

SURVEY DESIGN

ADDENDUM

I. Funding/sample size and characteristics

1. With current funding (R620,000 contributed by the research team, to cover the cost of the November 1990 pilot survey as well as the main survey), the sample size will be 6,900 for an interview of one hour in length.

2. Additional interviews will be conducted at R100 each, to a maximum of 7,700, contingent on additional funding.

3. Interviews in White, Asian, and Coloured areas will be conducted in the fall of 1991 (around March) and interviews in Black areas will be conducted in the spring of 1991 (around August or September).

4. At the discretion of HSRC, and at no additional cost, all interviews may be conducted as late as the spring of 1991 (around August or September).

5. Probability samples will be designed by Prof. Stoker according to the specifications agreed to in our meeting of December 2, that is, a sample of 7,700 designed in such a way as to be reduced to 6,900 if no additional funding is forthcoming, or to somewhere between 7,700 and 6,900 if only partial additional funding is available (see below).

6. The sample of White, Asian, and Coloured areas will be designed first, and the sample of Black areas will be designed subsequently. The sampling frame will be based on the the 1985 Population Census estimates. If the small area data from the 1991 Census (the C-tapes) are available in time, they will be used to design the Black sample; if the C-tapes are not available, other more recent data will be utilized insofar as possible.

7. There will be no separate charge by HSRC for one or two pilot studies in addition to the main--November, 1990--pilot study. These additional pilot studies, of no more than 30 interviews each, will be conducted early in 1991. The main pilot sample consists of 210 interviews (30 in White areas, 30 in Asian areas, 30 in Coloured areas, and 120 in Black areas, chosen to cover a variety of respondent types).

8. The sample breakdown worked out in the December 2 meeting is shown below, together with the sample breakdown to be used in the event that the sample size is reduced to 6,900:

Total	7,700	6,900
<u>Urban general</u>		
White areas	2,500	2,200
Asian areas	800	700
Coloured areas	800	700
Black areas	2,500	2,200

Rural

TVBC + homelands	400	400
Commercial farms	100	100
Laborers on commercial farms	100	100

Special black

Elite areas	400	400
Mine hostels	100	100

9. Sampling will consist of three steps:

a) Selection of a stand or, for areas without stands (squatter settlements), of a dwelling unit, as far as possible randomly with probability proportionate to size.

b) For areas with stands, selection of a dwelling unit on the stand, randomly. In Black areas, all dwelling units on the stand have an equally probability of being selected. In White, Asian, and Coloured areas, the selection grid is to be designed such that on one of five multiple-household stands both the main dwelling unit and the secondary dwelling unit (or one of the secondary dwelling units, chosen randomly) will be selected. The point of this is to produce small samples of servants and tenants.

c) Within each dwelling unit, selection of an individual to be interviewed, randomly, among adults age 20 and over, but with males sampled at twice the rate of females.

10. Post-hoc weighting. Post-hoc weighting of the data will be carried out using the C-tapes (small area estimates) of the 1991 South African census, which are expected to be available in June, 1991. Two weights will be created for each respondent: one to produce a representative sample of each population group, and the other to produce a representative sample for all population groups combined. In each case, the weights will be designed to produce representative samples for all of Southern Africa (that is, including the homelands and TVBC states).

II. Questionnaire and field work

1. The questionnaire will have a maximum average length of one hour (when administered in English or Afrikaans); it will include educational, residential, and work histories, and extensive questions about the economic activity of the respondent, spouse, siblings, parents, and grandparents. The questionnaire will be drafted by Treiman, in consultation with the other members of the research team and the MarkData staff.

2. Call backs. If the selected respondent is not available to be interviewed on the spot, three call backs will be made, after which the stand (or where there are no stands, dwelling) to the right or left will be substituted. The entire screening, selection, and interviewing process will be repeated, including three call backs. If this effort results in no interview, the effort will be abandoned. Special replacement procedures will be developed for the White area sample, to cope with the fact that refusals tend to be substantially higher than for other groups.

3. Language

a) The questionnaire is to be translated into three languages: Afrikaans, Southern Sotho, and Zulu.

b) The questionnaires are to be independently back-translated into English; the results compared with the original English; and the translations modified accordingly. Prof. Schlemmer will be

responsible for the back-translations. But the questionnaires must be provided to him in a timely way.

c) A multi-lingual interview schedule is to be designed, with the full questionnaire, including space for coding responses on the right hand page of each pair of facing pages and Afrikaans, South Sotho, and Zulu translations on the left hand page.

d) A multi-lingual coding booklet is to be created, with response categories to questions shown as necessary.

e) Responses to open-ended questions are to be recorded in English in the field or recorded in the language of the interview and translated into English by the interviewer.

4. Open-ended questions

a) Occupation/principal duties/type of organization. These questions are to be asked in an open-ended way, for respondent's current job and work history; spouse's job; respondent's and spouse's father's job; father's father's job; mother's job; and job of sibling closest in age. Each of these sets of responses will be captured verbatim on computer tape, to be coded by Treiman. That is, the exact wording recorded by the interviewer (or, in the case of responses recorded in languages other than English, an exact English translation of it) will be captured.

b) Residence. These questions to be asked in an open-ended way, for respondent's current place of residence and residential history; father's place of birth; mother's place of birth; father's father's place of birth; and current place of residence of sibling closest in age. The responses are then to be coded by the interviewers each evening, using the 1991 geographical coding manual of the Central Office of Statistics, modified as necessary. Geographical information will be recorded down to the level of the suburb or township, or, where there are no suburbs, to the level of the town or nearest town.

c) Other. There may be a small number (<10) of other open-ended questions. In addition, some questions include "Other (specify)" responses. All open-ended responses, identified by record number, will be provided to Treiman, verbatim or in a verbatim English translation, in a Word Perfect or ASCII file.

5. Instruction manual and interviewer training. Because of the complexity of the questionnaire, and the novelty of life history questions in HSRC's experience, Treiman will draft an instruction manual for interviewers (to reach HSRC by January 15, 1991). HSRC will undertake to provide exceptionally careful training of interviewers--including, in particular, practice in administering the life history questions. Full fledged training, using the instruction manual, will be provided interviewers prior to each of the additional pilot studies.

6. Data products (in addition to items mentioned above)

a) Treiman will be sent (by international express mail) copies of the completed interviews from each of the additional pilot studies, to enable him to analyze the results and modify the questionnaire.

b) Each of the three principal investigators (Treiman, Schlemmer, and Moeno) will receive one copy of the computer tape containing the data from the main data collection and one copy of the codebook and other documentation.