KWAZULU-NATAL INCOME DYNAMICS STUDY (KIDS)

THIRD WAVE, 2004

PUBLIC RELEASE VERSION 1.1

University of KwaZulu-Natal
University of Wisconsin-Madison
London School of Hygiene & Tropical Medicine
International Food Policy Research Institute

March 2006
(Release 1.1 - March 2007)

Acknowledgements

The 2004 data collection was a collaborative project between researchers at the University of KwaZulu-Natal, the University of Wisconsin, London School of Hygiene and Tropical Medicine, the International Food Policy Research Institute (IFPRI) and the South African Department of Social Development. In addition to support from these institutions, the following provided financial support: the UK Department for International Development, USAID, the Mellon Foundation, and National Research Foundation/Norwegian Research Council. Finally our sincere gratitude goes to the individuals and communities who generously gave their time to respond, once again, to our questions.

Julian May
Jorge Agüero
Michael R. Carter
Ian M. Timæus

1. Introduction

The KwaZulu-Natal Income Dynamics Study (KIDS) is a panel study that follows a random sample of households who lived in the South African province of KwaZulu-Natal (KZN) in 1993. These households and those who have split off from them were interviewed again in 1998 and 2004. This document summarizes the main features of the third wave of KIDS conducted in 2004. For further details see May et. al. (2007).

Users of the data must 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 between researchers at the University of KwaZulu-Natal, the University of Wisconsin, London School of Hygiene & Tropical Medicine, International Food Policy Research Institute (IFPRI), the Norwegian Institute of Urban and Regional Studies and the South African Department of Social Development. In addition to support from these institutions, the following organizations provided financial support: UK Department for International Development; the United States Agency for International Development (USAID); the Mellon Foundation; and National Research Foundation/Norwegian Research Council grant to the University of KwaZulu-Natal."

2. Disclaimer

All of us who have worked on this project encourage others to make use of the KIDS data. We must, however, emphasize that the attached data files are unit record or 'raw' data files. All information that would allow individuals to be identified has been deleted from the files and some edits have been undertaken to eliminate duplicate reports and records, in particular to the household roster files. However, all other information

collected remains unedited in the data files. While great effort was taken to obtain high quality data, they are made available 'as is'. The accuracy or reliability of the information are not guaranteed or warranted in any way and in no event are the collaborating institutions liable for any damages resulting from use of the KIDS data. The decision not to alter the contents of the data files means that the user of these files will need to take care in analyzing the data to allow for any missing observations, meaningless codes, outliers, and violations of logical consistency.

3. Household questionnaire

KIDS re-interviews the KwaZulu-Natal (KZN) sample of the 1993 nationwide survey known as the *Project for Statistics on Living Standards and Development* (PSLSD.) The original project was financed by the World Bank and had the characteristics of the Living Standard Measurement Surveys. Reflecting their origin, all three waves of fieldwork for KIDS—1993, 1998, and 2004—collected information on household composition, expenditure on food and on other durable and non-durable goods, education, health, agricultural production, employment, and additional sources of labor and non-labor income. To ensure comparability, the 1998 and 2004 questionnaires largely followed the 1993 version of the questionnaire, however, a few modules have been added and removed.

For example, the 1998 survey added sections on assets to marriage, economic shocks, and social capital and trust (see May *et. al.*, 2000). The 2004 version added a section on the Child Support Grant, a module on recent deaths of household members, and a module with learning tests for children between the ages of seven and nine. The 2004 version also

added a section on the work history of those aged between 24 and 30 at the time of interview.

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

Finally, in all three waves of KIDS, community surveys were taken through interviews with key informants in each of the survey clusters.[1] In 2004 the community questionnaire included new sections on local social networks in addition to sections on local economic activity, infrastructure, and prices.

4. Sample design

Due to the geographic concentration of African and Indian households, KIDS—unlike the PSLSD—limits its scope to African and Indian households. In the KwaZulu-Natal province, Africans represent 85 percent of the population and Indians represent 12 percent. Compared with their representation nationally, White and Coloured people are underrepresented in KwaZulu-Natal. Effectively, the numbers of White and Coloureds in the KwaZulu-Natal sample are too small, and too geographically concentrated in a few clusters, to permit meaningful inference. The KIDS study has thus been limited to the first two population groups.[2]

[1] As described below, the original 1993 survey was based on a multi-stage design that randomly selected enumerator districts, creating what we call survey clusters. We will also refer to these clusters as 'communities.' In using this designation, we do not mean to imply that these clusters of individuals who live in close proximity necessarily form a community in a sociologically meaningful sense.
[2] The 1993 KIDS data is then a subset of the 1993 PSLSD.

PSLSD was a survey of households. However, households are a complicated object to define, particularly in longitudinal studies. To transform KIDS from a single-round household survey into a longitudinal household panel study required a redefinition of the sampling unit. In 1998, a decision was made to follow the *core household members* with the intention of capturing the major decision makers within the household. A household member is a *core person* if he/she satisfied any of the following criteria:

- The self-declared head of household from the 1993 survey
- A spouse/partner of the self-declared head of household (from the 1993 survey)
- Lives in a three generation household and all of the following are true:
    - Child of the self-declared household head, son/daughter-in-law of the household head, or niece/nephew of self-declared head
    - At least 30 years old
    - Have at least one child living in household
    - Spouse/partner of person satisfying criterion.

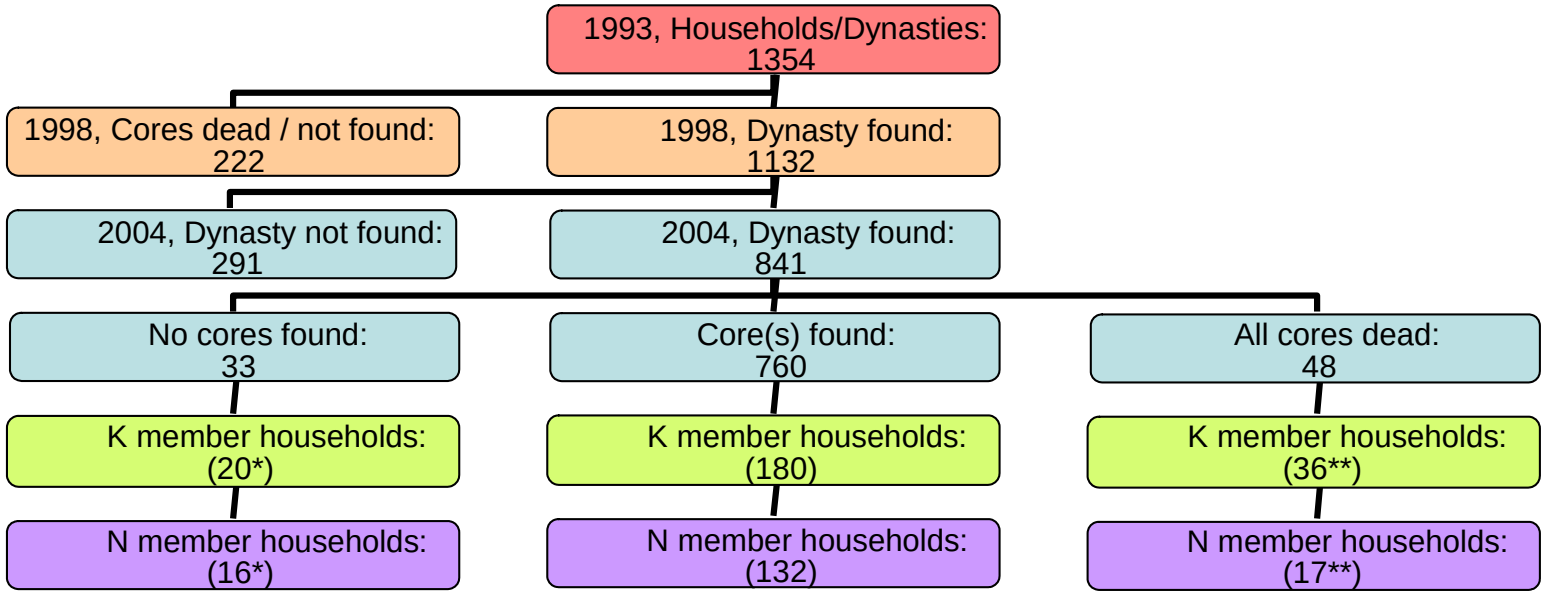
Thus all heads of households and spouses of heads are automatically classified as core and, in some three-generation households, adult children are also included in this cateogry. In this way, we can see the 1993 survey as the baseline information for a random sample of *dynasties*. The efforts of the 1998 and 2004 surveyors to find the location of the 1993 core members can then be seen as a way to keep track of the 1993 dynasties.

In 2004, due to the aging of the core members and the high prevalence of HIV/AIDS in South Africa, the study was extended in a complementary way to track and interview the households of the *children of the core* or the *next generation*. These are sons and

daughters of core members older than 18, who have established a "new" household since 1993 (labeled as “K”). By establishing a new household we mean that these children are now living away from their own parents with their own children, or with the children of their partner. Using the *next generation* to keep track of family “dynasties” provides a way of refreshing the panel and establishing a generational transition. In addition, due to our interest in the impact on children of the HIV/AIDS epidemic, the 2004 wave followed foster children to their new households. This group is defined as children aged less than 18 years old of core and next generation household members who no longer live with their parents i.e. no longer live in core or next generation households (labeled as “N”). As described in Appendix A, different questionnaire modules were administered in the core, next generation, and foster child households.

As the goal of the 2004 wave of KIDS was to find and interview the households of the children of the core and the foster children in addition to those of the regular core members, we had three ways in which we could contact the 1993 dynasties. The dynamics of the KIDS sample are shown in Figure 1 below.

In 1998, almost 84% of the 1993 dynasties were found as documented by May *et al*. (2000). From the 1132 dynasties interviewed in 1998, the 2004 wave found 841, yielding a response rate of 74%. Most of these dynasties were still composed of the original core members (760) however some of them were represented by the next generation of household members (K) or foster children (N). These K and N households are depicted in green and purple in Figure 1.



* Three dynasties where none of the original core members were found are represented by both K and N households in 2004
** Five dynasties where all the original core members have died are represented by both K and N households in 2004

Figure 1. Dynasty dynamics in KIDS: 1993-2004

5. Household ID

The records in each wave of data are identified by a household id variables, *HHID* (in 1993), *HHID* (in 1998) and *HHID2004*. The household ID for 2004 is a two-digit extension of the corresponding ID for 1998, which was itself a one-digit extension of the 6-digit 1993 ID. In other words, let X be a household ID in 1993. The 1998 ID was constructed by adding one extra digit Y={0,2,3} to allow for household splits, so that each 1998 household ID takes the form XY. When Y=0, it meant that at least one core member from the 1993 household was still living in the same location in 1998. If the core members had split up by 1998, the second household had Y=2, and a third split was represented by Y=3.

The household IDs for 2004 follow a similar rule. The new household ID is XYZ, where Z is a two-digit addition explained in the following table in terms of X and the Y.

Values for Z	Description
00	X0 household with X0 cores only
02	X2 household with X2 cores only
03	X3 household with X3 cores only
04	Further splits of XY
05	
06	
11	Recombined households (*e.g.*, an X0 recombined with an X2
20	K questionnaire (e.g., X020 would be a next generation household that emerged from the 1998 X0 household).
24	Further K questionnaires (i.e. more than one next generation started his/her family)
25	
26	
40	N questionnaire (e.g., X040 would be a foster child household where the foster child emerged from the 1998 X0 household).
44	Further N questionnaires (i.e. more than one foster child)
45	

Individuals can be uniquely identified across the three waves by a concatenation of the household ID (*HHIDQQ*) and the person code (*PCODE*). Persons with a *PCODE* <40 were first interviewed in 1993, persons with a *PCODE* of 40-72 in 1998, and anyone with a *PCODE* of 100+ in 2004. Thus a unique identifier can be calculated as follows (using *Stata* code):

```
gen double pid=hhid93*1000000+pcode if pcode<40
replace pid=hhid98*100000+pcode if pcode>=40 & pcode<100
replace pid=hhid2004*1000+pcode if pcode>=100
```

where *HHIDQQ* represents the household ID for years *QQ*={1993, 1998, 2004}. Note, however, that a few individuals with *PCODE*s in the range 40-72 became members of a different 1998 household (within the same 1993 dynasty) by 2004.

6. Correcting the 1998 data

The third and last release of the 1998 data (Version 3, January 2003) addressed the issue of fabricated households. Follow-up field work in May 2001 revealed evidence suggesting that the 1993 and 1998 interviews in some clusters had been fabricated by the enumerators. In 2004 we revisited the suspect clusters. In most of them, we were able to identify and interview at least some households from the earlier waves and establish the veracity of the earlier data. We have concluded that there is no evidence of fabrication except in clusters 217 and 218. These two clusters have been permanently removed from the sample. Hence the numbers reported in Figure 1 exclude these clusters.

7. Data and Supporting Documentation

To simplify working with the longitudinal data, we are releasing the 2004 data in an analogous way to the 1993 and 1998 versions. The three datasets can be downloaded from the KIDS website (http://sds.ukzn.ac.za). The release includes the data, the questionnaires, and the Stata code used to calculate income and expenditures. The data

files are in SPSS and Stata version 7 format. The codebooks list all of 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 of the data files for corresponding sections of the questionnaire in 1993, 1998 and 2004 are different, variables that respond to the exact same question in any of the periods have exactly the same variable name. New questions in 1998 are indicated on the questionnaire with the prefix "N" for 1998 and "Q" for 2004. Additional information on the 1993 PLSD is available from the World Bank (http://www.worldbank.org/html/prdph/lsms).

8. References

May, J., Carter, M.R., Haddad, L. and Maluccio, J. (2000), KwaZulu-Natal Income Dynamics Study (KIDS) 1993-1998: A longitudinal household data set for South African policy analysis, **Development Southern Africa,** Vol. 17(4), pp. 567-581.

May, J., Agüero, J., Carter, M. and Timæus, I. M. (2007), The KwaZulu-Natal Income Dynamics Study (KIDS) 3rd wave: Methodology, First Findings and an Agenda for Future Research, **Development Southern Africa** , in press.

Appendix A. Structure of the 2004 household questionnaires by sample unit

Section Number	Section Name	Questionnaire		
		Core (C)	Next generation (K)	Foster Children (N)
	Household Roster			
	1.1 Household members from 1993 and 1998	Yes	Yes	Yes
	1.2. New household members	Yes	Yes	Yes
1	1.3 Death of household members	Yes	No	No
	1.4 Other decision makers	Yes	Yes	No
	1.5 Education and household relationships	Yes	Yes	No
	1.6 Parents of the Core Persons	No	Yes	Yes
2	Household Services	Yes	Yes	Yes
3	Food Spending and Consumption	Yes	Yes	Yes
4	Non-Food Spending and Assets	Yes	Yes	Yes
5	Remittances	Yes	Yes	Yes
6	Household Income from Non-Employment Sources	Yes	Yes	No
7	Economic Shocks	Yes	Yes	No
8	Agriculture	Yes	Yes	No
9	Employment	Yes	Yes	No
10	Health	Yes	Yes	Only 10.1
11	Social Capital and Trust	Yes	Yes	Yes
12	Children	Yes	Yes	Yes
13	Tests of Learning and Anthropometry	Yes	Yes	Yes