Solution

"Let me look at similar neighborhoods where I DO have good data, and use that to make a better guess about your neighborhood."



# Understanding Small Area Estimation

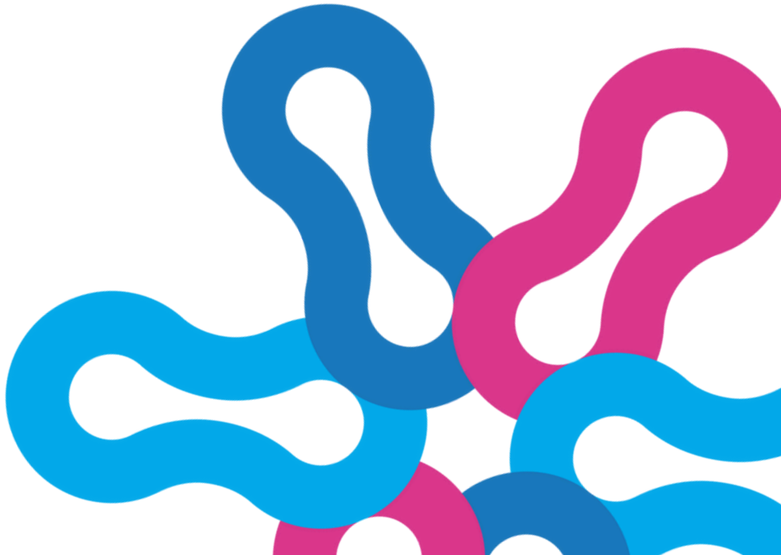
## The Statistical Foundation - "Borrowing Strength"

SAE uses hierarchical statistical models that combine:

- Direct estimates from sample data in each area
- Indirect estimates from related areas with similar characteristics
- Auxiliary variables (census data, satellite imagery, administrative records)
- Random effects to account for area-specific variations

### "Borrowing Strength" Principle

Use information from similar areas with comparable characteristics to strengthen estimates for areas with limited data



# Understanding Small Area Estimation

## Mathematical Framework

Basic SAE Model:

$$\theta_i = X_i\beta + v_i + e_i$$

Where:

- $\theta_i$ = true parameter for area i
- $X_i\beta$ = fixed effects (auxiliary variables)
- $v_i$ = random area effect
- $e_i$ = sampling error

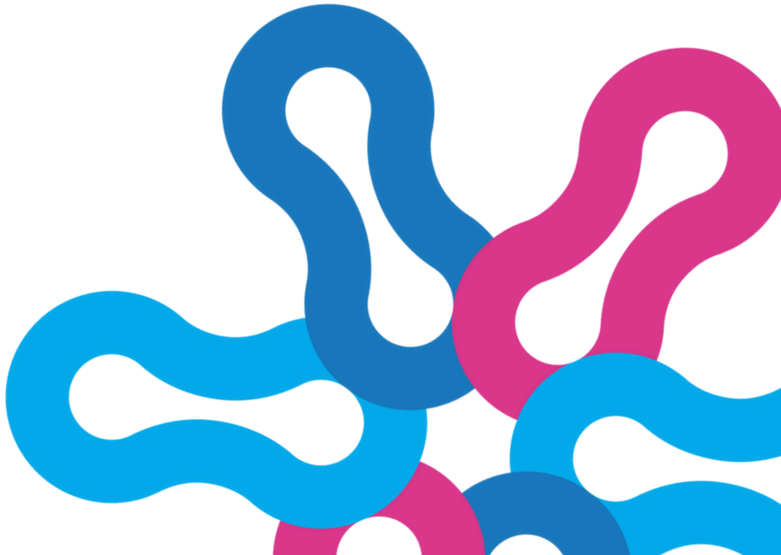
Plain English - **The Magic Formula:**

Your Area Estimate = (Your limited data) +

(What we learned from similar areas) +

(Statistical adjustments for local conditions)





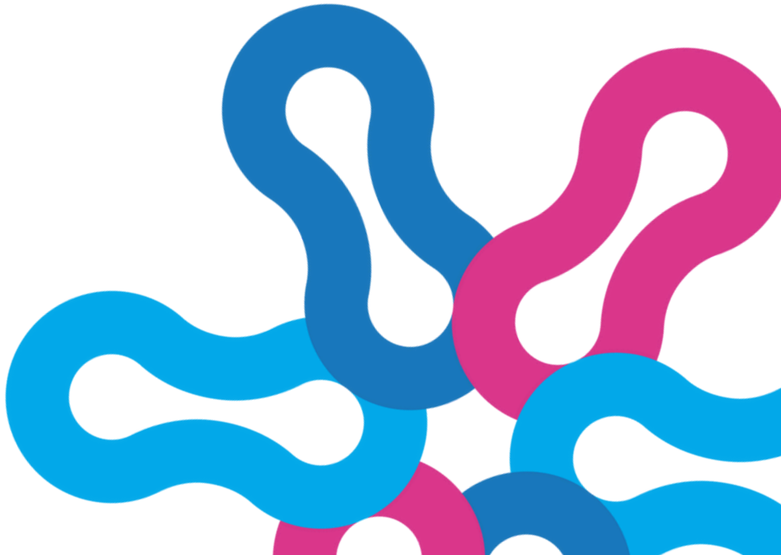
# Understanding Small Area Estimation

## Why This Matters for M&E

| Before SAE | With SAE | M&E Impact |
|---|---|---|
| "We improved health in District X" (2 million people) | "15% improvement in UC-A, 8% in UC-B, 23% in UC-C" (50,000 people each) | Know exactly where to scale |
| "Our program worked somewhere" | "Worked 23% better in areas with X characteristics" | Understand success conditions |
| "Should we scale this?" (We don't know) | "Scale model from UC-A, redesign for UC-B" | Precise scaling roadmap |

### The M&E Revolution

SAE transforms program evaluation from a yes/no question into a detailed roadmap for replication and scaling



# Understanding Small Area Estimation

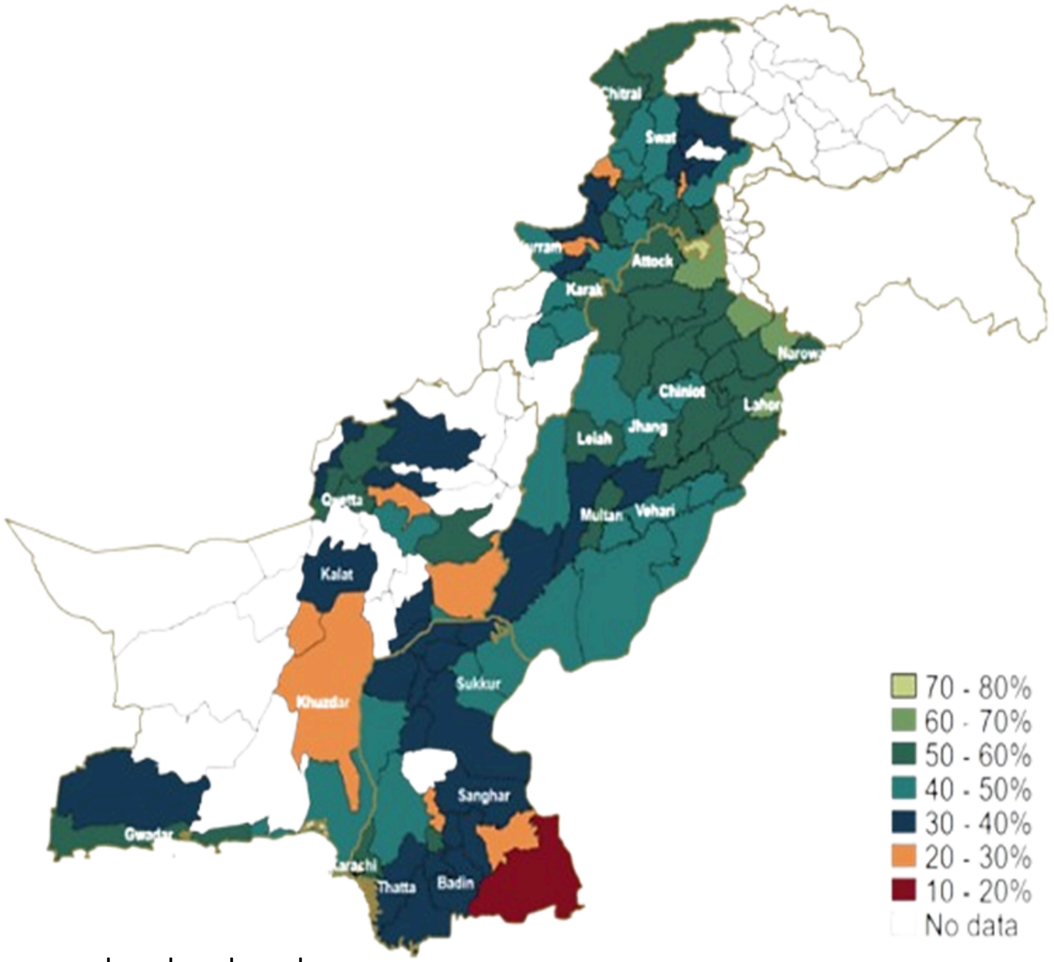
## Why This Matters for M&E

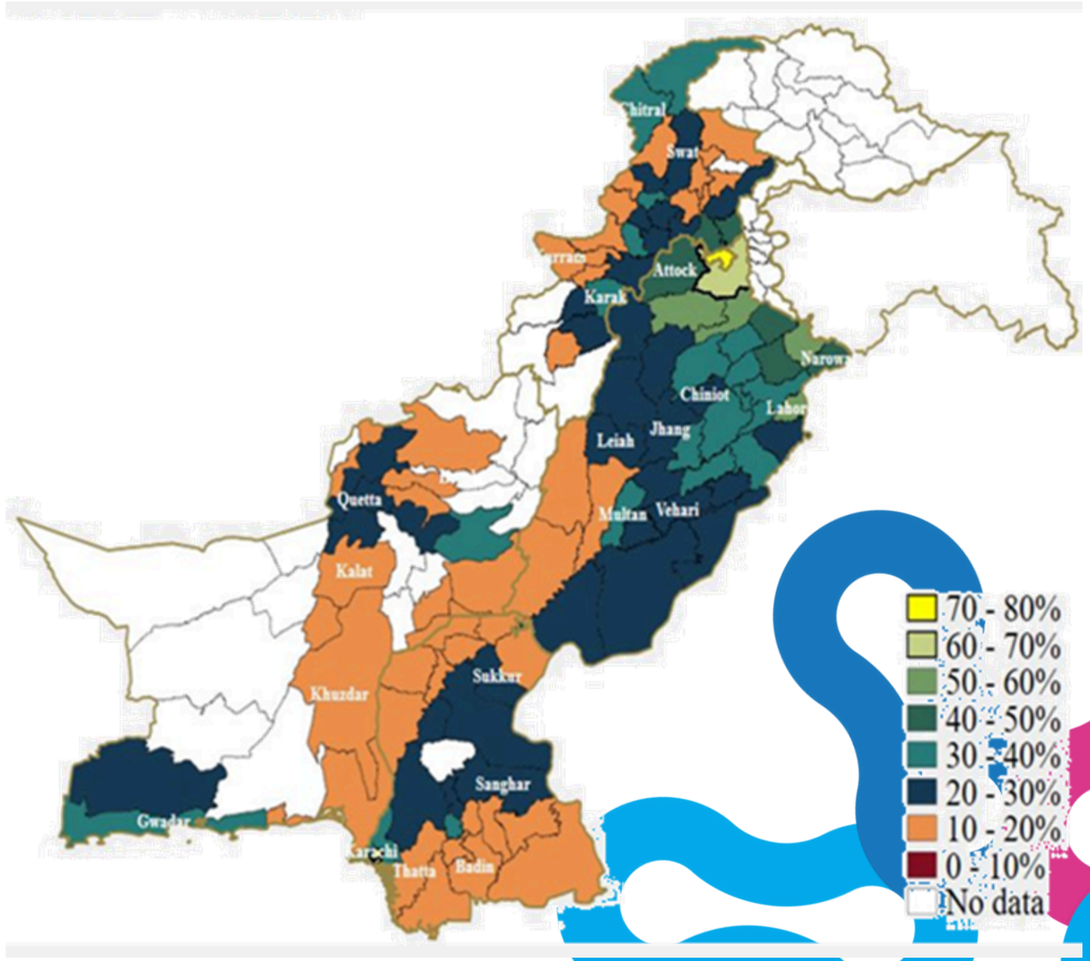
| Old M&E Question | New M&E Answers (with SAE) |
| --- | --- |
| "Did our program work?" | "It worked best when X, Y, Z conditions were present" |
| "Should we scale this?" | "It failed when A, B, C characteristics existed" |
| **"Where should we expand?"** | "Prioritize areas that look like our high-performing locations" |

SAE turns vague program evaluation into precise scaling instructions

# Understanding Small Area Estimation

## Why This Matters for M&E



# Understanding Small Area Estimation

## Technical Example: Pakistan HDI Data

**0.722**

Karachi East HDI

**0.685**

Karachi Central HDI

**0.107**

Tharparkar HDI

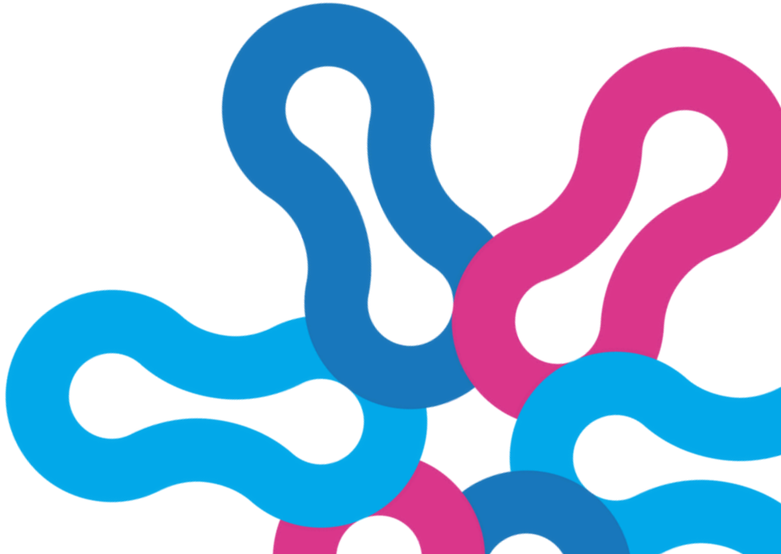
**Hidden**

Massive intra-district variation

**SAE-Enhanced Precision**

Instead of district averages, we get Union Council-level estimates with confidence intervals:

- UC-1: HDI 0.68 ± 0.03 (95% CI)
- UC-2: HDI 0.74 ± 0.04 (95% CI)
- UC-3: HDI 0.71 ± 0.03 (95% CI)



# Understanding Small Area Estimation

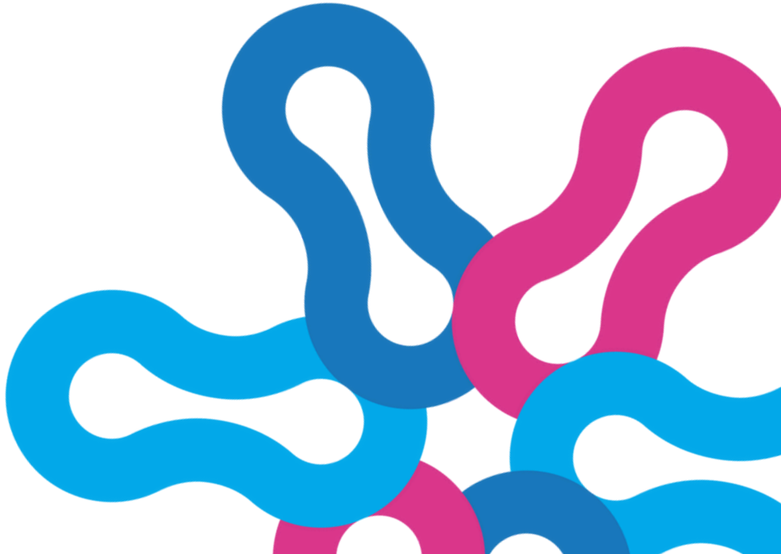
## Pakistan Context - From Methodology to Application

**Current Limitations:**

- District-level HDI ranges from 0.107 (Tharparkar) to 0.722 (Karachi East)
- Coefficient of variation >50% within districts
- Sample sizes insufficient for UC-level direct estimates

**SAE Technical Advantages:**

- Empirical Best Linear Unbiased Predictors (EBLUP): Optimal statistical properties
- Bootstrap MSE estimation: Robust uncertainty quantification
- Model diagnostics: Residual analysis, influence measures
- Benchmarking constraints: Ensures estimates sum to known totals



# Creating Cost-Effective Baselines

## Case Study: Child Malnutrition Estimates in Orangi

| | Sample Size | Cost | Time | Statisitical Power |
|---|---|---|---|---|
| Traditional Survey | 800+ households | $40,000-60,000 | 4-6 months | 80% (5% difference) |
| SAE + Geospatial Approach | **200 households** | **$8,000-12,000** | **6-8 weeks** | **85% (3% difference)** |

# Creating Cost-Effective Baselines

## SAE Technical Implementation

Step 1: Auxiliary Variable Construction

- Building density per hectare (from our enumeration)
- Economic indicators from satellite imagery (night lights, vegetation)
- Distance to health facilities (spatial analysis)
- Market access indices (road network analysis)
- Demographic composition (age structure from census)

# Creating Cost-Effective Baselines

## SAE Technical Implementation

Step 2: Model Specification

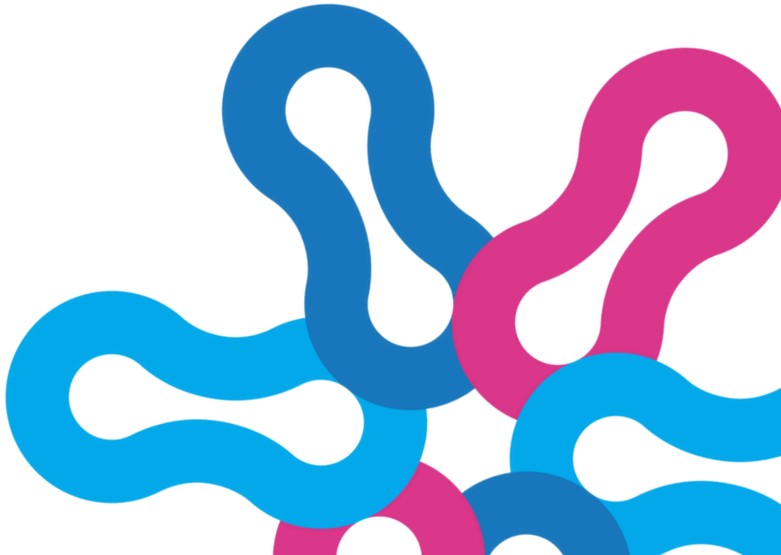
SAE Model Structure:

Level 1: $y_{ij} = \beta_{0j} + \beta_1(\text{building\_density}) + \beta_2(\text{night\_lights}) + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{health\_access}) + u_{0j}$

Where:

- $y_{ij}$ = malnutrition indicator for household i in area j
- Random effects: $u_{0j} \sim N(0, \tau^2)$
- Residuals: $e_{ij} \sim N(0, \sigma^2)$



# Creating Cost-Effective Baselines

## SAE Technical Implementation

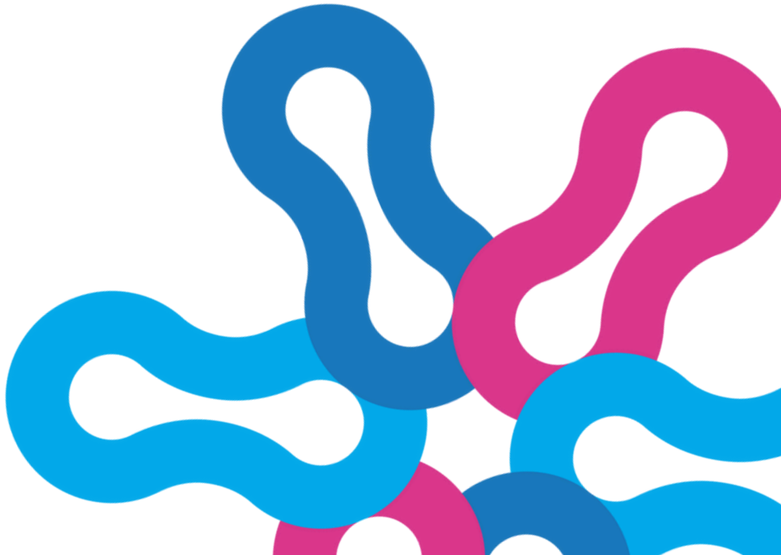
Step 3: Validation Framework

- Cross-validation: Leave-one-out area validation
- External validation: Independent survey comparison
- Goodness-of-fit: $R^2 = 0.73$, Root MSE = 2.1%
- Coverage assessment: 94% of true values within confidence intervals

# Creating Cost-Effective Baselines

## Multilevel Modeling Framework

- **Level 1:** Household characteristics (our GPS-tagged survey data)
- **Level 2:** Building characteristics (height, density, commercial mix)
- **Level 3:** Neighborhood characteristics (infrastructure, services)
- **Level 4:** Union Council characteristics (administrative, economic)



# Part 4: PHC Innovation - Putting It All Together

# The Aapi Model

## Digital System Innovation

What We're Doing: Reaching 450,000 people through systematic house-to-house primary healthcare delivery with complete digital integration.

**Digital-First Approach:**

- Every Household: GPS coordinates, unique identifier, demographic profile
- Every Person: Individual health record, service history, outcome tracking
- Every Interaction: Time-stamped, location-verified, outcome-recorded

Smart Workforce Strategy:

**Low-Literacy Women as Digital Vessels:**

- Community women trained on smartphone-based health protocols
- Voice-guided applications in local languages
- Picture-based diagnostic support
- Automatic data sync and backup

---

**Rahbar System for Men:**

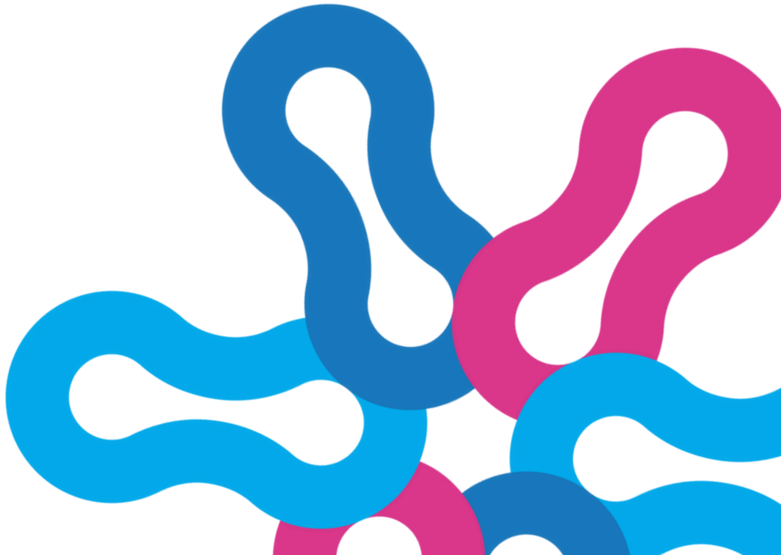
- Male community health workers for cultural appropriateness
- Integrated with women workers for comprehensive coverage
- Specialized training for men's health issues

# The Aapi Model

## Four Arm Testing

- Public Sector
- Private Sector - Algorithm Assisted
- Private Sector
- Telemedicine





# The Complete Integration Model

## Triple Innovation Framework

### Component 1
Building Population Data

- Every household GPS-tagged
- Real-time denominators
- 95% accuracy validated

### Component 2
SAE-Enhanced Targeting

- UC-level precision
- Cost-effective baselines
- Continuous monitoring

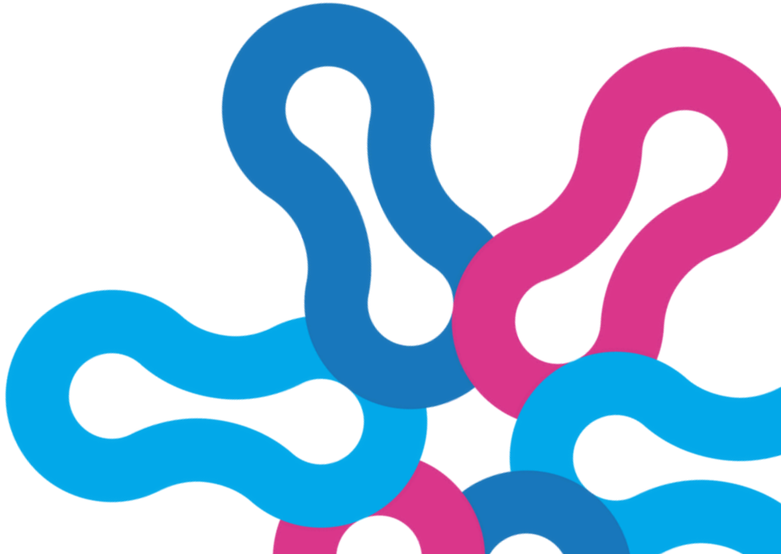
### Component 3
Digital PHC Delivery

- House-to-house reach
- Real-time tracking
- Multiple delivery models

Step 3: Validation Framework

- Cross-validation: Leave-one-out area validation
- External validation: Independent survey comparison
- Goodness-of-fit: $R^2 = 0.73$, Root MSE = 2.1%
- Coverage assessment: 94% of true values within confidence intervals





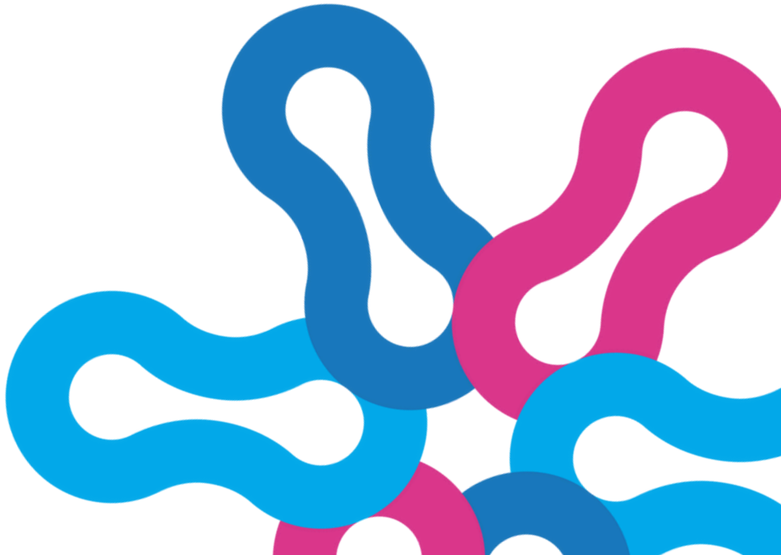
# Impact on M&E

## From Guesswork to GPS-Verified

- **Traditional Program Monitoring:** "We estimate we've reached 60% of the target population"
- **Our GPS-Verified Approach:** "We've reached exactly 267,543 individuals in 52,847 households across 184 Union Councils, with real-time coverage maps showing precise gaps"

Exact Coverage Rates:

- Household-level: "House #47 on Street 12 in UC-23 was visited on March 15th at 2:47 PM"
- Individual-level: "Zainab, age 2, received vaccination Penta #2
- Outcome-level: "Follow-up visit scheduled based on treatment response"



# That’s a Wrap

Key Takeaways

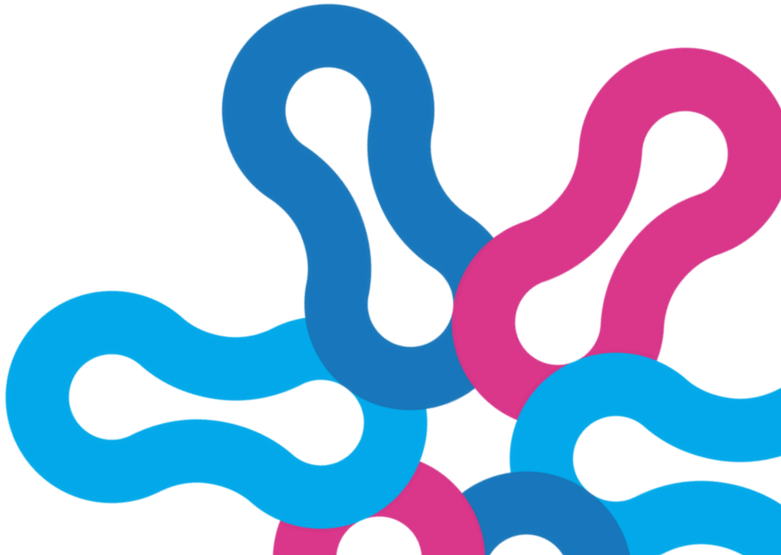
## The Triple Innovation:

- Building-based population estimation - From guesswork to GPS-verified denominators
- Small Area Estimates - From district averages to Union Council precision
- Digital PHC delivery - From coverage estimates to real-time verified reach

## The Paradigm Shift:

- From Estimates to Exactness: We know exactly who we've reached
- From High-Cost to High-Tech: Community workers become digital vessels
- From Assumptions to Analytics: Every decision backed by real-time data

**The Replication Promise**
This isn't just a Pakistan solution - it's a globally adaptable methodology for any context where precise population data and cost-effective service delivery matter.



# That's a Wrap

Reach Out!

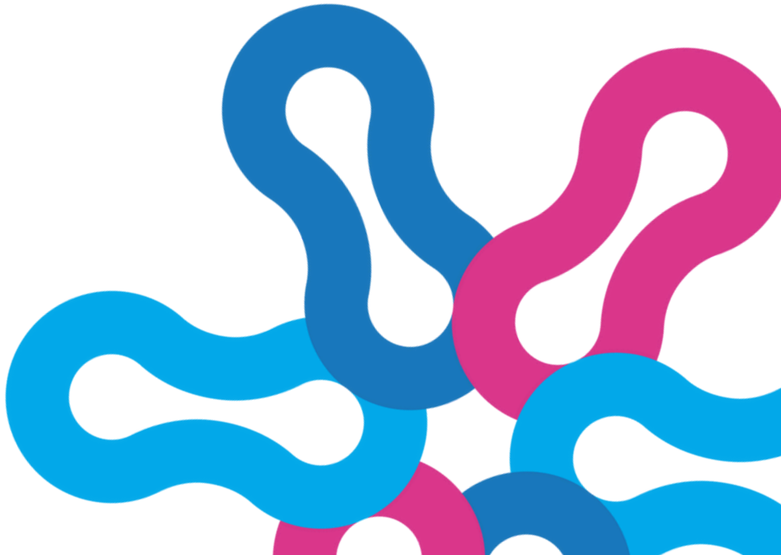
**Research and Development Solutions**

**Email:** info@resdev.org

**Website:** www.resdev.org

Syed Saddam Haider
saddam@resdev.org

Heena Tariq
heena@resdev.org



global **evaluation** initiative

# Thank you!

www.glocalevalweek.org