Gender-Based Vulnerability: Combining Pareto ranking and geostatistics to model gender-based vulnerability in Rohingya refugee settlements in Bangladesh

Erica Nelson (elnelson@partners.org)
Harvard Medical School https://orcid.org/0000-0003-1696-443X

Daniela Reyes Saade
Clark University

P.Gregg Greenough
Harvard Medical School

Research

Keywords: Rohingya, refugees, gender, open-source data, vulnerability index, spatial analysis, GIS, Pareto ranking, spatial autocorrelation

Posted Date: February 8th, 2020

DOI: https://doi.org/10.21203/rs.2.22973/v1

Version of Record: A version of this preprint was published at International Journal of Health Geographics on May 29th, 2020. See the published version at https://doi.org/10.1186/s12942-020-00215-3.

Title: Gender-Based Vulnerability: Combining Pareto ranking and geostatistics to model gender-based vulnerability in Rohingya refugee settlements in Bangladesh

Authors

Erica L Nelson[1,2*], Daniela Reyes Saade[3], P.Gregg Greenough[1,2,4]

[1] Harvard Humanitarian Initiative, Harvard University, Cambridge, MA, United States of America

[2] Department of Emergency Medicine, Division of Global Emergency Care and Humanitarian Studies, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, United States of America

[3] Graduate School of Geography, Clark University, Worcester, MA, United States of America

[4] Harvard T.H. Chan School of Public Health, Boston, MA, United States of America

*Corresponding Author:

Email: Elnelson@partners.org (EN)

ABSTRACT

Background: The Rohingya refugee crisis in Bangladesh continues to outstrip humanitarian resources and undermine the health and security of over 900,000 people. Spatial, sector-specific information is required to better understand the needs of vulnerable populations, such as women and girls, and to target interventions with improved efficiency and effectiveness. This study aimed to create a gender-based vulnerability index and explore the geospatial and thematic variations in gender-based vulnerability of Rohingya refugees residing in Bangladesh by utilizing pre-existing, open source data.

Methods: Data sources included remotely-sensed REACH data on humanitarian infrastructure, United Nations Population Fund resource availability data, and the Needs and Population Monitoring Survey conducted by the International Organization for Migration in October 2017. Data gaps were addressed through probabilistic interpolation. A vulnerability index was designed through a process of literature review, variable selection and thematic grouping, normalization, and scorecard creation, and Pareto ranking was employed to rank sites based on vulnerability scoring. Spatial autocorrelation of vulnerability was analyzed with the Global and Anselin Local Moran's I applied to both combined vulnerability index rank and disaggregated thematic ranking.

Results: Of the Rohingya settlements, 24.1% were ranked as 'most vulnerable,' with 30 highly vulnerable clusters identified predominantly in the *upazila* of Sadar. Five settlements in Dhokkin, Somitapara, and Pahartoli were categorized as less vulnerable outliers amongst highly vulnerable neighboring sites. Security- and health-related variables appear to be the most significant drivers of gender-specific vulnerability in Cox's Bazar. Clusters of low security and education vulnerability measures are shown near Kutupalong.

Conclusion: The humanitarian sector produces tremendous amounts of data that can be analyzed with spatial statistics to improve research targeting and programmatic intervention. The critical utilization of these data and the validation of vulnerability indexes are required to improve the international response to the global refugee crisis. This study presents a novel methodology that can be utilized to not only spatially characterize gender-based vulnerability in refugee populations, but can also be calibrated to identify and serve other vulnerable populations during crises.

Key Words

Rohingya, refugees, gender, open-source data, vulnerability index, spatial analysis, GIS, Pareto ranking, spatial autocorrelation

BACKGROUND

The Rohingya refugee crisis is a result of decades of systematic discrimination, statelessness, and violence exacerbated by the denaturalization of the Rohingya people by the then-Burmese state in 1977 [1]. Since that time, waves of Rohingya, now considered 'illegal' in their birth country, have fled into Bangladesh in a complex cycle of forced displacement and repatriation efforts [2]. In August of 2017, violent attacks perpetrated against the Rohingya triggered the largest and fastest mass displacement of refugees from Myanmar's Rahkine State to Bangladesh. As of March of 2019, there were over 909,000 Rohingya refugees in Ukhiya and Teknaf upazilas, representing an influx of over 745,000 people in less than two years [3].

Women and girls comprise over 52% of the population in Rohingya settlements in Bangladesh, with approximately one-sixth of families headed by a single mother [4]. Almost every woman and girl in the Rohingya refugee community in Cox's Bazar has either experienced or witnessed incidences of gender-based violence [5], and the crisis disproportionately affects women, girls, and other marginalized populations due to the perpetuation of pre-existing inequalities, violence, and discrimination. As of February 2018, the Inter-Agency Working Group on Reproductive Health in Crisis [6] identified women and girls as 'critically underserved' in the Rohingya Humanitarian Response, emphasizing a lack of access to sexual and reproductive services and gender-based violence care. According to the World Food Program, Rohingya refugee households in Bangladesh headed by women are disparately more vulnerable to food insecurity, with 45% defined as vulnerable or highly vulnerable [7]. Over one-third of Rohingya women report insecurity when collecting water or toileting [8], and over half of the female population lack appropriate menstrual supplies. Many women remain within their shelters due to a lack of clothing, insecurity, and concerns regarding cultural norms and dignity. Despite the humanitarian community's commitment to gender mainstreaming and women-targeted strategies [9] in response efforts, qualitative data from a 2018 Oxfam report demonstrated that the humanitarian response has yet to adequately meet the needs of women and girls in this community, failing to provide them with access to services and/or addressing gender-specific issues critical to preventing further harm [4].

As of October 2019, there were nearly 272 million United States Dollars (USD) in unmet funding requirements for the Rohingya Refugee Crisis Joint Response Plan for the year [10], making 2019 the year with the biggest relative deficit in funding, with over one-third of requirements being unmet. This funding-needs gap within humanitarian response is a global and protracted phenomenon, with nearly twenty-seven billion USD needed for funding humanitarian programming as of 2019. As climate change, protracted conflicts and social inequality continues to grow the need for humanitarian assistance, the

funding gap continues to hover around 40% [11], leaving nearly one-third of affected populations without essential aid.

As is the case in many humanitarian and disaster response environments, the sheer magnitude of the Rohingya crisis juxtaposed with inherent resource limitations compels donors and humanitarian actors to create mechanisms to more precisely characterize need and design targeted, cost-effective services. One such method of need assessment is through the lens of 'vulnerability' analysis, which articulates the reality that hazards and subsequent assistance impact various population groups in grossly heterogeneous ways [12]. Enumerable vulnerability indices have emerged in fields including infectious diseases [13, 14], environmental health [15, 16], disaster preparedness [17-19], refugee services [7, 20], and climate change [21-23] which incorporate diverse, cross-sectoral indicators such as socio-economic status, education, information access, mobility, health and morbidity, security, and geographic location.

Such vulnerability indices require significant data, which is often incomplete, untimely, and ambiguous in the humanitarian context. Designing and implementing primary data collection studies in humanitarian settings requires substantial resources that can be otherwise utilized for response activities and is further complicated by insecure or inaccessible environments. However, there exist tremendous data being produced by agencies already operating in these contexts, and a recent impetus to share these invaluable data has led to open source dissemination platforms. The Humanitarian Data Exchange (HDX) [24], managed by United Nations Office for the Coordination of Humanitarian Affairs' (OCHA) Center for Humanitarian Data, was launched just over five years ago and is an online, open source data platform aimed at making humanitarian data available for analysis and use by non-governmental organizations, governments, and United Nation agencies. As of April 2018, it houses over 6,000 data sets from nearly 1,000 sources in 245 locations [25].

The aim of this study was to harness the potential of pre-existing, open source data provided by the Humanitarian Data Exchange to create a gender-based vulnerability index and explore the geospatial and thematic variations in the gender-based vulnerability of Rohingya refugees residing in Bangladesh.

METHODS

This study designed a gender-based vulnerability analysis grounded in constructs derived from a literature review and modified secondary to data availability, executed a Pareto ranking as a method to avoid artificial weighting in aggregated index scores, and subsequently employed geostatistical methods of spatial autocorrelation and cluster/outlier analysis. This methodological framework (Figure 1) was applied

to open-source humanitarian data collected at Rohingya refugee settlements in Bangladesh directly after the 2017 refugee influx. Briefly, this methodology intends to capitalize on pre-existing spatial humanitarian data to identify those refugee settlements wherein women and girls are the most vulnerable, and create a statistically-grounded, spatial understanding of gender-based vulnerability that can lead to more relevant and better targeted programmatic design and resource allocation.

Figure 1. Methodological process for creating a gender-based vulnerability index and subsequent geostatistical analysis

Data collection

Three data streams were utilized for this study, including the International Organization for Migration (IOM) Needs and Population Monitoring Data from October 2017, in Bangladesh; REACH remotely sensed infrastructure location data, and United Nations resource availability data. These data were obtained from the open source, online Humanitarian Data Exchange between October 2017 and January 2018. A description of these data and data collection methodologies follow.

The International Organization for Migration conducted a Needs and Population Monitoring site assessment of Rohingya refugee communities at 28 collective sites and 99 locations within dispersed settings in host communities in Cox's Bazar, Bangladesh between the 30th of September and 9 October 2017 [26] with a final sample size of 157 sub-site assessments. The site assessment covers all locations where Rohingya refugees have been identified, notwithstanding the type of location or proximity to Bangladeshi host communities. Information regarding the demographics, distribution, and needs of the Rohingya refugee population spanned 12 areas of inquiry and 184 unique questions that were both quantitative and qualitative in nature. Information was collected by locally-recruited enumerators through key informant interviews using closed-ended questionnaires, and findings were triangulated at the field level through direct observation and spontaneous group discussions.

Distance to distribution centers, health care, nutritional facilities, and child- and women-friendly spaces were gleaned from remotely-sensed and ground-based, global positioning system (GPS)-enabled surveys conducted by REACH and the United Nations Population Fund (UNFPA) in cooperation with the Inter Sector Coordination Group in Bangladesh (ISCG), with datasets collected between November 2017 and February 2018. These data were retrieved from the Humanitarian Data Exchange as either geodatabases containing point data, shapefiles, and/or spreadsheets containing each facility's spatial coordinates.

Gender-Based Vulnerability Index

This study followed a multi-step process for developing a gender-based vulnerability score, founded on principles articulated in the Handbook on Constructing Composite Indicators [27]. After reviewing existing literature, a theoretical framework was developed to serve as a basis for selecting indicators into a composite score. By reviewing pre-existing data available through the HDX database, component themes and indicators were identified to include in the gender-based vulnerability index (Table 1).

Table 1. Component themes and indicators selected to contribute to the gender-based vulnerability index. All distances are in kilometers.

Component theme	Indicators
Demographics	Percentage of pregnant women, percentage of women lactating
Education	Presence of barriers to education for adolescent girls
Health	Access to antenatal care, access to psychiatric support services, access to vaccinations, distance to the nearest healthcare facility
Water, Sanitation, and Hygiene (WASH)	Perception of 'enough water for household needs,' recent outbreaks of diarrhea, and where women defecate at night (categorized as private facilities, communal, and open defecation, with open defecation being the most 'vulnerable').
Resource Availability	Access to food supplementation for pregnant and lactating women, distance to the nearest distribution center
Security	Access to gender-based violence services, access to incident reporting mechanism, access to police and courts, distance to the nearest women-friendly space, distance to the nearest child-friendly space.

Vulnerability, commonly defined in the humanitarian community as the diminished capacity of an individual or group to anticipate, cope with, and resist and recover from the impact of a hazard [12], is a relative and dynamic concept. It is multi-dimensional and differential, scale-dependent, and spatially heterogeneous. The definition of vulnerability suggests that it is referential to specific hazards. In the setting of Rohingya refugees, hazards are abundant, including vulnerability to food insecurity, interpersonal violence, diseases, and natural hazards such as cyclones, landslides, and flooding. In light of such reality, this study defines vulnerability in alignment with the Vulnerability Assessment Framework established by the United Nations High Commissioner for Refugees as the 'risk of exposure of … refugee households to harm, primarily in relation to protection threats, inability to meet basic needs, limited access to basic services, and food insecurity, and the ability of the population to cope with the consequences of this harm" [20].

Despite the fact that gender is a social construct, vulnerability is undoubtedly gender-differentiated. Variables, such as lack of access to and/or control over essential resources and lack of entitlements, compound women's vulnerability and undermine agency and adaptive and recovery capacities from hazards [28]. Many gender indices exist to examine the role of gender discrimination as it influences human development and wellbeing at the national level. These indices, including the Social Institutions Gender Index [29], the Gender Inequality Index [30], the Global Gender Gap Index [31], identify several dimensions that influence gender inequality. These range from health and protection parameters such as access to prenatal care and laws regarding gender-based violence, attainment of education, economic indicators, and freedom of movement, to discriminatory family codes, restricted civil liberties, representation in government, and empowerment within socio-economic realms.

However, previous gender-based vulnerability indices are limited in their utilization and reporting in academic literature. One such gender vulnerability index, conceived and implemented by Plan India, attempted to develop collective perspectives and generate a normative consensus on the status of girls and women in India. This index identified nine thematic indicators of gender-based vulnerability, including 1) safety and protection, 2) health and survival, 3) illiteracy, 4) poverty, 5) policy framework and implementation, 6) climate change and migration, 7) digitalization vulnerabilities, 8) cultural and social practices, and 9) urban/rural vulnerability [32]. Given the lack of data available to speak to each of these themes, they identified four dimensions that captured the predominant variables contributing to gender-based vulnerability: protection, health, education, and poverty. Similarly, while we acknowledge that many variables influence gender-based vulnerability in refugee populations, these data were not available in open-source, secondary data, and thus were not included in the analysis.

Incomplete Data: Probabilistic Interpolation

Within the IOM's Needs and Population Monitoring dataset, there were settlement assessment points that lacked data regarding specific indicators. For instance, gaps in gender specific food insecurity household coping strategies such as preferentially feeding boys over girls were not included. Rather than include these incomplete data points, access to food at distribution points was utilized as these were more complete in the dataset. Gaps in binary and categorical data were determined to be random in the geographic bounds of the data, and were imputed by assigning likelihoods of "0" or "1 or by category using a probabilistic interpolation method of indicator formalism. This required transforming the missing indicator binary or categorical variable to a probability for the category or binary value; the probability of

a category or binary value for a given indicator, *I*, at a location *u,* for a binary or categorical value z_k is the transformation:

$$I(u{:}\,z_k) = \{1, if\ Z(u)\ =\ z_k\}\ or\ \{0, if\ otherwise\}$$

Several sites did not include enough data to be included in the vulnerability analysis, leading to a total number of settlements evaluated being 145.

Quantifying and Normalization:

Quantification

Each variable was enumerated for incorporation into the vulnerability scoring, with higher scores indicating a higher contribution to vulnerability. Non-enumerated variables, such as binary variables that defined the presence or absence of service, were quantified as vulnerable=1 and not vulnerable=0. For example, if prenatal care was available, the site would be scored a zero, and if prenatal care was not available, the site was scored a one, given the assumption that a failure to provide prenatal care would make women and girls more vulnerable. Similarly, the places in which women defecate at night were reported as a generalized response of either 1) in private, 2) in communal latrines, or 3) open defecation. These were scored as 1-3 responses, with open defecation defined as the most vulnerable.

Distance to facilities or resources were measured in kilometers (Km) utilizing Euclidean distance and quantified within the scorecard with the equation:

$$Score = Distance\ (Km)\ to\ facility/Max\ Distance\ (Km)$$

thereby creating a vulnerability score that is relative to other refugee sites. Variables that were reported as percentiles (e.g., percent lactating women) were incorporated as decimals, normalizing the variable between zero and one.

Normalization

Prior to total vulnerability score calculation, each thematic score was normalized to prevent arbitrary weighting of individual thematic components utilizing the equation below, which has been employed by many studies and is recommended by the Humanitarian Development Index as a standardized method [33].

$$Normalized\ score = \frac{(x\ -\ xmin)}{(xmax\ -\ xmin)}$$

After quantification and normalization were completed, scorecards that summed each of the six thematic scores (i.e., education, health, sanitation, etc.) were utilized to create an aggregate 'vulnerability score' for each refugee settlement site to provide a preliminary understanding of gender-based vulnerability.

Aggregating thematic component scores and the role of Pareto ranking

Composite vulnerability indices have long been criticized for employing arbitrary calculations to aggregate indicator scores into a single metric [34]. Thematic component scores (e.g., health or security) may be simply averaged or a weights-matrix may be applied, but both of these mathematical decisions are foibled and more likely represent the researcher's biases than the realities of vulnerability. Simple averaging obscures extreme values when both high and low scores exist in different thematic components. Weighting schema are equally problematic, requiring researchers to judge and quantify the relative importance of different components either through Delphi procedures or qualitative or quantitative methods. Even qualitatively derived weights matrices will likely fail the test of external validation, as the importance of indicators is spatio-temporally heterogeneous.

Thus, this study utilized Pareto ranking, as implemented by Rygel et al. [35], to avoid the aforementioned challenge of aggregation and the subjective weighting of thematic components, and yet still create a composite vulnerability hierarchy that takes into account multiple variables. Pareto ranking is a method by which non-domination is applied to ascribe vulnerability ranks to each refugee site. Non-domination is achieved by a site when no other sites in the dataset are more vulnerable, i.e., if in each thematic component, that site scores at least as high or higher than all other refugee sites. The non-dominated set of sites includes all sites that are non-dominated in the dataset and represent the most vulnerable. These sites are then removed from the dataset, and the algorithm iterated to identify the next most vulnerable set of refugee sites. The process is repeated until all sites are ranked, thereby creating a hierarchical ranking of gender-based vulnerability amongst these sites. An illustration of this process is provided in Figure 2. Note that this illustration demonstrates each point as having only two thematic components that factor into its vulnerability ranking. In this study, Pareto ranking utilized the six thematic components identified above with identical logic and procedures applied to the higher-dimensional data.

Figure 2. A two-dimensional illustration of Pareto ranking in which each data point (in our study, settlement site) has two thematic component scores. Those data points represented as circles are considered the most vulnerable rank, in that each site is non-dominated. Those represented as 'X's are second-most vulnerable, and those represented by squares are least vulnerable.

The Pareto ranking algorithm was executed in Python.

Geospatial analysis

Spatial analysis has been extensively adopted for exploring spatial heterogeneity and cluster detection in public health research and environmental epidemiology [36-38] Ranging greatly in application and methodology, spatial autocorrelation and local indicators of spatial association (LISA) have been utilized to appreciate clusters of disease, morbidity, and mortality, and thereby contribute to a knowledge base surrounding exposure pathways, social determinants of health, and the intersection between socio-economic, environmental, and human wellbeing. Practically, cluster detection methods have provided a means by which health policy officials and public health personnel can better target interventions and further research. Through this lens of spatial epidemiology, this study undertook a geospatial analysis of gender-based vulnerability ranking to characterize the spatial heterogeneity that would lead to a better spatial understanding of gender-based vulnerability in this community and could potentiate better targeting of programs for women and girls within the Rohingya refugee settlements. Spatial autocorrelation analysis and LISA, utilizing the Global Moran's and Anselin Local Moran's I, were employed to identify large-scale trends of autocorrelation and clusters and outliers of refugee settlement sites that possessed high and low gender-based vulnerability scores.

Global Moran's I is an inferential, spatial statistical method that determines spatial autocorrelation based on feature location and attributes value, e.g., vulnerability score utilizing the equation:

$$I = \frac{N}{W} \frac{\Sigma i \, \Sigma j \, w_{ij} \, (x_i - \overline{x}) \, (x_j - \overline{x})}{\Sigma_i \, (x_i - \overline{x})^{\,2}}$$

Wherein N is the number of spatial units indexed by i and j, x and $xbar$ are the variables of interest and its mean, w_{ij} is a matrix of spatial weights, and W is the sum of all w_{ij}. With the null hypothesis being no spatial autocorrelation, the expected value of Moran's I is

$$E(I) = \frac{-1}{N - 1}$$

Wherein, as N approaches infinity, the expected value approaches zero [39]. Application of the Global Moran's I statistic allows for the characterization of the pattern of features and associated attributes to be

identified as clustered, dispersed, or random. Positive values of Moran's I with significant p-values suggest clustering of features more than expected given complete spatial randomness.

Anselin's Local Moran's I was then executed to characterize clusters of settlements that possessed high and low gender-based vulnerability scores and to identify those communities that had scores statistically higher or lower than neighboring settlements, i.e., outliers. The calculations below were applied to both the Pareto ranks as an aggregate gender-based vulnerability score and subsequently to disaggregated thematic scores to lend insight into which thematic variables, such as health or security, are potentiating high/low levels of vulnerability in those communities.

Local Moran's I statistic [40] of spatial association is given as:

$$I_i = \frac{x_i - \overline{X}}{S_i^2} \sum_{j=1, j \neq i}^{n} w_{i,j}(x_j - \overline{X})$$

Where X_i is an attribute for feature i, $Xbar$ is the mean of the corresponding attribute, $w_{i,j}$ is the spatial weight between feature i and j, and

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^{n} (x_j - \overline{X})^2}{n - 1}$$

with n equal to the total number of features.

In this study, parameterization of the model included the adoption of inverse-weighted distance as the spatial conceptual model, Euclidean distance as the measurement construct, and row standardization to address any sampling bias present in the IOM settlement assessments. These parameters were applied throughout both Global Moran's I and Anselin's Local Moran's I analyses.

All mapping and geostatistical analyses were done within ESRI ArcMap 10.6 software. The study was reviewed by the Harvard T.H. Chan School of Public Health Institutional Review Board and determined to be exempt from full review given that it involves open-source data and is not explicitly human subjects research.

RESULTS

Pareto Ranking of Gender-based Vulnerability

Only 145 of the 174 settlements sampled by the IOM's Needs and Population Monitoring Survey possessed enough data to be included in the vulnerability analysis. The Pareto rankings determined by this methodology ranged from 1-6, with one being the least vulnerable and six being the most vulnerable (Table 2). Approximately 24%, or 35 settlements, were deemed 'most vulnerable' based upon Pareto rankings of the thematic elements. Thirty-nine settlements were characterized as the second most vulnerable, comprising 26.9% of the refugee communities. Only one settlement was considered 'least vulnerable,' and 15 settlements were considered second to least vulnerable, comprising 10.3% of the analyzed encampments. The remaining 55 settlements have intermediate scores of relative gender-based vulnerability.

Table 2.

	Least Vulnerable					Most Vulnerable
	1	2	3	4	5	6
Number of Settlements	1	15	29	26	39	35
Percentage	<1%	10.3%	20%	17.9%	26.9%	24.1%

The spatial distribution of the relative gender-based vulnerability scores can be seen in Figure 3. Grossly, a large geographic spread of gender-based vulnerability is evidenced throughout the study area. Visuo-spatial patterns identify higher levels of vulnerability around Cox's Bazar and the north of the region, along the Myanmar/Bangladesh border near Gundum, and in the south-east near Teknaf. Those settlements with the lowest gender-based vulnerability appear to be dispersed, but predominantly located in the south-east of the region, with intermediate scores predominating much of Ramu upazila.

Figure 3: Total gender-based vulnerability index rankings of Rohingya refugee settlements in Bangladesh wherein 1=least vulnerable and 6=most vulnerable.

This visuo-spatial intuition of the clustering of higher levels of gender-based vulnerability is ratified by the Global Moran's I spatial autocorrelation analysis in which spatial clustering is significantly noted. With a Z-score of 4.5 and a P-value less than 0.001, the spatial distribution of high and low vulnerability scores demonstrate a spatial clustering more than would be expected given complete spatial randomness.

The Anselin Local Moran's I analysis demonstrated thirty settlements that occupy 'highly vulnerable' clusters, with only five settlements comprising 'less vulnerable' clusters (Table 3). Several communities in the northwest of the study area (Figure 4) around Cox's Bazar, delineated as statistically-significant, 'highly vulnerable' clusters are, strictly speaking, the cores of each 'highly vulnerable' cluster. Two settlements in Idgar, along the northernmost part of the study region, are spatially disparate cores of two high-vulnerability clusters. Less vulnerable clusters are demonstrated in the south-east region near Kutupalong, its expansion and surrounding settlements, and near Dakshin Nhila.

Figure 4. Cluster analysis with Anselin Local Moran's I of gender-based vulnerability in Rohingya Refugee settlements in Bangladesh

Table 3.

Analysis Outcome	Number of Sites
Highly Vulnerable Clusters	30
Highly Vulnerable Outliers	8
Less Vulnerable Outliers	3
Less Vulnerable Clusters	5
Not Significant	115

Outlier communities that display statistically significant higher or lower vulnerability scores than their neighbors are demonstrated in Figure 5. Given the evidenced generalization that gender-based vulnerability is higher in the Cox's Bazar metropolitan region, the outcomes of the outlier analysis are justified, with all low-level vulnerability outliers, the settlements of Dhokkin, Somitapara, and Pahartoli, existing in that vicinity. Conversely, those settlements that are statistically more vulnerable than surrounding neighbors reside within the relatively less vulnerable regions along the Myanmar/Bangladesh border near Kutupalong.

Figure 5. Outlier analysis with Anselin Local Moran's I of gender-based vulnerability in Rohingya Refugee settlements in Bangladesh

When disaggregated into thematic components of education, WASH, security, and health, the spatial heterogeneity of hot- and cold-spots of gender-based vulnerability remains (Figure 6). Gender-based

health and security variables display significant spatial dependence. Those previously classified, composite highly-vulnerable clusters in the northwestern regions near the metropolitan area of Cox's Bazar are now defined as the most-vulnerable settlements when considering health and security indicators. Conversely, security-dependent, gender-based vulnerability is lowest in refugee settlements along the border with Myanmar. And refugee settlements identified as the centroids of low levels of health-related vulnerability clusters are in the south-east, and specifically in Nhila and Sabrang Unions along the Naf River. Education variables are demonstrated to have very few statistically significant clusters, yet educational access appears to be best along ingress points near Gundum and problematic at the Kutupalong 2E refugee site. Water, sanitation, and hygiene clusters are similarly sparse, but cold-spots are identified in the most northern and southern refugee settlements, and in the Kutub Bazar community near Cox's Bazar, where composite least vulnerable outliers were demonstrated previously.

Figure 6. Cluster analysis of gender-based vulnerability in Rohingya refugee settlements as disaggregated by thematic components, including education, WASH, security, and health

DISCUSSION

Through the application of a vulnerability index and subsequent spatial analytics to open source humanitarian data, this study demonstrated significant spatial heterogeneity of gender-based vulnerability in Rohingya refugee settlements in the southwest of Bangladesh. Clusters of high levels of gender-based vulnerability were identified in or around the town of Cox's Bazar, with clusters of low levels of gender-based vulnerability residing predominantly near the Myanmar border in or near Kutupalong camp. This finding may be attributable to many variables, including the formality of settlements, longevity of those settlements, and/or relative urban/rural contexts.

Kutupalong and Nayapara in Teknaf are the two oldest official refugee camps in Bangladesh, having been established in 1992 [2]. While population size within Kutupalong has swelled exponentially since 2017, the formality and longevity of this vast settlement likely potentiate organizations to provide resources and services by providing preexistent infrastructure when compared to more informal, less established settlements. With over 600,000 refugees, Kutupalong and its surrounding expansion camps also possess significant notoriety, given its label as the largest refugee camp in the world [41]. This status has likely lead to larger degrees of funding and programming for Rohingya refugees, in general, but also for women and girls, specifically.

Settlements in and around Cox's Bazar town are predominantly designated as 'informal,' with Rohingya often living in and amongst the host community. With over 250,000 people in the metropolitan area, Rohingya refugees in this urban environment likely face similar hardships as nearly half the world's refugee population. Outside of formal, often rural, settlements, refugees confront unique challenges such as difficulty accessing services, uncertain legal status and subsequent harassment, obstacles regarding identifying and maintaining non-exploitive livelihoods, exclusion from social security systems and community support mechanisms, and discrimination [42]. Given that these threats to wellbeing affect all refugees, the finding of high levels of gender-based vulnerability surrounding Cox's Bazar town are predictable.

Outliers of vulnerability are understandably identified within clusters of high- or low-gender-based vulnerability. Sites identified as low vulnerability outliers, such as Dhokkin, Somitapara, and Pahartoli in the Cox's Bazaar township region, and the alternately high vulnerability outliers found in the settlements near Kutupalong and southwards, are in need of critical investigation. Identifying variables that influence access to resources, humanitarian programmatic impact, and security will undoubtedly shed valuable light on how these settlements have become outliers, and how best to improve gender-based vulnerability in both present and future programming.

In this study, security and health variables appear to be the predominant drivers of gender-based vulnerability in this region. Security was evidenced as a key component to both high and low gender-based vulnerability clusters in Cox's Bazar township and near the border, respectively. While it is feasible that the high level of security-based vulnerability in Cox's Bazar township is likely secondary to settlements existing within a metropolitan area, there are also hot spots of security-based vulnerability near Teknaf. And currently, Teknaf has the highest level of drug related violence within Bangladesh with kidnapping and violence noted specifically in refugee camps [43]. Notwithstanding explanatory variables, both gender-based vulnerability themes provide researchers and programmatic organizations with valuable data to better understand and target specific gender-based programming.

Limitations

As with all vulnerability indices, this framework is subject to the critique of being too positivistic and reductionist. It does not take into account the multitude of alternate variables such as societal norms and preexistent systems of power and privilege that Rohingya women and girls face. Limitations on their involvement in certain parts of civic and public life, 'purdah' or gender segregation, mobility constraints, and decision-making inequalities [44] are likely to exacerbate vulnerability in the refugee context. Given

these societal norms, many Rohingya refugee women are considered to be living in 'extremely vulnerable conditions of insecurity' as heads of households. Furthermore, Rohingya women are often subjected to harassment, economic deprivations, and psychological, physical, and sexual violence. While clearly important to understanding the vulnerability of Rohingya women and girls, disambiguated data were not available for many of these variables. According to the Women's Refugee Commission [9], gender equality and women's empowerment indicators are being monitored through the 2019 Rohingya refugee Joint Response Plan, but these data have yet to be made open-source, and there is 'currently no structured process for analysis.' The index and methodology utilized in this study have the capacity to include additional parameters given the availability of future data if it is spatially codified.

Upcycling data

The utilization of open-source data is fraught with challenges that undermine scientific rigor, accuracy, and external validation. Indicators used by organizations collecting data may be inappropriate, incomplete, or adjacent to those critical to the research question at hand. Data collection methodologies may not follow standardized or evidenced-based methods, be they sampling methodology, best practices for survey or interview conduct, or precise geo-coding. Data collection methodologies may include processes that lead to bias. Appreciating, accounting for, and being transparent about these challenges is critical to the utilization of any research that uses secondary data. Humanitarian data is certainly rife with these issues, and the imperfections of these data are recognized. However, upcycling pre-existent, humanitarian data not only minimizes the resources required for situational awareness and critical programmatic design, but also reduces the impact on the community and the potential for harm reduction [45]. In an attempt to capitalize on these resources, this study chose to utilize secondary data in an effort to demonstrate that actionable conclusions can be produced through a methodology that intentionally combines pre-existent, freely available, open-source data.

Validation

The validation of such a methodology will prove difficult. As previously discussed, no specific construct has been academically validated to quantify gender-based vulnerability. More prudent, however, is the impact of such a method on funding, programming, and the improvement of the lives of refugee women and girls. In alternative vulnerability indices, such as the Heat Vulnerability Index [15] and the Center for Disease Control's Social Vulnerability Index [17], success has been measured by including independent variables such as morbidity and mortality, or the ability to recover from external hazards, respectively. Unfortunately, there continues to be equipoise regarding indicators that should be utilized to gauge success regarding gender-based vulnerability within refugee communities. Future research is required to

understand how vulnerability indices, and specifically gender-based vulnerability indices in refugee populations, impact not only programming but the health and wellbeing of refugee women and girls.

CONCLUSION

In this study, we propose and implement a methodology for utilizing previously captured humanitarian data to identify and spatially characterize populations in need. While specifically targeted at women and girls in the Rohingya refugee community in Bangladesh, the application of a methodology that combines vulnerability indices, Pareto ranking, and spatial analysis represents a novel technique to upcycle precious humanitarian data that may be applied to many populations defined as ‘vulnerable '. While validation is critical and difficult, by leveraging such a methodology, researchers and humanitarian actors can harness the potential of open-source data to bypass the resource-intensive process of primary data collection and create actionable outcomes to target research and programming, monitor progress, and strengthen coordination and resource-allocation in many humanitarian spheres.

LIST OF ABBREVIATIONS

GPS: Global positioning system

HDX: Humanitarian Data Exchange

IOM: International Organization for Migration

ISCG: Inter Sector Coordination Group in Bangladesh

Km: Kilometer

LISA: Local Indicators of Spatial Association

NPM: Needs and Population Monitoring

OCHA: United Nations Office for the Coordination of Humanitarian Affairs

UNFPA: United Nations Population Fund

USD: United States Dollar

WASH: Water, Sanitation, and Hygiene

DECLARATIONS

Ethics approval and consent to participate

This research (Protocol IRB19-2073) has been reviewed by the Harvard T. H. Chan School of Public Health Institutional Review Board (FWA00002642) and was determined to be exempt from the IRB process given that it involves open-source data and is not explicitly human subjects research.

Consent for publication

Not applicable

Availability of data and materials

The datasets analyzed during the current study are available in the Humanitarian Data Exchange repository, https://data.humdata.org/dataset.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding was provided for this research.

Authors' contributions

EN designed this study, assisted in statistical analysis and interpretation of the data, and drafted and revised the majority of the manuscript. DRS assisted in refining methodology, and contributed to data acquisition, processing, and analysis. PGG assisted in refining methodological design, interpreting data and results, and assisted in the revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

REFERENCES

1. Medecins Sans Frontieres. Timeline: A visual history of the Rohingya refugee crisis. Doctors without Borders, USA. 2019. https://www.doctorswithoutborders.org/what-we-do/news-stories/news/timeline-visual-history-rohingya-refugee-crisis. Accessed 22 Nov 2019.

2. Assessment Capacities Project. Rohingya influx since 1978: Background on the context in Bangladesh. 2017. https://reliefweb.int/report/bangladesh/historical-review-rohingya-influx-1978. Accessed 11 Nov 2019 cited 2019.
3. United Nations Office for the Coordination of Humanitarian Affairs. Rohingya Refugee Crisis. https://www.unocha.org/rohingya-refugee-crisis. Accessed 11 Nov 2019.
4. Sang D. One Year On: Time to put women and girls at the heart of the Rohingya response. 2017. https://www.oxfamamerica.org/static/media/files/bp-one-year-on-rohingya-refugee-women-girls-110918-en.pdf. Accessed 9 Oct 2019.
5. UN Women. Gender Brief on Rohingya Refugee Crisis Response in Bangladesh. In: UN Women – Asia-Pacific. 2018. https://asiapacific.unwomen.org/en/digital-library/publications/2017/10/gender-brief-on-rohingya-refugee-crisis. Accessed 22 Nov 2019.
6. Inter-agency Working Group on Reproductive Health in Crisis. Women and Girls Critically Underserved in the Rohingya Humanitarian Response. 2018. http://iawg.net/resource/women-girls-critically-underserved-rohingya-humanitarian-response/. Accessed 9 Oct 2019.
7. World Food Programme. Rohingya Emergency Vulnerability Assessment (REVA) –Summary Report. 2017. https://docs.wfp.org/api/documents/WFP-0000050429/download/?_ga=2.5768243.1666492886.1580884116-445744275.1580884116. Accessed 20 Sep 2019.
8. Toma I, Chowdhury M, Laiju M, Gora N, Padamada N. Rohingya Refugee Response Gender Analysis: Recognizing and responding to gender inequalities. 2018 Aug. http://hdl.handle.net/10546/620528. Accessed 9 Oct 2019.
9. Women's Refugee Commission. "We Need to Write Our Own Names": Gender Equality and Women's Empowerment in the Rohingya Humanitarian Response in Cox's Bazar: Gender Operational Review Report. 2019. https://www.womensrefugeecommission.org/resources/233-

586 gender-issues/1828-we-need-to-write-our-own-names-gender-equality-and-women-s-
587 empowerment-in-the-rohingya-humanitarian-response-in-cox-s-bazar. Accessed 9 Oct 2019.

588 10. Financial Tracking Service. Bangladesh: Rohingya Refugee Crisis Joint Response Plan 2018.
589 https://fts.unocha.org/appeals/656/summary. Accessed 22 Nov 2019.

590 11. Humanitarian InSight: Global Humanitarian Overview. https://hum-insight.info/ (2019).
591 Accessed 11 Nov 2019.

592 12. International Federation of the Red Cross: What is vulnerability? https://www.ifrc.org/en/what-
593 we-do/disaster-management/about-disasters/what-is-a-disaster/what-is-vulnerability/ (2019).
594 Accessed 11 Nov 2019.

595 13. Dickin SK, Schuster-Wallace CJ, Elliott SJ. Developing a vulnerability mapping methodology:
596 applying the Water-Associated Disease Index to Dengue in Malaysia. PLoS One. 2013 May
597 10;8(5).

598 14. Bizimana JP, Twarabamenye E, Kienberger S. Assessing the social vulnerability to malaria in
599 Rwanda. Malar J. 2015;14(1).

600 15. Nayak SG, Shrestha S, Kinney PL, Ross Z, Sheridan SC, Pantea CI, et al. Development of a heat
601 vulnerability index for New York State. Public Health. 2018 Aug 1;161:127–37.

602 16. Shrestha R, Flacke J, Martinez J, Van Maarseveen M. Environmental health related socio-spatial
603 inequalities: Identifying “hotspots” of environmental burdens and social vulnerability. Int J
604 Environ Res Public Health. 2016 Jul 1;13(7).

605 17. Flanagan, B Hallisey, E, Adams, E., Lavery A. Measuring community vulnerability to natural and
606 anthropogenic hazards: The Centers for Disease Control and Prevention’s social vulnerability
607 index. J Environ Health. 2018;80(10):34–6.

18. Sankari TS, Chandramouli AR, Gokul K, Surya SSM, Saravanavel J. Coastal vulnerability mapping using geospatial technologies in Cuddalore-Pichavaram Coastal Tract, Tamil Nadu, India. Aquat Procedia. 2015;4:412–8.

19. Chan EYY, Huang Z, Lam HCY, Wong CKP, Zou Q. Health vulnerability index for disaster risk reduction: Application in belt and road initiative (BRI) region. Int J Environ Res Public Health. 2019 Feb 1;16(3).

20. United Nations High Commissioner for Refugees. Vulnerability Assessment Framework - Population Study 2019 - Jordan. Found in: ReliefWeb. 2019. https://reliefweb.int/report/jordan/vulnerability-assessment-framework-population-study-2019. Accessed 22 Nov 2019.

21. Alam MB, Rahman KMA. Women and climate change in Bangladesh: An analysis from gender perspective. Cross-Cultural Comm, 2017; 13(8):7-9.

22. Sugden F, De Silva S, Clement F, Maskey-Amatya N, Ramesh V, Philip A, et al. A Framework to Understand Gender and Structural Vulnerability to Climate Change in the Ganges River Basin: Lessons from Bangladesh, India and Nepal. Colombo, Sri Lanka: International Water Management Institute. 2014; 159.

23. Pandey R, Jha SK. Climate vulnerability index - measure of climate change vulnerability to communities: A case of rural Lower Himalaya, India. Mitig Adapt Strateg Glob Chang. 2012 Jun;17(5):487–506.

24. Humanitarian Data Exchange: Welcome page. https://data.humdata.org/. Accessed 23 Oct 2019.

25. Adams L. Sixth Annual Health Data Liberator Award Honors Work to Increase Use and Impact of Humanitarian Data. In: AcademyHealth. 2018. https://www.academyhealth.org/blog/2018-

04/sixth-annual-health-data-liberator-award-honors-work-increase-use-and-impact-humanitarian-data. Accessed 11 Nov 2019.

26. International Organization for Migration (IOM) Bangladesh, Displacement Tracking Matrix-Needs and Population Monitoring (NPM) Round 6 Site Assessment. In: Humanitarian Data Exchange. https://data.humdata.org/dataset/needs-and-population-monitoring-npm-round-6-for-rohingya-population-in-cox-s-bazar-district. Accessed 12 Jan 2018.

27. Organization for Economic Co-operation and Development. Handbook on Constructing Composite Indicators: Methodology and UserGuide. Paris, France. 2008.

28. Al Nahian M, Islam T, Bala SK. A new approach in gender vulnerability assessment using matrix framework. 2012. https://www.researchgate.net/publication/303666930_A_NEW_APPROACH_IN_GENDER_VULNERABILITY_ASSESSMENT_USING_MATRIX_FRAMEWORK. Accessed 11 Nov 2019.

29. Organization for Economic Cooperation and Development. Social Institutions and Gender Index. https://www.genderindex.org. Accessed 25 Nov 2019.

30. United Nations Development Programme. Gender Inequality Index (GII). In: Human Development Reports. http://hdr.undp.org/en/content/gender-inequality-index-gii. Accessed 25 Nov 2019.

31. World Economic Forum. The Global Gender Gap Report 2018 Insight Report. Geneva, Switzerland. 2018.

32. Plan International India. Plan for every child: Gender Vulnerability Index Report - 1. New Delhi, India. 2018;1–96.

33. Alkire S, Friedman M, Robles Aguilar G, Valenti S, Santos ME. UNDP Human Development Report Office Training Material for Producing National Human Development Reports. March 2015.

34. Adger N, Brooks N, Bentham G, Agnew M, Eriksen S. New indicators of vulnerability and adaptive capacity. Tyndall Centre for Climate Change Research. Technical report 7. Norwich, England. 2004.

35. Rygel L, O'Sullivan D, Yarnal B. A method for constructing a social vulnerability index: An application to hurricane storm surges in a developed country. Mitig Adapt Strateg Glob Chang. 2006 May;11(3):741–64.

36. Jerrett M, Gale S, Kontgis C. Spatial modeling in environmental and public health research. Intl J Env Research Pub Health. 2010; 7:1302–29.

37. Rushton G. Public health, GIS, and spatial analytic tools. Annu Rev Public Health. 2003;24:43–56.

38. Scribner RA, Simonsen NR, Leonardi C. The Social determinants of health core: Taking a place-based approach. Am J Prev Med. 2017 Jan 1;52(1):S13–9.

39. Getis A, Ord JK. The analysis of spatial association. Geogr Anal. 1992;24(3):189–206.

40. Anselin L. Local indicators of spatial association -LISA. Geogr Anal. 2010 Sep 3;27(2):93–115.

41. United Nations High Commissioner for Refugees. Camp profiles- Rohingya refugee response. Dec 2019. https://reliefweb.int/map/bangladesh/camp-profiles-december-2019-rohingya-refugee-response-bangladesh. Accessed 20 Sep 2019.

42. United Nations High Commissioner for Refugees. Designing appropriate interventions in urban settings: Health, education, livelihoods, and registration for urban refugees and returnees. Geneva, Switzerland. 2009.

674 43. Safety and security. In: Bangladesh travel advice. GOV.UK. https://www.gov.uk/foreign-travel-
675 advice/bangladesh/safety-and-security. Accessed 2 Dec 2019.

676 44. Tay, AK, Islam, R, Riley, A, Welton-Mitchell, C, Duchesne, B, Waters, V, et al. Culture, context
677 and mental health of Rohingya refugees: A review for staff in mental health and psychosocial
678 support programmes for Rohingya refugees. United Nations High Commissioner for Refugees.
679 Geneva, Switzerland. 2018.

680 45. Data Policy. In: The Centre for Humanitarian Data. https://centre.humdata.org/data-policy/.
681 Accessed 3 Dec 2019.

682
683
684
685

Figures

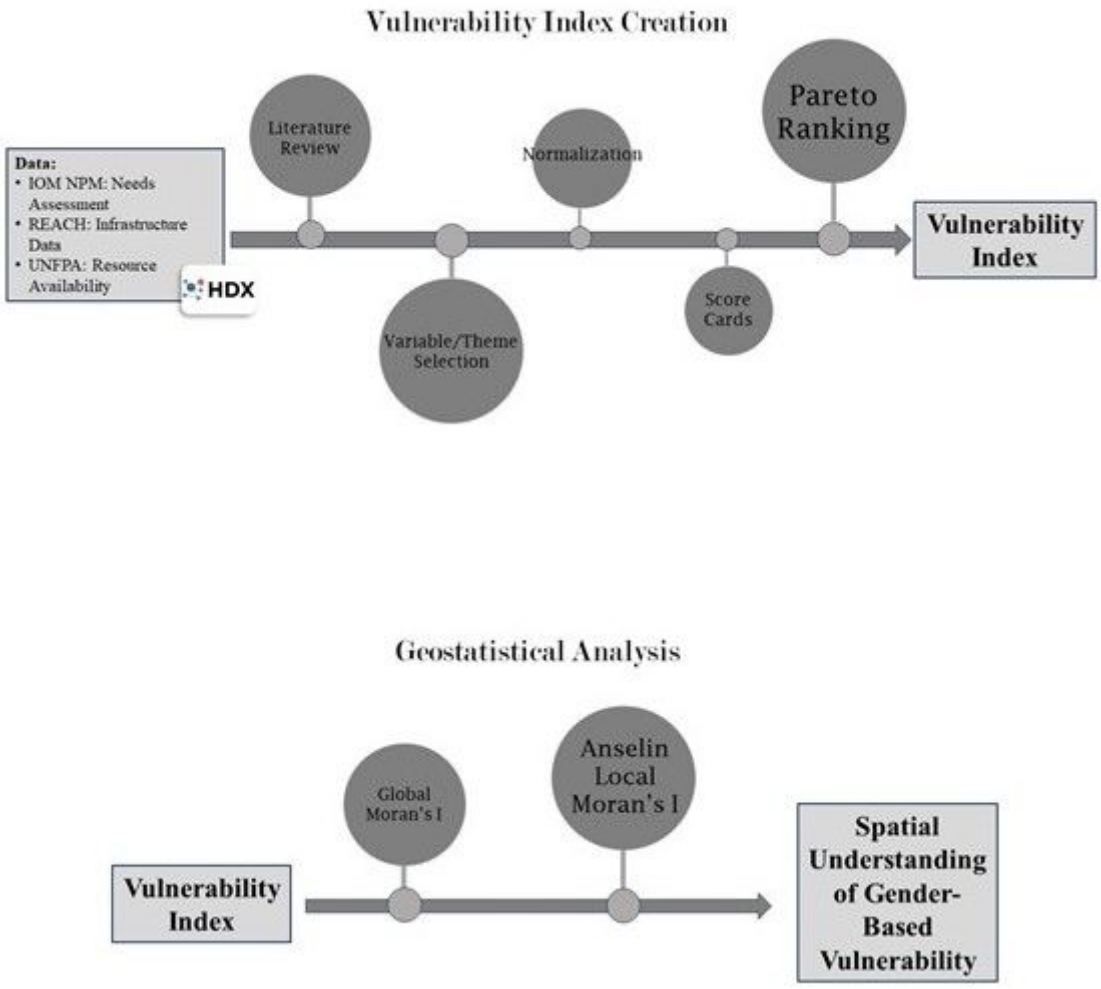

Figure 1

Methodological process for creating a gender-based vulnerability index and subsequent geostatistical analysis

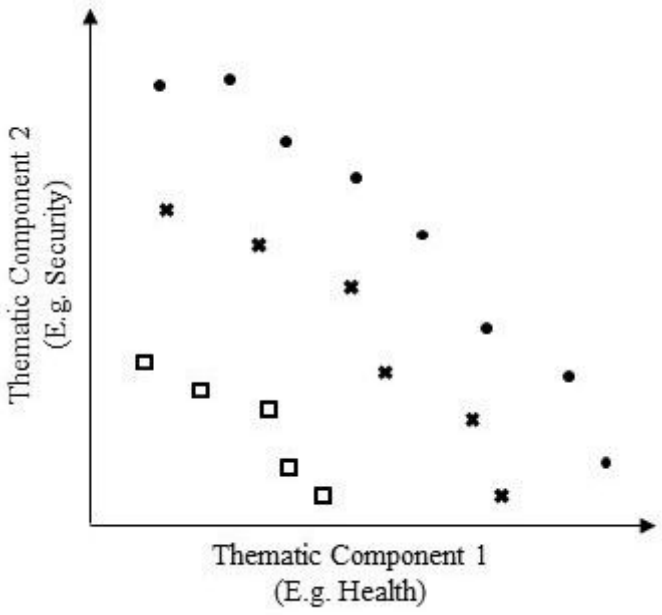

Figure 2

A two-dimensional illustration of Pareto ranking in which each data point (in our study, settlement site) has two thematic component scores. Those data points represented as circles are considered the most vulnerable rank, in that each site is non-dominated. Those represented as 'X's are second-most vulnerable, and those represented by squares are least vulnerable.

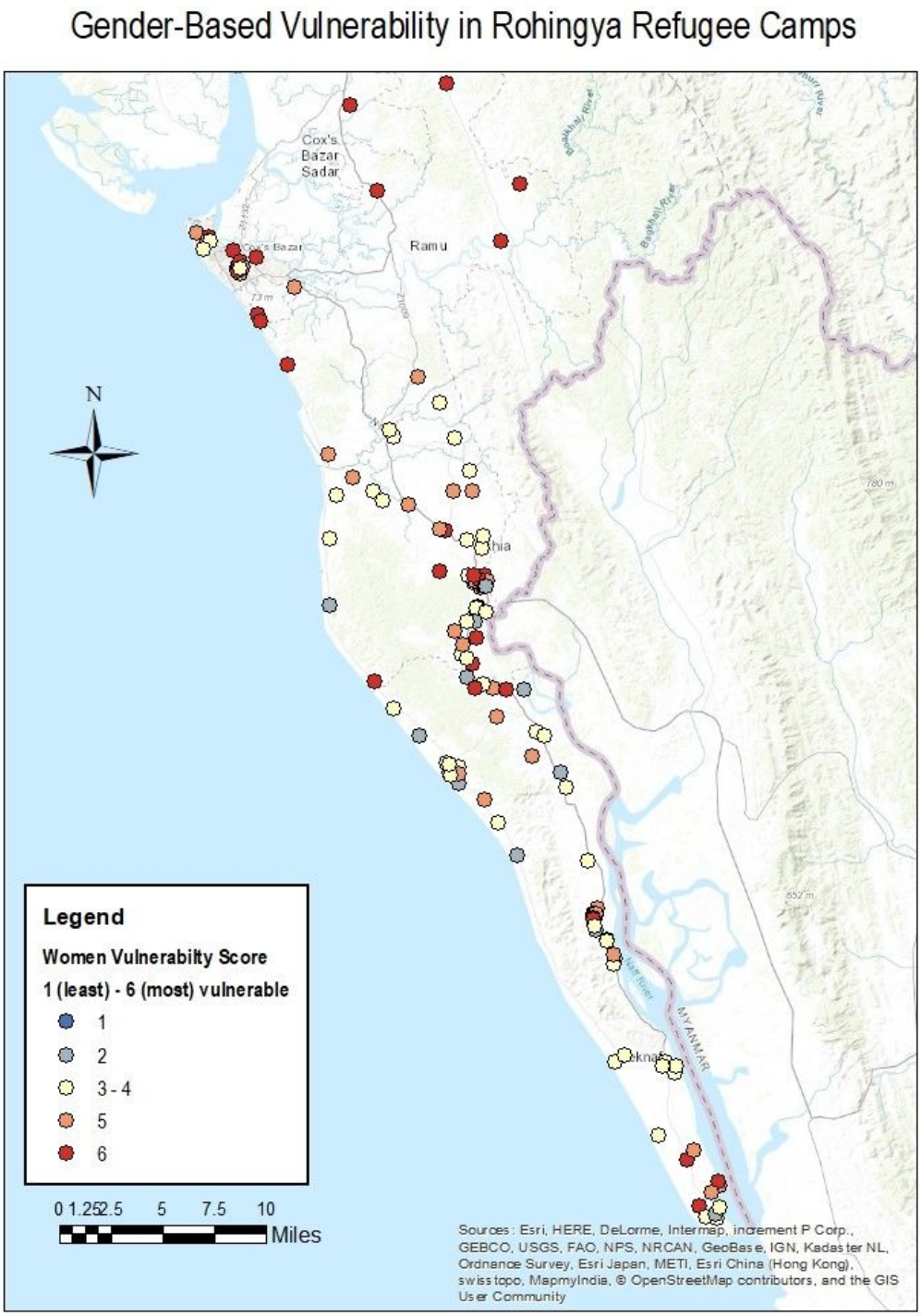

Figure 3

Total gender-based vulnerability index rankings of Rohingya refugee settlements in Bangladesh wherein 1=least vulnerable and 6=most vulnerable.

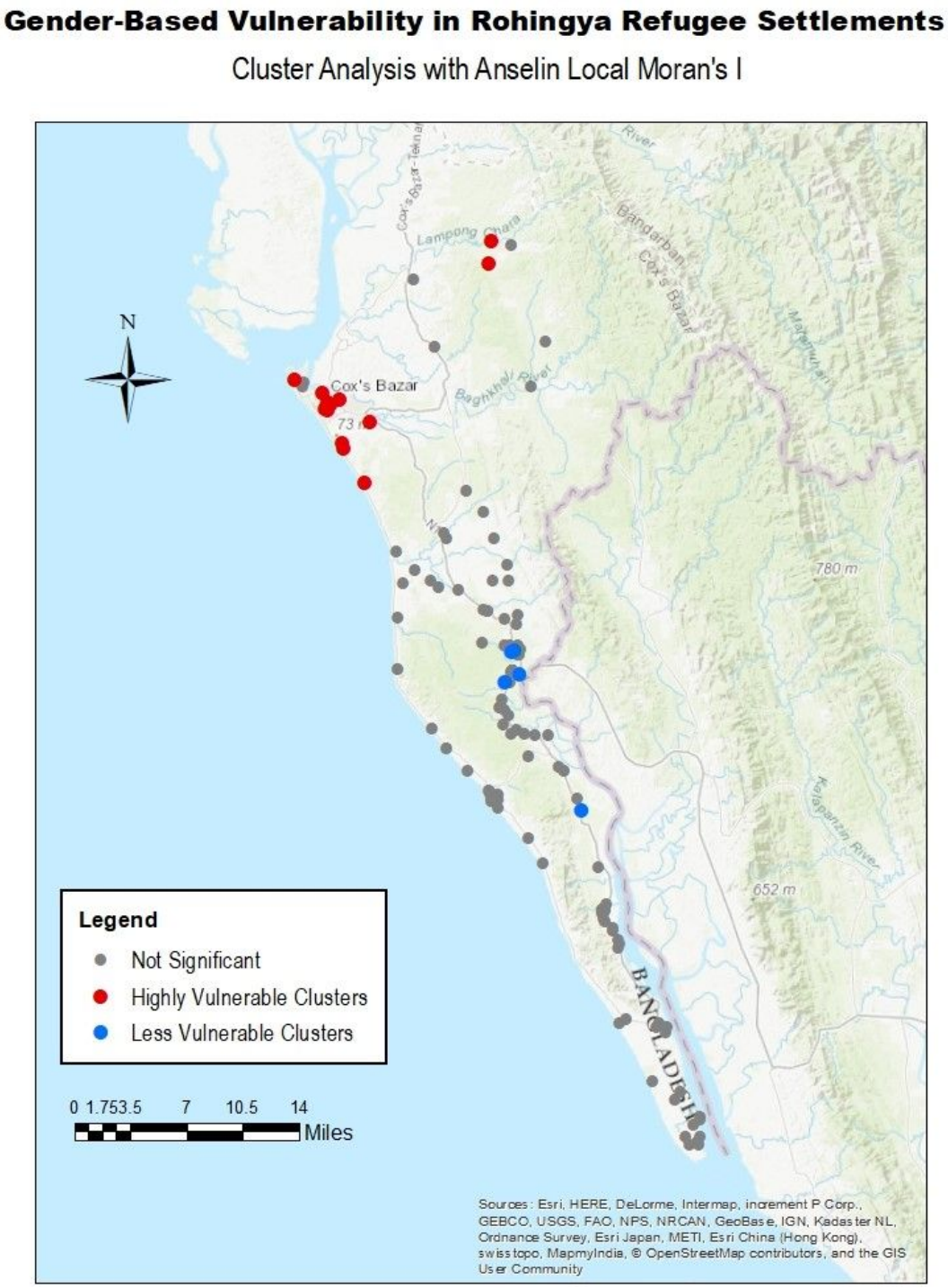

Figure 4

Cluster analysis with Anselin Local Moran's I of gender-based vulnerability in Rohingya Refugee settlements in Bangladesh

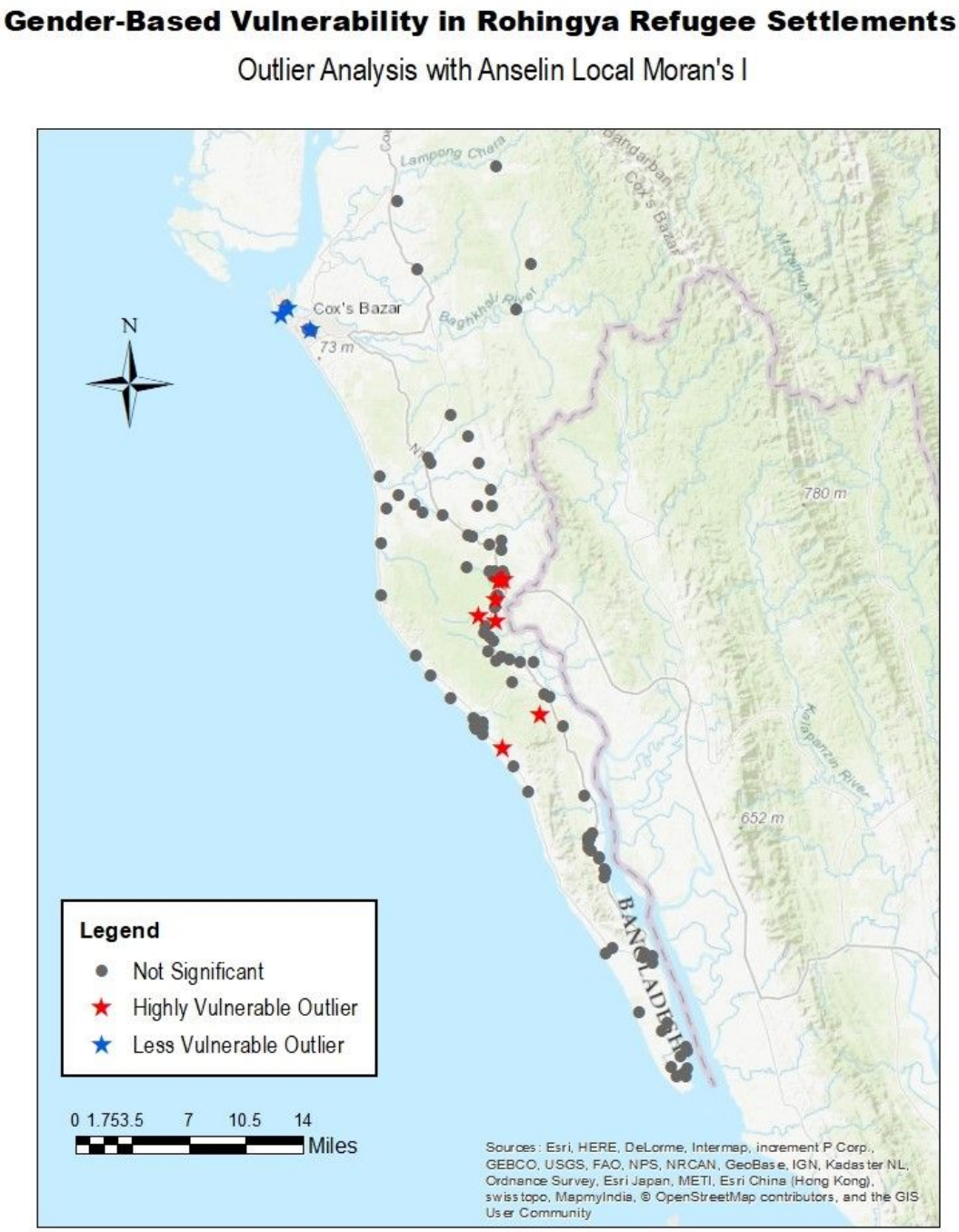

Figure 5

Outlier analysis with Anselin Local Moran's I of gender-based vulnerability in Rohingya Refugee settlements in Bangladesh

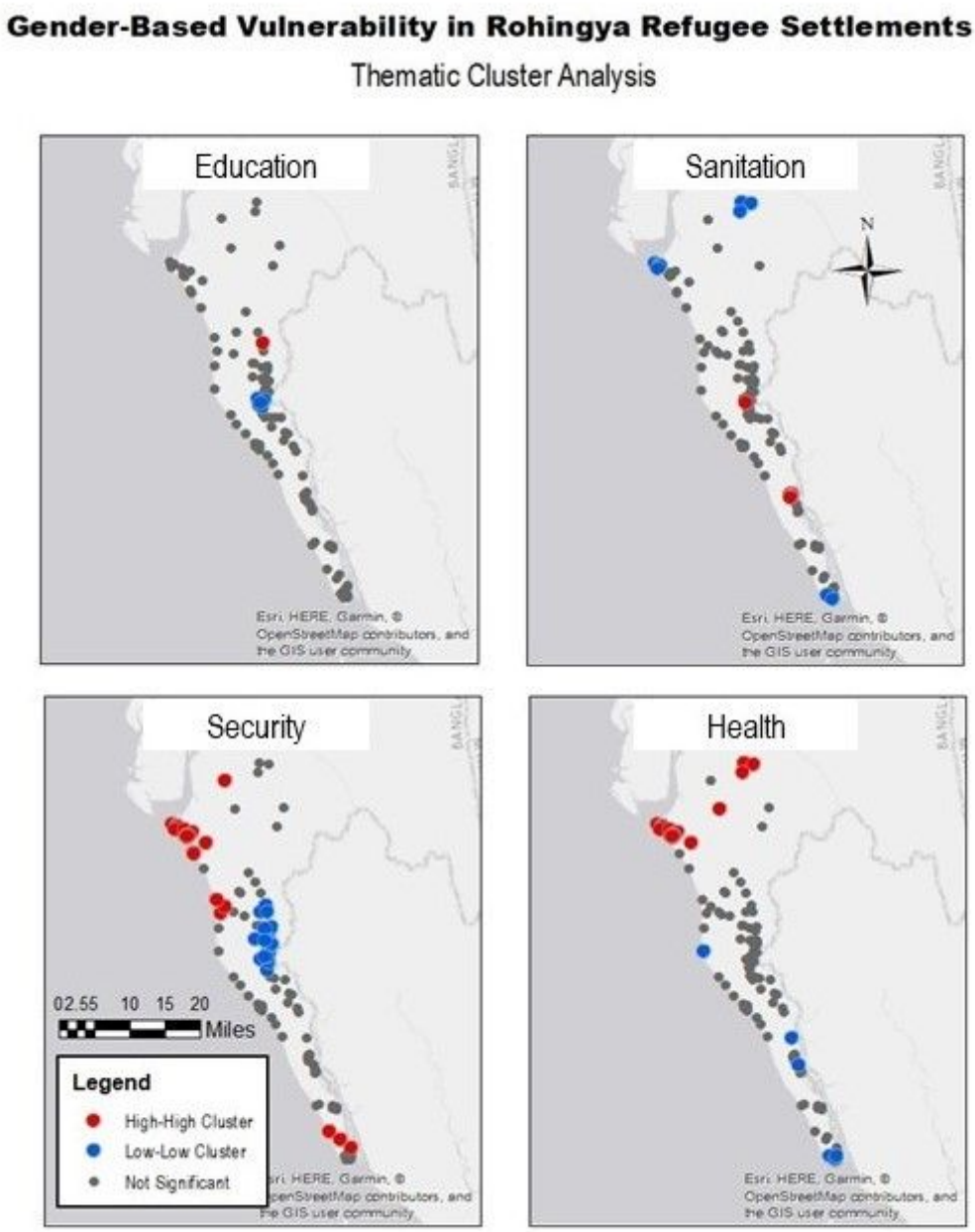

Figure 6

Cluster analysis of gender-based vulnerability in Rohingya refugee settlements as disaggregated by thematic components, including education, WASH, security, and health