1. **Data description**

Human\_data.csv

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Description** | **Categories** |
| HomesteadID | Homestead ID |  |
| HumanID | Individual ID |  |
| Age | Age group | 5 to 910 to 1415 to 1920 to 2425 to 2930 to 3435 to 3940 to 4445 to 4950 to 5455 to 5960 to 6465 to 6970 to 7475 to 7980 to 8485+ |
| Sex | Gender | MaleFemale |
| H\_Taenia\_Cyst | HP10 antigen ELISA result | 0 = negative1 = positive |
| EthnicGroup | Ethnic group | JapadholaKalenjinKikuyuKisiiKuriaLuhyaLuoMugandaPokotSaboatSamiaTesoTurkanaUgandan |
| Religion | Religion | BaptistMuslimPentacostalProtestantRoman CatholicOther ChristianTribal religionOtherNone |
| Education | Level of education attained | No formal educationPre-schoolPrimarySecondaryCollegeTertiaryVocational/technical schoolUniversity |
| EatBeef | Individual consumes beef | YesNo |
| EatPork | Individual consumes pork | YesNo |
| PorkFreq | Frequency of pork consumption | NeverOnly on special occasionsAt least once per yearAt least once per monthAt least once per weekDaily |
| LatrineFreq | Frequency that individual makes use of the latrine | AlwaysFrequentlySometimesNever |
| KeepPigs | Household keeps pigs | YesNo |
| PigN | Number of pigs owned by the household |  |
| WellDry | Does the household use well water during the dry season? | YesNo |
| RoofDry | Does the household use water collected from the roof during the dry season? | YesNo |
| RiverDry | Does the household use river water during the dry season? | YesNo |
| PipedDry | Does the household use piped water during the dry season? | YesNo |
| Dam\_pondDry | Does the household use dam or pond water during the dry season? | YesNo |
| BoreholeDry | Does the household use borehole water during the dry season? | YesNo |
| SpringDry | Does the household use spring water during the dry season? | YesNo |
| WellWet | Does the household use well water during the wet season? | YesNo |
| SpringWet | Does the household use water collected from the roof during the wet season? | YesNo |
| RoofWet | Does the household use river water during the wet season? | YesNo |
| RiverWet | Does the household use piped water during the wet season? | YesNo |
| PipedWet | Does the household use dam or pond water during the wet season? | YesNo |
| Dam\_pondWet | Does the household use borehole water during the wet season? | YesNo |
| BoreholeWet | Does the household use spring water during the wet season? | YesNo |
| VillageFlood | Recent village flooding | YesNo |
| LatrineCompound | Presence of a latrine in the compound | YesNo |
| LatrineType | Type of latrine present in the compound | Open pitPartially closedCompletely closed |
| EvidenceLatrineUse | Evidence the latrine is in use | YesNo |
| AgricGrass [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is crops and grassland |  |
| Flooding [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is flooding land |  |
| FloodingAgricGrass [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is flooding crops and grassland |  |
| Swamp [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is swamp |  |
| WoodlandShrub [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is woodland and shrubs |  |
| Vegetated [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is vegetated |  |
| Waterbody [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is waterbody |  |
| Sand [[2](#_ENREF_2)] | Sand content of soil (%) |  |
| pH [[2](#_ENREF_2)] | pH (measured in water) |  |
| TMean [[3](#_ENREF_3)] | Mean temperature (°C \* 10) |  |
| Precipitation [[3](#_ENREF_3)] | Precipitation (mm) |  |
| Elevation [[4](#_ENREF_4)] | Elevation (m) |  |
| PopDens [[5](#_ENREF_5)] | Population density (people per 100 m x 100 m cell) |  |

Pig\_data.csv

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Description** | **Categories** |
| homestead\_id | Homestead ID |  |
| pig\_id | Individual (pig) ID |  |
| P\_Taenia | HP10 antigen ELISA result | 0 = negative1 = positive |
| Origin | Origin of pig | Internal (bred in homestead)External (e.g. purchased elsewhere) |
| Gender | Gender | MaleFemale |
| Breeding | Breeding status | MaleYes (breeding female)No (non-breeding female) |
| AgeGrp | Age group | <4 months4 to 12 months13 to 24 months25 to 48 months |
| Sheep | Household owns sheep | YesNo |
| Goats | Household owns goats | YesNo |
| VillageFlood | Recent village flooding | YesNo |
| SellPiglets | Household sells piglets | YesNo |
| RaiseForMeat | Household raises pigs for meat | YesNo |
| Latrine\_Compound | Presence of a latrine in the compound | YesNo |
| Latrine\_Type | Type of latrine present in the compound | Open pitPartially closedCompletely closed |
| Latrine\_Use | Evidence the latrine is in use | YesNo |
| Latrine\_PigsScav | Evidence of pigs scavenging around latrine | YesNo |
| AgricGrass [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is crops and grassland |  |
| Flooding [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is flooding land |  |
| FloodingAgricGrass [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is flooding crops and grassland |  |
| Swamp [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is swamp |  |
| WoodlandShrub [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is woodland and shrubs |  |
| Vegetated [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is vegetated |  |
| Waterbody [[1](#_ENREF_1)] | Proportion of land within 1 km of homestead which is waterbody |  |
| Sand [[2](#_ENREF_2)] | Sand content of soil (%) |  |
| pH [[2](#_ENREF_2)] | pH (measured in water) |  |
| TMean [[3](#_ENREF_3)] | Mean temperature (°C \* 10) |  |
| Precipitation [[3](#_ENREF_3)] | Precipitation (mm) |  |
| Elevation [[4](#_ENREF_4)] | Elevation (m) |  |
| PopDens [[5](#_ENREF_5)] | Population density (people per 100 m x 100 m cell) |  |

1. **Data creation**
	1. **Parasitological survey, questionnaire and diagnostic methods**
		1. **Survey protocol**

A clustered sampling design was used, with 416 households selected randomly from the study region: the number of households selected per sub-location (the smallest administrative area in Kenya) was weighted by the cattle population of each sub-location (i.e. more households selected in sub-locations with more livestock), resulting in between one and ten households per sub-location. Spatial points were generated randomly within each sub-location, and the closest household was selected for inclusion in the study. The precise geographical coordinates for the study households were recorded using a handheld global positioning system. Following a procedure of informed consent, all humans older than 5 years (excluding women in the third trimester of pregnancy) and all pigs (excluding pregnant animals in their third trimester and suckling piglets) within the household were included as study participants and stool (for humans) and blood (for humans and pigs) samples were obtained. Questionnaires were also administered for all households relating to household level factors (e.g. water sources, livestock keeping; answered by a single individual per household), individual human level factors (e.g. demographic information) and individual pig level factors (e.g. age, sex).

* + 1. **HP10 Ag-ELISA for detection of cysticercosis infections**

Human and pig blood samples were tested using an HP10 ELISA, which detects circulating *Taenia* spp. antigens [[6-8](#_ENREF_6)]: individuals (humans and pigs) were recorded as having *Taenia* spp. antigens in sera (which is suggestive of cysticercosis) where the HP10 ELISA gave a positive result. For the purposes of analyses, the survey data were recorded as binary outcomes (positive or negative) for each outcome.

This procedure was carried out at the International Livestock Research Institute (ILRI), Nairobi. The HP10 antigen was detected by Ag-ELISA as described previously by Harrison *et al* [[9](#_ENREF_9)]. The procedure was as follows: 100µl of a 10µg/ml solution of 50% saturated (NH4) 2S04 precipitate McAb HP10 diluted in carbonate-bicarbonate buffer 9.6 pH (Sigma C3041) was added to each of the wells of a flat bottomed Immunlon® 4HBX ELISA plate and the plate was incubated overnight at 4°C. The wells were washed out twice with washing solution (0.9% NaCl with 0.05% Tween®20 (Sigma P1379)), 200µl of phosphate buffered saline (PBS) 7.3 pH (Sigma P4417) / 1% Bovine Serum Albumin (BSA) (Sigma A4503) / 0.05% Tween®20 was added to each of the wells and incubated at room temperature for 1 hour to block any non-reacted sites on the plate. The plates were then washed three times with washing solution. 100µl of undiluted sera was then added to each well, with each sample running in duplicate. The plate was incubated for 1 hour at 37°C, followed by emptying and washing three times. Biotinylated-McAb diluted 1:2,500 in PBS/BSA/Tween®20 was added at 100µl/well, covered and incubated for 1 hour at 37°C followed again by washing three times. Streptavidin Peroxiase (sigma S5512) conjugate diluted 1:10,000 (ie 0.1µg/ml) in PBS/BSA/Tween®20 added at 100µl per well, covered and incubated for 1 hour at 37°C. After a further three washes 100µl 3,3’, 5,5’- Tetramethylbenzidine substrate (Sigma T8665) substrate was then added to each well and incubated at room temperature for 15 minutes. The reaction was then stopped with 100µl of 0.2M H2SO4 per welland read at 450nm on an ELISA plate reader.

Five negative controls were run on each plate and these were obtained from an indoor farrow- finish unit in the UK, a country that is free from *T. solium* cysticercosis. For screening of human samples negative controls were obtained from the PI of this project. Two positive controls were also run for each plate and these were obtained from experimental infections of cattle with *T. saginata* provided by Dr Leslie Harrison.

A correction factor as provided by Dr Leslie Harrison (per. comms. LJSH) was used to correct for plate-to-plate and day-to-day variations, this correction factor refers to the first plate of a screening run performed with the same negative and positive control sera. Correction factor = P0-N0/Pt-Nt where

P0 =mean of positive control sera plate 1

N0  = mean of negative control sera plate 1

Pt = mean of positive control sera test plate

Nt = mean of negative control sera test plate

Applying the correction factor to every samples optical density (OD) provides a standardisation of samples across the full run.

Cut-off values were determined by using the mean corrected OD of all negative controls run during the full screen plus three standard deviations. Any sera sample with a corrected OD value over this cut off was counted as being positive.

* 1. **Additional variables**

A land cover map was derived from remote sensed satellite imagery (available from <http://eprints.soton.ac.uk/383135/>) [[1](#_ENREF_1)]. The land cover classification was carried out in a hierarchical manner, resulting in two overarching classes (a) vegetated land (versus built up and bare ground) and (b) flooding land; and four lower level classes (c) agricultural land and grassland, (d) swamp, (e) trees and shrubs, and (f) water bodies. An additional, aggregate class, (g) flooding agricultural land and grassland, was created via an overlay of the flooding land with the agricultural land and grassland classes. Buffers of 1 km were created around each study homestead and the percentage of each land cover type within the buffer was calculated using ArcMap 10 (ESRI, Redlands), to represent the landscape surrounding each homestead. This buffer size was selected to ensure the full landscape in which humans and pigs may be exposed to *Taenia* spp. eggs was included. Previous research has indicated that free-ranging pigs (in western Kenya) move an average of over 4 km per day [[10](#_ENREF_10)], and that human short-term daily average trip distances (in neighbouring Uganda) in a rural village setting are approximately 3 km [[11](#_ENREF_11)].

Temperature (minimum, maximum and mean) and precipitation data were obtained from the Worldclim dataset at a spatial resolution of 1 km [[3](#_ENREF_3)]: only mean temperature was included as the ranges of values present across the study area for minimum and maximum temperature were small. Interpolated soil properties (sand content and pH [measured in water]) were obtained from the International Soil Reference and Information Centre at a spatial resolution of 1 km [[2](#_ENREF_2)]. Population density data were obtained from the World Pop project at a spatial resolution of 100 m [[5](#_ENREF_5)] and elevation data were obtained from the Shuttle Radar Topography Mission (SRTM) at a spatial resolution of 90 m [[4](#_ENREF_4)]. Temperature, precipitation, soil information, elevation and population density data were linked to the survey households by geographical location using ArcMap 10 (ESRI, Redlands).

1. **Funder details**

The parasitological survey was conducted as part of the People, Animals and their Zoonoses project, funded by the Wellcome Trust (085308). Data processing was funded by the Medical Research Council, UK (project MR/J012343/1)

1. **References**

1. Wardrop NA: **Landcover classification: western Kenya, 2010.** Southampton: University of Southampton; 2015.

2. Hengl T, Heuvelink GBM, Kempen B: **Soil property maps of Africa at 1 km.** available from [www.isric.org:](http://www.isric.org:) ISRIC: World Soil Information; 2013.

3. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A: **Very high resolution interpolated climate surfaces for global land areas.** *International Journal of Climatology* 2005, **25:**1965-1978.

4. Jarvis A, Reuter HI, Nelson A, Guevara E: **Hole-filled seamless SRTM data V4.** available from <http://srtm.csi.cgiar.org:> International Centre for Tropical Agriculture (CIAT); 2008.

5. **Worldpop alpha version 2010 estimates of numbers of people per grid square.** <http://www.worldpop.org.uk/data/summary/?contselect=Africa&countselect=Kenya&typeselect=Population;> 2010.

6. Harrison LJS, Garate T, Bryce DM, Gonzalez LM, Foster-Cuevas M, Wamae LW, Onyango-Abuje JA, Parkhouses RME: **Ag-ELISA and PCR for monitoring the vaccination of cattle against Taenia saginata cysticercosis using an oncospheral adhesion protein (HP6) with surface and secreted localization.** *Tropical Animal Health and Production* 2005, **37:**103-120.

7. Harrison LJ, Joshua GW, Wright SH, Parkhouse RM: **Specific detection of circulating surface/secreted glycoproteins of viable cysticerci in Taenia saginata cysticercosis.** *Parasite Immunol* 1989, **11:**351-370.

8. Krecek RC, Michael LM, Schantz PM, Ntanjana L, Smith MF, Dorny P, Harrison LJS, Grimm F, Praeth N, Willingham AL: **Prevalence of Taenia solium cysticercosis in swine from a community-based study in 21 villages of the Eastern Cape Province, South Africa (vol 154, pg 38, 2008).** *Veterinary Parasitology* 2011, **183:**198-200.

9. Harrison LJP, R.M.: **Taenia saginata and Taenia solium: reciprocal models.** *Acta Leidensia* 1989, **57:**143-152.

10. Thomas LF, de Glanville WA, Cook EA, Fevre EM: **The spatial ecology of free-ranging domestic pigs (Sus scrofa) in western Kenya.** *Bmc Veterinary Research* 2013, **9**.

11. Bryceson DF, Mbara TC, Maunder D: **Livelihoods, daily mobility and poverty in sub-saharan Africa.** *Transport Reviews* 2003, **23:**177-196.