COST-EFFICIENCY CONSIDERATIONS FOR THE COMPLETION OF ROAD NETWORKS IN OPENSTREETMAP: A PRIORITY-BASED MAPPING APPROACH APPLIED IN PAKISTAN

Robert Banick, Manish Basnet, Lander SMM Bosch, and Moritz Meyer[1]

THE CASE FOR HIGH-RESOLUTION ACCESSIBILITY MODELLING

Limited access to services and opportunities is systematically linked to poorer development outcomes. In Pakistan, 22.8 million children between the ages of 5 and 16 – 44% of the total demographic – do not attend primary or secondary education, numbers which rise sharply with age. Distance to school and a lack of provision are two of the main reasons for not attending school in rural areas of Pakistan. Similarly, a shortage of and long distances to health facilities hinder the access of primarily rural and poor households to these critical services. This is compounded by a dearth of transportation options with poorly maintained roads, affected by unfavourable weather and unsafe driving conditions. Responses to the Pakistan Social And Living Standards Measurement Survey indicate that more than half of the households in most Pakistani districts have to travel over 2 kilometres to reach a health clinic or hospital, entailing unequal health outcomes.

Detailed spatial knowledge of disparities in accessibility to these services is crucial to designing targeted and cost-effective policies, investments, and projects to address them. While the existence of disparities is well-known and acknowledged, they are rarely measured at the administrative levels where service provision happens and investment decisions are made, hampering efficient interventions and resource allocation. The World Bank's Pakistan Poverty & Equity team, together with the Pakistan Transport team, therefore refined and applied a high-resolution method to measure and visualize accessibility disparities to services at the level of *tehsils* (third-level administrative units) in Pakistan. This accessibility model was applied to access to schools, healthcare facilities, and markets in Khyber Pakhtunkhwa (KP) Province.

WORLD BANK GROUP
Poverty & Equity

The methodology underlying the accessibility modelling is described in a recent poverty note and can be replicated in other contexts for access to any type of service, opportunity, or other points of interest through the toolkit and material available in the GitHub repository developed by the team.

ROAD NETWORKS AND POINT DATA FOR SERVICES: THE FOUNDATION OF ACCESSIBILITY MODELS

Accessibility is inherently multidimensional and requires a broad set of data inputs. At its core are two types of inputs: road network data, and geocoded point data of the location of people, services, opportunities, or other places of interest. The former is needed to calculate least-cost paths between settlements and points of interest and, together with information on the terrain, enable the estimation of speeds along the road network, enabling the computation of travel time accessibility. The latter pinpoint where people live, and where they can access services and opportunities that will benefit their development outcomes. Data on roads and the location of points of interest are usually collected by government partners and development peers. Where data gaps exist or these data cannot be shared, public data, such as OpenStreetMap (OSM) data, can be used in addition to manual data collection. This use of open and public OSM data aligns with the objective of the 2021 World Development Report to bolster and boost the potential of the fast-changing changing data landscape to improve the lives of poor people. The work feeds into the broader World Bank efforts to invest in open geospatial data for analytics including, but not limited to, the study of spatial patterns of accessibility, population modelling, development outcomes, risk assessments, and climate forecasting.

## DATA QUALITY ISSUES AND THE NEED FOR ROAD NETWORK IMPROVEMENT

The completeness and quality of inputs data are critical to model accuracy. While government road network databases have the advantage of being official and frequently include useful maintenance data, they are often not publicly accessible or are incomplete or outdated, particularly if different agencies manage separate parts of the road network. Freely available road network datasets from OSM are straightforward to access and employ in most accessibility models and could provide the required additional data to obtain accurate estimates of travel time and accessibility. These public data may, however, contain significant gaps in certain geographies, particularly in rural environments.

THE CHOICE FOR COST-EFFICIENT IMPROVEMENTS OF ROAD NETWORK DATA

These gaps in public and government-owned road network datasets need to be closed to improve the accuracy of accessibility models, yet manually mapping every missing road in a network is time- and resource-intensive. A balance must be struck between the exhaustive, full-scale manual mapping of missing roads to complete networks, and the selection of a limited set of additional roads to be mapped with priority that will considerably improve model accuracy, without compromising on the quality of results. While each incremental improvement of the road network through additional mapping will further improve modelled accessibility and bring results closer to the 'true' accessibility on the ground, the added value of mapping extra roads decreases as networks become more complete. Moving from unmapped units to mapping priority roads delivers a considerably higher return on investment – not only financially but also in terms of impact on results – than going from priority mapping to full-scale mapping. This can be thought of as a logarithmic curve of decreasing marginal returns, as per *Figure 1*.

The World Bank's Poverty & Equity Global Practice therefore developed and tested a cost-efficient approach to identify priority areas and roads to be included in the mapping. The methodology underlying this approach is detailed in *Box 1*, as developed for the case study of *tehsils* in KP Province, Pakistan. To assess its performance and usability, the modelled accessibility improvements obtained through this priority-based mapping approach were compared with results for *tehsils* that remained unmapped, and *tehsils* for which each missing road was mapped.

Figure 1. Curve of return on investment for full-scale and priority-based additional road mapping in OSM
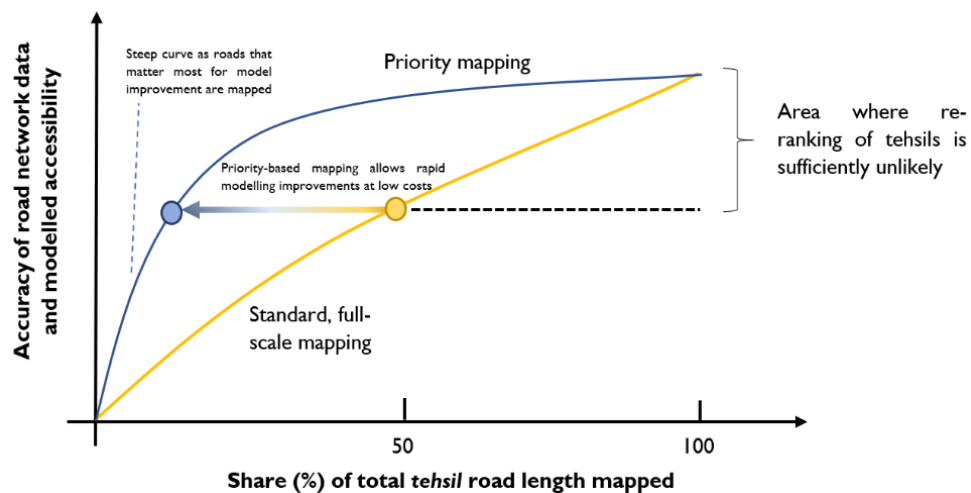
Box 1 – Measuring accessibility and identifying priority roads for additional mapping

An insight into the **change in modelled accessibility** due to the mapping of additional roads can be obtained by subtracting the results for accessibility (measured in hours of travel time to services or opportunities) after the investment in the mapping of additional roads from the accessibility results prior to this investment. A positive (negative) change in accessibility figures for *tehsils* implies an improvement (reduction) in modelled accessibility due to the mapping of additional roads as part of the OSM study. It is important to emphasize that this change in accessibility figures does not reflect any changes on the ground. Post-investment estimates of accessibility are simply the result of basing models on a more complete roads network dataset which better resembles the actual distribution and typologies of roads in each *tehsil*. Similarly, if modelled accessibility decreases, this is due to the inclusion of mislabelled roads in the road network database, which are allocated higher speeds than in reality.

The benefits of a priority-based versus full-scale mapping approach can be calculated in terms of the improvements or reductions in modelled accessibility (measured in hours of travel time to services or opportunities) per dollar invested in the mapping of additional roads. To allow for this comparison, the team used **three different approaches:**

- ❖ Approach 1: full-scale mapping
- ❖ Approach 2: priority mapping
- ❖ Approach 3: no mapping

Testing these approaches in KP Province, the OSM road dataset for KP Province as of January 1st, 2021 and a smaller roads dataset provided by the government were used as primary road network input, with Government data on the location of schools and health facilities and a World Bank-collected dataset on the location markets serving as inputs on the location of vital services and opportunities. The 116 *tehsils* of KP Province were randomly subdivided into three categories. 23 selected *tehsils* were fully mapped, with all roads not included in the OSM dataset manually added to obtain a comprehensive road network for those *tehsils*, spanning eight districts. 49 *tehsils* in thirteen KP districts were selected for priority mapping. Here, only major roads connecting to significant settlements that were missing from the OSM dataset were manually added, offering the cost-effective alternative to be tested. No mapping took place in the 44 remaining *tehsils* that fell outside the project area as instructed by the Government of KP Province. These *tehsils* served as controls.

For the **priority mapping of *tehsils***, we used the 2019 World Settlement Footprint dataset at 10m resolution, filtered by a 7-by-7 kernel majority filter, to remove minor hamlets. This identified discrete, major settlements. Next, we created buffers of 500 meters around existing roads in the OSM and government road network datasets. Settlements falling outside these 500-meter buffers were identified as significant populated areas disconnected from road networks, and priority areas for inclusion. For these priority areas, the project team mapped and integrated the most prominent road connecting each settlement to the road network into the OSM road network, if such road was visible on satellite imagery.

Tehsils, and the roads that run through them, differ strongly based on the terrain and remoteness, especially in a mountainous province like KP. It was therefore important to **differentiate the type of roads** that form part of the road network in *tehsils*, to explore the topography, and to distinguish primarily rural from mostly urban administrative units. In first instance, roads were classified based on their type, influencing modelled speeds and,

therefore, travel times to destinations. Road types were validated during mapping and, in case of misclassification, reassigned to the right category. The team scanned all major highways in KP Province, including those in the 44 unmapped *tehsils*, to ensure these were correctly tagged. This can impact the accessibility results also in unmapped *tehsils*, as misclassified highways can be assigned lower or higher speeds depending on the type of road in reality.

Tehsils were also categorized based on their **elevation**, to study whether accessibility improvements in OSM were dependent on the ruggedness of the terrain. 52 *tehsils* with an average altitude below 1,500m were categorized as low-elevation *tehsils*. 18 tehsils between 1,500m and 2,300m were categorized as medium-elevation, while 2 *tehsils* above 2,300m were categorized as high-elevation. As the number of medium and high elevation *tehsils* is small, these were grouped together.

A final subdivision of *tehsils* was based on their **degree of urbanization**. *Tehsils* were labelled urban if over 30% of their WorldPop 2020 modelled population fell within the urban extent of the Global Human Settlement Layer Settlement Model for 2019 (GHSL - SMOD). However, as KP is a primarily remote, rural Province, only ten *tehsils* were classified as urban. Of these ten, two *tehsils* were priority mapped and one was fully mapped, with seven remaining unmapped as they fell out of the primarily rural scope of interest of the KP Government.

OSM data is improved using freely available satellite imagery and editing software maintained by OSM's active volunteer community. The Poverty & Equity team, supported by Kathmandu Living Labs, invested in mapping missing roads in KP Province, and reclassified those who were assigned the wrong category in OSM (motorway, trunk, primary, secondary, tertiary, service, residential road, or unpaved rural roads). This results in a 'pre-investment' road network dataset prior to World Bank funding of additional mapping, and a 'post-investment' dataset in which additional roads are included. Not only does this deliver the best available road networks for accessibility modelling, the **OSM data created are also freely available to others** working on the area of study. This is illustrated below for the mapping in Garyum tehsil, KP Province.
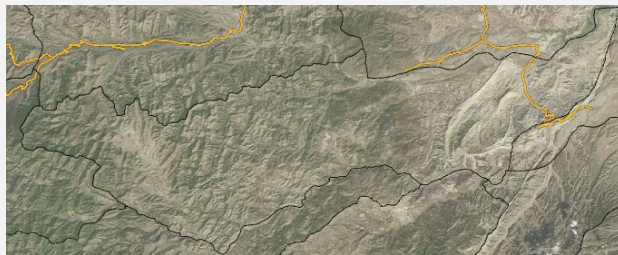
OSM road dataset (yellow) pre-OSM investment in priority-mapped Garyum *tehsil*, KP
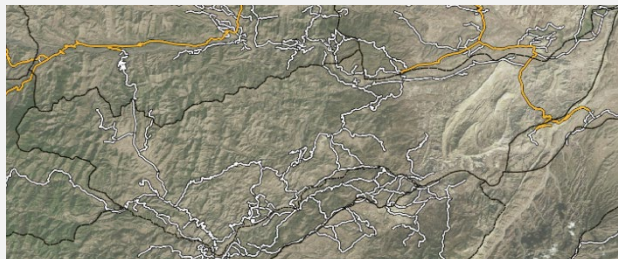
New roads (white) mapped post-OSM investment in priority mapped Garyum *tehsil*, KP

CHANGES IN ACCESSIBILITY FOR THE DIVERSE MAPPING APPROACHES AFTER INVESTMENT

While it is not possible to calculate how closely a priority-based approach of mapping additional roads approaches the 'right' results, the return on investment of such priority-based approach vis-à-vis the full-scale mapping of each missing road in the network can be assessed. Several results stand out.

Result 1: Considerable improvements in modelled accessibility are found for both approaches

The priority mapping approach ensured the inclusion of all significant settlements that were previously not connected to the OSM road network in the accessibility analyses. For the majority of fully and priority mapped *tehsils*, significant improvements in accessibility figures emerge due to the additional mapping of roads. *Figure 2* shows the percentage-wise improvements in modelled accessibility to markets (measured in travel time hours) for *tehsils*. Among the 72 fully and priority mapped *tehsils*, 51 showed absolute changes (both positive and negative) of more than 10%, and 25 of more than 20%. Similar findings are obtained for accessibility to education and healthcare.

Result II: Reductions in modelled accessibility are also possible

Several fully and priority mapped *tehsils* show negative changes, implying a decrease in accessibility (higher travel times) to services after additional OSM mapping. This is due to several minor roads in those *tehsils* being incorrectly labelled as 'primary highways' in the pre-investment dataset (more information on this investment in *Box 1*). Consequently, these roads were allocated unrealistically high speeds in the accessibility model. The project team appropriately relabelled these as minor roads. The consequent reduction in modelled speeds over these roads post-investment yielded higher, more realistic travel times to services, and emphasises the importance of data accuracy and data validation.

Result III: Exceptionally large improvements are seen in poorly mapped units

Similarly, some priority-mapped and unmapped *tehsils* showed exceptional positive changes, i.e., a reduction in modelled travel time accessibility by more than 30%. A comparison of input datasets showed that only a limited number of roads were mapped in these priority mapped *tehsils* prior to the investment in additional mapping. In addition, in the case of the unmapped *tehsils* that showed exceptional positive changes, some major roads that completed the network in priority and fully mapped *tehsils* passed through these unmapped *tehsils*. Therefore, OSM investment and subsequent road mapping led to considerable increases in modelled accessibility and associated decreases in travel times to services and opportunities in those *tehsils* as well.

Figure 2. Additional Road Mapping - Change in Market accessibility

% CHANGE IN ACCESSIBILITY TO MARKETS DUE TO ADDITIONAL ROAD MAPPING

□ Fully Mapped Tehsils   □ Priority Mapped Tehsils   □ Unmapped Tehsils
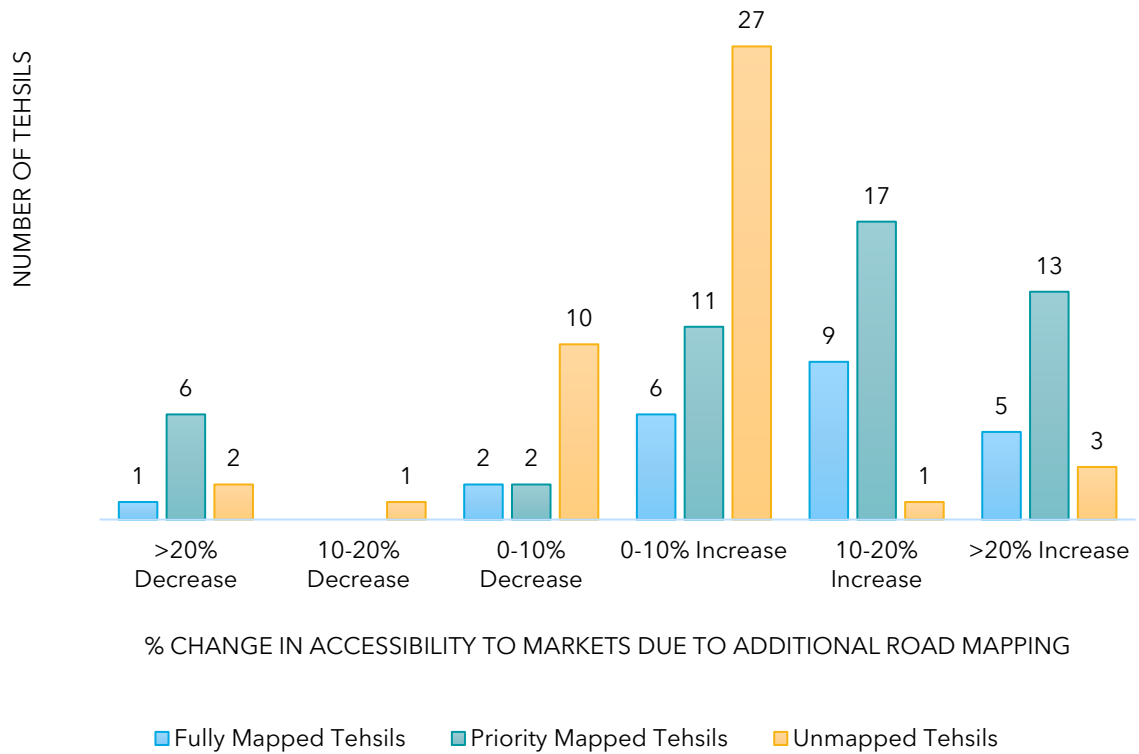
COMPARING ACCESSIBILITY CHANGES FOR PRIORITY AND FULL-SCALE MAPPING

The population-weighted ranking of the accessibility of KP Province *tehsils*, measured by the hours of travel time it takes a person to reach a service or opportunity, changes significantly when pre- and post-investment scenarios are compared, particularly for *tehsils* at the top of the list of those with the lowest accessibility – and therefore in highest need of investment. *Table 1* shows the ranking of the top-ten inaccessible *tehsils* based on the overall service accessibility index, combining access to education, healthcare, and markets. While *tehsils* with low accessibility are generally correctly assessed as such both before and after additional mapping, the order of ranking shifts significantly using pre- and post-investment figures. Three out of the top-ten *tehsils* with the lowest accessibility would not have been identified without the improvement of the road network through manual road mapping. This implies that resources may be misallocated if project investments are targeted based on accessibility metrics without improved road network datasets. The mapping of additional roads this pays off and is foundational to informing evidence-based decisions.

Table 1: Top ten least accessible *tehsils*, ranked before and after investment in additional mapping

KP Province Tehsil	Ranking prior to investment	Ranking after investment	Elevation	Mapping approach
Dassu	1	1	Medium	Completely Mapped
Kalkot	17	2	High	Priority
Data Khel	11	3	Medium	Priority
Toi Khulla	2	4	Low	Priority
Mastuj	9	5	High	Priority
Birmal	5	6	Medium	Priority
Pattan	6	7	Medium	Completely Mapped
Upper Orakzai	3	8	Medium	Priority
Chitral	20	9	Medium	Priority
Central Kurram	7	10	Medium	Priority

The average change in ranking for fully mapped *tehsils* is 2.9 spots, with a median change of 2. In contrast, the average change in ranking for priority mapped *tehsils* is 4.5, with a median of 3. This apparent contradiction, with fully mapped *tehsils* having smaller changes in ranking compared to priority-mapped *tehsils*, can be explained by the observation that priority mapped *tehsils* tended to be coincidentally less extensively mapped prior to investment. Measuring these differences in terms of percentage-wise changes in accessibility, full-scale mapped *tehsils* record an average positive change in accessibility to markets of 11.5%, whereas this is 5.5% for priority mapped *tehsils*. Other points of interest, such as schools and hospitals, show similar trends. Combining accessibility to primary, middle, and high school, the average positive change in access to education for fully mapped tehsils is 30.5%, while for priority mapped *tehsils* this is 24.8%. For public health services, accessibility improves by an average 12.5% for fully mapped *tehsils*, versus 13.4% for priority mapped *tehsils*. These changes in modelled accessibility bring measurements closer to the actual travel times in *tehsils*, given that they reflect a more complete picture of the road networks. However, no conclusive statement on the accuracy of these modelled accessibility levels can be made, as no data sources are available to ground truth real-time travel times.
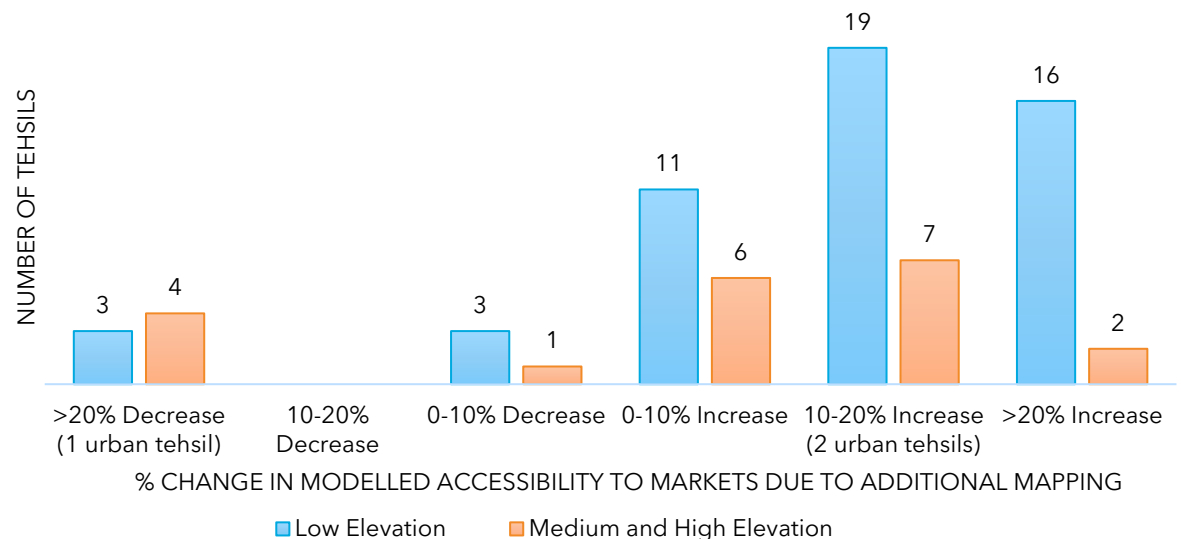
OSM mapping investments thus contribute significantly to improving the accuracy of accessibility figures for *tehsils* and aids the prioritizing of *tehsils* for future investments in road infrastructure, as well as services and opportunities. The case study does, however, not allow to compare priority and full-scale mapping in the same *tehsils*, which would provide an insight into their comparative improvements. Future work should aim to study the improvements for both approaches within the same spatial unit.

CONSIDERATIONS OF ELEVATION AND DEGREE OF URBANIZATION

As hypothesised, both the elevation of the terrain and degree of urbanisation can impact the mapping of road networks. The average change in accessibility for lower elevation *tehsils* is 7.9%, while for high and medium elevation *tehsils* is -4.6% (*Figure 3*, using markets as the exemplar). However, when *tehsils* with negative changes are discarded, the average change for lower elevation *tehsils* is around 13.0%, while this is 10.3% for higher and medium elevation *tehsils*. A desk review of the negatively changed high and medium elevation *tehsils* shows that only a few minor roads were mapped in those *tehsils*, and they were incorrectly labelled as 'primary highways' in the pre-investment dataset. This shows the importance of reviewing, mapping and correctly labelling the road networks, particularly in high and medium elevation *tehsils*. These *tehsils* would otherwise have been misrepresented in accessibility analyses, resulting in an overestimation of accessibility, and a lower prioritization in investment decisions.

Preliminary analysis of the improvements in accessibility in urban versus rural *tehsils* suggests that the degree of urbanization might impact the mapping outcomes (the improvement range of the urban *tehsils* is indicated in *Figure 3*). However, insufficient urban samples are available to deduce a clear trend, and no percentage changes are therefore included here for urban administrative units. For rural *tehsils*, an improvement in modelled accessibility of 8.1% on average is observed. Urban areas are likely better mapped and represented in the OSM dataset, and additional OSM mapping of rural areas is therefore particularly important for similar assignments with low degrees of urbanisation in different geographic contexts, central to the focus areas of the World Bank.

Figure 3: Modelled Accessibility - Elevation and Degree of Urbanization

Y-axis: NUMBER OF TEHSILS

X-axis categories:
>20% Decrease (1 urban tehsil): Low Elevation 3, Medium and High Elevation 4
10-20% Decrease: (none)
0-10% Decrease: Low Elevation 3, Medium and High Elevation 1
0-10% Increase: Low Elevation 11, Medium and High Elevation 6
10-20% Increase (2 urban tehsils): Low Elevation 19, Medium and High Elevation 7
>20% Increase: Low Elevation 16, Medium and High Elevation 2

% CHANGE IN MODELLED ACCESSIBILITY TO MARKETS DUE TO ADDITIONAL MAPPING

Legend: Low Elevation ■ Medium and High Elevation ■

COST-EFFICIENCY CONSIDERATIONS FOR SELECTION OF OPTIMAL APPROACH

The key advantage of the priority mapping approach is its reduced cost to map missing roads. Mapping the 49 priority mapped *tehsils* was considerably cheaper compared to exhaustively mapping the 23 fully mapped *tehsils*. In KP Province, the priority mapping process mapped 12,094km of roads in an area spanning 52,017km^2 at a cost of USD 23,433. Full-scale mapping covered 47,935km of additionally digitized roads in an area of 29,277 km^2 at a cost of USD 26,067. The cumulative cost of the priority mapping process thus was approximately \$0.45/km^2, while the cost for fully mapping process is approximately \$0.89/km^2. Moreover, for priority mapping, each dollar invested in additional mapping returned an improvement of the accessibility model of 8 seconds of travel time. For full-scale mapping, each invested dollar only returned a travel time improvement of 0.58 seconds. In addition, priority mapping is faster, with the completion of priority mapping taking around half as long as the full-scale mapping exercise. Priority mapping thus significantly improves the quality of accessibility analyses, yet is efficient and affordable, offering the highest cost-benefit approach for analytical work in areas with poorly mapped road networks.

IMPLICATION OF FINDINGS

To accurately model the accessibility of people to services, opportunities, and points of interest, the data quality and coverage of the road network underlying these models is crucial. Significant gaps in road networks do, however, exist, implying accessibility estimates will be far off the 'true' accessibility citizens experience in real life. Road network data should therefore be improved through the additional mapping of missing roads. The analytical results in this note show that full-scale, manual mapping of each missing road is time- and resource intensive and offers a rapidly and starkly decreasing return on investment. In contrast, an approach to additional mapping in OSM which first identifies and then digitizes priority missing roads results in a cost-effective improvement of accessibility estimates. This cuts the cost of mapping an additional square kilometre of roads in KP Province, Pakistan, by half. Additional mapping of roads beyond priority roads has diminishing marginal benefits with costs soaring.

While our findings suggest priority mapping is cost-effective and reliable, we cannot confidently state that the two methods yield equal results, given that the *tehsils* being compared are not fully equivalent units of analysis. Yet the advantages of priority mapping in terms of efficiency and affordability entail a recommendation for its use in analytical work in areas with poorly mapped road networks. However, the extent of additional mapping required is dependent on population density, road density, settlement extents, and the coverage of the roads in the pre-investment OSM map. Areas with higher population and road densities, larger geographic areas with

dispersed settlements, and areas with relatively low coverage of roads in the pre-investment OSM map will require relatively more extensive additional mapping. For projects where a full set of road data is quintessential, for instance in dense urban environments, the need for completeness might also outweigh the importance of cost-effectiveness. Mapping processes are thus not one-size-fits-all and teams will need to tailor methods to the project's budget, road network quality, and areal extent to be covered. For example, we found that mapping equivalent-sized areas in Pakistan was far more expensive than in Bhutan and Nepal due to the greater population size, and road and settlement density.

About The Authors

Robert Banick worked with the World Bank's Poverty and Equity Global Practice as a geospatial consultant.

Manish Basnet is a consultant working with the World Bank's Poverty and Equity Global Practice on the Bhutan, Nepal, and Pakistan Programs.

Lander SMM Bosch is the Regional Geographer for the South Asia Region in the World Bank's Poverty and Equity Global Practice and a World Bank Young Professional.

Moritz Meyer is Senior Economist in the World Bank's Poverty and Equity Global Practice and Task Team Leader of the Pakistan Poverty and Equity team.

CONNECT WITH POVERTY & EQUITY GLOBAL PRACTICE

www.worldbank.org/poverty   @WBG_Poverty

WORLD BANK GROUP
Poverty & Equity