Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Ettan Patel April 9, 2025

ML-based geographic sampling frames miss transitory populations in fragile regions

By

Ettan Patel

Stephen O'Connell

Adviser

Economics

Stephen O'Connell

Adviser

Neha Gupta

Committee Member

Chris Hansman

Committee Member

2025

ML-based geographic sampling frames miss transitory populations in fragile regions

By

Ettan Patel

Stephen O'Connell

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2025

Abstract

ML-based geographic sampling frames miss transitory populations in fragile regions

By Ettan Patel

Post-conflict environments often lack reliable survey data, complicating aid distribution. We studied 18 Iraqi communities, comparing machine learning and traditional methods for creating sampling frames. Using satellite imagery and Microsoft's GlobalMLBuildingFootprints, we validated and manually added points in QGIS. An on-site survey recorded building conditions and conducted interviews in inhabited residences. Across 210.20 km^2, we identified 61,603 valid buildings, visiting 1,225. Of these, 1,061 were inhabited. Comparing automated and manually-located building structures, 27.54% of manually added points were buildings inhabited by internally displaced persons (IDPs), while this rate was 20.62% for machine-learning-found structures. This statistically significant difference suggests machine learning methods for structure detection and its use to create geographic sampling frames may overlook transient populations, who are often the focus of aid and social assistance programs.

ML-based geographic sampling frames miss transitory populations in fragile regions

By

Ettan Patel

Stephen O'Connell

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2025

Acknowledgements

I would like to thank Dr. Stephen O'Connell (Emory University), who obtained project funding, for providing technical advice, supervising research, and proofreading the manuscript. I would like to thank Dr. Christopher Hansman (Emory University) and Dr. Neha Gupta (Emory University) for helping to proofread the manuscript and providing suggestions. I sincerely thank Andrea Caflisch (UCL) for establishing study conception and design, data cleaning, and technical advice. I am grateful to Dr. Daniel Masterson (UCSB), who obtained project funding and provided satellite imagery for the study. I would like to thank Guido Romero and Tianqi Zhang (Emory University) for their work in data cleaning.

Contents

1 Introduction **1**

2 Data **4**

2.1 ML-Based Building Footprints 4

2.2 Manually Placed Points 5

3 Methodology **6**

3.1 Footprints 6

3.2 Mapping 8

3.3 Geocoding 8

3.4 Sampling Method 9

3.5 Survey Administration 9

4 Results **10**

5 Conclusion **14**

ML-based geographic sampling frames miss transitory populations in fragile regions

Ettan Patel

April 9, 2025

1 Introduction

In post-conflict environments, the breakdown of administrative systems and large-scale displacement of populations create significant inefficiencies in aid distribution and economic recovery. Impacted regions suffer a loss of records and oftentimes do not have a complete enumeration of inhabitants, posing challenges for aid organizations that rely on demographic data to distribute aid and assess its impact. Vulnerable groups impacted by conflict are often displaced, leading to increased movement within an area and resulting in negative economic and social outcomes for those forced to move. A study conducted in Colombia by the Brookings Institution suggests that internally displaced people have fewer opportunities for work, often face barriers to housing, and may experience discrimination from the host community (Arredondo, 2011). Tracking displaced populations is particularly difficult because they are more mobile than other groups, often residing in camps or newly constructed areas. In developed countries, researchers can rely on established systems and telephone surveys for

sampling, but such resources are often unavailable in less developed, post-conflict regions. As a result, organizations must create sampling frames - lists of geographic points where people likely reside - before deploying survey teams. While manual sampling frame creation is time-intensive, advances in machine learning offer automated alternatives. However, concerns remain about whether these methods effectively capture vulnerable groups, such as displaced populations.

Existing epidemiology methodologies have previously used a Geographic Information System (GIS) paired with satellite imagery data to completely enumerate a population of interest to get more accurate measures of death counts or to evaluate the medical needs of refugee camps (Galway, 2012; Lin and Kuwayama, 2016). Past studies strove to diverge from the spin-the-pen method because of its tendency for bias (Lin and Kuwayama, 2016; Bauer, 2014). Additionally, manual creation of sampling frames can be costly, time consuming, and arduous, while the use of GIS involves much less time spent in the field, which is especially important for safety when conducting research in a conflict or post-conflict area such as Iraq.

When creating sampling frames using a GIS, past studies would create points containing the geographic information of each identified household structure (Escamilla, 2014), or trace a polygon over each identified building before reviewing the information and uploading it to a street map database for use by the survey team (Wagenaar, 2018). Previous methodologies also use Google maps as a base for creating their sampling frames, but indicate that they be checked with more recent satellite imagery when identifying buildings, as provided imagery may be out of date (Kamedjeu, 2009). Once sampling frames were created, survey teams were outfitted with GPS devices that would direct them to a randomly selected point from the sample frame to travel to. Upon reaching the point, the team would go to the closest domicile administer their survey. Points would be visited not in order of convenience, but by order generated, giving no consideration to distance between generated points due to possible bias that may arise.

Our team sought to capture the responses of Hosts, Returnees, and Internally Displaced

Persons (IDPs) in a randomized controlled trial (RCT) in post-war Iraq surrounding the impact of economic aid provided by the International Office for Migration (IOM). The use of satellite imagery was especially important to this study due to the recent conflict that took place in Iraq which displaced many civilians, leading to out of date census information that does not accurately describe the population. Our methodology allowed us to create an up-to-date sampling frame that may not have been possible if we had only relied on census methods or other methods that may have required our team or survey teams to spend more time in the field. We examined 18 communities within Iraq, with the set of communities ranging from rural farming communities to densely populated urban and suburban communities as well. Some communities were close to each other, while others were significantly separated from the nearest communities in the set. The use of our methodology is important as it leverages new technology which may hasten the process in which sampling frames are created, making research reliant on surveys in regions that have experienced or are experiencing conflict such as Iraq much easier to conduct.

Other researchers looking to survey economically challenged groups have faced the same problem of out-of-date census data and hard-to-reach populations in developing countries. One such methodology employed mobile phone data and satellite data concurrently to determine the distribution of poverty spatially (Steele, 2017). A more recent paper by some of the same researchers finds that this data paired with a machine learning model can identify which households are in need of aid as accurately as standard survey-based measures (Aiken, 2022). This raises the question of how machine learning might be used in a study that faces the same roadblock of insufficient census data but needs to administer a survey. We devised a method that builds on past use of GIS systems and satellite imagery by leveraging an open source machine learning model. This open-source model has already done work to identify domiciles all over the world, creating shape files and points over identified houses, allowing for faster identification of buildings for the sampling frame. We pair these pre-marked maps with imagery purchased from a satellite imaging company to create an up-to-date sampling frame of our 18 communities. The machine learning model was able to identify a large num-

ber of potential domiciles, and by concurrently using up-to-date satellite imagery to manually identify domiciles and correct points incorrectly placed by the machine learning model, we believe that we were able to accelerate the sample frame creation process. The question we seek to answer with our study is whether employing this machine learning model will allow for better identification of IDPs, who are more transient than other community members in the wake of conflict to determine whether researchers conducting studies distributing and observing the impact of aid, or conducting health studies in post-conflict environments.

The following sections will outline our methodology as well as the differences between machine-learning based creation of a sampling frame and the manual creation of a sampling frame. We will consider the implications of such differences and explore how studies based in post-conflict areas might use machine-learning aided sampling frame creation in the future.

2 Data

2.1 ML-Based Building Footprints

The footprints used for our machine learning-created model were sourced from an open source repository on GitHub (Microsoft, 2022). This repository was created by Microsoft in 2022 using imagery sourced from Bing Maps, Maxar, Airbus, and IGN France and is free for use under the Open Database License Agreement. The dataset uses maps from between 2014 and 2024, however at the time we began to use the repository, imagery was only available from 2014 to 2023, limiting how precise imagery was when trying to identify domiciles.

The repository includes scripts, examples, and images. Building extraction was done by Microsoft using neural networks to recognize pixels which may contain buildings and then converting those pixels into polygons which show up as a border around the detected building on GIS imagery. The dataset itself can be downloaded using the file: dataset-

links.csv (Microsoft, 2022). This dataset contains the country, Quadkey, url, and upload date of each polygon.

After opening this file within QGIS, our team created a Google Earth layer and footprints layer (Map Data ©2020 Google). Each identified building's polygon border was then converted to a point in another layer, preserving the original polygon border around each identified building, but giving an option of whether to view it or not while validating in GIS. Each point was then attached with information that could be edited if a user selected it, which includes the region, governorate, and whether that point had been marked for deletion.

All footprint points were initially loaded in with a green fill to denote that they had not been marked for deletion, but if a point was found to mark a location on the updated satellite imagery that did not contain a shelter, the point or set of points were selected and marked for deletion. This can clearly be seen in Figures 1 and 2, where small patches of red points denote recent developments in building state. Marking points for deletion turned them red which allowed for interesting observations to be made, such as how a large group abandoned what was once a large encampment, which can be seen in Figure 2. The manual validation of points was done to ensure that our survey team in the field did not visit a location that would definitely not contain a building. Every footprint in every community was validated in this way by a team member who would place points over structures which may have contained inhabitants as well.

2.2 Manually Placed Points

Manually placing our own points required that we use up-to-date satellite imagery rather than place points on the Google Earth layer as the Microsoft data overlaid onto the imagery supplied by the Google Satellite was not as up-to-date as was required; which is partially due to the movement of groups when displaced or in the face of conflict. The team was

sent imagery files containing multiple pictures of the community area that could be pieced together through multiple layers to cover our communities. We would then merge these imagery layers into a single raster which would then be renamed to the community that it covered in GIS.

With this up-to-date imagery we would place points on buildings or shelters that could be inhabited which were not already marked by the ML model. This meant screening out buildings which were partially destroyed, small buildings in close proximity to a stand alone home, as well as cars, trucks, and other vehicles. To do this we created a new layer in which we could place centroids in our GIS project. The points placed in this layer would be attached with a timestamp for when they were placed as well as the name of the team member that had plotted the point. This can be seen in Figures 1 and 2, where manually placed points are blue. Both figures contain regions with few manual additions as well as subsection in which the majority of points are blue, signaling recent development of an area. These points were transferred to our sampling set with the additional information attached to each point. Getting data on when our point was placed was of interest to us as we looked to audit the process and determine whether it might be a viable method of use for future researchers.

3 Methodology

3.1 Footprints

Using the GlobalMLBuildingFootprints repository on Github, we were able to convert data to a geoJSON format shapefile layer of buildings from all over our communities. While validating points generated from the Microsoft repository, validators would check that points correctly identified an intact building, and made sure that relevant buildings were not marked by the program multiple times, as it could increase the probability that a building was picked

Figure 1: **8Shibat Community**

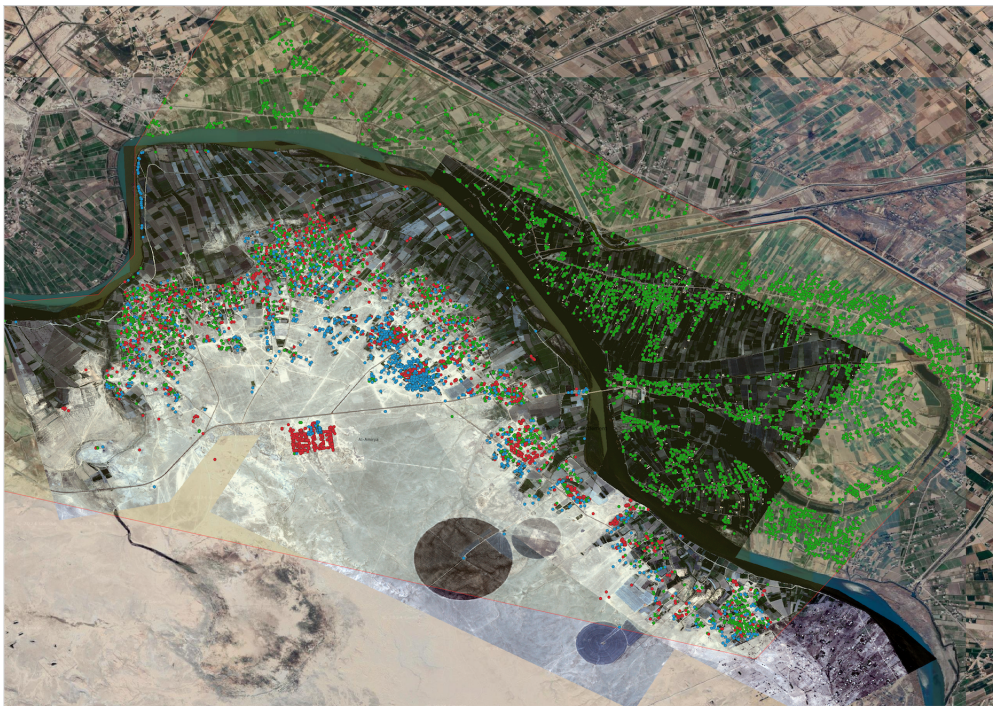

Figure 2: **BZBZ Informal Settlement**

for the survey team to visit. If points did not pass validation they would be marked for deletion, removing them from the possible pool of points for the survey team to visit.

3.2 Mapping

We used a Google Satellite imagery layer, provided by a web map service called EPSG. Additionally, we provided the coordinates as well as a radius around a central point of our communities to Planet Labs (San Francisco, California, USA), which sent us files containing an image or series of images of our communities we could then use to compare against our Google Earth layer and ML footprints, as the Google Earth and ML Footprints layers did not capture more recent developments in building construction and housing changes. Imagery files sent by Planet Labs were combined into a single raster before being used as a layer. This layer was overlaid on the Google Satellite imagery layer to help us identify new constructions as well as where footprint points were incorrectly identifying domiciles. Files and data were stored in DropBox to for ease of storage and collaboration between team members. Relevant files contained in the folders sent by Planet Labs were marked on a document in each community satellite imagery folder for record keeping, as some files contained irrelevant imagery that could be ignored.

3.3 Geocoding

The imagery provided by Planet Labs enabled the identification of buildings not identified by the Microsoft Footprints layer when overlaid on the Google Earth imagery. The updated imagery enabled us to mark buildings not included or marked in the map used by the machine learning model. We marked buildings with centroids in QGIS, and with each point included a time stamp, point number, and the name of the team member who placed the point. Additionally, we were able to check whether points laid by the Footprints layer were accurate when overlaid on our Planet Labs imagery, allowing the validation of points provided

by the Microsoft Footprints algorithm. Thus, all remaining points had a unique geospatial identification number containing longitude and latitude which would enable to survey team to arrive at a marked building or area close by.

3.4 Sampling Method

Our trial took place within an ongoing livelihoods trial funded by the IOM, which is operating a cash grant program for small-scale entrepreneurs. This is a RCT that gives a cash grant to a portion of citizens that have previously expressed interest in starting their own business. IOM contracted IIACSS, a research company based in Iraq and the only representative of GALLUP International in Iraq to conduct surveys in the field (IIACSS). After the creation of our sampling frames, randomly selected geographic points in a community of interest would be sent to IIACSS for them to visit and survey respondents. IIACSS would first visit a selected geographic point, record the state of the building represented by that point, and proceed with the survey if able.

3.5 Survey Administration

Upon visiting occupied residential buildings, the survey team used handheld Android devices to ask residents questions pertaining to gender, number of people in their household, and if they had applied for the IOM grant or if they knew anyone who had applied for the IOM grant. Responses were recorded on each enumerator's device and sent to the team, where they could review duration of interview and calculate other summary statistics to screen out any interviews that had been mistakenly recorded or to flag any responses of note.

4 Results

When creating our sampling frame, our team examined a combined area of 210.20 km^2. This area is composed of 18 communities, and initially contained 52,448 buildings identified by Microsoft's machine learning repository alongside 9,660 additionally buildings found manually. During our validation of Microsoft identified points, 505 of those of those buildings identified by Microsoft's program was marked for deletion, leading to a combined total of 61,603 viable points for our survey teams to visit. This can be seen in Table 1, which also notes that a total of 118.44 hours were spent adding and validating points.

Table 1:

Total Area	210.20 km^2
Starting Points	52,448
Deleted Points	505
Found Points	9,660
Ending Points	61,603
Hours Spent Adding Points	118.44

Total Area and Footprints. Reports the distribution of each type of geospatial point marked in the region. Data is sourced from the datasets containing additional attributes for found points, as well as datasets containing the extent of each community and new building information.

In identifying buildings in which possible respondents might live, we used machine learning and manual point validation concurrently. To quantify time spent placing points and validating points, we attached each manually placed point in QGIS with a timestamp to allow us to track the time spent in between placing points. The team member in charge of manually placing points was also in charge of validating points placed by Microsoft's ML system. This task demanded we be as thorough as possible in validating points as well as checking over communities that had been validated and checked to be sure we did not miss any buildings, which required going over communities multiple times to ensure that all points were placed correctly before being sent off to the survey group in the region. To account for this, we allowed for a maximum interval of 20 minutes between points placed to account

for time spent validating ML identified points while simultaneously manually adding points. It took around 33.6 minutes per km^2 to completely validate points placed by Microsoft's program and to manually identify buildings which the Microsoft model had not been able to mark. Time spent both placing and validating points totaled 7,106.24 minutes or 118.43 hours.

Table 2:

	t
Total Minutes Spent Placing Points	7106.24
Total Hours Spent Placing Points	118.44
Hours spent per km^2	0.56

Time spent Reports the time spent placing new points to identify buildings with a maximum of 20 minutes between point placement. Data is sourced from a dataset containing timestamps for each found point.

Of the 61,603 viable points we were left with after manually adding and validating points, 1,225 geocoded points were visited. Of the viable points visited: 23 were commercial buildings, 46 were construction sites, 23 were destroyed buildings, 31 were found to no longer have a building present, 41 were abandoned, and 1,061 were found to have residents who completed the survey. As can be seen in Table 3, the about 87% of points visited by the survey team were residential buildings. This means that our method of sampling allows for a high capture of actual domiciles which may not have been possible without the use of up-to-date satellite imagery.

Table 3:

Building State	**N**	**%**
Residential Buildings (Including formal or temporary shelters)	1061	86.61
Commercial, Industrial, or Administrative Buildings	23	1.88
Construction Sites (no one living inside)	46	3.76
Destroyed Buildings	23	1.88
Abandoned House	41	3.35
There is no building	31	2.53
Total	1225	

State of Buildings Visited. Reports the condition of buildings visited by each survey team. Data is sourced from the survey dataset recorded by each survey team.

In the set of points corresponding to willing participants, we can further divide our respondents by whether the point that was selected that led to their response corresponds to a ML marked point that was kept and not marked for deletion when validating the ML program or whether their geocoded location corresponds to a point that was found manually due to up-to-date satellite imagery. From this we can further split points by type of respondent, classifying whether the respondent is a host, returnee, or IDP. This is particularly important as our team sought to answer an overarching question of the impact of economic aid on IDPs in Iraq, and making sure that a sufficient number of IDPs were captured was imperative. From Table 4, we see that in the subset of machine learning marked buildings that are kept points, 28.64% of respondents are hosts, 50.74% of respondents are returnees, and 20.62% of respondents are IDPs. However, in the subset of found points, 16.95% of respondents are hosts, 55.51% of respondents are returnees, and 27.54% of respondents are IDPs. We see that ultimately, adding points manually led to a difference of 6.92% in IDP responses between the manually marked points and ML points. After conducting both a χ^2-test and linear regression on the different capture rates of IDPs between methods, we found that there is a statistically significant difference between the two methods. While the total number of IDPs identified by the manual method was lower, the percentage was higher at a statistically significant level, indicating that had our sampling frame been created solely through the validation of points without adding points, we would have had fewer respondents from our group of interest.

Table 4:

	Original and Kept Points (%)	**Found Points (%)**
Host	28.64	16.95
Returnee	50.74	55.51
IDP	20.62	27.54

Share of Respondent Types in Random Community Survey 1. Pearson's χ^2 test with Yates' continuity correction between found points and original and kept points for IDPs, where the number of original and kept points used is 810 and the number of found points used is 236. $\chi^2 = 14.362$, df = 2, p-value = 0.00076. 2. Linear Regression: The coefficient for Found points for IDPs has a point estimate of 0.06925 (SE = 0.03069). Statistical significance is denoted as $p < 0.05^*$.

Table 5 illustrates the percentage of responses from each individual community that corresponds to IDPs, as well as the differences between find rates between methods. Find rate difference is calculated by subtracting the percentage of original and kept points that correspond to IDPs from the percentage of found points that correspond to IDPs. The differences between find rates allows us to divide the communities into two groups, those with a high find rate and those with a low find rate. We see that each group contains nine of our 18 communities, with the differences in the high find rate group ranging from 5.71 to 50.76, while the differences in the low find rate group range from -11.20 to 1.15. A negative find rate difference means that a higher percentage of original and kept point led to the identification of IDPs than found points in a given community. An example of a community with a high find rate is 8 Shibat, in which all found points led to IDPs, while only 49.23% of original and kept points led to IDPs. Conversely, an example of a community with a low find rate is Rajm Hadid, in which 28.57% of found points led to IDPs, while 34.78% of original and kept points led to IDPs. By running two linear regressions, one for the communities with a high find rate and one for communities with a low find rate, we can see the impact of found points within each group and whether the differences between found and original points are significant. For the high find rate group, we see that the estimate for found points is 0.14798, with a statistically significant $p < 0.001^{***}$. This means that found points have a 14.798% higher chance of leading to an IDP than an original point in the set of communities with a high find rate. For the low find rate group, the estimate for found points is -0.08604, with a $p = 0.0531$, which is not statistically significant at the $p < 0.05^{*}$ level. Communities with a high difference between the percentage of IDPs identified through found versus original and kept points shows that found points have a statistically significant association with IDPs. In contrast, communities with a low or negative difference between the percentage of IDPs identified through found versus original and kept points do not show that found points have a greater association with IDP identification and find a similar amount compared to original and kept points.

Table 5:

Region	Original and Kept (%)	Found (%)	Find Rate Difference
Total	20.62	27.54	
8 Shibat	49.23	100.00	50.76
Markaz Tooz-Hay Al Askari	28.13	50.00	21.88
Tooz-Hay Al Teen	35.90	50.00	14.10
Al Tajneed Qtr	15.38	27.78	12.39
Nahda Sharqya	0.00	9.09	9.09
Rabeaa and Samma	7.04	14.29	7.24
Al Ajelyah Village	5.36	12.50	7.14
Hay Tal Baajah	8.77	15.38	6.61
BZBZ Informal Settlement	80.00	85.71	5.71
Al Shehabbi Village	3.85	5.00	1.15
Yaramjah and Hay Somar	49.12	50.00	0.87
Duguri	0.00	0.00	0.00
Al Teneraa	7.69	5.56	-2.14
Yangija Village	2.99	0.00	-2.99
Gaza	3.13	0.00	-3.13
Al-Obaidy 1	13.33	10.00	-3.33
Rajm Hadid	34.78	28.57	-6.21
Markez Baaj	25.49	14.29	-11.20

Percent of IDPs between each region. 1. Linear Regression (High Find Rate Difference): Coefficient for Found points is 0.14798 (SE = 0.04156). $p < 0.001^{***}$. 2. Linear regression (Low Find Rate Difference): Coefficient for Found points is -0.08604 (SE = 0.04437). $p < 0.1$. Data is sourced from a merged dataset combining survey data and manually added points.

5 Conclusion

Historically, using GIS in the creation of sampling frames has been advantageous when compared to traditional sampling methods, especially in environments impacted by conflict. The use of GIS limits observer selection bias as well as geographic sampling bias from clustering (Lin and Kuwayama, 2016). This means that the use of satellite mapping allows for houses not easily accessible to be surveyed, while the same cannot be said for traditional methods. Additionally, the use of a number generator to select geographic points allows for each building in a community to have an equal probability of being chosen while tra-

Figure 3: **Closeup of BZBZ Informal Settlement**
Left: Without up-to-date SkySat Imagery. Right: Including recent SkySat imagery

ditionally, more densely populated areas of our chosen communities might have had more surveys administered due to an already high likelihood of the survey team being sent to the area. While the use of GIS enables the creation of the sampling frame to be done in a safe environment, it also allows the deployment of the survey to be done rather quickly, as teams can be sent out to a community as soon as validation is complete, while this can take much longer in the field with numerous roadblocks.

Our percentage of residential buildings visited was high, at 87%. While we did validate the buildings marked by the ML program using up-to-date imagery, we see that only 505 of the 52,448 identified points were marked for deletion, which is less than 1% of ML identified points. We also see that manually placed points make up around 15% of total points. The additional manually added points are not insignificant as they contribute a higher percentage of IDPs in each region than the original and kept points. IDPs are a group of interest due to their historic economic outcomes compared to other groups such as hosts and returnees. With less access to housing and job opportunities, the aid distributed by our overarching study to start small businesses is important as they assimilate to their new communities and build a new life.

It is important to note why the 505 points marked for deletion were designated as such. Many points were marked for deletion due to a building being over-marked. This can clearly

be seen in the right image of Figure 3, where many red points are marking a building already marked by another point. Other points can be seen that do not accurately mark a building or mark tracts of land or roads near buildings. Much of the clustering on a single building can be seen in areas where buildings are close together, where people live in more compact spaces. This means that if points had not been validated using up-to-date imagery and the ML dataset had been sent to survey teams without any validation, there would have been problems with randomly selecting geographic points to visit as some buildings would have a higher probability of getting selected, which would compromise the integrity of the randomly controlled trial.

The advantages offered by our use of a machine learning model in the creation of our sample are that a great number of points are already identified from the outset, while past researchers had to manually place all of their points. Despite this, there is a statistically significant difference between the percentages of IDPs captured through manual versus ML-assisted sampling frame creation. The manual method of creating a sampling frame yields a higher percentage of IDPs, which is important as researchers strive to obtain responses from a group that transient at higher rates than others in a given community who might not be in the same location in a year, especially in the wake of conflict. This makes sense, as the manual method uses more up-to-date satellite imagery that can be replaced with relative frequency while the imagery used for the machine learning model may not be updated with as much frequency. Figure 3 perfectly illustrates this, as numerous new buildings, and thus respondents, were identified in just a small area due to a difference in just a few years.

This paper describes the use of machine learning in sampling frame creation and compares it to the traditional manual method to administer surveys for a RCT on the economic and social impact of monetary aid on communities in Iraq. We find that there is a significant difference between the capture rate of IDPs between methods. While this methodology is feasible and involves the use of multiple open-source tools, such as QGIS, a free GIS application, and the Microsoft Footprints repository, investigators should acknowledge that

due to the nature of some groups, for best rate of response it may be best to stick to traditional GIS methods. While conducted for economic evaluation, this methodology is not limited in application and can be used by a wide variety of disciplines that have a need for sample creation in potentially dangerous environments, such as for the tracking of malaria outbreak or administering medical assessments to refugees, though it is important to remember that points should be validated due to problems with the randomly controlled trial that might arise due to mistakes made by the ML algorithm, even in areas with less transient populations where there is not much movement into or within a community of interest.

References

Bellue S. Karlan D. Udry C. Blumenstock J. Aiken, E. Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, 2022.

Salcedo J. Lopez C. Arredondo, C. The effects of internal displacement on host communities. *The Brookings Institution*, 2011.

J. Bauer. Selection errors of random route samples. *Sociological Methods and Research*, 2014.

Emch M. Dandalo L. MIller W.C. Martinson F. Hoffman I. Escamilla, V. Sampling at community level by using satellite imagery and geographical analysis. *Bulletin of the World Health Organization*, 2014.

Bell N. Hagopian A. Al Shatari S. Burnham G. Flaxman A. Weiss W. Rajaratnam J.-Takaro T. Galway, L. A two-stage cluster sampling method using gridded population data, a GIS, and Google Earth^{TM} imagery in a population-based mortality survey in Iraq. *International Journal of Health Geographics*, 2012.

IIACSS. IIACSS website. https://iiacss.org/. Accessed: 2025-02-16.

R. Kamedjeu. Tracking the polio virus down the congo river: a case study on the use of Google Earth^{TM} in public health planning and mapping. *International Journal of Health Geographics*, 2009.

Y. Lin and D. Kuwayama. Using satellite imagery and GPS technology to create random sampling frames in high risk environments. *International Journal of Surgery*, 2016.

Microsoft. Globalmlbuildingfootprints. https://github.com/microsoft/GlobalMLBuildingFootprints, 2022. Accessed: 2025-01-22.

Steele J. Sundsoy P. Pezzulo C. Alegana V. Bird T. Blumenstock J. Bjelland J. Engo-Monsen K. de Montoye. Y. Iqbal A. Hadiuzzaman H. Lu X. Wetter E. Tatem A. Bengtsson L. Steele, J. Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface*, 2017.

O. Asbjornsdottir K. Akullian A. Manaca N. Chale F. Muanido. A Covele A. Michel-C. Gimbel S. Radford T. Girardot B. Sherr K. Wagenaar, B. Augusto. Developing a representative community health survey sampling frame using open-source remote satellite imagery in mozambique. *International Journal of Health Geographics*, 2018.