KWAZULU-NATAL INCOME DYNAMICS STUDY (KIDS) 1993-1998

OVERVIEW AND DESCRIPTION OF DATA FILES

RELEASE VERSION 2

International Food Policy Research Institute
University of Natal-Durban
University of Wisconsin-Madison

September 2001

Acknowledgements

The 1998 data collection was a collaborative project between researchers at the International Food Policy Research Institute (IFPRI), the University of Natal, and the University of Wisconsin. In addition to support from these three institutions, the following provided financial support: the United States Agency for International Development (Office of Women in Development, grant no. FAO-0100-G-00-5050-00, Strengthening Development Policy through Gender Analysis; the BASIS/CRSP project at the University of Wisconsin-Madison (LAG-A-00-96-90016-00), and a University Partnership Grant); the Ford Foundation; a Centre for Science Development population studies grant to the University of Natal; and the Development Bank of Southern Africa. The assistance of the Southern Africa Labour and Development Research Unit (SALDRU) was also crucial and is gratefully acknowledged, including the permission to redistribute the 1993 data.

Two other institutions contributed substantially to the survey design and implementation: Data Research Africa (DRA) and Policy and Praxis. The former group carried out the fieldwork with invaluable efforts from Aki Stavrou, Faith Slu Hlongwa, and Ben Roberts. The latter group was responsible for data entry and we thank in particular Juby Govender, Malani Govender, Priya Gayadeen and Stacey von Schalkwyk. Also, Ingrid Woolard of the University of Port Elizabeth undertook the task of constructing the aggregate measures. Others who contributed their time and ideas include Justine Barnes, Catherine Cross, Deon Filmer, Chris Gibson, and Duncan Thomas.

Finally our sincere gratitude goes to the individuals and communities who generously gave their time to respond, once again, to our questions.

Michael R. Carter
Lawrence Haddad
Julian May
John A. Maluccio

1. Introduction

An important adjunct of apartheid has been the absence of credible and comprehensive data on which policies, such as poverty reduction strategies, can be grounded; the data here present one of several efforts that are beginning to address this gap.

The first South African national household survey, the Project for Statistics on Living Standards and Development (PSLSD), was undertaken in the last half of 1993 under the leadership of the Southern Africa Labour and Development Research Unit (SALDRU) at the University of Cape Town (PSLSD, 1994). Households in KwaZulu-Natal Province, on the east coast, were re-surveyed from March to June 1998 for the KwaZulu-Natal Income Dynamics Study (KIDS) (May et al., 2000). Combining these two surveys yields a panel (or longitudinal) data set in which the *same* individuals and households have been interviewed at two points in time, in 1993 and again in 1998.

IFPRI requests that users of the data acknowledge the source of the KIDS data files in all publications, conference papers, and manuscripts with the following statement: "The KwaZulu-Natal Income Dynamics Study (KIDS) was a collaborative project of the International Food Policy Research Institute, the University of Natal-Durban, the University of Wisconsin-Madison, and the Southern Africa Labour and Development Research Unit at the University of Cape Town."

2. **Disclaimer**

IFPRI and its collaborators encourage the use of the KIDS data, but emphasize that the attached data files are unit record or 'raw' data files. While all the information that would allow individuals to be identified has been deleted from the files, all other information remains in the data files. The data are provided 'as is' and in no event shall IFPRI or the collaborating institutions be liable for any damages resulting from use of the data. While great effort was taken to obtain high quality data, the accuracy or reliability of the data is not guaranteed or warranted in any way. The decision not to alter the contents of the data files means that the user of these files will need to take care in handling missing observations, outlier values, and violations of logical consistency.

3. Household and Community Questionnaires

The 1993 survey was an integrated household survey similar in design to a World Bank Living Standards Measurement Survey. The main component was a comprehensive household survey that collected a broad array of information on the socio-economic condition of households. Among other things, it included sections on household demographics, household environment, education, food and non-food expenditures, remittances, employment and income, agricultural activities, health, and anthropometry (weights and heights of children aged six and under).

To ensure comparability, the 1998 household questionnaire largely followed the 1993 version, though there were some important changes. One of these was a greater

focus on individual (as opposed to household) ownership of assets and control over their use so that individual, generation, and gender-differentiated analyses are possible. A second difference was an expanded emphasis on those individuals not living in the household but economically linked to it. Finally, four new sections were added including economic shocks (both positive and negative), social capital (including group membership, kin networks, civic engagement, and trust), assets brought to marriage, and household decisionmaking.

The household questionnaire was necessarily quite involved and to ensure data quality, survey enumerators were trained for over two weeks including practice interviewing on non-sample households in the field and separate anthropometric training, the questionnaire took close to three hours on average to complete; often repeat visits were required in order to avoid respondent fatigue.

Finally, in both years community surveys were completed from interviews with key informants in each of the clusters. In 1998 the community questionnaire included new sections on economic shocks and social capital (replicating the corresponding household level sections to the extent possible) in addition to sections on local economic activity, infrastructure, and prices.

4. Sample Design

The 1993 sample was selected using a two-stage self-weighting design. In the first stage, clusters were chosen with probability proportional to size from census enumerator subdistricts (ESD) or approximate equivalents where an ESD was not available. In the

second stage, all households in each chosen cluster were enumerated and a random sample of them selected. (See PSLSD, 1994, for further details.)

In 1993, the KwaZulu-Natal portion of the PSLSD sample was designed to be representative at the provincial level, conditional on the accuracy of the 1991 census and other information used for the sampling frame, and contained households of all races. It was decided not to re-survey the small number of white and coloured households in 1998, however. While there were minor advantages to retaining these groups, the relatively small number of households in each group (112 white households and 53 coloured) would have precluded most comparative ethnic analyses. Moreover, the households in these ethnic groups were entirely located in a small number of clusters (due to the general lack of spatial integration of the population), undermining their representativeness. As a result, the 1998 sample includes only African and Indian households.

Given the various purposes for the study (e.g., income and asset generation, child health, etc.), the identification of "key" decision makers within households was very important. In 1993, the PSLSD recorded a household head for each household. While in many or even most instances it might be correct to assume that this “self-declared” household head corresponds to a main decision maker, given the cultural diversity and complexity of households in South Africa, this may not always be accurate. To capture some of these complexities, an expansion of self-declared headship was implemented and these persons, whom we call “Core” persons, were targeted for re-interview. May et al. (2000) and the 1998 survey documentation provide further details.

Another important aspect of the 1998 re-survey is that when possible we tracked, followed, and re-interviewed households that had moved. The combination of “Core”

persons, of which there were often more than one in an original 1993 household, and tracking movers meant that it was possible for original households to split and for the split-offs to remain in the sample. Thus, it is possible to analyze the sample as a panel of households (ignoring or possibly recombining the split-off households) or as a sample of "Core" persons.

Household Attrition

Here we describe attrition at the household level in the 1998 re-survey (ignoring split households). Patterns of attrition are very similar at the "Core" person level so those are not discussed.

The ensuing analysis is updated from the first version of the data release (April 2000) and excludes two clusters from both the 1993 and 1998 samples. During related follow-up field research in May 2001 it was discovered that all 39 household interviews in clusters 217 and 218 had been fabricated in both 1993 and 1998; these households are dropped in this analysis (and in the updated release of the data), leading to minor discrepancies between the figures concerning attrition reported here and those reported in previously published work, in particular, May et al. (2000). Also, cluster 206 is now coded as urban as it was incorrectly coded as rural in the 1993 released data. Note: Weights calculated by the World Bank and provided with the original data are NOT updated to reflect these changes.

The 1993 (and thus 1998 target) portion of the PSLSD sample included 1354 households (215 Indian and 1139 African, see Table) in 67 communities. Of the target sample, at least one core person from 1132 households (83.6%) was successfully re-interviewed.

Household Attrition Rates in the KIDS Survey (% of column)

Status	African (Non-Urban)	African (Urban)	Indian (All)	Total
Located and interviewed	688 83.4%	276 87.9%	168 78.1%	1132 83.6%
Moved, and could not be located	60 7.3%	15 4.8%	18 8.4%	93 6.9%
Not known in the area, refusal, or death	77 9.3%	23 7.3%	29 13.5%	129 9.5%
Total	825	314	215	1354

In most surveys of this type in developing countries, refusal rates are low. This is true in the KIDS survey: only eleven re-contacted households refused an interview.[1] Many surveys in developing countries do not attempt to track movers. Had we followed that strategy we would have missed 60 households who were successfully tracked and consequently re-interviewed.

Re-interview rates were highest for Africans in urban areas, where 87.9% of the target households were re-contacted. In metropolitan areas, part of the urban sample characterized by more permanent housing structures and street addresses, they were even higher (not shown). The KIDS was less successful in re-interviewing Indian households, however.

For more than 1/3 of the households not re-interviewed, information collected verified the household had moved but was not detailed enough to allow tracking to a new

[1] In addition, all household members died prior to the 1998 re-survey in four households.

residence. For the remaining households, however, there was simply no trace, i.e., no one approached in the community recognized the name of any household members when presented with the 1993 household roster. While the loss of the former group may be regarded as attrition, it is possible that at least part of the latter group reflects bogus interviews in 1993. Maluccio (2001) describes attrition in the sample in more detail.

Split Households

As described above, the process of following and interviewing "Core" persons led to "split" households in 1998. For 36 original 1993 households in the sample, two (or more) interviews were completed in 1998. Hence the total sample includes 1171 household level observations in 1998.[2] Household splits are identified in the file HHCOVER by the variables *split98* and *triple98*. The household identifying variable, *hhid,* is the same for the split-offs of the original household except that the root (or original location) *hhid* ends in "0" while the split-off ends in "2" (or "3" in the case of a triple split). For example, household 2220140 from 1993 split into three households in 1998 and they are identified as 2220140 (root), 2220142, and 2220143.

For individuals, those interviewed in 1993 retain their unique person identifier code, variable *pcode,* in 1998 regardless of whether they are in the root or split-off household. For completeness, each individual from 1993 (*pcode* < 40) is repeated in the roster for every split-off, even if they are not resident. Thus person 4 in household 2220140 is the same person as person 4 in household 2220142 and person 4 in household 2220143; basic information from the roster (e.g., age, education) is the same although clearly residential status may be different. For new household members (*pcode* >= 40),

[2] There were 33 original 1993 households that split into two in 1998 and 3 original 1993 households that split into three in 1998 yielding 33 + 6 + 1132 = 1171.

however, the person codes across split households do not refer to the same person. In other words, person 40 in household 2220140 is not the same as person 40 in household 2220142.

5. Data and Supporting Documentation

To simplify working with the longitudinal data, we are releasing both the 1993 and 1998 data together. The 1993 materials include the original PSLSD data for KwaZulu-Natal only (with a small number of corrections based on a re-examination of some of the original questionnaires), the 1993 expanded codebook and questionnaires, and the Stata software programs used to calculate 1993 income and expenditures. Additional information can be obtained from the World Bank (http://www.worldbank.org/html/prdph/lsms). The 1998 materials include the 1998 KIDS data, the 1998 expanded codebook and questionnaires, the Stata software programs used to calculate 1998 income and expenditures, and the 1998 fieldworker manual.

The data files are in both Stata version 6.0 (extension .dta) and SPSS for Windows (extension .sav) formats. The codebooks list all the data files and variable names for both household and community level variables. The questionnaires also list the data files and variable names (except for the 1993 community questionnaire). While the names for data files for corresponding sections of the questionnaire in 1993 and 1998 are different, variables that are in response to the *exact* same question in both periods have *exactly* the same variable name. New questions in 1998 are indicated on the questionnaire with the prefix "N".

6. Updates to data (Release Version 2, September 2001)

As described above, some important changes have been made for the release version 2 of the data (September 2001). Users can either copy directly the files provided here (files names have not changed from the first release but files can be distinguished from the earlier version by date and/or by examining whether they exclude observations from clusters 217 and 218) or applying the Stata or SPSS program patches located in the main directory with this document (stata_patch_v2.do and spss_patch_v2.sps will require slight modification to run, essentially changing the file locations) to the data from the first release.

Directory Structure for KIDS 1993-1998 Data and Documentation

1993: 1993 PSLSD – KwaZulu-Natal Only

Basic: Household and community raw data (from questionnaires)

Constrct: Constructed data (including income and expenditures)

Docs: Household and community questionnaires and expanded codebook

Programs: Stata programs for constructed data

1998: 1998 KIDS

Basic: Household and community raw data (from questionnaires)

Constrct: Constructed data (including income and expenditures)

Docs: Household identification form, household and community questionnaires, fieldworker training manual, and expanded codebook

Programs: Stata programs for constructed data, program to copy all files and to convert all files from Stata to SPSS

The complete set of materials can be obtained by contacting the Population and Poverty Studies Programme at the School of Development Studies, University of Natal (www.nu.ac.za/csds) or the International Food Policy Research Institute (www.ifpri.org).

REFERENCES:

Maluccio, J.A., 2001. "Using Quality of Interview Information to Assess Nonrandom Attrition Bias in Developing Country Panel Data," forthcoming in *Review of Development Economics*.

May, J., M. Carter, L. Haddad, and J.A. Maluccio. 2000. KwaZulu-Natal income dynamics study (KIDS) 1993-1998: A longitudinal household database for South African policy analysis. *Development Southern Africa*, 17(4) p. 567-81.

PSLSD, 1994. *Project for Statistics on Living Standards and Development: South Africans Rich and Poor: Baseline Household Statistics*. Cape Town: South African Labour and Development Research Unit, University of Cape Town. (Note: The introductory text from this document, including the description of sampling, is reproduced in the 1993 expanded codebook released here.)