APPENDIX

Survey Language:

The survey was administered in Kiswahili. The questionnaire was translated from English into Kiswahili and then back-translated into English by Muhimbili University of Health and Allied Sciences (MUHAS). The translation was reviewed and revised by survey team members who were fluent in both Kiswahili and English during the training for the pilot. The translations were further revised based on feedback from the pilot and interviewers who administered the survey. The final back-translated English version of the questionnaire was created based on the revisions made to the Kiswahili questionnaire during pilot testing and data collection.

Data Quality Notes:

Sample

Interviews were collected from 199 enumeration areas (EA) rather than 200 EAs because 1 female EA on Mainland Tanzania was inaccessible due to bad weather. Also, data collection was interrupted and only partially completed in 3 EAs (one mainland female EA, one mainland male EA, and one Zanzibar female EA) due to sudden inaccessibility due to weather or security issues. If an EA was larger than 200 households, a randomly selected segment of the EAs between 100 to 200 households in size was sampled.

A total of 8,693 households were visited during the study. The households visited for each of the four groups were: 2,104 households for females on Mainland Tanzania, 2,338 households for males on Mainland Tanzania, 2,020 households for females on Zanzibar, 2,231 households for males on Zanzibar. The household response rates varied from 97.0% to 98.3%.

Overall, 3,739 interviews were conducted and were divided across the four groups in the following manner: 908 females on Mainland Tanzania, 891 males on Mainland Tanzania, 1,060 females on Zanzibar, and 880 males on Zanzibar. The individual response rate varied slightly across the four groups: 95.4% for females on Mainland Tanzania, 96.0% for males on Mainland Tanzania, 94.6% for females on Zanzibar, and 93.9% for males on Zanzibar. The overall response rate ranged from 91.1% to 93.8%.

Data Entry and Cleaning

The data were entered into Epi Info version 3.5.1. There were no unique identifiers in the database that could ever be linked to a respondent. In order to ensure accurate data entry, the first 100 surveys were entered twice. After entering the first 100 surveys and verifying a low rate of data entry error, 20% of the remaining surveys were randomly selected for double-entry. Finally, when random checks of surveys consistently revealed two or fewer data entry errors per survey, 10% of all remaining surveys were randomly selected for double-entry. Overall, 12.7% of the surveys were entered twice. Discrepancies in surveys that were entered twice were resolved by reviewing the original hard-copy survey.

In addition to entering surveys twice, survey data were reviewed to detect mistakes in skip patterns, survey sections that were skipped and out-of-range values. During this process, 1.3% of the surveys were found to have data problems. These surveys were reviewed and entered again to ensure these

problems did not result from data entry. Because not every survey entered was reviewed for accuracy, some incorrect entries can be found in the dataset.

During data cleaning, attention was given to variables corresponding to the measures presented in the Tanzania report. Data in other variables were not cleaned and therefore are still raw.

SAS (version 9.3) was used for data management, further cleaning, and preparation of public use datasets.

Calculating Denominators for Analysis

The VACS includes complex skip patterns, therefore users should keep track of the multiple skip patterns to ensure the correct denominator is being used for analysis. Since this was a paper-and-pencil survey, discrepancies may arise when calculating denominators because interviewers may or may not have accurately followed the skip patterns when conducting the interviews.

When conducting data analysis, users must be aware that their results may differ from those presented in the Tanzania report due to data cleaning and analytical decisions made by Tanzania report authors.

Missing Data

Most measures have very low percentages of missing, less than 5%, and a portion of this missing data is associated with interviewer and data entry error; thus, it is assumed that the missing data are randomly distributed.

A higher percent of respondents (≥10%) did not provide responses to questions related to context of sexual violence such as descriptions of the perpetrators, location, and timing of sexual violence. This data cannot be considered missing at random because there are compelling reasons why a respondent may have chosen not to discuss the perpetrator or context of sexual violence with an interviewer.

Caution: Time of day sexual abuse occurred

Caution is advised when analyzing the set of questions that asked respondents about the time of day they experienced sexual abuse (Questions F328, F431, F530, F633 for females and M328, M431, M528, M631 for males). An inconsistency in the Kiswahili translation used in the field and the English questions on the protocol was detected after data collection and analyses were completed. Specifically, the questions asked respondents whether the incident occurred: 1) between the middle of the night and morning, 2) in the morning, 3) during daytime, 4) in the evening, or 5) in the night, and this was inconsistent with hours of the day associated with each category. For example, 1am to 6pm was listed as morning (category #2), instead of 7:01AM to Noon. Because this issue was not identified in the pilot testing or fieldwork, the exact interpretation of how the question was asked and interpreted by respondents cannot be determined. Monitoring found that interviewers did not report issues with these set of questions which suggests respondents did understand the questions. We suggest either not using these questions or providing the broad categorizations of between the middle of the night and morning, morning, daytime, evening, and night, with a footnote indicating that results should be interpreted with caution.

Additional Text Variables

The "*Additional Variables*" files are companion documents to the male and female codebooks and contain a list of all text variables corresponding to "Other" response categories and open-ended questions included in the Tanzania VACS. These text variables are included in the public use dataset and can be identified by the "_OK" at the end of their names. The textual information in these variables is presented in Kiswahili.

Additional Information:

Further details on the Tanzania VACS methodology and results can be found in the country report '*Violence Against Children in Tanzania: Findings from a National Survey, 2009'* available at: http://www.cdc.gov/violenceprevention/VACS/VACS-Reports.html