## Gro for GooD – Household survey data collection methodology and processes

**Ethical Procedures**

Prior to data collection, research permits and approvals were obtained from the Government of Kenya’s National Council of Science and Technology and Oxford University’s Central University Research Ethics Committee. Disclosure assessment was conducted as part of the ethical procedures by Oxford University.

In terms of ethical considerations, participation in the data collections campaigns was voluntary. Information about the study and its objectives was provided in all cases. Consent was obtained verbally (data collection campaigns were conducted in areas of low literacy in Kwale County). Consent discussion contained information about 1) the right to decline participation or withdraw participation at any point in time, 2) confidentiality procedures, 3) purposes of the research study structure, and 4) contact details for follow-up information. All data collected have been and will be treated confidentially and anonymised for archive.

## Monitoring socio-economic status

Based on a data collected during the waterpoint census, a longitudinal panel study was started in October 2013. A sample of 531 handpump locations was used as a sampling frame for three rounds of household surveys in 2013/14 (November-January), 2015 (March-May) and 2016 (September-November). GSM-enabled transmitters (Thomson et al., 2012) were installed on 300 operational handpumps to provide daily usage data. The survey generated a comprehensive dataset capturing information on a) demographic characteristics, b) socio-economic status of the household, c) household health status, d) main and secondary household water sources, e) waterpoint management, f) water payment, g) water resources management as well as h) governance and political engagement for each household.

### Sampling Strategy and Data Collection

For the first survey, a stratified random sample of households was selected within the service area of each of the 531 handpumps. In total, 3,361 households were surveyed. An average of six households was randomly selected in the vicinity of each pump (4.6 residents per household). Typically, between six and ten households were interviewed at handpumps that were functional at the time of interviewing or had been functional at some point in the previous 12 months. Typically, four to five households were interviewed at handpumps that had been non-functional for more than one year. In order to randomly select participating households, a sketch map of all dwellings within the estimated waterpoint service area was first drawn by an enumerator in consultation with a local community member. Each household was allocated a number, and the households were then chosen using a random number generator application installed on a tablet device. All the households surveyed were geo-referenced for mapping purposes. The survey/questionnaire took between 45 minutes to one hour to complete.

The sample sizes are as follows: wave 1: n=3,361; wave 2: n=3,567; wave 3: n=3,542. Attrition was due to households moving away and non-responses or refusal to participate a second time, though the majority (97%) of the households were successfully resampled. After data quality checks and cleaning a core sample of 3,234 surveys were analysed. Given the variation in livelihood systems from inland, remote communities to the more densely-populated coastal strip we report both on aggregate results and on a simplified typology of three economic geographies in the area: (1) the southern, coastal belt with people living within a 5 km strip of the sea (52% of the sample size), their main socio-economic activity is fishing; (2) inland and more remote areas below the Shimba Hills and away from the coastal margin (37% of the sample size), their main socio-economic activity is farming; and (3) the small town of Ukunda/Diani which largely serves the tourism industry along Diani beach (10% of the sample size).

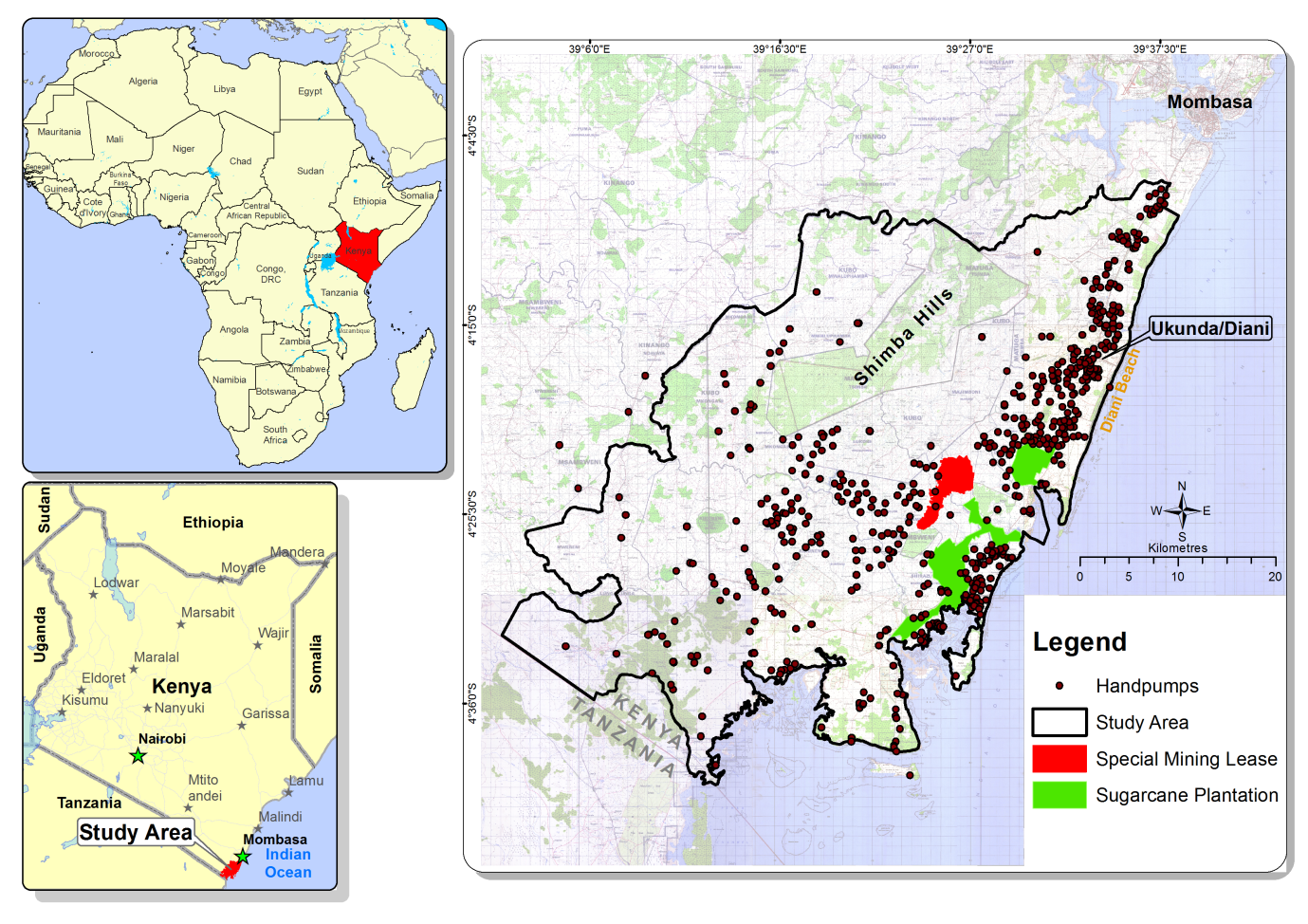


Figure 1 Study area with handpumps indicating sampling activities

### Enumerator Training and Piloting

We recruited between 19 and 25 local enumerators spanning the study area for the three rounds, who demonstrated experience in survey work and had completed secondary education or had a college degree. One key criterion was that they were able to conduct the survey in the local languages (Swahili, 53.8%; Digo, 42.6%; Duruma, 2.1%; other, 1.5%). The survey instrument was translated into Swahili. Due to local circumstances (a Muslim dominated culture), the majority of enumerators were male.

For each survey wave, the enumerator training had several components: a) providing a background about the purpose of the research, b) discussing all survey questions in detail to ensure general agreement among the enumerators, c) translating the survey questions into the tribal languages to ensure cohesion for the delivery, d) training usage of the electronic tablets, and e) discussing sampling strategy.

Enumerators were split into the three groups listed above – each of which was led by a team leader. These team leaders were trained separately to a) manage survey logistics, b) ensure the sampling strategy was followed, c) oversee survey delivery and d) conduct water quality analysis.

One area of the wider study area was designated as the pilot area and any issues with the survey instrument or the sampling strategy were addressed then. A follow-up training was conducted and then the delivery of the survey began. At the beginning of each wave, a repeat training and piloting of the instruments were conducted.

### Data management and quality control

For the delivery of the survey the software [doforms](http://www.doforms.com) was used, which allowed the survey forms to be uploaded to an online platform and managed from Kenya and remotely. All surveys conducted throughout the day were uploaded every evening (on average around 100) to avoid data loss. The team examined all collected data on a daily basis to ensure the quality of data entry and responded immediately if any patterns of data inconsistency arose. These were discussed at weekly meetings with the enumerators. Incentives for best performance were provided.

Post-survey data management and cleaning included deleting cases if they were incomplete, inconsistent, or there was a suspicion of data falsification. Through this process less than five per cent of the sample was removed from the final dataset.

Data are shared on ESRC ReShare site (UK Data Archive) once the project is completed.