HIV/AIDS household impact study in Free State province (2001-04): Background and notes

This research project is jointly sponsored by the UNDP and the foreign development agencies of Australia (AusAID), the United Kingdom (DFID) and the United States (USAID) and administered by the Joint Economics, AIDS and Poverty Programme (JEAPP) of the African Asian Society under a subcontract from Nathan Associates Inc. Other research support includes a research grant from the National Research Foundation (NRF) and a Mellon Foundation grant from SALDRU, University of Cape Town.

1. Background

The HIV/AIDS epidemic poses a severe threat to the economies of developing countries, and those on the African continent in particular. South Africa, which is being affected fundamentally by the epidemic, is no exception. The estimated adult prevalence of HIV amongst 15-49 year olds in 2001 was 20.1 percent (UNAIDS, 2002), while the ASSA2000 model put adult prevalence amongst 20-65 year olds at 24.1 percent (ASSA, 2003). A recent national household survey in turn has put the 2002 estimate of adult prevalence amongst those older than 25 years at 15.5 percent (HSRC, 2002). These infected individuals all belong to individual households and their deaths will have a significant impact on their families. From an economic point of view, moreover, the primary impact of the disease manifests mainly among individual economic agents, i.e. individuals and households. An assessment of the socio-economic of HIV/AIDS therefore has to start on this micro-level of analysis. Aspects of such assessment, amongst other things, include determining how the disease affects the economic decisions and position of individuals and households over time, i.e. how they generate, save, invest and spend income, and how their quality of life is affected.

2. Aims and objectives

The larger project had the following broad objectives:

(a) develop and test a methodology for assessing the socio-economic impact of HIV/AIDS at the (b) individual and household level in both an urban and a rural setting;

(c) identify and capture the standard minimum criteria and indicators to be incorporated into the methods of methodologies of studies of this nature;

(d) describe and evaluate the impact of different informal coping strategies and support systems adopted by individuals, households and communities, as well as that of formal HIV/AIDS-related interventions of national and provincial government departments and local authorities (TLCs), in terms of their impact over time on the quality of life of affected households living in both urban and rural areas;

(e) inform economic growth analyses and studies on the macroeconomic impact of HIV/AIDS by projecting information about the microeconomic impact of the disease onto trends in labor market participation, spending, savings and investment; and

(f) propose a framework for national 'best practice' for improving the quality of life of affected households in urban and rural communities based on existing macro- and micro-, as well as formal and informal responses to HIV/AIDS.

3. Methodology

The household impact of HIV/AIDS was assessed by means of a cohort study of households affected by the disease. The survey was conducted in two local communities in the Free State province, one urban (Welkom) and one rural (Qwaqwa), in which the HIV/AIDS epidemic is particularly rife. Welkom and Qwaqwa, situated in the Lejweleputswa and Thabo Mofutsanyane districts of the Free State province, in 2001 faced HIV prevalence rates amongst antenatal clinic attendees of 41.1 and 27.8 percent respectively. According to the report entitled Measuring Poverty published by Statistics SA early in 2000, the Welkom magisterial district is the third richest in the Free State province, with a headcount poverty ratio of 0.34 and

average monthly household expenditure of R2364. The magisterial district of Witsieshoek, which is within the boundaries of the former Qwaqwa, is the poorest in the Free State province and also ranks amongst the poorest in the country. The headcount poverty ratio in this district is 0.69, while average monthly household expenditure amounts to R807. Thus, the selection of study sites also allows one to compare the household impact of HIV/AIDS between communities that differ substantially in terms of the general standard of living (Statistics South Africa, 2000).

In the research outputs that have emanated from this project, comparisons are drawn between so-called affected and non-affected households. Affected households were sampled purposively via NGOs and other organizations involved in AIDS counselling and care and at baseline included at least one person known to be HIV-positive or known to have died from AIDS in the past six months. Informed consent was obtained from the infected individual(s) or their caregivers (in the case of minors). In order to explore the socio-economic impact on affected households of repeated occurrences of HIV/AIDS-related morbidity or mortality, a distinction is made between affected households in general and affected households that have experienced morbidity or mortality more frequently. Non-affected households represent households living in close proximity to affected households. These households at baseline did not include persons suffering from tuberculosis or pneumonia. The incidence of morbidity and mortality is considerably higher in affected households. The morbidity and mortality experienced by affected households exhibit a classic HIV/AIDS pattern, with large numbers of adults (i.e. those aged 15-49 years) having experienced illness or having died. Between 70 and 80 percent of morbidity and mortality in affected households can be attributed to HIV/AIDS or related infectious diseases and opportunistic infections (Bachmann & Booysen, 2003; Booysen et al., 2003). Due to the sampling design and small sample size, the findings from this household impact study cannot be generalised to households across South Africa, but pertain largely to the experience of poor, African households that utilise public health care services (Booysen et al., 2003). Thus, the research is indicative only (but nevertheless telling) of the socio-economic impact of the HIV/AIDS epidemic, a characteristic shared by most other HIV/AIDS household impact studies (Booysen & Arntz, 2003). Furthermore, the classification of households employed in the earlier analysis, albeit useful for the purposes of our analysis, belies the fact that

HIV/AIDS affect entire communities and affect various households directly or indirectly at different stages of the epidemic, rather than affect select groups only of households that directly experience morbidity and mortality.

Households were defined in terms of the standard definition employed by Statistics South Africa in the October Household Survey (OHS), i.e. 'a person or a group of persons who live together at least four nights a week at the same address, eat together and share resources'. A survey on the quality of life and household economics was conducted, using the household questionnaire of which copies are included in the documentation of this dataset. Slight changes were made to the questionnaire, while certain questions were deleted and others added to the instrument. These changes to the questionnaires are described in a Word file with the documentation. Interviews were conducted with one key respondent only, namely the 'person responsible for the daily organisation of the household, including household finances'. The first four rounds of interviews were completed in May/June and November/December of 2001 and in July/August and November/December of 2002. Rounds five and six of the study were completed in July/August 2003 and May/June 2004 respectively.

During the first wave of interviews a total of 404 interviews were conducted. During the second wave of data collection, interviews were conducted with 385 households, which translates into an attrition rate of 4.7% (19 households). During wave III, a total of 354 households were interviewed, with 31 households not being re-interviewed (7.7% of the original sample). In wave IV, 55 new households were recruited into the study, with particular emphasis on an effort to recruit child-headed households into the survey insofar as the sample to date did not include any such households. During waves IV, V and VI a total of 3, 13 and 9 households respectively could not be re-interviewed. The payment of a minimal participation fee (R150 per household per survey visit) to those households interviewed in each wave, following the interview and distributed in the form of food parcels, contributed to ensuring sustainability of the sample over the three-year period. The dataset includes data for 331 households interviewed in each of the six rounds of interviews. In almost 90 percent of cases the reasons for attrition are related to migration, given that this study did not intend to follow those households that move outside of the two immediate study areas, i.e. Welkom and Qwaqwa. In the majority of cases, attrition can be

ascribed to the failure to establish the current whereabouts of the particular household during follow-up, while in a third of cases it could be established that the household had moved to another country, another province, or another town in the Free State province. Less than ten percent of households had refused to participate in subsequent waves. The reasons for attrition in the original sample illustrate the manner in which migration and the disintegration of households, which are important effects of the epidemic, can act to erode the sample population.

Standards of living were measured at the household rather than the individual level, given that the focus here is on the household impact of HIV/AIDS. During the survey, data were collected from one key informant regarding the employment income, non-employment income and receipts of remittances for the members of the particular household. An estimate of total monthly household income was derived from these figures by adding up the various component items. Likewise, fieldworkers collected expenditure-related data from the household member in charge of household finances. This include estimates of household expenditure on specific items such as food, education, health care, transport, monthly repayments of debt, and clothing, as well as remittances made to persons not living with the household. As in the case of income, an estimate of total monthly household expenditure was calculated by adding these items together. Remittances and the consumption of own produce were values based on quantities and/or local prices for similar goods. The income-based estimates of household welfare in the case of this study exceeded the expenditure-based estimates. Normally, one would expect the opposite, with expenditure-based estimates exceeding income-based estimates of household welfare. This may be because the one informant that was interviewed (i.e. the person in charge of household finances) generally has a better idea of the employment status and average earnings of other members of the household. (In fact, the person during the interview often verified this information with other household members.) This person is unlikely to be knowledgeable about the manner in which each member of the household spends their income on a range of consumption categories. In fact, individuals and/or households have been found to rarely record expenditure data in detail (Woolard and Leibbrandt, 1999: 23). Expenditure, therefore, in this case most likely reflects only that amount of resources of household members that is spent on communal household needs and does NOT represent a complete estimate of total household expenditure. As a result,

household income is preferred to expenditure as measure of poverty in the current research outputs, although we report in more detail on the differences in the level and composition of income and of expenditure in affected and non-affected households. Users may want to use the highest estimate of income or expenditure as measure of household welfare, a practice adopted where expenditure estimates are not perfect, but income may give biased results due to the seasonality of earnings. In addition, income, expenditure, savings and borrowing were not summed and balanced in any way in collecting the data. As such, these data represent broad estimates of aggregate household income, household debt, household savings and regular household expenditure.

Households with the same level of income do not necessarily enjoy the same level of welfare. The larger the household, the lower the level of welfare at similar levels of household income or expenditure. Measures of equivalent income and expenditure are employed to allow for these differences in standard of living related to household characteristics (Lipton and Ravallion, 1995; Burkhauser et al., 1997). Estimates of household income and expenditure were here adjusted for differences in household size by dividing total monthly income by n^{α}, where n represents the number of household members and α an adjustment for household economies of scale (Filmer and Pritchett, 1998: 13). According to Lanjouw and Ravallion (1995), a α coefficient of 0.6 represents an adequately robust and reliable adjustment for household economies of scale. The household dataset includes these equivalence scales.

Unfortunately, specific consumer price indices for the two study sites are not available at present. Income and expenditure estimates, as well as other monetary aggregates, were converted into real values using the most recent CPI estimates (2000=100) published by Statistics South Africa (2004) for the Free State province, with both the nominal and real values being reported in the dataset. All file including monetary values also include these deflators.

4. Challenges and lessons learned

One major lesson learned from this study is that the methodology employed in this research, which corresponds closely to that employed in many earlier studies of this

nature (Booysen and Arntz, 2003), is severely limited in terms of allowing researchers to draw conclusions about the impact of the epidemic that can be generalised to a national or regional level. This ultimately requires population-based sampling using robust epidemiological tools to distinguish between affected and non-affected households. It is hoped that this research, which was conceived as and remains a pilot study, can inform such research efforts.

A second major challenge, and a way in which the above dilemma can partly be addressed, is to revisit the current classification of households as affected as opposed to non-affected on a retrospective basis with the aid of the data gathered during the survey. Such exercise is required in the absence of HIV-testing and due to the fact that households and communities are affected in various ways by the epidemic. Users of the data are therefore encouraged to develop their own methodologies for arriving at appropriate classifications of the households and individuals included in the survey.

5. Data

The dataset comprises 7 separate data files. All files are available in SPSS 11.0 and Stata 7 format. All files are LONG format panel data files and each file includes the necessary unique identifiers to reshape the data into LONG or WIDE format and/or to MERGE variables in different data files, either in LONG or WIDE format. The type of information contained in each of these data files is described below. The number of households or individuals observed in each wave (N) and the number of times each household or individual was observed is noted in the corresponding tables (n).

- ***Households*:** household-level information on a range of social and economic variables, including household finances

Number of households observed in waves 1-6		**Number of households observed 1-6 times**	
1	404	1	24
2	385	2	31
3	354	3	53
4	406	4	10
5	393	5	11
6	384	6	331
Total (N)	***2326***	***Total (n)***	***460***

- ***Individuals***: key socio-demographic information (all ages) and labour force participation information (ages 15 plus) for those individuals that belonged to interviewed households

Number of individuals observed in waves 1-6		**Number of individuals observed 1-6 times**	
1	1900	1	408
2	1794	2	386
3	1681	3	298
4	1902	4	229
5	1879	5	188
6	1812	6	1173
Total (N)	***10968***	***Total (n)***	***2682***

- *HeadofHousehold*: key socio-demographic information (all ages) and labour force participation information (ages 15 plus) for those individuals that headed the interviewed households

Number of individuals observed in waves 1-6		**Number of individuals observed 1-6 times**	
1	402	1	114
2	385	2	81
3	354	3	69
4	406	4	49
5	393	5	43
6	382	6	238
Total (N)	***2322***	***Total (n)***	***594***

- *Morbidity*: key information about the nature and consequences of illness episodes experienced by household members during the month preceding the interview

Number of individuals observed in waves 1-6		**Number of individuals observed 1-6 times**	
1	255	1	299
2	157	2	117
3	114	3	42
4	117	4	25
5	135	5	16
6	127	6	11
Total (N)	***905***	***Total (n)***	***510***

- *Mortality*: key information about the nature and consequences of deaths experienced by households during the six month preceding the interview

Number of individuals observed in waves 1-6		Number of individuals observed 1-6 times	
1	44	1	186
2	34	2	-
3	24	3	-
4	22	4	-
5	35	5	-
6	27	6	-
Total (N)	***186***	***Total (n)***	***186***

- *Outmigration*: key information about persons that had left the household since the time of the previous interview

Number of individuals observed in waves 1-6		Number of individuals observed 1-6 times	
1	-	1	481
2	96	2	22
3	106	3	1
4	41	4	-
5	147	5	-
6	138	6	-
Total (N)	***528***	***Total (n)***	***504***

- *Inmigration*: key information about persons that had joined the household since the time of the previous interview

Number of individuals observed in waves 1-6		Number of individuals observed 1-6 times	
1	-	1	478
2	Question not asked	2	6
3	132	3	-
4	52	4	-
5	183	5	-
6	123	6	-
Total (N)	***490***	***Total (n)***	***484***

The Excel file *VariableLists* includes a description of the variables in each of the data files. The Word file *CodingList* includes the codes over and above the codes printed on the questionnaire, including codes used for open-ended questions. These codes, which in many cases are numerous, were in most cases are not included as data labels in the dataset and users are encouraged to recode these variables as required and to assign the necessary labels to the new variables. Changes to the questionnaires are described in a Word file included with the documentation.

In addition, data users should take note of the following general notes and comments regarding the dataset:

- **IMPORTANT: The utmost care has been taken in preparing this dataset following quality controls implemented in the field and consistency checks conducted following data capture. Where questions or coding formats have changed between waves, the data included in this dataset were recoded to reflect the coding formats in the wave 6-version of the questionnaire. In addition, key information such as gender, age and race were compared across time for the same individuals to ensure consistency. However, due the fact that the same or a different respondent were asked questions pertaining to other household members at different points in time, certain information, such as educational qualification for example, in some cases vary considerably over time. In some cases these changes obviously reflect improvements in educational attainment, while in other cases these changes reflect inconsistent recall bias. The fact that most of the information are self-reported and are not verified with reference to documentation also contributes to these inconsistencies. For this reason, the onus rests on the user of this data to also perform further quality and consistency checks where this may be required, e.g. comparing variables across time to determine consistent patterns and recoding inconsistencies.**

- The recall periods for certain questions changed following the baseline survey. At baseline, some questions were asked with reference to a 12-month reference period, which changed to 6 months in follow-up interviews. Monetary values were

recoded to reflect a similar reference period, but the dataset includes the original data should users wish to employ these data. Therefore, care should be taken in comparing responses to questions asked at baseline and those asked at follow-up.

- A number of variables include codes representing reasons for missing data, which are described in the coding list. These are coded as minus (-) values. Users should take care in setting these values to missing before computing averages and other statistics based on these variables.

- The datasets include responses to almost all the questions in the questionnaire. However, in some cases, notably where questions were asked in one or two waves only or where the format of questions changed significantly between waves, the datasets do not include this information. For example, the available dataset excludes the responses to the special module on social grants included in the 6th round of interviews. Furthermore, only certain key information from the section on social support mechanisms (section 19 in the latest version of the questionnaire) was coded into the household file. Due to the fact that this relatively complex section was included right at the end of the questionnaire, not all the questions were always consistently completed and parts of this information remain missing. Should users or prospective researchers require this information for the purposes of their analysis, these additional files can be obtained from the project coordinator (see contact details below).

- Questions about familial relationships, used to document household structure and composition, such as the relationship of persons to the head of household (v8, v165, v183, v601) and whether or not the person's mother or father was alive (v12 and v13) are self-reported and reflect biological and/or socio-cultural relationships. For example, the mother of a child adopted by a household will be recorded as being alive, although this child's biological mother may be deceased. Therefore, as noted elsewhere, these and other self-reported data should be interpreted with care, given the complexity and fluidity of household structure and composition.

- In a few cases, households included more than one household head. This was the result of headship not being defined, but self-reported. Here the age and gender of the household heads were compared across the six waves and the *headofhousehold* file includes the information for the one person that was consistently recorded as household head across time.

- Question related to public services reflect access to these services (v27-v36), but do not in all cases reflect whether or not the service was available or working at the time of the interview. In some instances, albeit the exception, water or the telephone for example was disconnected.

- In the case of aggregate monetary variables contained in the household file, missing values were all set to zero.

- In the baseline interview, no distinction was made between savings on retirement annuities, life insurance and burial policies (v95). In subsequent interviews this distinction was made, with the sub-totals coded as v95.1 to v95.3, while v95 represents the total value of these types of savings.

- The repayments of debt (v115 to v127) reflect actual planned payments rather than required payments and include the interest charged on credit. The value of new borrowing (v147) reflects the amount borrowed and excludes interest charged on this amount.

- The responses to questions on the ownership and presence of assets (v128 to v139) do not always reflect ownership. In some cases, a household member or the household may possess such asset, but it actually belongs to a person outside of the household. Therefore, these questions and the asset index reflect access to assets rather than ownership of assets.

- The responses to the questions on main energy sources for heating, cooking and lighting (v28-v30) reflect seasonal variation. For example, in summer some

respondents indicated that heating was not necessary and no energy source was used for this purpose.

- The responses to the question about the main reason for in- and out-migration (v606 and v616) reflect general rather than specific motivations for migration. For example, where education was noted as the main reason for migration, it is not possible to distinguish whether the person moved to attend school or came back from attending school elsewhere. Likewise employment may reflect a person losing employment and returning to the household or a person finding employment here and therefore joining the household.

- As with most surveys, many responses remain subjective. So, for example, the response to the question of whether the person has recovered from the illness (v158) in some cases reflects that a person suffering from chronic diseases such as hypertension or tuberculosis have actually recovered temporarily.

- In the morbidity section of the questionnaire, the question on the performance of daily activities by ill household members (v159) in most cases were answered as "no" for ill children, giving that the daily activities listed here excludes playing or other child activities and were understood to relate to work and other adult activities.

- At baseline, deceased persons were not allocated a number on the household roster, which means that the unique identifier for these 24 individuals is missing. In the case of deaths that occurred subsequent to baseline, which refer to individuals that belonged to the household at a previous point in time, unique identifiers are available and the information in the *mortality* file can be merged with the socio-demographic and labour force participation details of these individuals contained in the *individuals* file.

- In waves I to IV, the question on the employment status of the deceased (v199) included a skip instruction asking the interviewer to skip to the question on lump-sum receipts (v203) if the person was not employed prior to their death. This skip

instruction was removed in wave V. Therefore, the estimate of loss of income (v200) excludes the loss of income from social grants in the case of data for waves I to IV, but includes loss of income from social grants for wave V and VI.

- Only in later waves (waves V and VI) have persons that have died, but that have joined the household since the previous interview been recorded as in-migrants. The sections on in-migration for the most part therefore include details about persons that left or joined the household since the previous interview and that were alive at the time.

- A coded response of 1 to the questions on the medical costs in the past month (v176-v181) or prior to death (v209-v214) in some cases reflect no consultation having taking place (v153=0), given the "nothing" label ascribed to the variable.

6. Contact

Enquiries pertaining to the research project and dataset can be directed to:

Professor Frikkie Booysen
Department of Economics and Centre for Health Systems Research & Development
PO Box 339
University of the Free State
Bloemfontein
9300 SOUTH AFRICA
Tel: +27 51 401 2623
Fax: +27 51 447 8274
Email: BooysenF.EKW@mail.uovs.ac.za