**Supplementary file 1: Details of HIPSS sampling methodology**

The additional details below on the methodology for the 2014-2015 survey of the HIV Incidence Provincial Surveillance System (HIPSS) were provided in an appendix to the article: Grobler A, Cawood C, Khanyile D., Puren A, Kharsany AB. Progress of UNAIDS 90-90-90 targets in a district in KwaZulu-Natal, South Africa, with high HIV burden, in the HIPSS study: a household-based complex multilevel community survey. *The Lancet HIV* 2017;4(11):e505-513. http://doi.org/10.1016/S2352-3018(17)30122-4

**“Methods**

**Sampling methodology**

We used multi-stage sampling to randomly select households and recruit a household representative sample of men and women. The enumeration area was the primary sampling unit. The sampling frame for the number of households and the number of persons in each enumeration area was created from three sources; a) the Census undertaken in 2011, b) a Community survey undertaken in 2007 and c) aerial imaging of dwellings supplied by Geo Terra Image.

The study area consists of an estimated 100,018 households with a total of 243,115 females

and 201,870 males. A total of 244,699 individuals in the age range 15-49 years are estimated

to live in the study area (based on the 2011 Census).

From a total of 600 enumeration areas, all 591 enumeration areas with more than 50

households were included in the sample. Of these, 221 enumeration areas were drawn

randomly. Within an enumeration area the households were drawn systematically with a

random start. Study staff identified households using the Global Positioning Systems receiver

to record the geographic coordinates of each randomly selected household. Sampling was

designed to obtain approximately 10,000 individuals. In instances where a selected household

was abandoned, refused to complete the composition form or the members away for an

extended period of time, then the household on the right side of the selected house, when

facing the entrance of the selected household, was used as a replacement.

Once a household was selected, and the head of the household provided consent for the

household to participate, a handheld personal digital assistant was used to compile a list of

the individuals residing in the household. The eligibility of each individual was determined by the personal digital assistant. These individuals were numbered and the handheld device

selected one of these individuals at random to be included in the study. If this individual

refused participation a second individual was selected randomly. Only one individual per

household was selected and enrolled in the study. Should the second individual selected also

refuse the household was replaced.

**Weighting**

Weights were calculated taking into account the probability of selecting the enumeration area, the probability of selecting the household in the enumeration area and the probability of

selecting the individual in a household, and was adjusted for non-response. The weights were

then revised to reflect the size of the population in the study area. The weights were calculated in three stages. In the first stage the probability of the enumeration area being selected was calculated. The probability of selecting an enumeration area was the number of enumeration areas selected divided by the total number of enumeration areas.

In the second stage the probability of a household being selected was calculated. The

probability of a household being selected was the number of households selected in the

enumeration area divided by the total number of households in the enumeration area. Within

each enumeration area the number of households that were found to be not eligible or who

refused participation or could not be contacted was calculated. The weight for a household

was then calculated as the reciprocal of the probability of selection multiplied by the probability of responding, thus inflating weights for non-response. Household weights above the 97.5th percentile and below the 2.5th percentile were truncated to remove extreme weights.

The third stage involved calculating the probability of selecting an individual within a

household. The probability of selecting an individual is the reciprocal of the number of eligible household members. The proportion of individuals of a certain age and gender who refused participation was calculated. The individual weight was then calculated as the reciprocal of the probability of selection multiplied by the probability of responding. This inflated the weights for individuals of the same age and gender as individuals who refused participation.

The weight for each individual was calculated as the product of the three weights calculated

in the three stages (enumeration area weight, household weight and individual weight).

Individual weights above the 97.5th percentile and below the 2.5th percentile were truncated to

remove extreme weights.

The final step was to benchmark these weights against census data by age group and gender

to ensure that the weights of the participants sampled added up to the population estimates

for the area. This means that the weight of an individual can be thought of as the number of

people in the community that the selected individual represents.”