Supplementary material

Appendix A: Statistics on selected districts in the immunization services unit cost estimation sub-study

Table 1 shows the salient features of the sampled districts. The criteria used for selection was a broad range in the variables: EPI coverage, distance from provincial capitals, and percentage of the district population in urban locations. The districts that are larger and predominantly urban (≥ 40% of the district population) are labelled as "Large" and generally have high slum populations, while the districts labelled "Other" are predominantly rural, situated away from provincial capitals and have lower EPI coverage rates.

Table 1: population, urban share and selection criteria of sampled districts.

Province and Sampled Districts	FIC 2014-15 (% children 12-24 months)	Population in 2017 Census (N)	Urban Population (%)	Selection criteria
Punjab (PB)				
Lahore	89.45	11,119,985	100.00	Provincial capital
Rawalpindi	92.44	5,402,380	55.64	Large district
Multan	91.52	4,746,166	43.38	Large district
Khushab	86.55	1,280,372	27.57	Other district
Layyah	23.65	1,823,995	17.61	Other district
Sindh (SD)				
Karachi City	63.66	16,024,894	92.88	Provincial capital
Sukkur	79.59	1,488,372	48.43	Large district
Kashmore	73.35	1,090,336	23.26	Other district
Umer Kot	7.51	1,073,469	22.69	Other district
Khyber Pakhtunkhwa (KP)				
Peshawar	94.80	4,267,198	46.16	Provincial capital
Swat	88.81	2,308,624	30.14	Other district
Kohat	83.53	992,427	27.20	Other district
Mansehra	67.48	1,555,742	9.31	Other district
Balochistan (BA)				
Quetta	64.63	2,269,473	44.04	Provincial capital
Khuzdar	60.53	798,896	34.52	Other district
Pishin	49.46	736,903	19.38	Other district
Overall sample	**Average 76.1**	**N = 56,979,232**	**Average 66.7**	

Appendix B: Estimating Pakistan's under-5 sub-populations at the district-level by geo-type

Understanding the breakdown of Pakistan's under-5 population into target sub-groups at the district-level is foundational to determining the most effective means for delivering critical interventions. Sub-populations living in rural areas, urban areas, slums, and fragile & mobile groups were all factored in. Urban areas are defined by administrative criterion coming under the municipal corporations, town committees and cantonments. Rural areas are defined as revenue village (also called revenue mauza), which is the smallest unit of revenue collection and administration in Pakistan as provided by union council level micro maps. Slums are based on the UN-Habitat definition for slums and are more commonly known as Katchi Abadis (squatter settlements). Their key feature is lack of security of tenure and location on marginalized lands with minimum amenities managed by the community itself. Fragile and mobile groups are defined as people living in areas dominated by internally displaced, mobile and/or conflict-affected communities. Mostly, these areas are now part of KP province.

The data used to estimate under-5 sub-populations living in rural areas, urban areas, slums, fragile & mobile groups is summarized in Table 2.

Table 2: data used in the micro population estimation study.

Data	Source	Data type
Population under 5	WorldPop	Rasters
Pakistan's district admin boundaries	GADM	Polygons (.shp)
Urban & rural extent	GRUMP	Polygons (.shp)
Slums	Mansueto Institute	Polygons (geoJSON)
Mobile and fragile population	High Risk Mobile Population Working (HRMP) group	Excel table
Health facilities	Humanitarian Data Exchange (HDx)	Points (.shp)
	Open Street Map (OSM)	Points (.shp)

The micro population estimation was performed according to the following steps:

1. ***Raster algebra***: Using R, we performed local raster algebra on individual age- and sex-specific population rasters to produce one single under-5 population raster.
2. ***Prepare slum data***: Using R, we used the subset of slum polygons as those with a K-index score of 6 or higher (this is the cut-off recommended by the Mansueto Institute that consistently characterizes a slum)
3. ***Reconcile slums and urban areas***: Using ArcGIS desktop, we clipped overlapping area between urban extent polygons and slum polygons such that overlap was removed from urban polygons to prevent counting the same people twice.
4. ***Calculate total under-5 population***: Using QGIS, we extracted population sums from under-5 population raster to district administrative boundaries to produce total under-5 population counts by district
5. ***Calculate under-5 population in urban areas and slums***: Using R, we created masked population rasters for slums and urban areas then extracted population sums from urban and slum masks and added as attributes to district administrative boundaries layer. Masking

the population raster sets all raster values that fall outside the polygons of interest to NA, so when extracting values to districts, only the cells with values are tallied.

6. ***Calculate under-5 population in rural areas***: Using excel, we calculated the rural population as the difference between the total population and the sum of urban and slum populations.
7. ***Prepare mobile & fragile groups data***: Using R, prepared mobile & fragile data for joining to table of other sub-population groups by geographic identifiers of district. To do this, we created a geo key to map HRMP geographies to GADM official district boundaries.
8. ***Join & reconcile***: Using R, we joined mobile and fragile population estimates to the main table and removed mobile and fragile population evenly from 3 other groups (conditional upon each group having at least [1 + mobile and fragile]/[number of groups meeting conditions]) and reconciled total population.

The faceted maps of Pakistan's under-5 sub-populations are shown in Figure 1.

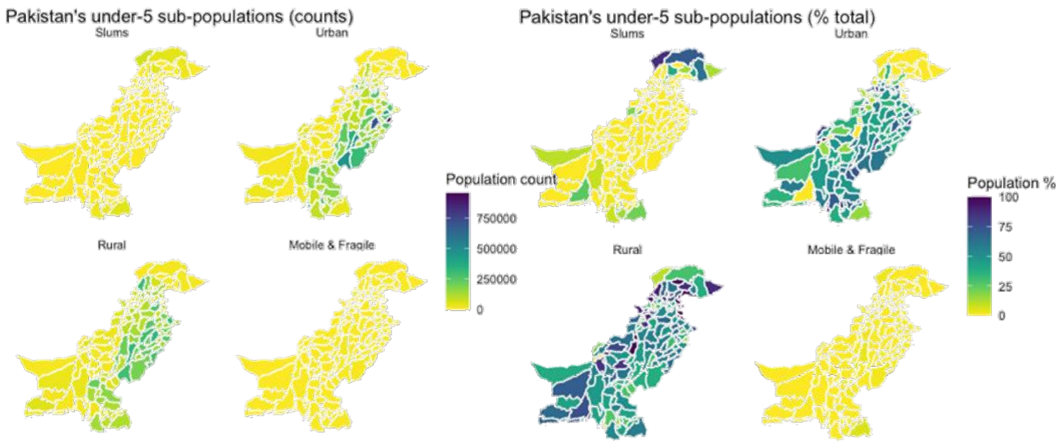


Figure 1: faceted maps of Pakistan's under-5 sub-populations count (left) and percent of the total (right).

Appendix C: Extrapolation of unit cost estimates to 113 districts

Primary technical and allocative efficiency analyses were conducted for the 16 districts for which primary unit cost data were collected. However, in order to inform provincial and national allocative efficiency assessments, unit costs were inferred for the remaining 113 districts using a generalized linear regression model with a gamma distribution and a log link function. The following predictive covariates were used: percentage of population which is urban, population size, health expenditure per capita, and average wage per capita. It was found that the unit cost was most strongly correlated with percent urban followed by average wage. There was also a strong correlation between percent urban and average wage, health spending and population size. For these reasons, the regression model included interaction terms between the variables and percent urban; and was calculated in R using the formula:

```
glm(unit_cost ~ delivery_modality + urban_population + population_size + health_expenditure_pc + average_wage_pc + urban_population*population_size + urban_population*health_expenditure_pc + urban_population*average_wage_pc, family=Gamma(link="log"), data = data)
```

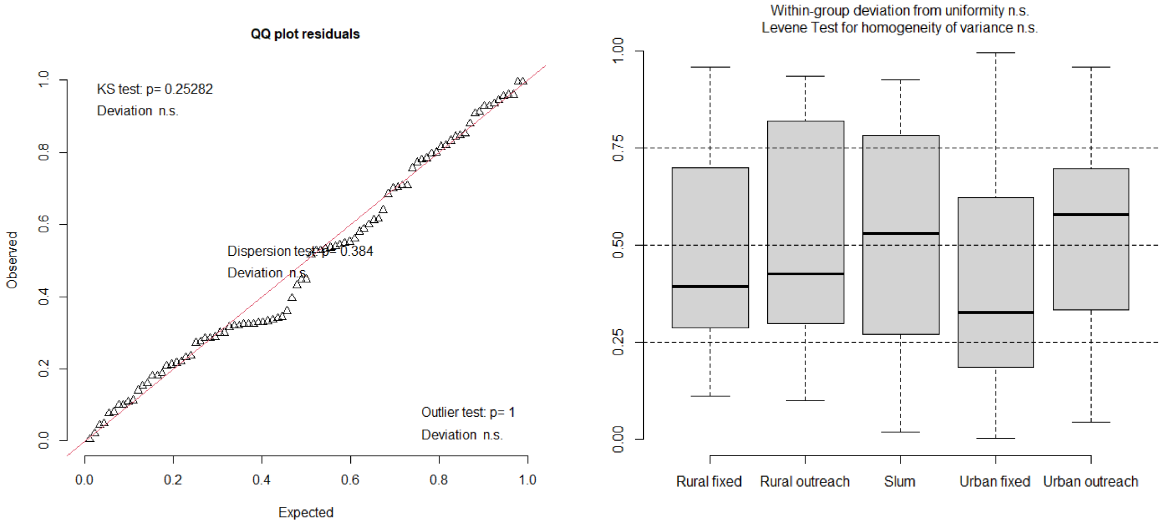
giving the following coefficients, standard errors, t- and p-values:

	Coefficient	Std. Error	t-value	p-value
urban_population	-3.45E+00	1.57E+00	-2.206	0.030311
population_size	-4.42E-07	7.90E-08	-5.587	3.16E-07
health_expenditure_pc	3.23E-04	7.84E-05	4.125	9.09E-05
average_wage_pc	-5.12E-05	5.84E-05	-0.877	0.383121
urban_population: population_size	3.15E-07	9.06E-08	3.482	0.000814
urban_population: health_expenditure_pc	-3.52E-04	8.70E-05	-4.042	0.000122
urban_population: average_wage_pc	3.53E-04	1.27E-04	2.783	0.006736

The R package 'DHARMa' was used to run goodness-of-fit tests on the scaled residuals (Figure 2a). The uniform QQ-plot shows a correct overall residual distribution, and the residuals box plot shows an insignificant deviation from the 0.25, 0.5 and 0.75 quantiles, as confirmed by the Kolmogorov–Smirnov test and Levene's test for homogeneity of variances.

Further tests were carried out to validate the model predictions by comparing the median, minimum, maximum, first quartile and third quartile with the sampled data. The extrapolated data relative to the sampled data can be seen in Figure 2b. The variance of the extrapolated data decreases when compared to the sample, but the mean for each modality remains close to the same (with a maximum variation of 7% in rural areas and 4% in urban modalities).

(a)



(c)

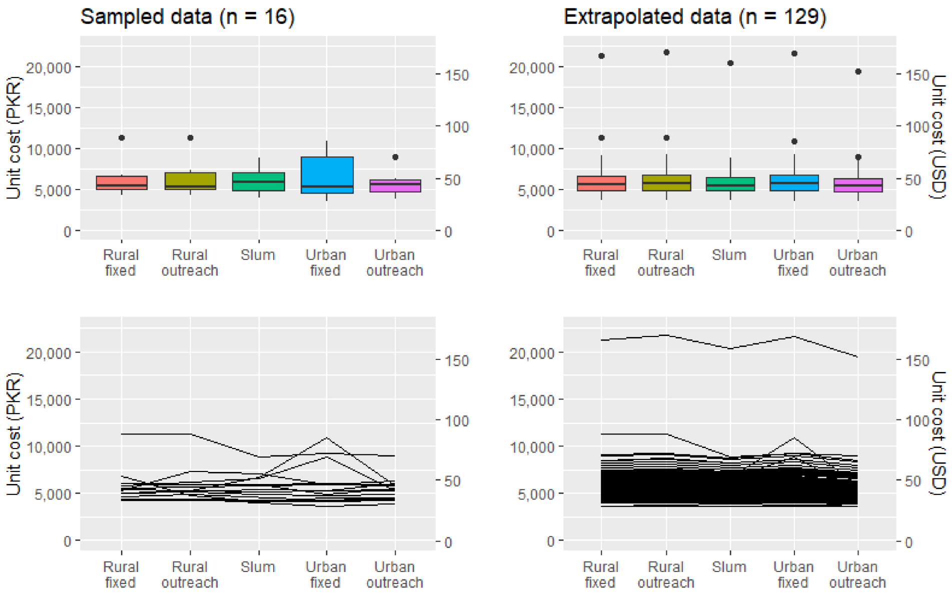


Figure 2: (a) residual QQ-plot (left) and residual plot (right) of regression model, (b) plots showing extrapolated unit costs (n = 129, right) compared with the sampled data (n = 16, right).

Appendix D: Allocative efficiency optimization – model schematic

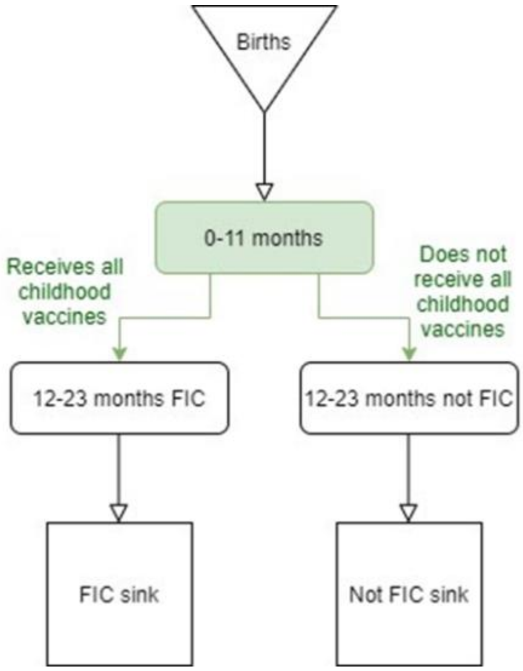


Figure 3: schematic of the compartment-based optimization model. The green compartment can be targeted by the vaccination modalities.

Appendix E: Cost components for potential efficiencies

Figure 4 shows the cost components of the current spending and the inefficient spending as computed by the DEA.

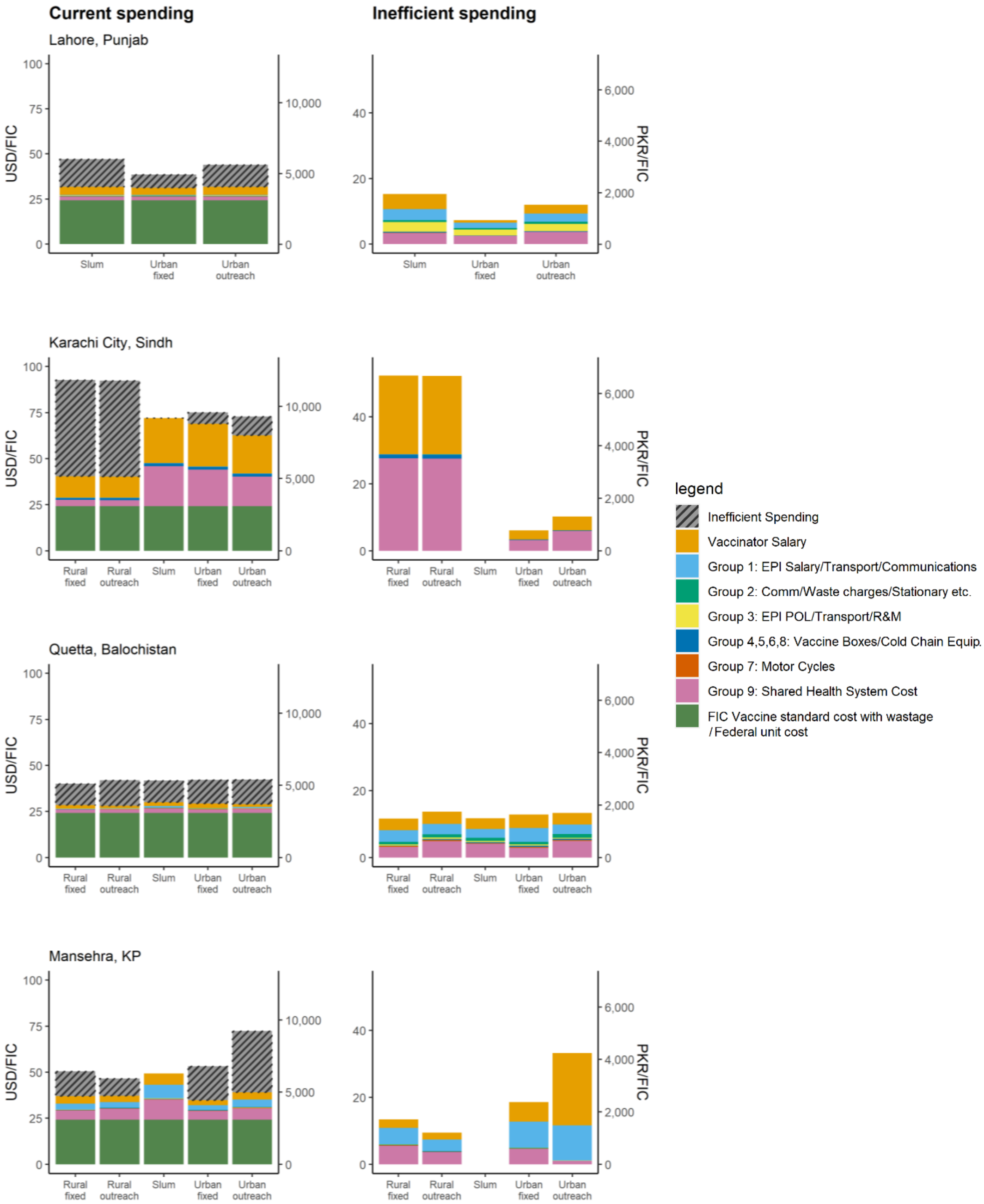


Figure 4: cost components in the exemplar districts of Lahore, Karachi City, Quetta and Mansehra. The graphs on the left show the cost components of the current spending with inefficient spending calculated in the DEA, while the right-hand side graphs show the cost components of the inefficient spending.

Appendix F: Maximizing equity or coverage with incremental budgeting in Lahore, Karachi City, Quetta and Mansehra

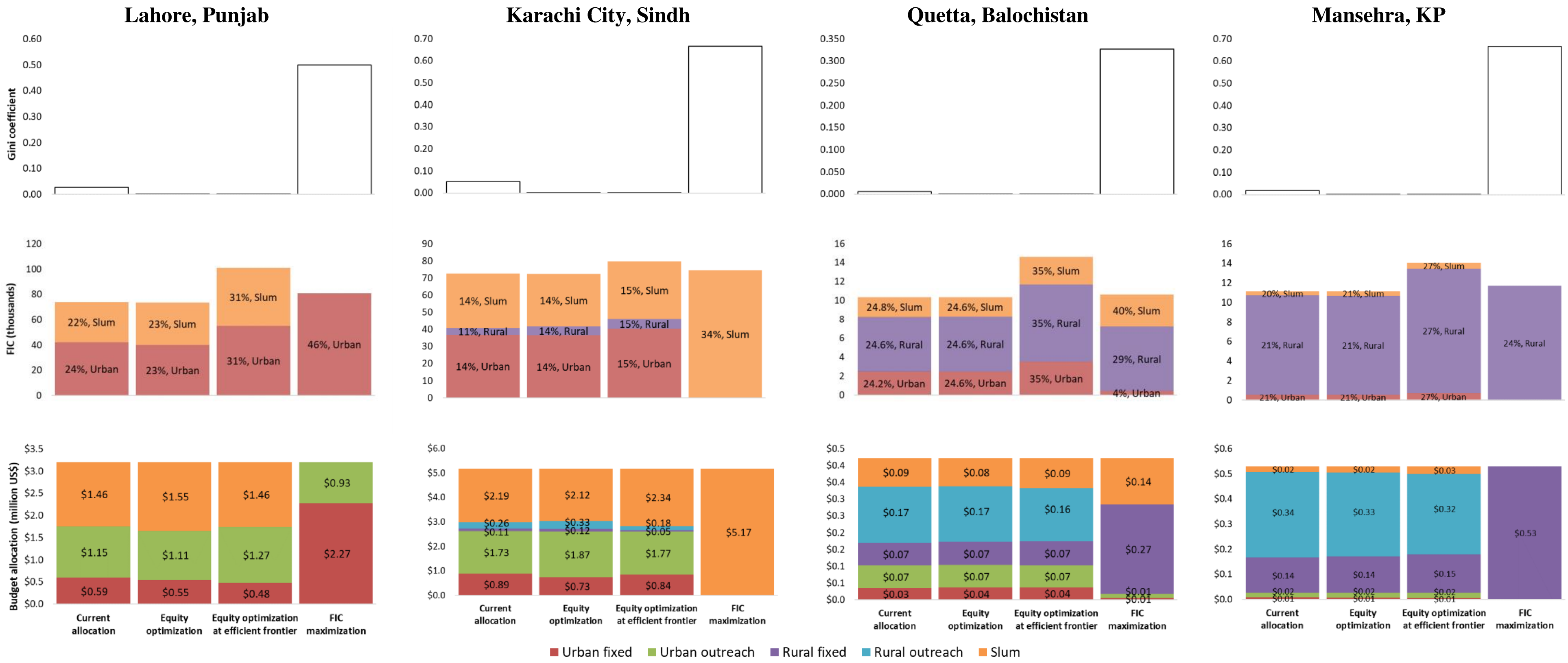


Figure 5: optimization outcomes to maximize equity and FIC at concurrent and optimal unit costs in the exemplar districts.

Top row: Gini coefficient as a function of health spending per child.

Middle row: FIC as a function of health spending per child divided by sub-population.

Bottom row: budget allocation as a function of health spending per child. The y-axis shows the total health expenditure.

Appendix G: Optimized operational planning: example for Quetta, Balochistan

An example of how operational planning can be optimized for equity is shown in Table 3 and Figure 5 for Quetta.

Table 3: example of how operational planning can be optimized for equity for Quetta.

	Current	Optimal (at $ per child)		
	~$20 per child	$10	$20	$30
Total number of vaccinators	152	74	147	222
Rural fixed	12	14	27	40
Rural outreach	27	27	55	82
Slum	42	15	29	44
Urban fixed	45	8	14	24
Urban outreach	26	11	23	32
Number of motorcycles	76	39	78	118
Vaccine boxes / cold chain equipment	250	115	230	346

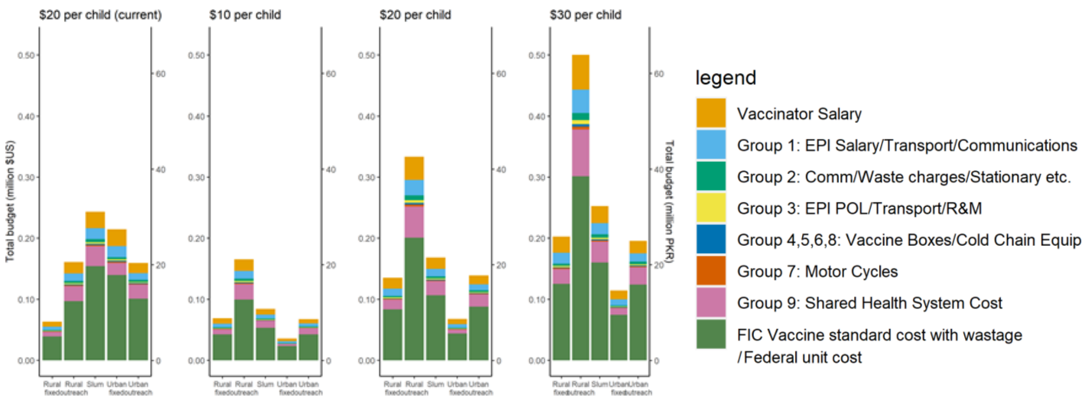
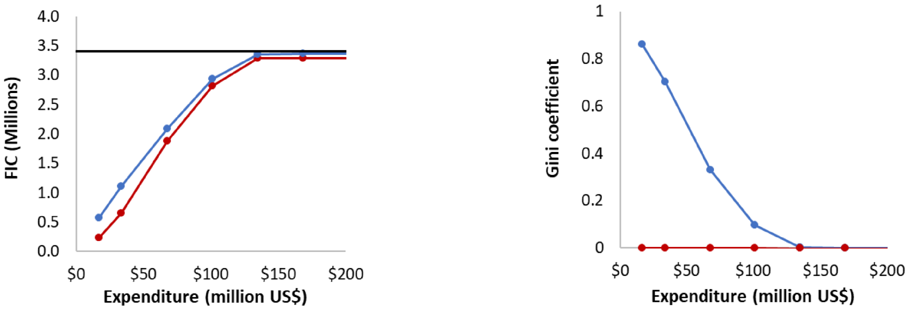


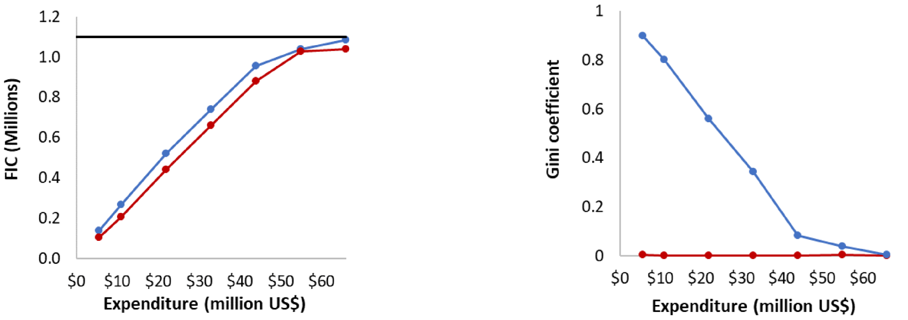
Figure 5: distribution of costs by component and service delivery modality and with increased resourcing in Quetta.

Appendix H: Geospatial optimization across districts

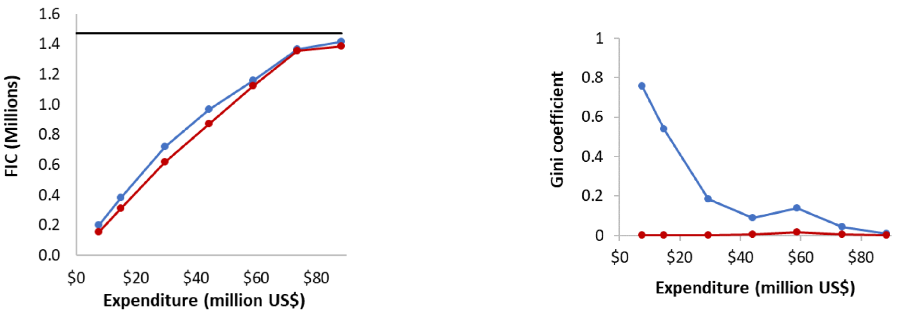
Punjab



Khyber Pakhtunkhwa



Sindh



Balochistan

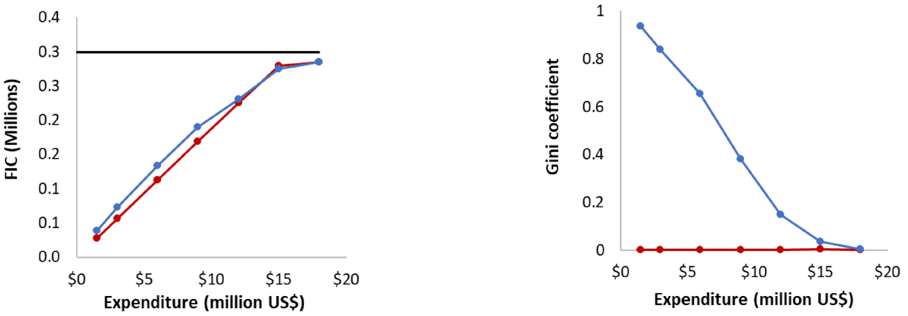


Figure 6: FIC and equity outcomes of the geospatial optimizations in all four provinces.

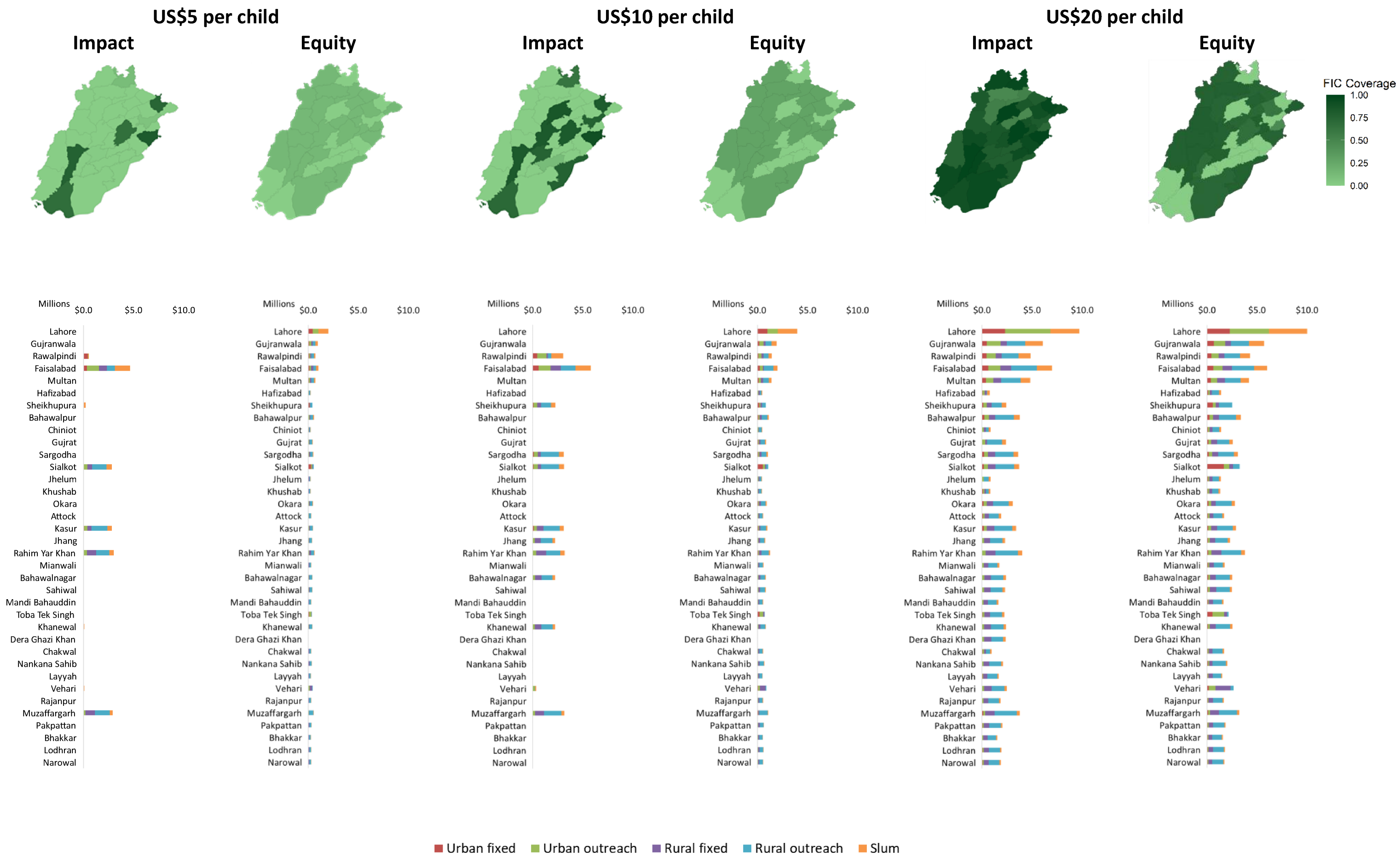


Figure 7: coverage maps and district-wise budget reallocations of the geospatial optimizations for Punjab at three budget levels.

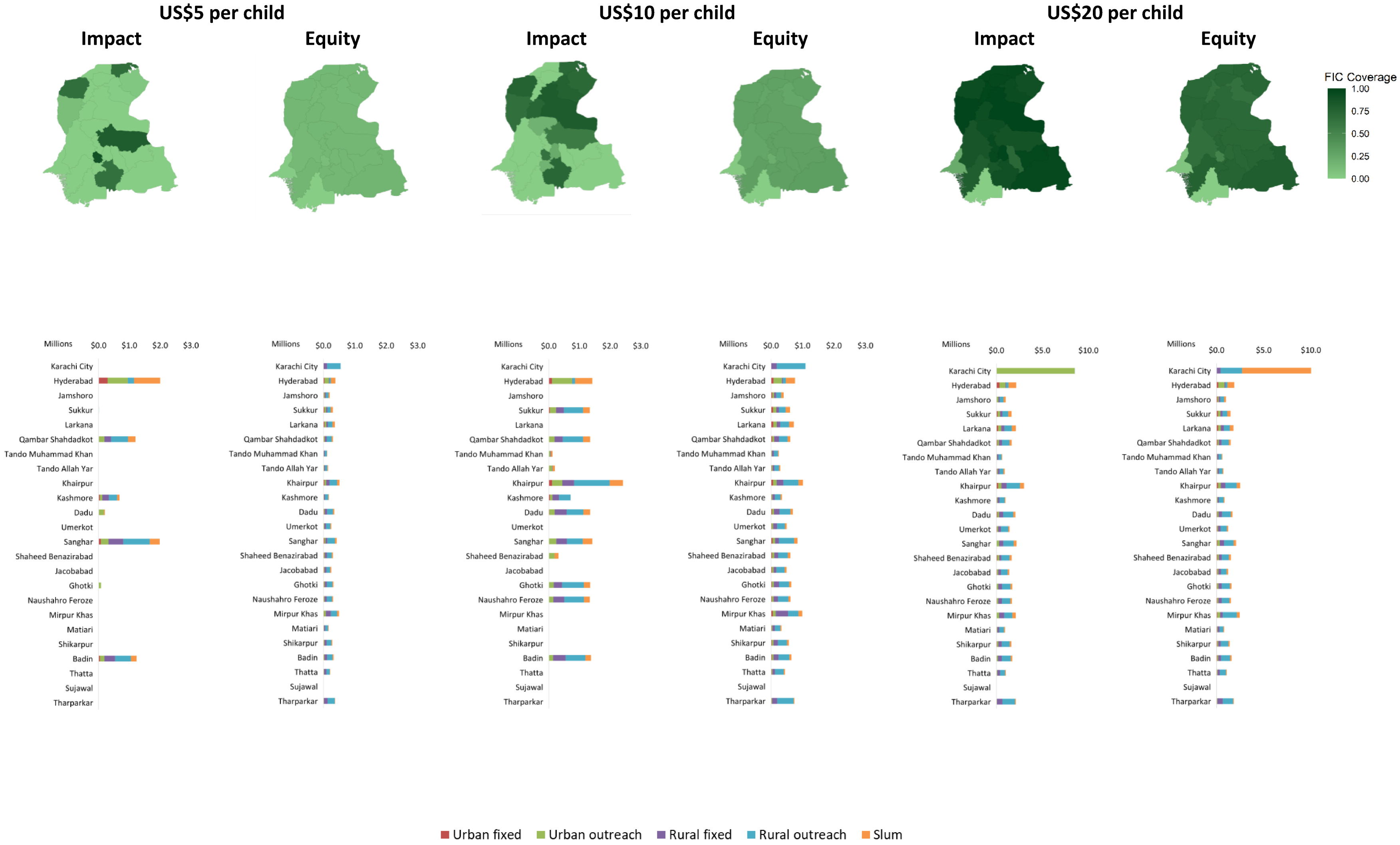


Figure 8: coverage maps and district-wise budget reallocations of the geospatial optimizations for Sindh at three budget levels.

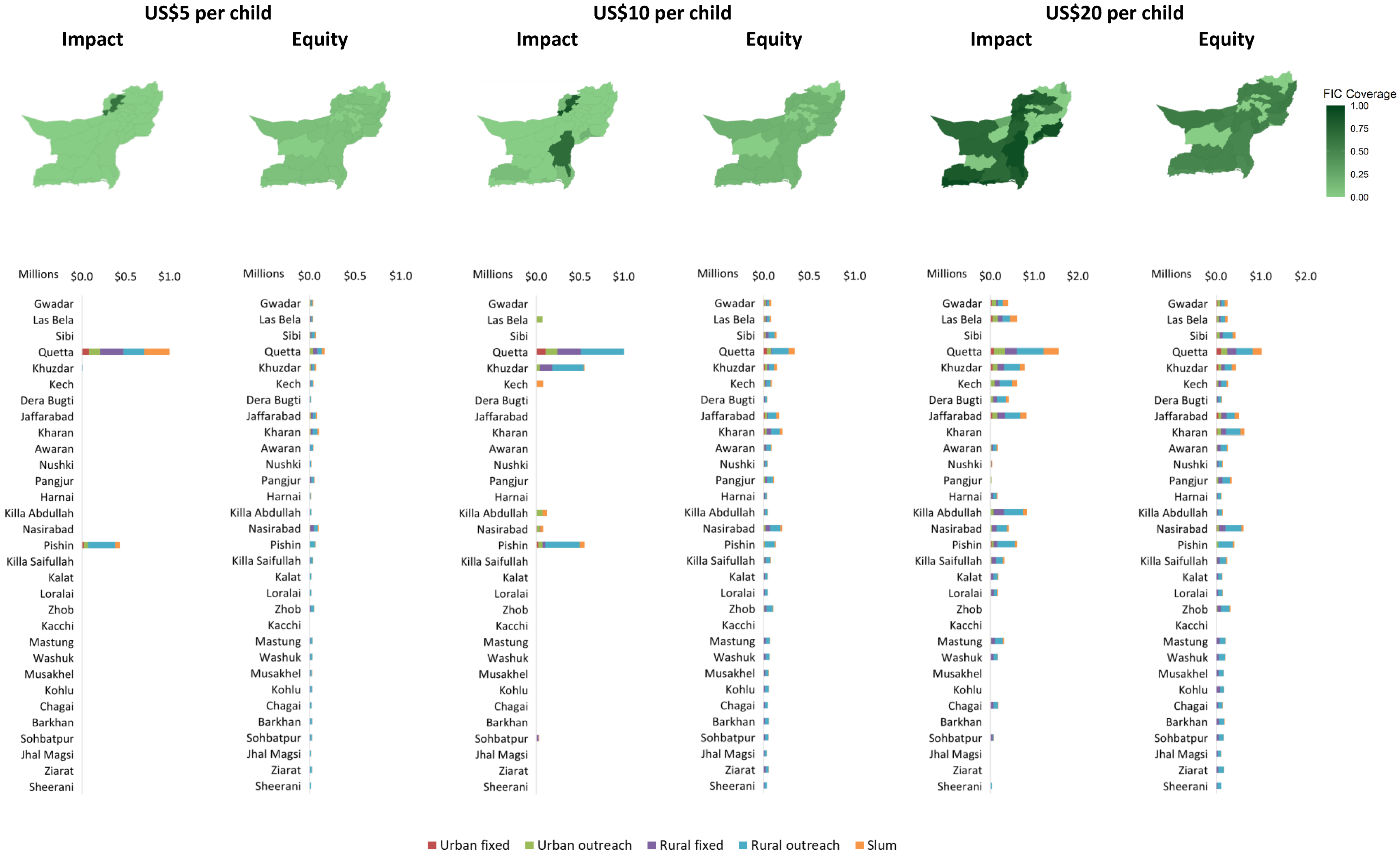


Figure 9: coverage maps and district-wise budget reallocations of the geospatial optimizations for Balochistan at three budget levels.