Republic of Namibia   **National Planning Commission**

Central Statistics Office

The 1993/1994 Namibia Household Income and Expenditure Survey

NHIES

ADMINISTRATIVE AND TECHNICAL REPORT

Table of Contents

PREFACE

The 1993/94 Namibia Household Income and Expenditure Survey (NHIES) is the first module of the National Household Survey Programme endorsed by the Government in 1993. This programme is an integrated part of A Five-Year Development Plan of Statistics in Namibia (see chapter 1).

The reporting and dissemination of statistics from the NHIES follows a plan described in section 2.4 of this report. The present report is the ADMINISTRATIVE AND TECHNICAL REPORT from the 1993/94 Namibia Household Income and Expenditure Survey (NHIES).

There are two main objectives with this report:
The first aim is to document as much as possible of the NHIES survey activities and experiences to facilitate similar surveys in the future.
The second aim is to document the NHIES and its data files to ensure that the material is such that it lends itself to archiving. The documentation should make possible simple and independent reuse and further use of the NHIES statistics and microdata.

The report is produced by the Survey and Cartographic Unit of the Central Statistics Office. The preparation and production of this report was supported by technical assistance from the United Nations Development Programme (UNDP) and from the Swedish International Development Agency under the executing agency of Statistics Sweden.

The NHIES has been supported by various donors through bilateral and multilateral arrangements. On behalf of the Government of Namibia, I take this opportunity to thank the UNDP and the Government of Sweden for their valuable technical and financial support towards this project. I would also like to express my deep appreciation to our Government for its financial and material support.

Finally, I wish to thank all those who contributed to the success of the NHIES project, and in particular, the user/producer reference group of the NHIES, the CSO staff who were involved in the NHIES activities as well as the households of the NHIES sample living all over Namibia without whose support and co-operation the NHIES would never have been possible.

Sarah Kuugongelwa
Director General
National Planning Commission

March 1996

Chapter 0. ADMINISTRATIVE INFORMATION

0.1 Survey name

The 1993/94 Namibia Household Income and Expenditure Survey, NHIES

0.2 Responsible staff

Vekondja Tjikuzu	Survey and Cartographic Unit
John de Wet	Survey and Cartographic Unit
Staffan Wahlström	Survey and Cartographic Unit
Alvis Weerasinghe	Survey and Cartographic Unit
Ottillie Iiyambo-Mwazi	Data Processing Unit
Onno Amutenya	Data Processing Unit
Bertil Georgsson/Soe Paing	Data Processing Unit
Sammy Kinyanjui	Data Processing Unit

0.3 Respondent responsibilities

Respondents had to answer survey questions according to the Statistics Act of 1976.

0.4 Confidentiality rules

Survey data are fully confidential in accordance with the Statistics Act of 1976.

0.5 Further documentation

Annexes to this report
1. The NHIES Forms and Questionnaires.
2. The NHIES Manuals.
3. NHIES Data Set Descriptions
4. Selected SAS programs used in NHIES.

References

(1) Development of Statistics in Namibia, A Five Year Plan 1993/94 - 1997/98, NPC/Central Statistics Office, 1993.
(2) Development of a national household survey infrastructure in Namibia - Report from a first short-term mission to the Namibia Central Statistics Office, March 20 - April 4, 1992, J.Eklöf, H.Pettersson, NAMSTAT 1992:1, Statistics Sweden.
(3) Development of a national household survey infrastructure in Namibia - Report from a second short-term mission to the Namibia Central Statistics Office, September 1992, J.Eklöf, H. Pettersson, NAMSTAT 1992:2, Statistics Sweden.
(4) Development of a national household survey infrastructure in Namibia - Report from a third short-term mission to the Namibia Central Statistics Office, February 1993, H.Pettersson, NAMSTAT 1993:1, Statistics Sweden,.

(5) Development of a national household survey infrastructure in Namibia - Foundation for the NHIES 1993-1994, March 1993, J.Eklöf, NAMSTAT 1993:2, Statistics Sweden.

(6) The Namibia 1993/94 Household Income and Expenditure Survey. Pilot survey report. NPC/Central Statistics Office. August 1993.

(7) The Namibia Household Income and Expenditure Survey. Report from a short-term mission to the Namibia Central Statistics Office, April 5 - May 13, 1993, V.Säfström, NAMSTAT 1993:7, Statistics Sweden.

(8) Development of a national household survey infrastructure in Namibia - Report from a fourth short-term mission to the Namibia Central Statistics Office, March 1994, H.Pettersson, NAMSTAT 1994:4, Statistics Sweden.

(9) The Namibia Household Income and Expenditure Survey - Report from a short-term mission to the Namibia Central Statistics Office, September 6 - October 16 and November 20 - December 2, 1993, V. Säfström, NAMSTAT 1993:10, Statistics Sweden.

(10) Technical Advice for the Namibia Household Income and Expenditure Survey (NHIES) 1993/94 - Report from a mission to CSO, Windhoek, April 19 - May 5, 1994, L. Lundgren, NAMSTAT 1994:5, Statistics Sweden.

(11) Education and Training in the Statistical Analysis System (SAS) at the Namibia Central Statistics Office in Windhoek - Report from a short-term mission, February 9 - 22, 1994, L. Johnson, K. Rydenstam, NAMSTAT 1994:1, Statistics Sweden.

(12) Workshop on the Statistical Analysis System (SAS) - Report from a short-term mission to Central Statistics Office (CSO), Windhoek, March 20 - April 7, 1995, L. Johnson, K. Rydenstam, NAMSTAT 1995:5, Statistics Sweden.

(13) Cartography at CSO: Classification, organization and storage of maps. Production of Census Atlas. Cartographic Aspects of the Namibian Master Sample. - Report on a short-term mission, May 1994, M. Sidindi, A. Sporrek, NAMSTAT 1994:6, Statistics Sweden.

(14) Improvement of map documentation of Namibian master sample - Report on a mission to CSO, Windhoek, March 27 - May 5, 1995, M. Sidindi, O. Manyame, NAMSTAT 1995:8, Statistics Sweden.

(15) Improvement of map documentation of Namibian master sample - Report on a short-term review mission to the Central statistics Office, Windhoek, Namibia, December 4 - 22, 1995, M. Sidindi, NAMSTAT 1995: ?, Statistics Sweden.

Chapter 1. BACKGROUND

1.1. Introduction

Prior to the independence of Namibia no statistical organization existed in Namibia. A local section of the South African Statistical Service was responsible for the collection of statistics as required by the Office in Pretoria. Data processing, analysis, interpretation and dissemination were carried out in Pretoria.

After the independence of Namibia, the Central Statistics Office of Namibia (CSO) was established within the National Planning Commission (NPC) to provide users of statistics with priority data in a timely and efficient manner. A Five-Year Plan of Development of Statistics in Namibia - reference (1) in section 0.5 - has been formulated with emphasis on the generation of statistics required to monitor the progress towards the achievement of the Transitional Development Plan and in order to provide a firm statistical basis for the First National Development Plan .

The organizational and personnel structure of the CSO decided on by the Government is being gradually implemented. The United Nations Development Programme (UNDP) provides technical support for the establishment of the CSO. A statistics cooperation project has been initiated between the Governments of Sweden and United Kingdom on the one hand and the Government of Namibia on the other in order to assist with the building of CSO. Zimbabwe is also involved in the Swedish part of the cooperation project. Support towards setting up an infrastructure for national household surveys constitutes one of the central themes in this cooperation.

1.2. The Namibia Household Survey Programme

There is an acute need of policy oriented information concerning the level and change in the living standard of the Namibian population. The lack of accurate and timely information is especially disturbing concerning the people living in communal areas. Such quantitative information is needed for a multitude of purposes related to policy formulation, monitoring and reviewing of national policies, for macro economic analysis as well as for more research oriented studies. One of the major operational demands of improved statistical information is for the follow-up of the Transitional Development Plan, and in order to constitute an appropriate statistical basis for the First National Development Plan. This calls for integrated and representative household based information in addition to data from mainly administrative sources.

Some earlier efforts to collect crucial sets of data by sector ministries have not always turned out cost effective, or even viable.

The National Household Survey Programme of Namibia, which is endorsed by the Government, was discussed on a user-producer workshop held during the week of 30 August - 3 September 1993 in Windhoek. During the workshop a set of general recommendations for the development of statistics in Namibia was adopted. The need for cooperation and coordination between different producers and between users and producers as well as the central role of the CSO was emphasized - see reference (1) in section 0.5.

The National Household Survey Programme is a natural follow-up of the 1991 Population and Housing Census and represents one more step in providing useful statistics for charting and assessing the socio-economic development of the Namibian society.

The 1993/94 Namibia Household Income and Expenditure Survey (NHIES) is the first module of the National Household Survey Programme. **In 1995 the first Agricultural Census** was carried out which will be followed by **Annual Agricultural Surveys as from 1996**. **For 1996 the Intercensal Demographic Survey** is planned. **The Labour Force Survey will take place in 1997** and finally **in 1998 the Level of Living Survey/Health and Nutrition Survey/ Equal Opportunity Survey** (the subject-matter contents of the 1998 national survey will be further specified). These national surveys constitute the first round of a National Household Survey Programme of Namibia. A second round should start from 1999 and should be decided on not later than 1997 in connection with the planning of the next Population and Housing Census.

The National Household Survey Programme will generate statistical information for:

      i.   Overall national planning and monitoring.
     ii.   Sector planning, follow-up and evaluation.
    iii.   Design of development projects and programs
    iv.   As a basis for research and analytical studies relating to the Namibian society and economy at large.

The National Household Survey Programme will generate competence and capability for country-wide survey operations and set up standards of comparison between surveys.

One important aspect of the National Household Survey Programme is to make sure not to duplicate efforts, while at the same time allow for comparison between surveys.

Chapter 2. OVERVIEW

2.1 Objectives

Based on discussions with a broad group of users from the Government, Non-Government Organisations (NGOs), the Private Sector and Research Institutions, the following objectives were specified for the NHIES:

- Provide information for follow-up and evaluation related to the 1991 - 1993 Transitional Development Plan of Namibia, and constitute a baseline for the First National Development Plan.
- Offer new weights and proposed basic aggregates for updating and extending the calculation of consumer price indices. Price indices are to be computed for the entire nation as well as separately for geographical areas and income groups.
- Supply data needed for estimation of household final consumption expenditure, owner occupied dwelling and related quantities for the national accounts.
- Provide data for studies of income and expenditure patterns and related issues.
- Provide basic information for poverty mapping and targeting.
- Provide baseline data for project analysis in various phases of implementation.
- Constitute a platform in order to be able to specify a set of basic welfare indicators for Namibia (economic and social) to be monitored at regular intervals.

2.2 Contents

The 1993/94 Namibia Household Income and Expenditure Survey was designed to give urgently required information on income and expenditure patterns among **the Namibian private households**. In addition, data relating to employment and housing, as well as possession of and access to land, livestock, water, sanitation and transport services are included in the survey structure. Further, some information on demographic particulars and education was collected.

The target *population* of the NHIES is the private households of Namibia. So called **institutional households** (like hospitals, hostels, barracks and prisons) are **not** included in the NHIES. The size of the institutional population is about 100 000 persons according to the 1991 Population and Housing Census.

Basically the NHIES covers the following subject matter areas:

- Particulars of the household composition
- Education particulars of household members
- Economic activity status (employment particulars and labour market participation) of "adult" members of the household
- Availability of physical and social services and facilities
- Housing information

- Income and expenditure of domestic households (in cash and in kind)
- Household business activities(farm and non-farm activities)
- Household opinions concerning improvement of well-being of the household

2.3 Design

The sample design of the survey is covered in chapter 4.

In essence, the design of the survey was as follows:

- a twelve-month survey approach was used in order to cover seasonal variation
- a representative sample of private households was selected for participation in the survey
- the sample was divided into twelve monthly subsamples
- each selected household was requested to inform about their income and expenditure during one full month
- field interviewers were used for carrying out the interview task including weekly visits for income and expenditure compilation
- each household was supplied with a record book in order to facilitate successive recording of all household transactions
- initial editing and validation of recordings was carried out by interviewers and supervisors in the field
- manual editing and coding at the CSO office in Windhoek preceded data entry into the computers
- data processing was done by using micro computers and mainly the software programmes IMPS ,WORD PERFECT 5.1 and LOTUS 1-2-3 3.1 for DOS and EXCEL 5.0, SAS 6.08 and WORD 6.0 for WINDOWS.
- The Survey and Cartographic Unit at CSO has been responsible for most of the NHIES survey operation including tabulation and report writing.
- The Data Processing Unit has been responsible for data-entry and validation programmes as well as the actual data-entry into the computers. Also the running of the computer validation programmes has been the responsibility of the Data Processing Unit while the Survey and Cartographic Unit has dealt with the outcoming error lists.

Before, during and after the survey month a number of visits to selected households were carried out by the interviewers. This was done in order to collect basic data from the household on the survey topics.

The data collected were:
- Basic demographic data about every member of the household.
- Particulars relating to economic activity by all household members 10 years and over.
- Information about housing conditions and location factors of the household.
- Household possession of durable goods and access to facilities.
- Purchase of durable commodities and certain annual transactions.
- Information on monthly and annual income of household members.

All the above information was recorded to one questionnaire **Form I: Particulars on Individuals and Households** .

- A **Daily Record Book** (Form II) was distributed to each selected household before the start of the survey month in order to assist the household in the recording of all household transactions in cash (*transactions in cash are transactions where money is involved*) and transactions in kind (*transactions in kind are transactions where no money is involved*) on a daily basis during the survey month.
- Transactions in cash recorded by the households were transferred by the interviewers on a weekly basis to **Form III:Cash disbursement and Receipts**. The interviewers also regularly probed the households for cash transactions which they might have forgotten to record in the Daily Record Book.
- Transactions in kind recorded by the households were transferred on a weekly basis to **Form IV:Transactions in Kind**. The interviewers also probed the households for kind transactions which they might have forgotten to record in the Daily Record Book.
- The Daily Record Book was given to the head of household or other responsible member of the household (Primary Respondent) at the first visit. Special provisions were made if there was no literate person in the household capable of taking notes for the Daily Record Book. This situation called for more frequent visits to the household by the interviewers.
- The Record Book was translated into the major local languages.
- The last part of the questionnaire set **Form V: HOUSEHOLD OPINIONS** consists of a module about household opinions concerning how to improve the economic well-being of the households. Also this form was translated into the major languages.
- All forms but the Daily Record Book were kept by the interviewer throughout the month the household participated in the survey.
- Each interviewer was responsible for 8 households during a survey month. After completing the survey for all 8 households during the full month the interviewer checked the entries and delivered the entire batch to the supervisor, and after further scrutiny it was sent to the head office.

A full set of QUESTIONNAIRES used during the survey is included in annex 1 to this report (section 0.5).

For the field work the INTERVIEWER's INSTRUCTION MANUAL and the SUPERVISOR's INSTRUCTION MANUAL were developed. See annex 2 to this report (section 0.5).

For editing and coding in the CSO head office an INSTRUCTION MANUAL FOR MANUAL PROCESSING UNIT was developed. See annex 2 to this report.

2.4 Survey output

Basic demands on the NHIES were to provide good quality estimates of the survey variables (see section 2.2-3 and annex 1) for the nation as a whole as well as for urban and rural areas respectively. The NHIES should also provide estimates of a reasonable quality for the 13 regions.

A. NHIES REPORTS.

<u>Published reports</u>

- Living Conditions in Namibia 1993/94. Economic and Social Indicators from Namibia Household Income and Expenditure Survey. Preliminary Results for November 1993 - January 1994. CSO October 1994.
- The Distribution of economic resources in the population of Namibia. Some highlights based on the 1993/94 Namibia Household Income and Expenditure Survey. CSO November 1995.
- The 1993/94 Namibia Household Income and Expenditure Survey. Administrative and technical report. CSO March 1996.

<u>Planned reports</u>

- Living Conditions in Namibia 1993/94. Basic description with highlights. The 1993/1994 Namibia Household Income and Expenditure Survey. Main report.
(to be published by CSO in June 1996)
- Household consumption and expenditure. Detailed disaggregation. Table report on diskette.
(to be published by CSO in June 1996)
Early proposals for the design of the NHIES reports were developed during a short-term mission from Statistics Sweden in early 1994 (see reference 10 in section 0.5). The final design of the NHIES Main Report was agreed on in a reference group of users which started to meet regularly from February 1995. Draft tables of the Main Report above were discussed in detail during a three-day user-producer workshop at the end of October 1995.

B. AD HOC TABULATIONS

On request from Government Institutions, NGOs, the Private Sector and Research Institutions CSO since July 1994 regularly produces especially designed tables based on the NHIES.

C. DISSEMINATION OF MICRODATA

To advanced users CSO on request disseminates anonymized datafiles including microdata from the NHIES. To warrant the confidentiality aspect CSO has developed a special policy for this type of dissemination of data.

2.5 Time schedule

The first planning period. During this time-period four short-term missions from Statistics Sweden took place. Basic proposals for the NHIES design were developed including sample design, questionnaires and manuals. The developing of a sample design for the NHIES also included the developing of a proposal for a Namibian master sample. See the references (2), (3), (4) and (5) in section 0.5.
From February 1993 a long-term survey advisor from Statistics Sweden was connected to CSO. UNDP provided a UN volunteer for the survey work already from 1992.

The building up of a Survey and Cartographic Unit at CSO started.

April - September 1993

The pilot survey period. An NHIES pilot survey took place during April-June 1993 and was evaluated during July-August 1993. 12 temporary staff members were recruited as interviewers during the pilot survey.
Major revisions of the originally proposed questionnaires took place based on pilot survey experiences as well as on other considerations. The revised survey design was finally discussed at the User/Producer Workshop in August/September 1993 (see reference 1 in section 0.5) and a final NHIES survey design was decided upon.
The pilot survey is documented in reference (6) of section 0.5.

September 1993 - November 1994

The field work period. The NHIES survey period is a twelve months' time period from November 1993 - October 1994. The preparations started already in September/October 1993 by the training of field staff (9 field supervisors and 55 interviewers) and the listing of all households in selected geographical areas (primary sampling units) for the first survey month - November 1993. Listing of households in selected geographical areas, sampling of households for the NHIES and data collection went on from October 1993 to November 1994. The data collection was finalized during the first two weeks of November 1994.

December 1993 - August 1995

The editing, coding, data entry and data cleaning period. Editing, coding and data entry of the data collected in the field was conducted by temporary staff at the Data Processing Unit and the Survey and Cartographic Unit of the CSO Head Office in Windhoek. Further data cleaning and preparation of the data files was conducted by the professional team at the Survey and Cartographic Unit.

<u>July 1994 -</u>

The reporting period. The reporting of statistics from the NHIES started from July 1994. During the period July 94 - August 1995 the reporting was based on the first three survey months i.e. November 1993 - January 1994. From September 1995 the reporting is based on data from the full survey year i.e. November 1993 - October 1994. The CSO reporting from the NHIES will continue for a number of years. All planned CSO publications from the NHIES should be ready not later than June 1996.

Chapter 3. THE PLANNING PHASE

3.1 The first planning period

The first planning period went on from March 1992 to March 1993.
During this time-period four short-term missions from Statistics Sweden took place.
Basic proposals for the NHIES design were developed including sample design,
questionnaires and manuals. The developing of a sample design for the NHIES also
included the developing of a proposal for a Namibian master sample. See the references
(2), (3), (4) and (5) in section 0.5.
From February 1993 a long-term survey advisor from Statistics Sweden was connected
to CSO. UNDP provided a UN volunteer for the survey work already from the second
half of 1992.

The building up of a Survey and Cartographic Unit at CSO started. When the planning
of the NHIES started in March 1992 no professional local staff was yet in place to be
counterparts to the Swedish consultants. From the second mission in September 1992
one statistician and two statistical clerks without any prior survey experience were
assigned to the project.

The contents of the NHIES set of questionnaires were mainly decided on by the
statistical user community. The first user-producer meetings took place during
March/April 1992 during the first short-term mission of two Swedish consultants. Users
from the following institutions took part in these meetings: the National Planning
Commission, the Ministry of Finance, the Ministry of Trade and Industry, the Ministry
of Labour, the Ministry of Local Government and Housing, the Ministry of Lands,
Resettlement and Rehabilitation , the Bank of Namibia, the Namibian Economic Policy
Research Unit (NEPRU) and the Social Sciences Division of the University of
Namibia.

The user-producer meetings were followed up during the second short-term mission by
Swedish consultants in September 1992 and again in February /March 1993 during the
third short-term mission of Swedish consultants and in connection with the arrival of
the long-term survey advisor from Statistics Sweden.

In addition to formal user-producer meetings continuous deliberations took place in the
Survey and Cartographic Unit with or without participation of potential users. Also user
needs from units within CSO like the National Accounts section and the Consumer
Price Index section were highlighted. These deliberations were necessary to clarify what
type of data was needed as well as possible to collect. The main idea was to find a
balance between the requested information from the users and the response burden of
the households participating in the NHIES. The quality of the collected information
from the sampled households is very much correlated to the length and complexity of
the survey questionnaires.

At the same time as the set of questionnaires and field work manuals were drafted the
development of a Namibian master sample took place supported by the short-term
missions from Statistics Sweden.

At the end of March 1993 there was a draft set of NHIES questionnaires and field work manuals ready as well as a proposal for the sample design of the NHIES.

3.2 The pilot survey period

The pilot survey period went on from April - September 1993.
The field work of the pilot survey took place from May 8 to June 18, 1993. The main objective of this pilot survey was to test all stages of the survey machinery as far as possible. The questionnaires, the field procedures as well as the data-processing procedures were tested. Also logistic aspects like printing of questionnaires and manuals, payment of field staff and organizing of transport and accommodation for field staff were tested in the pilot survey. Other important aspects tested were how to handle the language question (many different languages are spoken in Namibia), how to organize the public relation activities and how to reward the participating households.

The preparations for the pilot survey started in April 1993. The preparations included recruitment of 12 interviewers on a temporary basis and training of these interviewers in the different aspects of the NHIES field procedures.

12 geographical areas in different parts of Namibia were selected for the pilot survey, 6 in rural and 6 in urban areas. The pilot survey was designed in such a way as to maximize the chances of detecting problems, inconsistencies and inadequacies in the survey operation. The intention was not to obtain any useful statistical information to be disseminated to the users, only to obtain data which could be used to critically test all steps in the survey process.

For a detailed documentation of the pilot survey and of proposals for the NHIES main survey based on pilot survey experiences, see the pilot survey report. (reference 6 in section 0.5). During the pilot survey period another short-term consultant from Statistics Sweden assisted CSO. See reference 7 in section 0.5 for a report from that mission.

Based on the pilot survey experiences and other considerations, revisions of the NHIES design took place during July - August 1993. Major changes were made in the questionnaires. The questionnaire design and basic survey definitions like the definition of "a private household" were finally decided on with broad user-participation during a half-day session of the user-producer workshop in August/September 1993 (see reference 1 in section 0.5). But the last questionnaire - Form V - dealing with households' opinions concerning how to improve the economic well-being of the households - was added to the set of NHIES questionnaires on request from the Government after the user-producer workshop (the final set of NHIES questionnaires is documented in annex 1 to this report).

Chapter 4. SAMPLE DESIGN

4.1   Introduction

The sample design is an important part of the design of the data collection of a survey. Other important parts of the data collection design are the design of the survey questionnaires and the procedures used for the data collection. The design of the NHIES questionnaires is documented in section 2.3 (overview) and annex 1 of this report (see section 0.5) while the data collection procedures are dealt with in section 2.3 (overview) and chapter 6.

The *population* of the NHIES is the private households of Namibia. So called **institutional households** (like hospitals, hostels, barracks and prisons) are **not** included in the NHIES. The size of the institutional population is about 100 000 according to the 1991 Population and Housing Census.

In countries like Namibia where no complete administrative listing of population or households exists the proper and efficient way of sampling households is basically to use *a two-stage sample design*. In the first stage a sample of geographical areas is selected. In the second stage a sample of households is selected within the selected geographical areas. In both stages randomized statistical methods are used which means that in principle all private households in Namibia have a known and positive probability to be included in the sample.

The basic demands on the sample design of the NHIES were to provide good quality estimates for the nation as a whole as well as for urban and rural areas respectively. The sample design should also provide estimates of a reasonable quality for the 13 regions.

4.2 The master sample frame

To be able to select a random sample of geographical areas in Namibia it was necessary to create a sample frame of geographical areas. For this purpose a frame of geographical areas - so called Primary Sampling Units (PSUs) - was created. The areas on the average contain about 100 households and most of the areas are in the range 80 -200 households. The areas are built from the enumeration areas (EA) that were prepared for the 1991 Population and Housing Census. Small EAs were combined with adjacent EAs to form PSUs of a sufficient size. The rule applied was that the number of households in a PSU according to the 1991 Population and Housing Census should be at least 80 households.

Table 4.1. Number of PSUs in the master sample frame by region and stratum within region.

REGION	STRATUM	NUMBER OF PSUS
Caprivi	**Total**	**110**
	Rural	93
	Urban	17
Erongo	**Total**	**82**
	Rural	25
	Small urban	27
	Urban	30
Hardap	**Total**	**91**
	Rural	44
	Small urban	22
	Urban	25
Karas	**Total**	**89**
	Rural	32
	Small urban	26
	Urban	31
Khomas	**Total**	**227**
	Rural	30
	Urban - High income	64
	Urban - Middle/Low income	133
Kunene	**Total**	**85**
	Rural	58
	Small urban	17
	Urban	10
Ohangwena	**Total** - Rural	**199**
Okavango	**Total**	**113**
	Rural	92
	Urban	21
Omaheke	**Total**	**63**
	Rural	45
	Small urban	7
	Urban	11
Omusati	**Total** - Rural (216) + Small urban (3)	**219**
Oshana	**Total**	**140**
	Rural	110
	Small urban (3) + Urban (27)	30
Oshikoto	**Total**	**147**
	Rural	125
	Urban	22
Otjozondjupa	**Total**	**130**
	Rural	64
	Small urban	19
	Urban - High income	13
	Urban - Middle/Low income	34
NAMIBIA	**Total**	**1695**
	Rural	1133
	Small urban	124
	Urban	438

About 1300 of the 1695 PSUs are made up of single EAs from the 1991 Population and Housing Census while about 400 PSUs are formed by joining two or more EAs (see reference 8 in section 0.5).

The 1695 PSUs - covering the whole of Namibia **except the Walvis Bay area** - were stratified into strata of PSUs by region and by rural, small urban and urban areas according to table 4.1. The stratification into rural, small urban and urban areas was based on a classification of enumeration areas conducted during the preparations of the 1991 Population and Housing Census. (Note: A **different** definition of rural and urban areas is used in the statistical **reporting** from the 1991 Population and Housing Census and the NHIES.) As seen in table 4.1, the urban areas in the Khomas region and some

urban areas in the Otjozondjupa region were further stratified into high income and middle/low income areas. In this way 32 strata were created for the sampling of PSUs and the number of PSUs in each stratum is presented in table 4.1.

As a result of the way the PSUs are created, the number of PSUs of the master sample frame in each region and in each stratum is roughly proportional to the number of households in the region and in the stratum respectively as estimated in the 1991 Population and Housing Census.

Within the 32 strata the PSUs were listed in a geographical order. In urban and small urban areas the PSUs were also listed by the income level of the areas. First in the list of PSUs in urban and small urban strata the high income areas are to be found and they are followed in the list by the middle/low income areas.

CSO has prepared EXCEL-files of the master sampling frame of PSUs. The frame contains for each PSU the following information :
-    Region
-    A unique PSU number
-    Rural/Urban information
-    Income level (only for urban areas)
-    District (old administrative regions)
-    EA number(s)
-    Number of households as from the Census
-    Accumulative number of households by stratum
-    Population by sex according to the Census
-    Master sample usage (whether the PSU is in the master sample or not)
-    Master sample PSU number (only for PSUs in the master sample)
-    Weights (raising factors) (only for PSUs in the master sample).

There is one EXCEL-file for each region and within each EXCEL-file the PSUs are grouped by stratum according to table 4.1. The EXCEL-files are available at the Survey and Cartographic Unit.

4.3 The master sample

The selection of the master sample from the master sample frame is the first sampling stage of the two-stage sample design introduced in section 4.1. This fact is the reason why the geographical areas in the master sample frame are called *Primary Sampling Units* (PSU). The sample of geographical areas is called a *master* sample because the idea is to use this sample of geographical areas not only for one survey but for several surveys in the National Household Survey Programme and also for other surveys during a certain number of years (5-10 years).

Such a master sample should be looked upon as an important part of the statistical infrastructure of Namibia. Properly used the master sample will provide a coordinating mechanism for Namibian survey activities. The construction of a master sample is also the starting point for the building up of a *field organisation* for national sample surveys. The recruitment of field staff - field supervisors and interviewers - is to a large extent monitored by the geographical situation of the PSUs in the master sample.

The first survey in which the master sample was used was the NHIES.

The proposed *overall* master sample design - i.e the combined sample design of the first and second sampling stage - to be used in Namibian national household surveys is a standard *Probability Proportional to Size* (PPS) sample design (see section 0.5 reference 2 and 8). The design employs PPS-selection of PSUs in the first stage and a fixed-sized equal probability sample of households in each selected PSU in the second stage. The design has certain advantages:

- The design has approximately optimal sampling efficiency for national estimates in the situation where no information on differences between strata in variance and costs are available.
- There will be a comparatively equal workload in each PSU.
- Under certain conditions the sample will be approximately "self-weighing" (- a characteristic of the sample design which will simplify the analytical work. But the NHIES sample is **not** self-weighing. For clarification see section 5.3.)

The size of the master sample was set to be 192 PSUs. The rationales of this sample size are that the number of PSUs is big enough to
- allow for disaggregation of the statistics by region and by rural and urban areas
- allow for several national sample surveys over a certain number of years up to a sample size of 6-7000 households per survey without too much of respondent fatigue

Initially the master sample was proportionally allocated over the strata in the master sample frame according to the number of households in the 1991 Population and Housing Census. However, some modifications of the allocation were made based on the following:
- The variation between households in income level seems to be generally larger in the urban areas than in the rural areas.
- The survey costs are considerably lower in the urban areas.
- There should be at least 10 PSUs sampled from each region to allow for reasonably good statistics from each region.

Therefore it was necessary to have a slight oversampling in urban areas and in one region (Omaheke). A proportional allocation of the 192 PSUs over urban/rural areas gave 66 urban and 126 rural PSUs. But taken into account the above mentioned facts, the selection of the master sample resulted into 81 urban and 111 rural PSUs. (See reference 8 in section 0.5.)

The allocation over regions and rural/urban areas is shown in the table 4.2 on the next page:

Table 4.2. The Namibian master sample distributed by region and rural/urban areas.

REGION	Total number of PSUs in Urban and Small urban areas. *	Number of PSUs in Rural areas	Total number of PSUs
Caprivi	2 (0)	10	12
Erongo	8 (3)	3	11
Hardap	6 (2)	4	10
Karas	7 (3)	3	10
Khomas	27 (0)	2	29
Kunene	4 (3)	6	10
Ohangwena	0	18	18
Okavango	3 (0)	10	13
Omaheke	3 (1)	7	10
Omusati	0	19	19
Oshana	7 (0)	11	18
Oshikoto	4 (0)	11	15
Otjozondjupa	10 (3)	7	17
Total	81 (15)	111	192

* The first figure is the total number of PSUs in Urban **and** Small urban areas while the figure in brackets is the number of PSUs in Small urban areas only.

Probability sampling of PSUs was used. The sampling procedure can be described as stratified systematic PPS-sampling where the size measure for each PSU in the master sample frame was the number of households according to the 1991 Census.

The 32 strata in the master sample frame were used (table 4.1) for stratification. Furthermore an "implicit" stratification was carried out according to the geographical dimension in rural areas and the "income level" in urban and small urban areas. This implicit stratification was a combined effect of the way the PSUs were listed within each stratum (see section 4.2) and the use of systematic sampling. With such an implicit stratification no new strata are formed but the result is an improvement of the sample design because the sample is spread all over the stratum.

CSO has prepared an EXCEL-file of the master sample of PSUs. The variable contents are the same as for the EXCEL-files of the master sample frame. Besides, there is one variable for the NHIES survey month in the PSU (section 4.5) and another variable for the number of listed households in the PSU just before the NHIES survey month (section 4.4).

See reference (8) in section 0.5 for a more extensive documentation of the Namibian master sample.

4.4 Selection of households

The selection of households is the second stage of the two-stage sample design introduced in section 4.1. That is why the households are called the *Secondary Sampling Units* (SSU). A list of households for each PSU was prepared during a separate listing exercise. The listing was carried out as closely as possible to the start of the data collection in a certain PSU i.e. the month before the NHIES survey month of the PSU. The listing was done in order to catch all the households of the PSU with their respective addresses. Information such as the address of the housing unit, name of the head of the household and the number of individuals in the household was collected during listing.

The list of households in the PSU was used as the sampling frame for the selection of households and a systematic equal probability random sample of 24 households from each PSU was drawn.

4.5 Special remarks

Large PSUs

Some PSUs in the master sample frame and also in the master sample are very large and comprise more than 300 households according to the 1991 Census. If such a large PSU was selected in the master sample it was decided on a slight modification of the sampling procedure.

Before the listing of households in the PSU, the PSU was devided into two approximately equal-sized parts and one of these parts was randomly selected with the probability 0.5. Then the second-stage sampling procedure including listing and sampling of households only took place in the selected part of the PSU.

This procedure was applied for 9 PSUs of the master sample. (By mistake this modified sampling procedure was not applied for one PSU in the master sample including 330 Census households and therefore the whole PSU was listed.)

Twelve subsamples

The whole sample of households excluding Walvis Bay comprises 192 * 24 = 4608 households. To be able to study the seasonal variation of household consumption and income this sample was divided into 12 monthly equal-sized subsamples. This was achieved by randomly allocating the 192 PSUs to one of the 12 survey months under the restriction that each survey month should cover 16 PSUs (16 * 12 = 192) and 16 * 24 = 384 households.

Because of flooding problems in the northern regions during the survey year, the first selected month had to be changed to another month for 8 PSUs of the master sample. This reallocation of months was made still under the restriction that each month should cover 16 PSUs and 384 households.

Stratification of households

No explicit stratification of households took place within the PSUs of the master sample. But also in this case an implicit stratification was achieved via the combined effect of the order the households were listed before sampling and the use of systematic sampling of households.

Because the households were basically listed in a geographical order the sampled households are well spread over the geographical area of the PSU which improves the accuracy of the statistics.

In PSUs comprising commercial farming areas an intentional relisting of the households took place before the sampling of households in such a way that the households of the commercial farmers were listed before the non-farmer households to increase the probability to select the households of commercial farmers. The income levels of the households of commercial farmers and the households of - for example - the farm workers are normally very different and therefore such a relisting also makes the final sample more efficient from the estimation point of view.

4.6 Sampling in Walvis Bay

Walvis Bay was not integrated in Namibia until 1 March 1994 and could therefore not be included in the NHIES before that date. For planning and logistic reasons Walvis Bay was included in the survey somewhat later - from May 1994. This means that Walvis Bay was included in the NHIES during the last six months of the survey year.

The sampling procedure was different from the rest of Namibia. In Walvis Bay the municipality authorities have for administrative purposes created computerized registers of the households in all the three main town areas - Central Walvis Bay (incl. Langstrand), Kuisebmund and Narraville. Most of the households in Walvis Bay are covered by these registers.

But there are some areas/population groups which are not covered by the administrative registers. These areas are the hostel areas of Walvis Bay and the area along the Kuiseb river where the Topnaar population lives. To cover also these population groups, CSO conducted listing of the households in these areas. **For security reasons all the hostel areas could not be listed** but some areas had to be excluded from the NHIES. The number of listed households in the hostel areas was 99 and probably about the same number of households was not listed. The number of Topnaar households listed was 73.

Altogether 144 households from Walvis Bay were selected by mainly a *stratified one-stage sample design* (see next paragraph for an exception) to be included in the NHIES sample. 24 households were sampled during each of the six months (6 * 24 = 144). Separate strata were defined for Central Walvis Bay (incl. Langstrand), Kuisebmund (excl. the hostel areas), Narraville, the hostel areas and the Topnaar population and altogether 36, 54, 36, 6 and 12 households respectively were sampled from each stratum.(During May-July 1994 10 households were selected each month in Kuisebmund excluding the hostel areas. During August-October 1994 only 8 households were selected in Kuisebmund excluding the hostel areas while 2 households were selected in the hostel areas.)

For 5 households in the sample of Walvis Bay the households were in fact sampled in two stages. There was a rule in the design for the sampling of households in Walvis Bay that in case of more than one household staying in the house/compound of a household selected from the municipality registers, a second sampling stage should take place. All households in the house/compound should be listed and in a second sampling stage one of the listed households was randomly selected for participation in the NHIES. The probability to be selected in the second stage was the same for each household in the house/compound. This rule was introduced to take care of so called backyard squatting. The rule was applied only in 5 cases in Kuisebmund.

Chapter 5. ESTIMATION

5.1 Introduction

The estimation procedure describes the mathematical approach to the compilation of statistics from the NHIES. The estimation procedure is of fundemental importance for the dissemination of statistical information based on the NHIES.

The purpose of the NHIES is to estimate characteristics of the Namibian household population or characteristics for specified groups of the household population. The characteristics are defined from the variables included in the NHIES questionnaires (see section 0.5 annex 1).

For example,

- How many are unemployed in Namibia ?
- How many are unemployed in the Oshana region ?
- What is the percentage of households headed by females in the Karas region ?
- What is the average annual income in households where Otjiherero is the main language ?
- What is the rate of food consumption in Namibian households ?
- What is the average annual food consumption in rural households ?

In a statistical vocabulary these characteristics are called *parameters* of the private household population. Basically we are interested in parameters which can be defined as totals or arithmetic means. A percentage is an arithmetic mean - the value of which is between 0 and 1 - multiplied by 100.

In this chapter we will mathematically describe these parameters as well as the estimators used to estimate them by means of the NHIES sample.

We will also introduce mathematical formulas used to estimate an important aspect of the quality of the estimators used - the so called precision of the estimators or random variation of the estimators. The quality of the NHIES statistics is further dealt with in chapter 9, where figures concerning the precision of the NHIES statistics are presented.

5.2 Population parameters to be estimated by means of the NHIES

In mathematical terms *a total* can be written

(5.1) $T = \Sigma X$

where
the Σ-sign means summation over all units in the survey population and
X is the survey variable of interest.

If the purpose is only to estimate characteristics of a certain part of the population - e.g. the Oshana region - the Σ-sign might symbolize summation of the X-values of the

households of the Oshana region only. However, this is equivalent to a summation including all households in Namibia if the X-values of the households outside the Oshana region are defined as 0.

A general formula for *an arithmetical mean* AM is a ratio between two totals

(5.2) $AM = \Sigma X / \Sigma Y$

where the summation takes place over the same units (individuals or households) of the population in the nominator and denominator. Y is equal to 1 for all units to be included in the arithmetical mean. So in case of an arithmetic mean, the ratio (5.2) can be written

(5.3) $AM = \Sigma X / N$

where
$N = \Sigma Y$ = the number of units of the population to be included in the arithmetic mean.

The arithmetic mean can be a national average, a regional average or an average for a specified population group of households or individuals e.g. all individuals in households where the sex of the head of household is a female.

The general formula for *a percentage* P can be written as a modification of (5.2)

(5.4) $P = 100 * \Sigma X / \Sigma Y$

In many cases formula (5.4) can be written

(5.5) $P = 100 * N_c / N$

where
$N_c = \Sigma X$ = the number of units of the population to be included in the mean which have a characteristic C
$N = \Sigma Y$ = the number of units of the population to be included in the mean.

X is a binary variable taking the values 1 or 0 depending on the unit having the specified characteristics C or not.
Y is 1 for all units of the population to be included in the mean.

C can be the characteristic to be a farm worker, to be unemployed, to be a resident of the Khomas region, to be a household where German is the main language etc.

With few exceptions the population *parameters* we estimate in the published tables of the NHIES can be defined by using the formulas (5.1) - (5.5).

5.3 Estimating totals

What is presented in the tables and diagrams of the NHIES reports is not exact values of the parameters defined in the former paragraph but estimates based on the data collected from the households of the NHIES sample. Such estimates can be of a better or worse quality. A mathematical function used to estimate a total like (5.1) is called an *estimator* of the total. Such an estimator is

(5.1e) $t = \Sigma f*x$

where the summation takes place over all units in the *sample*. In (5.1) the summation takes place over all units in the *population* i.e. also over those units which are not included in the NHIES sample. (For reasons of distinction we use lower case for the sample and upper case for the population.) We can not calculate (5.1) as we do not know X for all units in the population, but we can calculate (5.1e) as we can use the information collected from the NHIES sample.

f is called the *raising factor* or *weight*. Such a raising factor (weight) can be defined for each unit (household or individual) in the NHIES sample by using statistical theory. **From statistical theory for probability sampling it can be proven that if the raising factor f is defined as the inverse of the probability for the unit (household or individual) to be included in the NHIES sample then the formula in (5.1e) provides an estimate of the total (5.1) which - according to certain statistical criteria - is a good estimate.** One such criteria is that the estimator (5.1e) is *unbiased* (from the sampling point of view) which means that "on the average" such an estimate is equal to the population total (5.1). Using another terminology from the statistical theory we can say that "the expected value" of the estimator (5.1e) is equal to the total (5.1).

But even if an estimator is unbiased, it might still differ from the population total (5.1) because of random variation. The value of (5.1e) varies depending on which units are selected for the NHIES sample and this is the reason for the random variation of the estimator (5.1e). The random variation or the precision of the estimator (5.1e) will be estimated by means of the so called *standard deviation* which will be dealt with in section 5.3.

Compilation of raising factors and inclusion probabilities

From what is stated above, f should be defined by the formula

(5.6) $f = 1/pb$

where pb is equal to the probability for a unit in the sample (household or individual) to be selected. pb is called the *inclusion probability*. **This probability might differ between units in the NHIES sample.**

The probability for a specific Namibian private household **outside** Walvis Bay to be included in the sample is decided by the probababilities of the following occurrances:

a) the probability for the PSU where the household is staying to be included in the master sample of 192 PSUs (see section 4.2 and 4.3)
b) in case the PSU is selected in the first sampling stage and the PSU is a large PSU according to section 4.5 then the household must stay in the part of the PSU which is selected for listing
c) the probability for the household to be selected in the second sampling stage among the listed households in selected PSUs (see section 4.4)
d) we can only calculate the value of (5.1e) for response households and therefore only households in the selected sample of households which really participate in the NHIES can be included. Non-response households have to be excluded from the compilations.

The probability pb for household to be included in the final NHIES sample is equal to the product of the four probabilities for a),b),c) and d)
i.e.

(5.7) $$pb=pb_a*pb_b*pb_c*pb_d$$

pb_a depends on to which stratum of PSUs in the master sample frame the PSU of the household belongs to. More specific pb_a is decided by the number of PSUs sampled from that stratum, the number of households in the PSU according to the 1991 Census and the number of Census households in the whole stratum of PSUs.
The formula for pb_a can be written

(5.8) $$pb_a= m_h* (A_{hi}/A_h)$$

where

m_h = the number of PSUs sampled from stratum h (one of the 32 strata in the master sample frame described in section 4.2 table 4.1)
A_{hi} = the total number of Census households in the i-th PSU of stratum h
$A_h = \Sigma_i A_{hi}$ =the total number of Census households in stratum h of the master sample frame
(Σ_i means that the summation only takes place over the PSUs of the master sample frame included in stratum h.).

pb_b = 1 or 1/2. If the selected PSU is a large PSU according to section 4.5
pb_b= 1/2 in other case pb_b=1.

pb_c depends on the number of listed households in the PSU a couple of weeks before the month of the data-collection.

(5.9) $$pb_c = 24/N_{hi}$$

where
N_{hi}= the number of listed households in the i-th master sample PSU of stratum h
24 = the number of households sampled from each PSU in the second sampling stage.

pb_d is the probability for a sampled household to be a response household. **Such a probability is of a different type.** It is not depending on a controlled sampling stage as the earlier probabilities but on different factors deciding whether a household will participate in the NHIES or not. Such a probability might differ between households for example because of differences in willingness to participate in the NHIES. As a simplification we assume that pb_d is the same for all households from a sampled PSU and can be estimated by the formula for simple random sampling.

(5.10) $$pb_d = n_{hi}/24$$

where
n_{hi} = number of responding households in the i-th master sample PSU of stratum h

We can now rewrite formula (5.7) in the following way:

(5.11) $$pb = m_h*(A_{hi}/A_h)*pb_b*24/N_{hi}*n_{hi}/24$$

(5.12) $$= m_h*(A_{hi}/A_h)*pb_b*n_{hi}/N_{hi}$$

where pb_b= 1/2 or 1 depending on the PSU being a large PSU or not according to section 4.5.

<u>Four remarks</u>

I. n_{hi} is equal to 24 if all sampled households in the PSU are participating in the survey i.e. if there is no unit non-response of sampled households.

II. **The inclusion probability for a household member is the same as the inclusion probability for the household as all household members in a selected household are included in the NHIES**
III. pb is the same for all households or individuals in the population if the following criteria are fulfilled:

> 1) A_{hi}/N_{hi} = constant for all (h,i) i.e. for all PSUs in the master sample frame. This is the case if the relative change of the number of households since the 1991 Census is the same for all PSUs.
> 2) n_{hi} = constant. This is the case if the non-response rate is the same for all PSUs.
> 3) pb_b=1 for all PSUs which is the case if there are no large PSUs according to section 4.5.
> 4) m_h/A_h=constant. This is the case if the number of PSUs selected from each of the 32 strata in the master sample frame is proportional to the number of households in the stratum according to the 1991 Census.

These criteria are **not** fulfilled in the NHIES and therefore the inclusion probability pb and also the raising factor f vary between households from different sampled PSUs. **This means that the NHIES sample is not *self-weighing*.** In fact the raising factor f varies a lot. In the whole NHIES sample the variation is between 6.08 and 245.3. If Walvis Bay is excluded f varies between 10.5 and 160.1.

IV. The sampling design in Walvis Bay is described in section 4.6. As no sampling of PSUs took place in **Walvis Bay**, the formula for the inclusion probability of households in the Walvis Bay sample is somewhat simpler than for the rest of the country. For most households in Walvis Bay, formula (5.12) can be simplified to

(5.13) $$pb_w = n_h/N_h$$

where
n_h = the number of response households in the sample of stratum h in Walvis Bay
N_h = the number of households in stratum h in Walvis Bay according to administrative registers or performed listing .

Formula (5.13) corresponds to pb_c*pb_d of formula (5.7).

For those 5 households which were selected in a second sampling stage from a list of all households staying in the house/compound of the originally selected household, formula (5.13) has to be changed to

(5.14) $$pb_w = n_h/N_h * 1/L$$

where

L = the number of households staying in the house/compound of the originally sampled household from the municipality register.

5.4 Estimating arithmetic means and percentages

In the statistical theory it has been proven that a good estimator of the arithmetic mean parameter (5.2) is

(5.2e) $$am= \Sigma f*x / \Sigma f*y$$

where the summation takes place over all units in the *sample*. In (5.2) the summation takes place over all units in the *population* i.e. also over those units which are not included in the NHIES sample. (For reasons of distinction, we use lower case for the sample and upper case for the population.) We can not calculate (5.2) as we do not know X and Y for all units in the population but we can calculate (5.2e) as we can use the information collected from the NHIES sample.

f is defined in exactly the same way as for the estimator (5.1e) in section 5.3. It is evident that (5.2e) is a ratio between two totals t_x and t_y defined as in formula (5.1e).

Two remarks

I. As the percentage parameters (5.4) only means multiplying the parameter (5.2) by 100 a good estimator of (5.4) is

(5.4e) $$p = 100*(\Sigma f*x / \Sigma f*y)$$

II. **If** the raising factor was the same for all units (households and individuals) in the NHIES sample the formula for the estimator (4.2e) could be simplified to

(5.15)   am= Σx / Σy

i.e. there is no need to pay attention to the raising factors when calculating arithmetic means or percentages in the tables. This fact is the reason why a sample design where the raising factor is the same for all units is called *self-weighing*.

As we have seen in section 5.3 the raising factors are varying a lot in the NHIES **why the simplified formula (5.15)** *cannot* **be used** (see section 10.2)**.**

5.5 Standard deviation and confidence intervals

The value of estimators like (5.1e) and (5.2e) varies depending on which units are selected for the NHIES sample. This is called *the random variation of the estimator* . The lower the random variation the better the estimate, and on the contrary, the higher the random variation the worse the estimate. Lower random variation means higher *precision* and the higher the random variation the lower the precision of the estimate. Due to the fact that we are using probability sampling, it is possible to estimate the size of the random variation of estimators like (5.1e), (5.2e) and (5.4e). There are different measures of the random variation which can be used, but the most common measure is the *standard deviation*.

The standard deviation can be used to compile so called *confidence intervals* for the estimates of the NHIES. The formula for a 95 % confidence interval for a total "t" is

(5.16) $t \pm 1.96 * s(t)$

where s(t) = the estimated standard deviation for the estimator t.

In the same way the formula for a 95 % confidence interval for an arithmetic mean "am" is

(5.17) $am \pm 1.96 * s(am)$

where s(am) = the estimated standard deviation for the estimator am.

And the formula for a 95 % confidence interval for a percentage "p" is

(5.18) $p \pm 1.96 * s(p)$

where s(p) = the estimated standard deviation for the estimator p.

Confidence intervals are important indicators of the quality concerning random variation or precision of published statistics. The "confidence" is 95 % that the population parameter (T, AM and P respectively), which we want to estimate, is

situated **within** the 95 % confidence interval. Other percentages than 95 can be used when compiling confidence intervals but 95 % is a common standard for compiling confidence intervals which we will use in the NHIES. Of course, the narrower the compiled 95 % confidence interval is the better is the precision in our estimate.

To be able to compile confidence intervals from the NHIES sample we must define formulas for the standard deviation s(t), s(am) and s(p) respectively.

From statistical theory it can be deduced that s(t) can be compiled by means of the following formula:

(5.19) $$s^2(t) = \Sigma_h 1/m_h(m_h-1)\Sigma_i(z_{hi}-am_i(z_{hi}))^2$$

where
Σ_i means that the summation only takes place over the master sample PSUs included in stratum h (stratum h is one of the 32 strata in the master sample frame described in section 3.2)
Σ_h means that the summation takes place over the 32 strata in the master sample frame
m_h = the number of PSUs sampled from stratum h
$z_{hi} = (A_h/A_{hi})*(1/pb_b)*(N_{hi}/n_{hi})*x_{hi}$
$am_i(z_{hi}) = \Sigma_i z_{hi}/m_h$
x_{hi} = the sum of the survey variable x for all responding households of the i-th master sample PSU in stratum h
A_{hi} = the total number of Census households in the i-th master sample PSU of stratum h
A_h = the total number of Census households in stratum h of the master sample frame
pb_b = 1 or 1/2. If the selected PSU is a large PSU according to section 4.5
pb_b = 1/2 in other case pb_b=1.
N_{hi} = the number of listed households in the i-th master sample PSU of stratum h
n_{hi} = number of responding households in the i-th master sample PSU of stratum h

The second power of s(t) is called the *variance* of the total "t".

The formula (5.19) is valid for the whole of Namibia **except** Walvis Bay. The sample design and the estimator for Walvis Bay is described in section 4.6 and 5.3 respectively. For **Walvis Way** the corresponding formula to 5.19 is

(5.20) $$s_w^2(t) = \Sigma_h ((N_h-n_h)/N_h)(n_h/(n_h-1)) \Sigma_i(z_{hi}-am_h(z_{hi}))^2$$

where
Σ_h means that the summation takes place over the 5 strata in Walvis Bay
$z_{hi} = (N_h/n_h)*x_{hi}$
N_h = the number of households in stratum h in Walvis Bay according to administrative registers or performed listing
n_h = number of responding sampled households in stratum h
$am_h(z_{hi}) = \Sigma_i z_{hi}/n_h$

Σ_i means that the summation only takes place over the responding sampled households in stratum h

x_{hi} = the value of the survey variable x for i-th responding sampled households in stratum h.

There is no sampling of PSUs in Walvis Bay but the sampling of households is conducted directly from the 5 strata of Walvis Bay (section 4.6).

The variance for a total "t" for the whole of Namibia **including** Walvis Bay is

$$(5.21) \qquad\qquad s^2(t) = s^2(t) + s_w^2(t)$$

where $s^2(t)$ on the right hand side of the equal-sign is decided by the formula (5.19) and $s_w^2(t)$ by the formula (5.20).

The standard deviation for a total "t" for the whole of Namibia **including** Walvis Bay is decided by the (positive) square root of $s^2(t)$ on the left hand side of the equal-sign in formula (5.21).

<u>Two remarks</u>

I. For those 5 households in Walvis Bay which were selected in a second sampling stage from a list of all households staying in the house/compound of the originally selected household (see section 4.6 and remark in section 5.3) x_{hi} in (5.20) should be exchanged to $L* x_{hi}$ (see formula (5.14) for an explanation of L).

II. From formula (5.2e) and (5.4e) it is evident that the estimators of an arithmetic mean "am" and a percentage "p" are basically ratios between two totals "t_y" and "t_x". From statistical theory it can be deduced that a good variance estimator of such a ratio is

$$(5.22) \qquad s^2(t_x/t_y) = (t_x/t_y)^2(s^2(t_x)/t^2_x + s^2(t_y)/t^2_y - 2*cov(t_x,t_y)/ t_xt_y)$$

where t_x and t_y are defined by formula (5.1e) and $s^2(t_x)$ and $s^2(t_y)$ are defined by formula (5.21). $cov(t_x,t_y)$ is also defined in an analogical way as formula (5.21) but instead of an expression powered to two like $\Sigma_i(z_{hi}-am_i(z_{hi}))^2$ the crossproduct between x and y defined as

$$(5.23) \qquad\qquad \Sigma_i(z^x_{hi}-am_h(z^x_{hi})) (z^y_{hi}-am_h(z^y_{hi}))$$

is used where the first component of the product relates to x and the second to y.

The standard deviation of the estimator "am" (formula 5.2e) is the (positive) square root of (5.22) $s(t_x/t_y)$ and the standard deviation of the estimator "p" is

$$(5.24) \qquad\qquad s(p) = 100*s(t_x/t_y)$$

Chapter 6. THE FIELD WORK

6.1 The field work routine

<u>Field staff</u>

Initially, the field staff were divided into eight (8) teams of seven (7) interviewers and one (1) supervisor each. The seventh interviewer in each team did not work as a regular interviewer but served as an assistant to the supervisor. This task rotated between the interviewers in a team on a monthly basis.
In the case of the Caprivi and Okavango regions the team organisation became a bit different because of special circumstances concerning languages and distances. One original team of eight individuals was split into two smaller teams consisting of three (3) interviewers with one (1) supervisor operating in each of these regions. Therefore, even though originally eight teams were organised, during most of the survey year nine teams worked in the field of which two smaller teams in the Caprivi and Okavango regions. Altogether the field staff comprised nine supervisors responsible for a group of 55 interviewers countrywide.

With the reintegration of Walvis Bay in 1994, it was decided that this area should also be included in the NHIES. Hence, starting from May 1, 1994, a new team of 3 interviewers and one supervisor was established to cover Walvis Bay. This brought our total of interviewers and supervisors to fifty-eight (58) and ten (10) respectively.

All the field staff were temporarily employed by CSO.

The supervisors were selected from the enumerators of the 1991 Population and Housing Census who had a good performance during the Census field work.

The interviewers were mainly recruited via the Office of the Governor in the 13 regions. Important criteria for the recruitment were

- at least grade 10 school education
- earlier experiences as interviewer/enumerator
- feasible language background (it is necessary that the different languages of Namibia are spoken among the field staff)

Three (3) permanent staff members of CSO at principal statistical clerk level were appointed as "senior supervisors." The task of the senior supervisors was to serve as the constant link between the field staff and the professional staff at the head office, particularly the head of the Survey and Cartographic Unit. These senior supervisors were also responsible for the transportation of new FORMS from the head office to the teams in the field and the collection of completed/finished FORMS from the field for transportation to the CSO head office.

<u>Transport</u>

Each team of 1 supervisor and 7 interviewers was provided with two cars for transport. The smaller teams in Caprivi and Okavango were provided with one car each. No professional drivers were used to transport the field teams. But supervisors and to a certain extent also interviewers with driver's licences were used as drivers.

In total, 16 cars of which about 10 were 4x4 cars were allocated to the field staff. Another 3 cars were allocated to the CSO head office staff for the monitoring and following-up of the field work.

<u>Training</u>

Training of the field staff took place in September and October 1993.

The field supervisors were trained during one week in September 1993. All field supervisors had participated as interviewers in the pilot survey in April-June 1993 and therefore they already had a good knowledge of the NHIES field work when the field supervisor training started.

Training of interviewers took place during the first two weeks of October 1993. During the first week more than the necessary number of interviewers participated in the training. After one week some of the trainees were excluded from the training to get the needed number of interviewers during the second week of training. The exclusion of trainees from the first week was based on the results of a written test at the end of the first week as well as the general performance during the first week of training. One important aspect was the performance of a standardised interview.

(Many of the excluded trainees from the first week were later retrained and employed as interviewers because of drop-outs during the NHIES field work period.)

The professional staff of the Survey and Cartographic Unit acted as teachers during the interviewer training supported by one short-term consultant from Statistics Sweden. Also the senior field supervisors and the field supervisors acted as resource persons during the training.

The training took place in three places: Oshakati, Tsumeb and Windhoek.

For further details about the training of the NHIES field staff see reference (9) in section 0.5.

<u>The listing operation</u>

To select the households in the NHIES sample from each PSU an up-to-date list of all households in the PSU should be available. Preparing such a list of households for the selected PSUs in the master sample (see section 4.2-3) is known as the listing operation. The list of households serves two main purposes namely

(i) For the selection of households to be included in the NHIES sample

(ii) To extract the necessary information to compile the raising factors in the estimation formulas (see section 5.3).

The listing operation was conducted the month before the survey month in each sampled PSU and therefore listing of households was a regular activity along the whole field work period. The listing operation was conducted by the field supervisor supported by the assisting interviewer.

A special form - Form L - was used during the listing operation. (See annex 1 of this report, section 0.5).

Included in the listing operation was the selection of the 24 households in the NHIES sample from the listed households in each PSU. Random systematic sampling was used for this selection. The selection was performed by the field supervisors.

In PSUs comprising commercial farming areas, an intentional relisting of the households took place before the systematic sampling of the households in such a way that the households of the commercial farmers were listed before the non-farmer households. This was done to increase the probability for households of commercial farmers to be included in the NHIES.

The data collection

Each team was allocated two Primary Sampling Units (PSUs) per month. Three interviewers were responsible for the data collection in each PSU. Each interviewer was given eight (8) consecutive sampled households and was responsible for the data collection from these households for a period of one month.

The supervisor supplied each interviewer with a Sample Household Sheet for PSU -FORM S - (see annex 1 of this report, section 0.5) containing information on the 8 households for that particular interviewer. The information in Form S was transferred from Form L and contained the necessary information to identify and find the selected households in the field.

The different forms used in the collection of information from the selected NHIES households are documented in annex 1 of this report (see section 0.5). For a summary see section 2.3.
The main part of **Form I: Particulars on Individuals and Households** was filled in at the *first interview visit* which normally took place the week before the survey month.

In order to assist in the systematic recording of all daily transactions a **Daily Record Book - FORM II** - was given to the household at the first visit. The Daily Record Book was translated into the major local languages. The household was urged to record all their transactions on a daily basis in the book. If no literate person was available in the household or its proximity, frequent visits had to be paid by the interviewer. The first interview visit was followed by weekly visits to the household for collecting data on transactions.
Transactions in cash recorded by the household were transferred by the interviewers on

a weekly basis to **Form III:Cash disbursement and Receipts**. The interviewer also regularly probed the households for cash transactions which they might have forgotten to record in the Daily Record Book.

Transactions in kind recorded by the household were transferred on a weekly basis to **Form IV:Transactions in Kind**. The interviewer also probed the households for kind transactions which they might have forgotten to record in the Daily Record Book.

The last part of the questionnaire set **Form V: HOUSEHOLD OPINIONS** consists of a module about household opinions concerning how to improve the economic well-being of the households. The Form V interview took place at *the last interview visit* after the survey month when the data collection concerning all other forms was ready. Form V was also translated into the major languages in Namibia.

All forms but the Daily Record Book were kept by the interviewer throughout the month the household participated in the survey.

The supervisor visited his/her team regularly during the month to "spot-check" their performance.

After finishing the survey in one month, the interviewers carried out the concluding field check procedure for each set of questionnaires, signed and delivered them to the supervisor in charge. The supervisor in turn carried out his/her checks and signed the questionnaires.

The questionnaires - the FORMS - were then collected by the senior field supervisors around the middle of the month after the survey month for the transport to the Central Statistics Office (CSO) head office in Windhoek .

A detailed documentation of the listing and data collection operations is to be found in the NHIES Interviewer's Instruction Manual and Supervisor' s Instruction Manual (see annex 2 of this report, section 0.5).

6.2 Supervision and field trips

Regular visits were paid to the field by the head office personnel during the field work period from October 1993 - November 1994. During these visits spot-checks of the performance of interviewers and supervisors were carried out.

During the month of January 1994, the Swedish survey adviser to the CSO visited all the teams in their respective areas of operation for that month. Besides the aforesaid checks, the adviser discussed some of the problems detected in the office from the previous months' work already received. He also listened to the teams about the common problems they were encountering. This interchange of views took place in refresher training sessions for supervisors and interviewers.

The Chief Technical Adviser to the CSO also undertook a trip, specifically to some teams operating in the northern part of the country. He visited some of the PSUs together with the teams and observed them at work. This "familiarisation" trip gave the adviser some insight in the survey activities on the ground.

A two-day seminar with all the field supervisors was held in Windhoek by the professional staff of the Survey and Cartographic Unit during the month of April 1994, to discuss both technical and logistical problems as observed at the head office.
In addition to these, the head of the Survey and Cartographic Unit - assisted by one clerk from the manual-processing team - conducted further refresher training sessions during the months of July/August 1994. The training was held in three different venues and teams were grouped together. The teams from Khomas, Hardap/Karas, Omaheke and Erongo regions were trained in Windhoek. Those operating in Oshana, Ohangwena, Omusati and Oshikoto regions were trained in Oshakati while those from Caprivi, Okavango, Kunene and Otjozondjupa regions were trained in Grootfontein.

6.3 Some experiences during the NHIES field work.

One main problem during the field work period was car accidents and broken cars. The field work was interrupted or slowed down several times during the survey year for lack of properly working cars. The worst car accident happened in November 1993 when 3 persons - of which 2 interviewers - were killed and 4 persons - of which 1 senior supervisor, 1 supervisor and 2 interviewers - were more or less seriously injured. Besides the individual tragedies in this terrible car accident, in a moment it wiped out almost a full field team which had to be replaced urgently to keep the survey going. A short term consultant from Sweden conducted the training of new interviewers together with one of the professional staff members of the Survey and Cartographic Unit (see reference 9 in section 0.5).

The field staff experienced some problems in the data collection from certain sceptical households. A great part of the resistance came from affluent households. Many households in this group clearly expressed their fear that the data might be used against them due to the socio-economic standing in the society. But on the whole the household response rate was high in the survey - 92.5 percent.

In one region, the field staff encountered problems with the residents of one constituency who accused the interviewers of working for the local authorities while they did not consider themselves as part of the region but rather a neighbour region. Their resistance had both tribal as well as political undercurrents.

The survey operations in the north were almost disrupted when some instigators started to spread rumours that the CSO interviewers were a clandestine group working for the official opposition party. All these rumours started when one interviewer, due to her Christian faith, gave some foodstuff to one of the households she was working with for that particular month. She also gave old clothing and footwear materials to other household members who had begged her to help them. This, the instigators said, was to bribe those households and individuals to vote for the official opposition party at the detriment of the ruling party. The problem was further aggravated by the fact that the vehicle the interviewers used in this particular PSU was not the famous white 4x4 landcruisers with government (GRN) registration numbers. Instead they were using a red car with a white canopy -two colours which are associated with the official opposition party- but with a GRN registration number. The problem was thus political and the rumours spread to many parts in the northern regions via radio.

After the Christmas-New Year holidays, the survey experienced another type of problems. Some interviewers started to stay away from duty for more than a week or two while others resigned either to go school or take up permanent job appointments.

Because of flooding problems in the northern regions during the survey year, the first selected month had to be changed for 8 PSUs of the master sample. This reallocation of months was made under the restriction that each month should still cover 16 PSUs and 384 households.

Despite these problems and obstacles, it must be said that the field work on the whole ran well and for a greater part according to plan - especially in the view that this was the first time the newly organised Namibian CSO conducted a large national statistical survey.

Chapter 7. THE DATA PROCESSING

7.1 Introduction

The processing of the data of the NHIES was done in the following stages:

- Reception of questionnaires from the field
- Manual editing and coding
- Data entry
- Data cleaning - phase I
- Data cleaning - phase II
- Tabulation

Out of these stages, the first two and the last two were handled by the Survey and Cartographic Unit while the data entry stage and the data cleaning-phase 1 stage were handled by the Data Processing and Cartographic Unit.

7.2 Data system overview

I. SURVEY AND CARTOGRAPHIC UNIT

Questionnaires from the field

I.1. Checked in

I.2. Manually edited

I.3. Manually coded

Manually processed questionnaires transferred to

II. DATA PROCESSING UNIT

II.1. Data entry using IMPS
    - Automatic range checks according to the IMPS Data Dictionary
    - Additional validation checks via Concor program in IMPS
    - Any errors detected, corrected or forced in for later correction.

II.2. Verification
    - Independent data entry matching the first data set. Errors of the first data entry corrected.
II.3. Validation (Data cleaning - Phase I)
    - Validation checks via Concor program in IMPS
    - Error printout A : Forced in values at data entry printed on paper
    - Error printout B: Errors thrown out from the validation run printed on paper

Error printouts A and B, entered batches of questionnaires and the verified IMPS batch files in diskette transferred to

III. SURVEY AND CARTOGRAPHIC UNIT

III.1.Errors recorded in the printouts checked against questionnaires and the batch files.

III.2 Corrections recommended on the error printouts and the relevant questionnaires also corrected accordingly.

Batches of questionnaires stored in Survey and Cartographic Unit for future reference.

Error printouts A and B with the recommended corrections and the batch files in diskette transferred to

IV. DATA PROCESSING UNIT

IV.1. Batch files corrected according to the recommendations.

IV.2. Rerunning of the validation program for the corrected batch files. The correction procedure repeated until no errors occurred.

Finally cleaned IMPS batch files transferred to

V. SURVEY AND CARTOGRAPHIC UNIT

V.1. Cleaned IMPS batch files converted to SAS files

V.2. Data cleaning - Phase II in SAS

V.3. Finally cleaned SAS files

V.4. File handling to prepare the files for tabulation

V.5. Compilation of NHIES tables for dissemination to users using SAS and EXCEL.

7.3 Manual editing and coding

After enumeration had been completed in each PSU, completed questionnaires and other relevant documents were checked by the interviewers and field supervisors to ensure that entries made were consistent and that there were no missing entries. The questionnaires and any relevant survey materials were then submitted to the senior field supervisor, who in turn checked them before bringing them to the head office.

The first stage of the manual processing in the head office was the receiving and check-in of questionnaires. The documents for each PSU were received, checked-in and stored by the head of the Survey and Cartographic Unit to ensure that all documents sent to the field had been returned. It was expected that all five main questionnaires from each of the 24 households as well as the PSU Listing Form would be returned (see annex 1 to this report, section 0.5). If there were blank questionnaires, an explanatory note should be attached. This did not work out all the time. In cases where fewer than the expected

24 households from each PSU were received without any explanation, the field supervisor responsible for that PSU was requested to provide an explanation. Only the number of filled-in forms received were recorded in the especially designed receipt and check-in book. There was a column for each of the forms expected and hence whenever one of the forms was missing for a particular household, it was indicated in the receipt and check-in book.

After the check-in procedure was ready the real manual editing and coding could start. The manual editing and coding was conducted by 8 temporary staff members. 10 temporary staff members were trained for the work but 2 of them were transferred to the Data Processing Unit for data-entry work (see section 7.4). The manual editing and coding team was monitored in their daily work by a supervisor who is a permanent statistical clerk at the Survey and Cartographic Unit.

The questionnaires were organized in batches - one batch for each PSU. The editing and coding was conducted in one process and the questionnaires were distributed in batches to the different members of the manual editing and coding team. After the manual editing and coding of a batch was ready, another member of the team verified the conducted editing and coding of that batch.

The detailed instructions for the manual editing and coding are documented in the NHIES Instruction Manual For Manual Processing Unit in annex 2 to this report (section 0.5).

7.4 Data entry into computers

Once the questionnaires were manually edited and coded in the Survey and Cartographic Unit, they were transferred to the Data Processing Unit for data entry into computers. The maximum number of households expected to be completed in the field during a survey month was 384 (16 PSUs x 24 households). After the completion of the field work for a particular month the questionnaires underwent the following steps:
- Field work is finished at the beginning of the month after the survey month.
- Questionnaires reach the head office by the middle of the month after the survey month.
- Manual processing of these questionnaires takes about one month after receipt.
- Questionnaires reach the data entry phase about 11/2 months after the completion of the survey month.
- Data entry phase takes about 3 to 4 weeks.

Hence the data entry into computers of all the survey questionnaires pertaining to a particular survey month was completed in 2 to 2 1/2 months after the survey month.

Data entry machines

Two 286 PCs were used for the data entry most of the time.

Software

The IMPS (Integrated Microcomputer Processing System) package developed by the
US Bureau of Census was used for Data Entry and for Data Cleaning - phase I.
It has the following modules.

1.	CENTRY	-	Data Entry Module
2.	CONCOR	-	Data Cleaning Module
3.	QUICKTAB	-	Tabulation Module (simple tables)
3.	CENTS	-	Tabulation Module (advanced tables)

Staff

The responsible staff of the Data Processing Section is presented in section 0.2. Two
temporarily employed data entry operators were involved in entering data most of the
time. These operators were transferred from the editing staff of the Survey and
Cartographic Unit. The arrangement was to transfer persons to data entry from the
editing team depending on the number of edited questionnaires available at a certain
point of time. The pace at which the editing team produced their output was well
matched by the working speed of the two data entry operators and this pattern prevailed
till the NHIES data entry activity was finished.
In the daily work the data entry operators were supervised by the permanently employed
statistical clerk of the Data Processing Unit. Management and program development
was the responsibility of the professional staff of the Data Processing Unit (see section
0.2).

Preparation Procedures

The following preparations were fulfilled before the actual data entry:

1. Developing the Data Dictionary in IMPS.
This process defines the structure of the IMPS data set. The development of the Data
Dictionary involved the following activities :

- Creation of record types
In total there were 11 record types defined. Part A, B, C and D in Form I were put into
different record types. Part E and F were divided into 2 record types each. Form III
(Cash transactions), Form IV (Kind transactions) and Form V (Household opinions)
were each put into a separate record type. The questionnaires are documented in annex 1
of this report (section 0.5).

- Creation of common items
Eight common items were defined and they occur in all records. They occupy the first
18 positions of each record.

- Creation of variables
Under each record type the variables were defined according to the items in the relevant
part of the NHIES questionnaires. Variable names were allocated and the lengths were
defined for each variable. The values or range of values for each variable were then
defined along with the value names.
The IMPS Data Dictionnary is documented in annex 3 of this report (section 0.5).

2. Developing the data entry program in IMPS.
The NHIES Data Entry Program was developed using the CENTRY module of IMPS.
The screen format for data entry looked very similar to the questionnaire itself. This
simplified the work of the data entry.
The NHIES Data Entry Program is available at the Data Processing Unit.

3. Validation checks during data entry through the CONCOR module of IMPS.
Some validation checks were incorporated to the data entry phase. Especially it was
checked that only valid values of each variable were entered into the computers.
The IMPS/CONCOR program for validation checks during data entry is available at the
Data Processing Unit.

Data entry

After the above preparations the actual entry of NHIES data into the computers could
start. At an early stage the NHIES pilot survey data were entered to check the program.
Several shortcomings in the program were checked and eliminated in this way.

The set of household questionnaires relating to one PSU (maximum number = 24) were
taken as one **batch**. Transfer of questionnaires from the Survey and Cartographic Unit
was made on the basis of batches. Data entry was also done on the same basis.

Data entry was organized in the following way. Two batches of questionnaires were
entered at the same time by the two operators. Then they exchanged the batches among
themselves and entered them independently so as to verify the first data entry. In this
way most of the errors introduced by the first operator should have been corrected by the
second operator. But also after this second data entry some errors remained which were
rectified later during the data cleaning -phase 1.

During the data entry, the program threw out the values punched outside the ranges
defined in the IMPS Data Dictionary and also the values which were not consistent
according to the checks in the CONCOR program. These errors were corrected by the
two operators with support from their supervisor at the Data Processing Unit as well as
from the editing staff of the Survey and Cartographic Unit. Some errors were not
corrected at this stage because of their complexity. They were allowed to remain in the
data files. At the data cleaning-phase 1 they were identified again and corrections were
decided on by the professional staff of the Survey and Cartographic Unit.

7.5 Data cleaning - Phase I

Verified batch files from data entry were copied to one of the machines of the professional staff of the Data Processing Unit. Backup copies were also kept.
As the next step the professional staff of the Data Processing Unit produced frequency tables for all the variables and printed them on paper. IMPS QUICKTAB module was used for producing these simple frequency tables. The tables were then examined by the Survey and Cartographic Division staff. A number of errors were disclosed by these tables ("macro-editing").

The error sources were of the following types:
1. Errors made during the interviewing in the field but not captured during the manual editing process.
2. Errors made during the manual editing process.
3. Errors made during the data entry into computers.

A large number of tables were produced and printed. A lot of paper was used for printing. After some time it became evident that it was not necessary for the Data Processing Unit to produce these tables. Instead, only the batch files were transferred to the Survey and Cartographic Unit and the tables were produced by this unit. The printing of the tables was no longer necessary. The tables were scrutinized on the computer screen by the professional staff of the Survey and Cartographic Unit. Then the corrections were recorded on paper and sent to the Data Processing Unit for correction of the data files. These corrections were made on the verified batch files by the data entry supervisor.

This approach for the data cleaning-phase 1 was used for the batches from the first three survey months. The NHIES data files used for the Preliminary Report (see section 2.4) relied on this data-cleaning approach. The approach had some disadvantages:

1.   It was tiresome to watch the computer screen for a long time to detect these errors. Due to this some of the errors could be bypassed.
2.   Professional staff of the Survey and Cartographic Unit had to use a lot of their time for this data-cleaning purpose.
3.   Corrections had to be written on paper clearly right from the beginning and carefully so that Data Processing Unit staff could understand them.

It became very evident that this approach was not possible to follow for all the batches pertaining to all the 12 months. Instead a separate validation program was written for the data-cleaning-phase 1. From the earlier data-cleaning approach as described above, the most common type of errors had been identified. Using this knowledge the Survey and Cartographic Unit indicated the type of checks which had to be carried out by a new IMPS validation program. Based on these proposals a validation program was written by the Data Processing Unit using the IMPS CONCOR module. Certain automatic imputations were also included in the program. These automatic imputations dealt with codes for non-response and not-applicable.

Once the data entry of one month's data (number of batches = 16) was completed then these batches went through the new IMPS validation program for data-cleaning-phase 1.

The errors thrown out were printed on paper. The data set involved in the check was also downloaded into the error printout with the error message. Similarly the forced in values at the time of data entry were also printed on a separate paper. The Data Processing Unit sent these two error printouts and the batches of questionnaires together with the batch files in diskette to the Survey and Cartographic Unit.

At the Survey and Cartographic Unit every error was checked against the questionnaires. Sometimes it was necessary to look at the IMPS batch files to see what had been entered. In this way it was possible to identify how the error occurred. Then the correction was recommended on the error printout itself by changing the existing values of the variables which were already printed on the printout. The relevant questionnaires were also corrected if necessary.

Then these error printouts with proposed corrections and the diskette with the batch files were again transferred to the Data Processing Unit. The batches of questionnaires were stored in the Survey and Cartographic Unit. The data entry supervisor of the DataProcessing Unit made the corrections. Once the corrections were completed for a certain data set it went back for the validation run again. If any errors were disclosed, they were printed out again and the same correction procedure was repeated.

The IMPS validation program for data-cleaning-phase 1 was upgraded about four times at the beginning when new general errors were disclosed.

All batches from the full NHIES sample (also the batches from the first three survey months) were checked through the final version of the IMPS validation program before the data-cleaning-phase 1 was finished.

After all the corrections were completed the final corrected batch files were transferred to the Survey and Cartographic Unit. The Data Processing Unit kept back-ups of all the files.

7.6 The SAS software package

Both the Data-cleaning-phase 2 and the production of NHIES tables for dissemination to users were done by the Survey and Cartographic Unit using the SAS (Statistical Analysis System) software package. The SAS version used was SAS 6.08 for Windows. The CSO staff was trained in-house to use this package for data cleaning and tabulation purposes. The training was conducted by two consultants from Statistics Sweden. The training took place during two workshops with a time-period of about one year in between. The first workshop took place in early 1994 before the data-processing for the production of the Preliminary Report started. The second workshop took place in early 1995 before the production of tables based on the full NHIES sample started (see references 11 and 12 in section 0.5).

The first step was to convert the final IMPS files from the data cleaning - phase I to SAS files. First SAS files were created for each record type (see section 7.4) and each survey quarter excluding Walvis Bay. Separate SAS files were created for the Walvis Bay sample.

Then the SAS files were manipulated and combined in different ways during the data-cleaning - phase II and the tabulation work.

For example, files with individual information were combined with files with household information, quarterly files were merged to annual files and files with cash and kind transactions were aggregated and transposed to household variables.

All SAS programs for these file manipulations are stored at the Survey and Cartographic Unit. Examples of SAS programs are documented in annex 4 of this report (see section 0.5).

Two basic SAS files, used for tabulation purposes, are presented in detail in chapter 8.

7.7 Data cleaning - Phase II

During the data cleaning - phase II certain checks were conducted and households and individuals were excluded from the data files based on certain criteria:

- Households having no transactions concerning domestic consumption were excluded as unit non-response.
- Individuals who were recorded as absent from their household during the whole survey month were excluded as members of the household.

The value of certain variables was checked and some changes of recorded values were performed:

- If the code for "Other means of income" (8) or "Non Response" (9) was recorded for the variable "main source of income for the household" in Form I Part D it was tried to deduce some other code for this variable from the other collected information in Form I.
- Special checks were conducted concerning itemcodes and values for transactions in cash and in kind from Form III and Form IV. These checks resulted in several changes of itemcodes and values and even deletion of certain recorded transactions due to duplication.
- In some cases the "estimated market value of rent" from Form I Part C was evidently unrealistic and a more realistic value was imputed.

New variables were created and values imputed for "imputed rent in cash" and "imputed rent in kind". These imputations were based on certain standard assumptions as well as collected information in Form I Part C. Sometimes the collected rent values were looked upon as unrealistic and more realistic values were imputed based on information collected from the household itself and from other households in the same area. See chapter 8 for further information on the imputation of rent in cash and in kind.

In general, the checking, correction and imputation procedures during data cleaning - phase II were conducted thoroughly and individually. Corrections were decided on after scrutinizing the values in the SAS files carefully along with the questionnaires. For each type of check a small program was written using SAS language. When this was submitted, data sets with potential errors were separated and listed. After scrutinizing was ready the SAS files were rectified by means of SAS programs. All SAS programs

are stored at the Survey and Cartographic Unit. See annex 4 of this report for examples of SAS programs (section 0.5).

7.8 Tabulation

NHIES tables and charts for dissemination to users were produced using

- SAS PROC FREQ PROCEDURE
- SAS PROC TABULATE PROCEDURE
- A MENU DRIVEN SAS PROGRAM

The menu driven SAS program is developed at Statistics Sweden to simplify the compilation of tables, percentiles, lorenz curves and gini coefficients by means of SAS.

All the three mentioned approaches to produce NHIES tables and charts have been used by the professional staff of the Survey and Cartographic Unit.

At the initial stages, especially for the Preliminary Report (section 2.4), the tabulations were done using the Proc Freq Procedure. After the second SAS workshop in early 1995 the tabulation work for the NHIES Main Report was mainly performed using the Proc Tabulate Procedure. The menu driven SAS program has mainly been used for quick ad hoc tabulations and for producing lorenz curves and compiling gini coefficients.

When producing tables, an important aspect is to have the variables required in the data file. In many cases the already existing files have all the variables required. Then that file can be used directly to produce the table. If there is no file with the required variables then a new file has to be constructed using one or more existing files. This file handling in SAS is one of the most important aspects of the production of the tables. All file handling and tabulation programs are stored at the Survey and Cartographic Unit. Examples of SAS tabulation programs used in the NHIES are documented in annex 4 of this report (section 0.5).

SAS is not the best software for the final editing of the tables before publishing. For that purpose Microsoft EXCEL 5.0 and Word 6.0 for Windows have been used. This means that the SAS tables are first imported to EXCEL for further editing. Finally the EXCEL tables are imported to the NHIES report file which is written in Microsoft Word for Windows. To a certain extent Wordperfect 5.1 for DOS has also been used as a word processor. Wordperfect 5.1 files can easily be imported to Word for Windows.

To sum up: IMPS for DOS, SAS 6.08, EXCEL 5.0 and WORD 6.0 for Windows are the main software programs used to produce the NHIES reports.

Chapter 8. DATA SET DESCRIPTIONS

8.1 Introduction

The main purpose of this chapter is to document in detail the contents of the two basic data sets (files) used for producing the NHIES Main Report Part 1. One of the data sets is based on households as units (objects) while the other one is based on individuals. The first data set is used for producing statistics on the private households of Namibia while the second data set is used for producing statistics on the individuals of the private households of Namibia (the private household population).

Some derived variables used to produce the NHIES Main Report Part 1 are not included in the two basic data sets for households and individuals. These variables are dealt with in section 8.5 "Derived variables and data sets".

Most of the collected data in the 1993/94 NHIES are included in the two data sets but certain subject areas are not included. Income information collected in Form I Part F is excluded as well as household transactions in Form III dealing with Domestic Cash Receipts and Household Business (i.e. cash transactions from itemcode 575 and onwards in Form III are excluded). All information collected in Form V Household Opinions is also excluded from the data sets.

8.2 Definition of survey units

Private households constitute the population of the NHIES.

Definition of private household

A private household consists of one or more persons, related or unrelated, who live together in one or part of one or more than one housing unit/dwelling *and have common catering arrangements*.

In general, a household consists of husband and wife , their children with or without other relatives, domestic servants, boarders and lodgers.

A person who lives alone and caters for himself/herself forms a one person household.

For further details, see Interviewer's Instruction Manual in annex 2 of this report, page 19-20.

Definition of household member

A household member is a person *who slept in the household during the last 24 hours before an interview* which took place before or during the survey month *and who had common catering arrangements with the household* during these 24 hours.

For further details see Interviewers' Instruction Manual in annex 2 of this report, page 29.

8.3 The household data set

In this section we will in detail describe the basic NHIES household dataset used to produce the NHIES Main Report.

At the CSO the data set is stored as a SAS data set. The SAS file name is nf.hhfckts and the DOS file name is c:\sas608\sasdlib\sasexcf\hhfckts.sd2.

On request the SAS data set or part of the SAS data set can be transformed into other software program formats.

The basic name of the file is *hhfckts* where

- *hh* stands for household
- *f* stands for full survey year
- *c* stands for cash transactions
- *k* stands for kind transactions
- *t* stands for total of cash and kind transactions
- *s* stands for aggregates or sums of transactions.

All variables of the data set *hhfckts* are listed in alfabethical order based on the short names of the variables of the data set (see annex 3 of this report, section 0.5). Annex 3 contains a **codebook** of the dataset *hhfckts*. Most of the codes used for the different values of the variables in hhfckts are explained in the code book. Further explanations are given in this section.

How to read the code book

1. The first column presents the *short name* of each variable in the data set. This name is used for example in SAS programs when referring to a certain variable.
2. The second column classifies the variables in *numeric or character* (alfa-numeric) variables. Most variables in the data set are numeric.
3. The third column provides a *long name* of the variable which provides a better description of the variable. The long name can for example be used in tables.
4. The fourth column provides the *source* for the variable information. The source is important to get a full understanding of the meaning of a variable.
Mostly there is a reference to one of the NHIES forms. For example, the source of the variable BICYCLE in the first row is F1PD1m which means Form 1 Part D item 1m. In several cases the word "derived" or "decided" is written in the source column. This means that referring to an item in a form is not enough to explain the variable but further clarifications are necessary. Such clarifications are provided below in this section.
5. The fifth column provides the *value sets* of the different variables of the data set. This column is the real code book - it provides the meaning of all codes and values attached to each variable.

Clarifications on variables in the code book

BUSCODE: The code list of the industrial classification is included as annex 3 of the Instruction Manual For Manual Processing Unit (see annex 2 of this report, section 0.5).

C001 - C567: All monthly values in Form III have been recalculated to annual values through multiplication by 12.

CDMCNSF1: Total domestic consumption in cash is derived by adding all domestic consumption items in Form III.
However, for motor vehicles and furniture annual data collected in F1PE1a and F1PE1b have been used instead of the corresponding items in Form III. For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.
See also CDOMCONS below.

CDMCNSF1 = CDOMCONS + PRMV - C475 + PRFURN - (SC141 - C376 - C377).

CDMCSF12: Total domestic consumption in cash is derived by adding all domestic consumption items in Form III.
However, for motor vehicles and furniture annual data collected in F1PE1a and F1PE1b have been used instead of the corresponding items in Form III. For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.
And for school fees and insurances annual data in F1PE2- a,b,c,d and g have been used instead of the corresponding items in Form 3. For details, see SAS-program c:\nhies\ms\fullyear\anfee2.prg in annex 4.
See also CDOMCONS below.

CDMCSF12 = CDMCNSF1 + PREFEES + PRIFEES + SECFEES + HIEDFEES + OTHFEES - C505 - C506 - C507 - C508 - C551.

The variable CDMCSF12 is the variable used when estimating "total private consumption in cash" in the NHIES Main Report.

CDOMCONS: Total domestic consumption in cash is derived by adding all domestic consumption items in Form III i.e. all consumption items under heading 1. Domestic Consumption Expenditure.
But the following items are excluded as they are defined as investments or savings: C351, C353, C354, C450, C451, C452, C453, C455, C456 and C550.
Also the variables C340 and C350 are excluded. Instead the derived variable CRENT is used. See below for further details re CRENT.

COUNT: This variable is used for counting each record, for example when producing frequency tables.

CRENT: The values of CRENT, KRENT and TRENT have been decided by means of the variables HTYPE in F1PC1, LIGHT in F1PC4c, MVORENT in F1PC9, RENT in F1PC8, ROOMS in F1PC2, TENACY in F1PC7, TOILET in F1PC5 and WATER in F1PC6.

51

Certain general rules have been followed when deciding CRENT, KRENT and TRENT for a household. In case of non-response or in case of unrealistic values reported, rent values have been imputed. The imputed values are based on the variables mentioned above. Item values from the household itself and from neighbouring households in the same PSU have been used to decide values to be imputed for TRENT. For certain housetypes, standardized TRENT values have been used.

First all rent values were expressed as monthly values.

General rules which have been followed are:

FOR TRENT
- if tenancy = 4 then trent=rent;
- if tenancy = 1,2 or 3 then trent=mvorent;
- if htype=8 and rooms<=5 then trent=50;
- if htype=8 and rooms>5 then trent=100;
- if htype=10 and rooms<=5 then trent=20;
- if htype=10 and rooms>5 then trent=50;

In many cases the first 2 rules could not be followed because of non-response or unreasonable values reported. In such cases values for trent were imputed as indicated above. All such imputations are documented in separate files at the Survey and Cartographic Unit. The rules dealing with tradional houses and improvised houses have been followed without exceptions.

FOR CRENT and KRENT
- trent=crent+krent;
- if tenancy=4 then crent=trent;
- if tenancy=4 then krent=0;
- if tenancy=3 and rent<trent then crent=rent;
- if tenancy=3 and rent<trent then krent=trent-rent;
- if tenancy=3 and rent>=trent then crent=0.5*trent;
- if tenancy=3 and rent>=trent then krent=0.5*trent;
- if tenancy=1 or 2 then crent=0;
- if tenancy=1 or 2 then krent=trent;

The rules for CRENT and KRENT have been followed without exceptions.

Finally all rent values were multiplied by 12 to get annual rent values.

DISTRICT: The district codes were used already in the 1991 Population and Housing Census. The districts are named and defined according to the disaggregation in districts used before Independence.

HHNR: HHNR is a concatenation of PSUNO and SHHNO. This number is a unique identifier of each household in the hhfckts data file.

HHSIZE: HHSIZE (household size) is derived by counting all household members using the basic individual file. See section 8.4. See also SAS-program c:\nhies\ms\nodeja1\hhsize1.prg in annex 4.

HHTYPE: The variable HHTYPE is based on the variable "Relation to head of household" in F1PA3. If the head of household has no spouse, the household is a "Single" household and if the head of household has a spouse, the household is a "Couple". Only children to the head of household (code = 03 in F1PA3) are counted as children in the HHTYPE definition.
"Extended family" is a household which includes at least one of the following categories: spouse of child, grandchild, parent or spouse's parent and other relative (code = 04 - 07 in F1PA3). The "extended family" household may or may not include own children of the head of household. But the "extended family" household only includes members who are relatives of the head of household.
A household " with non-relatives" is a household which includes at least one member who is a non-relative of the head of household (code = 09 in F1PA3) .
Note:The occurrence of domestic employee(s) (code = 08 in F1PA3) does not influence the HHTYPE classification.

HIEDFEES: If HIED=NO or 9 then the value 0 is imputed for HIEDFEES.
If HIED=YES but non-response for HIEDFEES then a value is imputed for HIEDFEES based on corresponding fees paid by households in the same or neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\anfees2.prg in annex 4.

K001 - K567: Values and/or quantities were collected for all kind-transactions. In case of only quantities collected, values were imputed based on the reported quantities and knowledge of the local prices. This knowledge was provided by other households in the same PSU or by interviewing shop owners in the same PSU. Sometimes the data editing staff decided the values based on their knowledge of the local prices.
All monthly values in Form IV have been recalculated to annual values by multiplication by 12.

KDOMCONS: Total domestic consumption in kind is derived by adding all domestic consumption items in Form IV i.e. all consumption items under heading 1. Domestic Consumption Expenditure (see also Form III for detailed item codes).
But the following items are excluded as they are defined as investments or savings:
K351, K353, K354, K450, K451, K452, K453, K455, K456 and K550.
Also the variables K340 and K350 are excluded. Instead the derived variable KRENT is used. See CRENT for further details about KRENT.

KRENT: See CRENT

LHH: This variable describes the number of listed households in each PSU of the master sample and the number of households of the artificial PSUs defined in Walvis Bay. See PSUNO.
In the three main town areas of Walvis Bay i.e. Central Walvis Bay (incl. Langstrand), Kuisebmund and Narraville the lists were identical with computerized administrative registers of the households. These computer files were provided by the Municipality.

LOANPAY: If REPLNCR=NO or 9 then the value 0 is imputed for LOANPAY .
If REPLNCR=YES but non-response for LOANPAY then a value is imputed for
LOANPAY based on corresponding values paid by households in the same or
neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\anfees2.prg in
annex 4.

NUR: The Local Authorities Act No:23, 1992 has proclaimed 15 municipalities and
12 towns in Namibia. These areas are defined as "Urban". Also the Walvis Bay area is
defined as "Urban" except the area of the Topnaars which is defined as "Rural". The
rest of the country is defined as "Rural".

Municipalities: Swakopmund, Windhoek, Gobabis, Grootfontein, Karibib,
Karasburg, Keetmanshoop, Mariental, Okahandja, Omaruru, Otavi, Otjiwarongo,
Outjo, Tsumeb, Usakos, Walvis Bay

Towns: Hentiesbaai, Lüderitz, Okakarara, Ondangwa, Ongwediva, Opuwo, Oshakati,
Rehoboth, Rundu, Katima Mulilo, Khorixas, Arandis

*NUR is used for disaggregation in rural and urban areas in all tables from NHIES
which include a rural/urban disaggregation.*

OTHFEES: If OTHERINS=NO or 9, then the value 0 is imputed for OTHFEES .
If OTHERINS=YES but non-response for OTHFEES, then a value is imputed for
OTHFEES based on corresponding values paid by households in the same or
neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\anfees2.prg in
annex 4.

PENFEES: If PENINS=NO or 9, then the value 0 is imputed for PENFEES .
If PENINS=YES but non-response for PENFEES, then a value is imputed for
PENFEES based on corresponding values paid by households in the same or
neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\anfees2.prg in
annex 4.

PRBUILDM: If BUILDMAT=NO or 9, then the value 0 is imputed for PRBUILDM .
If BUILDMAT=YES but non-response for PRBUILDM then a value is imputed for
PRBUILDM based on corresponding values paid by households in the same or
neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in
annex 4.

PRCART: If CART=NO or 9, then the value 0 is imputed for PRCART .
If CART=YES but non-response for PRCART, then a value is imputed for PRCART
based on corresponding values paid by households in the same or neighbouring PSUs.
For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.
The variables CART and PRCART were not included in Form I during the first survey
month.

PREFEES: If PRESCHL=NO or 9, then the value 0 is imputed for PREFEES .

If PRESCHL=YES but non-response for PREFEES, then a value is imputed for PREFEES based on corresponding values paid by households in the same or neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\anfees2.prg in annex 4.

PRFURN: If FURN=NO or 9, then the value 0 is imputed for PRFURN .
If FURN=YES but non-response for PRFURN, then a value is imputed for PRFURN based on corresponding values paid by households in the same or neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.

PRHHANIM: If HHANIM=NO or 9, then the value 0 is imputed for PRHHANIM .
If HHANIM=YES but non-response for PRHHANIM, then a value is imputed for PRHHANIM based on corresponding values paid by households in the same or neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.

PRHOUSE: If HOUSE=NO or 9, then the value 0 is imputed for PRHOUSE .
If HOUSE=YES but non-response for PRHOUSE, then a value is imputed for PRHOUSE based on corresponding values paid by households in the same or neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.

PRIFEES: If PRISCHL=NO or 9, then the value 0 is imputed for PRIFEES .
If PRISCHL=YES but non-response for PRIFEES, then a value is imputed for PRIFEES based on corresponding values paid by households in the same or neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\anfees2.prg in annex 4.

PRLAND: If LAND=NO or 9, then the value 0 is imputed for PRLAND .
If LAND=YES but non-response for PRLAND, then a value is imputed for PRLAND based on corresponding values paid by households in the same or neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.

PRMV: If MV=NO or 9, then the value 0 is imputed for PRMV .
If MV=YES but non-response for PRMV, then a value is imputed for PRMV based on corresponding values paid by households in the same or neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.

PSUNO: The PSUs of the master sample are numbered from 001 to 192.
The master sample does not include Walvis Bay. In Walvis Bay artificial PSU-numbers are used defined by survey month and survey stratum in Walvis Bay. Also in Walvis Bay the PSUNO is a 3 digit number where the first digit defines the survey month, the second digit is 0 and the third digit defines the stratum. The first digit varies from 2 to 7 corresponding to the survey months May to October. The third digit varies from 0 to 4 where
0 = Kuisebmund excl single quarter areas
1 = Narraville
2 = Central Walvis Bay
3 = Kuisebmund-Single quarter area

4 = Topnaar area.

The Kuisebmund - Single quarter area was only surveyed from August-October 1994 while the rest of Walvis Bay was surveyed from May to October 1994.
See section 4.6 for further details re Walvis Bay.

Q: For households outside Walvis Bay this variable informs about the survey quarter of a household:
1 st survey quarter = Nov93 - Jan94
2nd survey quarter = Febr94 - April94
3rd survey quarter = May94 - July94
4th survey quarter = Aug94 - Oct94
All Walvis Bay households have the code Q=5.

RECTYPE: Record type is an important concept in the NHIES files of the IMPS program. See chapter 7 and annex 3. In the SAS-files this variable is redundant and of no use.

REGION: These are the thirteen political regions as proclaimed by the Delimitation Commission, Acts 22, 23, 24, and 27 of 1992. The Walvis Bay area has been given a separate code = 14.

RESCAT: RESCAT (Response category) indicates whether the household is the originally sampled one or a substituted one or, where applicable, it was impossible to collect data from this household.

RF: For households outside Walvis Bay the raising factor RF is derived according to the formulas (5.6) - (5.12) in chapter 5 of this report.
For households within Walvis Bay the raising factor RF is defined according to the formulas (5.6) , (5.13) and (5.14) in chapter 5 of this report.

RF1: For households outside Walvis Bay RF1 is derived according to the inverse of formula (5.8).
For households within Walvis Bay RF1 = 1.

RF4: For households outside Walvis Bay RF4 = 4*RF.
For households within Walvis Bay RF4 is not defined and therefore there is a missing value.

SC2:   = SUM of C560 - C567. See also Form III.
SC11:   = SUM of SC111-SC119, SC1110, SC1111, SC1112nb.
The different components of the sum are defined below.

SC12:   = SUM of SC121- SC124.
The different components of the sum are defined below.

SC13:   = SUM of CRENT, C342-C349,C352. See also Form III.
SC15:   = SUM of C465 - C469. See also Form III.
SC18:   = SUM of C540 - C544. See also Form III.

SC19:   = SUM of C551 - C555. See also Form III.
SC111:   = SUM of C001 - C015. See also Form III.
SC112:   = SUM of C020 - C039. See also Form III.
SC113:   = SUM of C045 - C051. See also Form III.
SC114:   = SUM of C055 - C064. See also Form III.
SC115:   = SUM of C070 - C075. See also Form III.
SC116:   = SUM of C080 - C104. See also Form III.
SC117:   = SUM of C110 - C117. See also Form III.
SC118:   = SUM of C120 - C154. See also Form III.
SC119:   = SUM of C160 - C162. See also Form III.
SC121:   = SUM of C245 - C263. See also Form III.
SC122:   = SUM of C270 - C290. See also Form III.
SC123:   = SUM of C300 - C322. See also Form III.
SC124:   = SUM of C330 - C334. See also Form III.
SC141:   = SUM of C360 - C378. See also Form III.
SC142:   = SUM of C385 - C396. See also Form III.
SC143:   = SUM of C400 - C414. See also Form III.
SC144:   = SUM of C420 - C441. See also Form III.
SC145:   = C454. See also Form III.
SC161:   = SUM of C475 - C484, C486. See also Form III.
SC162:   = SUM of C495 - C497. See also Form III.
SC171:   = SUM of C505 - C511. See also Form III.
SC172:   = SUM of C520 - C534, C341, C485. See also Form III.
SC1110:   = SUM of C170 - C173. See also Form III.
SC1111:   = SUM of C180 - C204. See also Form III.
SC1113:   = SUM of C235 - C240. See also Form III.
SC1112nb: = SUM of C215 - C219. See also Form III.
SC1112b: = SUM of C220 - C230. See also Form III.
SC141F1:   = PRFURN + C376 + C377. See CDMCNSF1 and CDMCSF12.
SC161F1:   = SC161 + PRMV - C475. See CDMCNSF1 and CDMCSF12.
SC171F1:   = SC171 + PREFEES + PRIFEES + SECFEES + HIEDFEES -
C505 - C506 - C507 - C508. See CDMCSF12.
SC19F1:   = SC19 + OTHFEES - C551. See CDMCNSF12.
SC2F1:   = SC2 + LOANPAY - C563.

SCHOOL: The 4 digit codes for primary schools of Namibia have been decided on
by the Ministry of Basic Education and Culture. See annex 4 of the Instruction
Manual for Manual Processing Unit where the codelist is presented. (The manual is a
part of annex 2 of this report, section 0.5.)

SECFEES: If SECSCHL=NO or 9, then the value 0 is imputed for SECFEES .
If SECSCHL=YES but non-response for SECFEES, then a value is imputed for
SECFEES based on corresponding values paid by households in the same or
neighbouring PSUs. For details, see SAS-program c:\nhies\ms\fullyear\anfees2.prg in
annex 4.

SEXHH: SEXHH is derived from the variables F1PA3 and F1PA5.

SHH: Number of response households in the PSU to which the household belongs. In Walvis Bay SHH is equal to the number of response households in the stratum to which the household belongs (see section 4.6).

SHHNO: This number is a sequential number for the sampled households within each PSU (also the artificial PSUs in Walvis Bay - see PSUNO).

The variables SK2 - SK1112NB below are based on only kind-transactions. They correspond to the variables SC2 - SC1112NB above which are based on only cash-transactions:

SK2:   = SUM of K560 - K567.
SK11:   = SUM of SK111-SK119, SK1110, SK1111, SK1112nb.
The different components of the sum are defined below.

SK12:   = SUM of SK121- SK124.
The different components of the sum are defined below.

SK13:   = SUM of KRENT, K342- K349, K352.
SK15:   = SUM of K465 - K469.
SK18:   = SUM of K540 - K544.
SK19:   = SUM of K551 - K555.
SK111:   = SUM of K001 - K015.
SK112:   = SUM of K020 - K039.
SK113:   = SUM of K045 - K051.
SK114:   = SUM of K055 - K064.
SK115:   = SUM of K070 - K075.
SK116:   = SUM of K080 - K104.
SK117:   = SUM of K110 - K117.
SK118:   = SUM of K120 - K154.
SK119:   = SUM of K160 - K162.
SK121:   = SUM of K245 - K263.
SK122:   = SUM of K270 - K290.
SK123:   = SUM of K300 - K322.
SK124:   = SUM of K330 - K334.
SK141:   = SUM of K360 - K378.
SK142:   = SUM of K385 - K396.
SK143:   = SUM of K400 - K414.
SK144:   = SUM of K420 - K441.
SK145:   = K454.
SK161:   = SUM of K475 - K484, K486.
SK162:   = SUM of K495 - K497.
SK171:   = SUM of K505 - K511.
SK172:   = SUM of K520 - K534, K341, K485.
SK1110:   = SUM of K170 - K173.
SK1111:   = SUM of K180 - K204.
SK1113:   = SUM of K235 - K240.
SK1112nb: = SUM of K215 - K219.

SK1112b: = SUM of K220 - K230.

SMONTH: The variable SMONTH provides the sequential number of the calendar month corresponding to a specific survey month. For example SMONTH = 11 for the first survey month which was November 1993. And SMONTH = 01 for the third survey month which was January 1994. And SMONTH = 10 for the last survey month which was October 1994.

The variables ST2 - ST1112NB below are based on aggregated cash- and kind-transactions. They correspond to the variables SC2 - SC1112NB and SK2 - SK1112NB above which are based on only cash- and kind-transactions respectively. The transactions T001 - T567 below are based on aggregated cash- and kind-transactions. They correspond to the variables C001 - C567 and K001 - K567 above which are based on only cash- and kind-transactions respectively.:

ST2:   = SUM of T560 - T567.
ST11:   = SUM of ST111-ST119, ST1110, ST1111, ST1112nb.
The different components of the sum are defined below.

ST12:   = SUM of ST121- ST124.
The different components of the sum are defined below.

ST13:   = SUM of TRENT, T342- T349, T352.
ST15:   = SUM of T465 - T469.
ST18:   = SUM of T540 - T544.
ST19:   = SUM of T551 - T555.
ST111:   = SUM of T001 - T015.
ST112:   = SUM of T020 - T039.
ST113:   = SUM of T045 - T051.
ST114:   = SUM of T055 - T064.
ST115:   = SUM of T070 - T075.
ST116:   = SUM of T080 - T104.
ST117:   = SUM of T110 - T117.
ST118:   = SUM of T120 - T154.
ST119:   = SUM of T160 - T162.
ST121:   = SUM of T245 - T263.
ST122:   = SUM of T270 - T290.
ST123:   = SUM of T300 - T322.
ST124:   = SUM of T330 - T334.
ST141:   = SUM of T360 - T378.
ST142:   = SUM of T385 - T396.
ST143:   = SUM of T400 - T414.
ST144:   = SUM of T420 - T441.
ST145:   = T454.
ST161:   = SUM of T475 - T484, T486.
ST162:   = SUM of T495 - T497.
ST171:   = SUM of T505 - T511.
ST172:   = SUM of T520 - T534, T341, T485.
ST1110:   = SUM of T170 - T173.

ST1111:   = SUM of T180 - T204.
ST1113:   = SUM of T235 - T240.
ST1112nb: = SUM of T215 - T219.
ST1112b: = SUM of T220 - T230.
ST141F1:   = ST141 + PRFURN - (SC141 - C376 - C377).
ST161F1:   = ST161 + PRMV - C475.
ST171F1:   = ST171 + PREFEES + PRIFEES + SECFEES + HIEDFEES -
C505 - C506 - C507 - C508.
ST19F1:   = ST19 + OTHFEES - C551.
ST2F1:   = ST2 + LOANPAY - C563.

T001 - T567: These variables are aggregates of the corresponding cash- and kind-
transactions C001 - C567 and K001 - K567.

TDALTF12: = TDMALTF1 + PREFEES + PRIFEES + SECFEES + HIEDFEES +
OTHFEES - C505 - C506 - C507 - C508 - C551.
This variable has been compiled in a different way than TDMCSF12 for checking
purposes but TDALTF12 = TDMCSF12. See below.

TDMALTF1: = TDOMALT + PRMV - C475 + PRFURN - (SC141 - C376 - C377).
This variable has been compiled in a different way than TDMCNSF1 for checking
purposes but TDMALTF1 = TDMCNSF1. See below.

TDMCNSF1: Total domestic consumption is derived by adding all domestic
consumption items in Form III and Form IV.
However, for motor vehicles and furniture annual data collected in F1PE1a and
F1PE1b have been used instead of the corresponding items in Form III. For details,
see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.
See also TDOMCONS below.

TDMCNSF1 = TDOMCONS + PRMV - C475 + PRFURN - (SC141 - C376 - C377)

TDMCSF12: Total domestic consumption is derived by adding all domestic
consumption items in Form III and Form IV.
However, for motor vehicles and furniture annual data collected in F1PE1a and
F1PE1b have been used instead of the corresponding items in Form III. For details,
see SAS-program c:\nhies\ms\fullyear\durable2.prg in annex 4.
And for school fees and insurances annual data in F1PE2- a,b,c,d and g have been
used instead of the corresponding items in Form III. For details, see SAS-program
c:\nhies\ms\fullyear\anfee2.prg in annex 4.
See also TDOMCONS below.

TDMCSF12 = TDMCNSF1 + PREFEES + PRIFEES + SECFEES + HIEDFEES +
OTHFEES - C505 - C506 - C507 - C508 - C551.

*The variable TDMCSF12 is the variable used when estimating "total private
consumption" in the NHIES Main Report.*

TDOMALT: TDOMALT = TDOMCONS. See below. TDOMALT has been compiled in a different way than TDOMCONS for checking purposes.

TDOMCONS: Total domestic consumption is derived by adding all domestic consumption items in Form III and Form IV i.e. all consumption items under heading 1. Domestic Consumption Expenditure.
But the following items are excluded as they are defined as investments or savings: T351, T353, T354, T450, T451, T452, T453, T455, T456 and T550.
Also the variables T340 and T350 are excluded. Instead the derived variable TRENT is used. See CRENT above for further details re TRENT.

TRENT: See CRENT.

UR: This variable defines the distribution of PSUs in the master sample as well as in the master sample frame in rural and urban PSUs. UR was only used when designing the master sample. No NHIES statistics has been published based on UR. The code UR = U is the same for Urban and Small urban PSUs. See section 4.2-3 of this report and the variable NUR above. See also section 10.1.

8.4 The individual data set

In this section we will in detail describe the basic NHIES individual dataset used to produce the NHIES Main Report.

At the CSO, the data set is stored as a SAS data set. The SAS file name is nf.indf and the DOS file name is c:\sas608\sasdlib\sasexcf\indf.sd2.

On request the SAS data set or part of the SAS data set can be transformed to other software program formats.

The basic name of the file is *indf* where

- *ind* stands for individual
- *f* stands for full survey year

All variables of the data set *indf* are listed in alfabethical order based on the short names of the variables of the data set (see annex 3 of this report, section 0.5). Annex 3 contains a **codebook** of the dataset *indf.* Most of the codes used for the different values of the variables in *indf* are explained in the code book. Further explanations are given in this section.

How to read the code book

1. The first column presents the *short name* of each variable in the data set. This name is used for example in SAS programs when referring to a certain variable.
2. The second column classifies the variables in *numeric or character* (alfa-numeric) variables. Most variables in the data set are numeric.
3. The third column provides a *long name* of the variable which provides a better description of the variable. The long name can for example be used in tables.
4. The fourth column provides the *source* for the variable information. The source is important to get a full understanding of the meaning of a variable.
Mostly there is a reference to one of the NHIES forms. For example, the source of the variable AGE in the first row is F1PA4 which means Form 1 Part A item 4.
In several cases the word "derived" or "decided" is written in the source column. This means that referring to an item in a form is not enough to explain the variable but further clarifications are necessary. Such clarifications are provided below in this section.
5. The fifth column provides the *value sets* of the different variables of the data set. This column is the real code book - it provides the meaning of all codes and values attached to each variable.

AGEFR15: All individuals have been classified into two major categories i.e children under 15 years, adults 15 years and above. For further details, see SAS-programme c:\nhies\ms\fullyear\employdf.prg in annex 4 of this report (section 0.5).

AGEGR5: Individual ages have been classified in groups of 5 years each. For further details, see SAS-programme c:\nhies\ms\majuju\agegr5.prg in annex 4 (see section 0.5).

COUNT: This variable is used for counting each record, for example when producing frequency tables.

DISTRICT: The district codes were used already in the 1991 Population and Housing Census. The districts are named and defined according to the disaggregation in districts used before Independence.

ECINAC: An individual who is 10+ years of age and not part of the labour force is defined as economically inactive. The variables used in this category are similar to those in F1PB7. Code 4 includes individuals having code = 2, 5, 6, and 7 in F1PB7. For further details, see SAS-programme c:\nhies\ms\fullyear\employdf.prg in annex 4 of this report (see section 0.5).

EMPLOYED: For individuals who are 10+ years of age EMPLOYED is derived from the variables WORKED (F1PB4), AVAIL (F1PB5), and LOOKED (F1PB6). For further details, see SAS-programme c:\nhies\ms\fullyear\employdf.prg in annex 4 of this report (see section 0.5).

EMPLST1: EMPLST1 is derived from EMPLOYED (see above) and EMPLSTAT (F1PB10). For further details, see SAS-programme c:\nhies\ms\fullyear\employdf.prg in annex 4 of this report (see section 0.5).

HHNR: HHNR is a concatenation of PSUNO and SHHNO. This number is a unique identifier of each household in the *indf* data file.

HHSIZE: HHSIZE (household size) is derived by counting all household members using the basic individual file. See SAS-program c:\nhies\ms\nodeja1\hhsize1.prg in annex 4 of this report (see section 0.5).

HLCOMPL: The code list of educational classification is included as annex 1 of the Instruction Manual For Manual Processing Unit (see annex 2 of this report, section 0.5).

INDNR: INDNR is a concatenation of PSUNO, SHHNO and PERSONNO. This number is unique identifier of each individual in the *indf* data file.

INDUSTRY: The code list of the industrial classification is included as annex 3 of the Instruction Manual For Manual Processing Unit (see annex 2 of this report, section 0.5).

LABOURF: LABOURF is derived from EMPLOYED (see above). For further details, see SAS-programme c:\nhies\ms\fullyear\employdf.prg in annex 4 of this report (section 0.5).

MAINOCC: The code list of the occupational classification is included as annex 2 of the Instruction Manual For Manual Processing Unit (see annex 2 of this report, section 0.5).

NUR: The Local Authorities Act No:23, 1992 has proclaimed 15 municipalities and 12 towns in Namibia. These areas are defined as "Urban". Also the Walvis Bay area is defined as "Urban", except the area of the Topnaars which is defined as "Rural". The rest of the country is defined as "Rural".

Municipalities: Swakopmund, Windhoek, Gobabis, Grootfontein, Karibib, Karasburg, Keetmanshoop, Mariental, Okahandja, Omaruru, Otavi, Otjiwarongo, Outjo, Tsumeb, Usakos, Walvis Bay

Towns: Hentiesbaai, Lüderitz, Okakarara, Ondangwa, Ongwediva, Opuwo, Oshakati, Rehoboth, Rundu, Katima Mulilo, Khorixas, Arandis

NUR is used for disaggregation in rural and urban areas in all tables from NHIES which include a rural/urban disaggregation.

PSUNO: The PSUs of the master sample are numbered from 001 to 192.
The master sample does not include Walvis Bay. In Walvis Bay artificial PSU-numbers are used defined by survey month and survey stratum in Walvis Bay. Also in Walvis Bay the PSUNO is a 3 digit number where the first digit defines the survey month, the second digit is 0 and the third digit defines the stratum. The first digit varies from 2 to 7 corresponding to the survey months May to October. The third digit varies from 0 to 4 where
0 = Kuisebmund excl single quarter areas
1 = Narraville
2 = Central Walvis Bay
3 = Kuisebmund-Single quarter area
4 = Topnaar area.

The Kuisebmund - Single quarter area was only surveyed from August-October 1994 while the rest of Walvis Bay was surveyed from May to October 1994.
See section 4.6 for further details re Walvis Bay.

Q: For households outside Walvis Bay this variable informs about the survey quarter of a household:
1 st survey quarter = Nov93 - Jan94
2nd survey quarter = Febr94 - April94
3rd survey quarter = May94 - July94
4th survey quarter = Aug94 - Oct94
All Walvis Bay households have the code Q=5.

RECTYPE: Record type is an important concept in the NHIES files of the IMPS program. See chapter 7 and annex 3. In the SAS-files this variable is redundant and of no use.

REGION: These are the thirteen political regions as proclaimed by the Delimitation Commission, Acts 22, 23, 24, and 27 of 1992. The Walvis Bay area has been given a separate code = 14.

RESCAT: RESCAT (Response category) indicates whether the household is the originally sampled one or a substituted one or, where applicable, it was impossible to collect data from this household.

RF: For households outside Walvis Bay, the raising factor RF is derived according to the formulas (5.6) - (5.12) in chapter 5 of this report.
For households within Walvis Bay, the raising factor RF is defined according to the formulas (5.6) , (5.13) and (5.14) in chapter 5 of this report.

RF1: For households outside Walvis Bay, RF1 is derived according to the inverse of formula (5.8).
For households within Walvis Bay RF1 = 1.

RF4: For households outside Walvis Bay RF4 = 4*RF.
For households within Walvis Bay RF4 is not defined and therefore there is a missing value.

SEXHH: SEXHH is derived from the variables F1PA3 and F1PA5.

SHH: Number of response households in the PSU to which the household belongs. In Walvis Bay SHH is equal to the number of response households in the stratum to which the household belongs (see section 4.6).

SHHNO: This number is a sequential number for the sampled households within each PSU (also the artificial PSUs in Walvis Bay - see PSUNO).

SMONTH: The variable SMONTH provides the number of the calendar month for a specific survey month. For example SMONTH = 11 for the first survey month which was November 1993. And SMONTH = 01 for the third survey month which was January 1994. And SMONTH = 10 for the last survey month which was October 1994.

UR: This variable defines the distribution of PSUs in the master sample as well as in the master sample frame in rural and urban PSUs. UR was only used when designing the master sample. No NHIES statistics has been published based on UR. The code UR = U is the same for Urban and Small urban PSUs. See section 4.2-3 of this report and the variable NUR above. See also section 10.1.

8.5 Derived variables and data sets

All variables needed to produce the tables of the NHIES Main Report are included in the two basic data sets *hhfckts* and *indf* which are described in detail in section 8.3 and 8.4 above. But for feasibility reasons, not all <u>derived</u> variables are stored in the two basic data sets but in separate data sets derived from these two basic data sets. In this section the most important of such variables used when producing the tables of the NHIES Main Report will be presented.

The headings below contain long names and short names of the variables presented.

All the variables presented below are stored in the SAS-file nf.chap89 at the Survey and Cartographic Unit.

<u>Two supporting variables</u>

rfind = rf*hhsize
rfinv = count/rf

<u>Highest level of educational attainment for head of household (HIED)</u>

Warning: The short name HIED is used for quite a different variable in nf.hhfckts.

Codes: 1 = No formal education
               2 = Primary education
               3 = Secondary education
               4 = Tertiary education
               9 = NR

For definitional details, see SAS-program chap8.prg in annex 4 of this report (section 0.5) as well as section 8.4.

<u>Number of full-time employment equivalents (TIMEEQV)</u>

Codes:
0 = No economic activity in household
1 = 0 - 0.5   full-time employment equivalents in household
2 = 0.5 - 1.0   full-time employment equivalents in household
3 = 1.0 - 1.5   full-time employment equivalents in household
4 = 1.5 - 2.0   full-time employment equivalents in household
5 = above 2.0 full-time employment equivalents in household
9 = NR

For definitional details, see the SAS-program chap8.prg in annex 4 of this report (section 0.5) and section 8.4 above.

Percentile groups - 90/10 division - based on domestic household consumption (PERA)

Codes:
1 =   Annual private household consumption less than the 90th percentile of TDMCF12
2 =   Annual private household consumption equal to or above the 90th percentile of TDMCF12

For definitional details, see SAS-program chap8.prg in annex 4 of this report (section 0.5) and section 8.3 above.

Percentile groups - 25/50/75/90/95/99 division - based on domestic household consumption (PERB)

Codes:
1 =   Annual private household consumption less than the 25th percentile of TDMCF12
2 =   Annual private household consumption equal to or above the 25th percentile of TDMCF12 but less than the 50th percentile of TDMCF12
3 =   Annual private household consumption equal to or above the 50th percentile of TDMCF12 but less than the 75th percentile of TDMCF12
4 =   Annual private household consumption equal to or above the 75th percentile of TDMCF12 but less than the 90th percentile of TDMCF12
5 =   Annual private household consumption equal to or above the 90th percentile of TDMCF12 but less than the 95th percentile of TDMCF12
6 =   Annual private household consumption equal to or above the 95th percentile of TDMCF12 but less than the 99th percentile of TDMCF12
7 =   Annual private household consumption equal to or above the 99th percentile of TDMCF12

For definitional details, see SAS-program chap8.prg in annex 4 of this report (section 0.5) and section 8.3 above.

Food consumption rate excl. alcoholic beverages and tobacco (FCONRATE)

FCONRATE = (ST11/TDMCSF12)*100

For further definitional details, see section 8.3.

Food consumption rate incl. alcoholic beverages and tobacco (FCRTNDSH and FCRTNSH2)

ST11NDSH = ST11+ST1112b+ST1113;
FCRTNDSH = (ST11NDSH/TDMCSF12)*100
FCRTNSH2 = (ST11NDSH/TDOMCONS)*100

For further definitional details, see section 8.3.

Variables presented in the first column of the tables 9.25 - 9.32 of the NHIES Main Report

Long names, short names and formulas are presented in this order. The variables on the right hand side of the formulas are available in the *hhfckts* file (see section 8.3).

FOOD EXPENDITURE(CFOOD)	CFOOD=SC11
Bread and cereals(CBREDC)	CBREDC=SC111
Meat(CMEAT)	CMEAT=SC112
Fish(CFISH)	CFISH=SC113
Milk, cheese and egg(CMILK)	CMILK=SC114
Oil and fat(COILF)	COILF=SC115
Vegetables, potatoes(CVEG)	CVEG=SC118
Fruits and nuts(CFRNUT)	CFRNUT=SC116+SC117
Sugar(CSUGAR)	CSUGAR=SC119
Non alcoholic beverages(CNONALC)	CNONALC= SC1110+SC1112NB
Other food(COTHF)	COTHF=SC1111-C202-C203
Meals(CMEALS)	CMEALS=C202+C203

OWN PRODUCED FOOD OR RECEIVED IN KIND(KFOOD)	KFOOD=SK11
Bread and cereals(KBREDC)	KBREDC=SK111
Meat(KMEAT)	KMEAT=SK112
Fish(KFISH)	KFISH=SK113
Vegetables, potatoes(KVEG)	KVEG=SK118
Fruits and nuts(KFRNUT)	KFRNUT=SK116+SK117
Other food(KOTHF)	KOTHF=SK114+SK115+SK119 +SK1110+SK1111+SK1112NB

TOTAL FOOD(TFOOD)	TFOOD=ST11
CLOTHING AND FOOTWEAR(TCLOTH)	TCLOTH=ST12
HOUSING(THOUSE)	THOUSE=ST13
FURNITURE AND UTENSILS(TFURN)	TFURN=ST141F1+ST142+ ST143+ST145
HOUSEHOLD OPERATIONS(THHOP)	THHOP=ST144
MEDICAL CARE(TMEDC)	TMEDC=ST15
TRANSPORT AND COMMUNICATION(TTRACOM)	TTRACOM=ST161F1+ST162
EDUCATION(TEDUC)	TEDUC=ST171F1
PERSONAL CARE(TPCARE)	TPCARE=ST18
RECREATION(TRECRE)	TRECRE=ST172
ALCOHOL AND TOBACCO(TALTOB)	TALTOB= ST1112B + ST1113
OTHER(TOTH)	TOTH=ST19F1

TOTAL CONSUMPTION (TCONS)	TCONS=TDMCSF12

- Own produced goods or

received in kind(KOWNG)   KOWNG=KDOMCONS-
                                                 KRENT

- Imputed rent(KIMRENT)   KIMRENT=KRENT

+ Remittances in cash
given away(CREMIT)   CREMIT=C561

+ Housing investments(CINV)   CINV=PRBUILDM+
                                                 PRLAND+PRHOUSE

+ Domestic animal investments(CDMANIM)   CDMANIM=PRHHANIM

+ Savings and other investments(CSAVE)   CSAVE=PENFEES+C562+
                                                 LOANPAY+C564+C565

+ Income tax and other
  wage/salary deductions(CTAX)   CTAX=C560

+ Non-consumption expenditure(CNONCEXP)   CNONCEXP=C566+C567

TOTAL EXPENDITURE(TOTEXP)   TOTEXP=TCONS-KOWNG-
                                                 KIMRENT+CREMIT+CINV+
                                                 CDMANIM+CSAVE+
                                                 CTAX+CNONCEXP

For further definitional details, see section 8.3.

Total household income (TINCOME)

TINCOME=TCONS+CREMIT+CINV+CDMANIM+CSAVE+CTAX+CNONCEXP

Total annual household income (TINCOME) is defined as total annual domestic household consumption plus certain further transactions mainly related to remittances given away in cash, household savings in cash, household investments in cash and income tax payments.

Percentile groups - 90/10 division - based on total household income (PERAI)

Codes:
1 =   Annual private household consumption less than the 90th percentile of TINCOME
2 =   Annual private household consumption equal to or above the 90th percentile of TINCOME

Percentile groups - 25/50/75/90/95/99 division - based on total household income (PERBI)

Codes:
1 =   Annual private household consumption less than the 25th percentile of TINCOME
2 =   Annual private household consumption equal to or above the 25th percentile of TINCOME but less than the 50th percentile of TINCOME
3 =   Annual private household consumption equal to or above the 50th percentile of TINCOME but less than the 75th percentile of TINCOME
4 =   Annual private household consumption equal to or above the 75th percentile of TINCOME but less than the 90th percentile of TINCOME
5 =   Annual private household consumption equal to or above the 90th percentile of TINCOME but less than the 95th percentile of TINCOME
6 =   Annual private household consumption equal to or above the 95th percentile of TINCOME but less than the 99th percentile of TINCOME
7 =   Annual private household consumption equal to or above the 99th percentile of TINCOME

Total household income 2 (TINCOME2)

TINCOME2=TINCOME+K351+K450+K451+K452+K453+K455+K456

TINCOME2 is used in the income tables of the NHIES Main Report Part 1.

The difference between TINCOME and TINCOME2 is that among the investment transactions the latter variable also comprises building materials and household animals received by the household as a kind transaction.

Percentile groups - 90/10 division - based on total household income 2 (PERAI2)

Codes:
1 =   Annual private household consumption less than the 90th percentile of TINCOME2
2 =   Annual private household consumption equal to or above the 90th percentile of TINCOME2

Percentile groups - 25/50/75/90/95/99 division - based on total household income
(PERBI2)

Codes:
1 = Annual private household consumption less than the 25th percentile of TINCOME2
2 = Annual private household consumption equal to or above the 25th percentile of TINCOME2 but less than the 50th percentile of TINCOME2
3 = Annual private household consumption equal to or above the 50th percentile of TINCOME2 but less than the 75th percentile of TINCOME2
4 = Annual private household consumption equal or above the 75th percentile of TINCOME2 but less than the 90th percentile of TINCOME2
5 = Annual private household consumption equal to or above the 90th percentile of TINCOME2 but less than the 95th percentile of TINCOME2
6 = Annual private household consumption equal to or above the 95th percentile of TINCOME2 but less than the 99th percentile of TINCOME2
7 = Annual private household consumption equal to or above the 99th percentile of TINCOME2

Chapter 9. QUALITY

9.1 Introduction

The quality of the statistics produced from a sample survey will depend upon:
- Careful and thorough planning before the implementation of the survey - the quality dimension should be an integrated part of the survey plan ,
- Carrying out the survey activities according to the decided plan and
- Adjusting the drawn plan if different conditions which were not considered during the planning stage are observed at the time of implementation.

Section 3.1 provides the general facts about the planning phase of NHIES. During the planning stage most technical and logistical aspects of the NHIES were discussed at length. For example, all concepts and definitions were formulated starting with existing international/UN definitions. Then they were adjusted wherever necessary, taking into consideration local conditions and experiences. The sampling procedure was adapted to the conditions for national household surveys in Namibia and to the user requests for regional and rural/urban disaggregation of the NHIES statistics. Several aspects were discussed in detail with the user community during the planning of the NHIES.

In this chapter special findings observed during the survey process, which affect the NHIES database as well as the quality of the compiled NHIES statistics, will be discussed.

9.2 Coverage

The question of coverage concerns to which extent the sampling frame(s) coincide with the NHIES target population and to which extent the units sampled have provided the requested information. The latter part of the coverage question is often referred to as the non-response problem.

Coverage can be discussed under the following items.
- Coverage of first stage units i.e. PSUs.
- Coverage of second stage units i.e. households.
- Coverage of the items of the questionnaires.

Coverage of the PSUs

The master sample frame coincides with the target area of the NHIES i.e. the whole of Namibia.
All the originally selected 192 PSU's were covered except one. The problematic PSU was observed in the urban area (Katima Mulilo town) of the Caprivi region. One selected PSU was found to be empty of households. Since this fact was discovered before the start of the field work it was decided to adjust the frame accordingly. In the adjustment of the frame the empty PSU was combined with an adjacent PSU and this "new" combined PSU was selected.

Coverage of the households

A. Boundary problems during the listing operation

One very important aspect in the listing operation was to identify the boundaries of the selected PSUs. The way this should be done was explained in the Interviewers' Instruction Manual (see section 0.5 annex 2). During this exercise it was found that especially in the northern regions (communal areas) the maps and the other available materials did not supply sufficient information to identify the boundaries of the PSUs. In these areas it was not easy to find different landmarks and physical references to identify a place. Even if they are available it is very easy for someone to make misjudgments. For this reason the supervisors had to follow the list of localities to identify the localities or part of localities inside the PSU and then list the households. The problem with this method is that the exact area of the PSU might not be covered. Hence the household listings in these areas may have some duplications and omissions.

A mapping project was undertaken to improve the maps as well as the boundary and other descriptions of the PSUs. This is a project which is aimed to be completed before the next Population and Housing Census. As a first step it was decided to carry out this activity in the PSUs of the master sample and then to extend it to other PSUs.

During the first phase of this activity 11 PSUs were mapped and the household listings of the NHIES were checked. In 3 of these PSUs the boundaries followed during the listing procedure very much deviated from what was indicated on the map. Another two of them had minor deviations and the rest followed the map. This project was continued and it was found in general that the number of PSUs with large deviations between the NHIES listing and the mapping exercise were rather few (see reference 13, 14 and 15 in section 0.5).

B. Undercoverage of the households in Walvis Bay

Part of the single quarters in the Kuisebmund area was not listed for security reasons (see section 4.6). Hence this part is not included in the survey population. Therefore the estimates in the NHIES Main Report will exclude the representation of this area.

According to section 4.6 it is possible to make the assumption that the part which was not listed is approximately of the same size as the part listed with respect to the number of households. Using this information it is possible to make adjustments to the estimated statistics for Walvis Bay.

C. Households not present.

The listing of the households within each PSU was carried out 2 to 3 weeks prior to the actual survey for that PSU. Hence the chances of not finding a selected household were reduced to a minimum. But still there were some selected households which could not be found at the listed location during the time of the survey.

This could happen due to the following:

- The household which was present at the time of listing had moved out of the PSU temporarily . For e.g. left for a vacation.
- The household had vacated the former residence and moved to some other place within the PSU or outside the PSU.

These specific cases were treated in the following ways:
If the household is temporarily away, then the interviewer should wait for them. If the household turns up at least during the last survey week, then the household should be included in the survey.
If they did not return before the survey period was over, then such a household was treated as a unit non-response household from the field.
If the household had moved within the PSU then the interviewer should follow the household to the new location and carry out the interview.
If the household had moved out of the PSU then such a household was substituted by a different household from the listing form. This new household was selected randomly from a subset of 10 households listed around the one to be substituted. *This is the only time that the substitution of a household was allowed. The decision to select a substituted household was only to be taken by the supervisor in collaboration with the head office.*
Out of the 4397 households in the basic NHIES household file (see chapter 8) , 72 or 1.6 % are substituted households.

D. Refusing households

Refusing households are households which refused to participate in the NHIES when they were approached by the CSO interviewer. These cases were treated as unit non-response from the field.

E. Partially participating households

This category covers the case where households participated in the survey but the questionnaires were only partially completed. This might depend on a number of reasons.

The two most common reasons are:

- The household refused to respond to certain sections of the questionnaires. For example concerning income transactions.
- After the first visit or some time during the early part of the survey month the household moved out of the PSU temporarily or permanently so that the interviewer could not continue with the interviewing. Again missing information will include income and expenditure transactions.

Such households were scrutinized during the data cleaning-phase II and some were excluded for lack of information (see section 7.7).

The table below shows the particulars about the NHIES unit non-response. The total household non-response in the NHIES is 7.5 percent.

Table of unit non-response

Reduction of the selected NHIES sample in the field work and in the data cleaning.	Namibia excluding Walvis Bay		Walvis Bay		Total	
	Number	%	Number	%	Number	%
1 Number of households selected	4608	100.0	144	100.0	4752	100.0
2 Number of households responding completely or partially in the field	4410	95.7	134	93.1	4544	95.6
3 Unit non-response from the field (1 - 2)	198	4.3	10	6.9	208	4.4
4 Number of households responding completely or partially in the field	4410	-	134	-	4544	-
5 Number of households in the final analysis	4266	-	131	-	4397	-
6 Number of households excluded during the data cleaning-phase II (4 - 5)	144	-	3	-	147	-
7 Number of households selected	4608	100.0	144	100.0	4752	100.0
8 Total unit non response (3 +6)	342	7.4	13	9.0	355	7.5
9 Number of households in the final analysis (7 - 9)	4266	92.6	131	91.0	4397	92.5

Coverage of the items in the questionnaires/item non-response

Exclusion of the whole set of questionnaires/households due to incompleteness was dealt with in the earlier paragraph.
The item non-response in the income sections of the questionnaires is higher than in the other sections. In general, the coverage in most other sections of the questionnaires is satisfactory although there were some non-responses.
A question of interest is to examine the undercoverage of the transactions in Form I Part E and F and in Form III and IV. The size of this undercoverage is not easy to find out from the questionnaires when they are back in the head office because there are no hard rules for guidance. Great care should be exercised while the field work is in progress to monitor the coverage of these transactions.
Once the estimated values for different transactions or groups of transactions are calculated from all the survey data, then it is possible to compare them with other existing sources. Such a comparison is made in section 9.5 below.

See also next section concerning the difficulties to cover all transactions.

9.3 Shortcomings during the data collection phase

Most of the supervisors and the interviewers of this survey were involved in the 1991 census of population as interviewers or processing staff. But they lacked the experience of working in a complex national survey like the NHIES. This was the first time such a survey was undertaken in Namibia. Hence a thorough training was necessary for them to understand the concepts, definitions and the methodology involved.

The details about the training programme can be found under section 6.1. In general it was felt that the classroom training of theory and practice was adequate. However, it is well known that the interviewers get their problems and experience only when they start the field work. This survey was no exception. Ideally after the first month's survey work the head office staff should visit them and go through all the problems and provide the necessary instructions and support. Then this process should be continued regularly during the year of field work.

In this survey it was not possible to follow this procedure to a full extent due to lack of staff. See section 6.2 for a review of field supervision by CSO head office staff.

Problems related to concepts and definitions

One area where a lot of difficulties were encountered by the field staff was in the labour force module of Form I Part B. Generally it took a fair amount of time for them to understand the different concepts involved and also the use of terms such as 'work' (see annex 3 of the Interviewer's Instruction Manual in annex 2 of this report, section 0.5). It was confusing to them to explain that some of the household duties were considered as 'work' while others were not. The concept 'available for work ' was a difficulty while 'looking for work' was better understood. It could be clearly seen that questions based on a lot of concepts were not well understood. Even after some experience there were interviewers without a clear understanding of such questions. For these reasons the responses to the question related to 'available for work ' may not be very reliable, especially in the early part of the survey period.

Lots of time should be devoted for such concepts during the training and their application in the fieldwork should be monitored from the beginning. Field staff should be trained to work according to the concepts and definitions of the survey without confusing them with their own notions.

Coverage of the transactions

Another important aspect is to know how the transactions are covered. This is not very straight forward. If some transactions are entered in the record book it is difficult to measure the amount of coverage. Here it is quite necessary that the interviewer checks the household very frequently for the entries in the record book and probes the household. Probing should be done for almost all the groups of transactions whether entered in the record book or not . Especially, probing should be conducted for transactions in kind. This is the only way to ensure that all the items are covered. It is the supervisor who should involve the interviewers in this process.

By scrutinizing and probing certain omissions could be detected.

One such example is as follows:

In a farm worker's household (commercial farm), no meat was consumed during a certain week. It was thought to be strange and was probed. It was found that the farm owner put the meat produced in the farm itself in a freezer and delivered it to the farm workers from time to time. Other food items were also given by the farmer but the way of distribution was different. The household recorded the other food items but not the meat. By probing this error was detected. It is only what the household understands that the interviewer communicates to the questionnaires. There is no other way to find out if the household understands except by scrutinizing and probing.

Another common occurrence in the upper income households was to attach a receipt from the cash register to the record book. These households buy most of their goods weekly or at other intervals of time and lots of items are covered. What they can provide to the interviewer is the receipt from the cash register which they received from the supermarket. They are not ready to break it down to the individual items and enter them in the record book. Nor can the interviewer do anything about it because the items in the receipt are coded under groups. For e.g. "veg" for "vegetables". There were several cases of this nature and all of them were included under the item 'combined groceries' in Form III.

Missing entries

At the beginning there were a fair number of missing entries in the filled in questionnaires. Field staff were instructed many times about this and the number went down. Missing values were imputed wherever possible by using information from the same household or questionnaires from other households in the same PSU. The entries where no imputation was possible were treated as item non-response and coded accordingly. Whenever the item non-response is possible to quantify and is of any significance it is presented explicitly in the tables of the NHIES Main Report.

9.4 Shortcomings during the data processing phase

Manual processing

Manual processing staff carried out the editing and coding of the questionnaires. This included checking the questionnaires for missing entries, wrong entries, inconsistent entries and coding of the relevant fields.
During the early part of the process this staff also had the same difficulties as the field staff to understand the concepts. But since they were in the head office it was possible for them to overcome this difficulty quickly by asking professional staff for clarifications.
The observations worthwhile noting re problems in the work of the manual processing staff are
- bypassing of errors
- wrong corrections
- deterioration of the quality of work as well as the speed towards the last days of employment.

With proper supervision and management these problems could have been minimised. But the professional supervision was not possible to do regularly for lack of senior staff.

Computer processing

This part of the processing was completed smoothly except for data entry errors. It looks as if some parts of the data entry work could have been done in a more careful way. Procedures should be developed to minimise these errors as far as possible.

9.5 Comparisons with other sources

Comparisons with the 1991 Census figures

The aim of this section is to make some possible comparisons between the related items of the 1991 Census and the NHIES to receive an indication of the quality of NHIES data.
The items that can be compared directly are the number of private households and the population in these private households. The figures are shown below.

Source	Persons in private households	Private households
1991 Census	1 318 935	254 389
NHIES excl Walvis Bay	1 367 788	240 611

When comparing these figures it is important to keep in mind that the Census was conducted in 1991 October/November and the NHIES in November 1993 to October 1994.
The first observation to make is the reduction of the number of households and the increase of the population in the NHIES compared to the Census. This means the average household size reported by the survey is higher than in the Census. A reasonable explanation can be found in the way the household members were listed in the NHIES. Each responding household participated for one month. Every week, if there were new persons present in the household (other than the ones listed earlier), who satisfied the conditions to be members of the household (i.e. slept in the household and had common catering arrangements with the household during the last 24 hours) then they were included as household members. Due to this, the number of household members will be higher in a household in the NHIES. On the other hand family members were taken away from a household in the NHIES only when they were absent for all the 4 weeks. Hence this could result in having more members in a NHIES household than in a Census household.
Strictly the population figure (Persons in private households) should be compared with the population projection for 1994. The following table shows this comparison:

Source	Total population	Household population
1991 Census	1 409 921	1 318 935
Population projection 1994 [1]	1 545 848	1 446 090
NHIES excl Walvis Bay	-	1 367 788

[1] Based on the Provisional Population Projections - July 1994.

The NHIES figure is less by about 78 000 persons from the projected household population. The two figures are not significantly different (see section 9.6 below) although the difference may be due to a slight undercoverage of the households considering the fact that the number of private households were also a bit low in the NHIES.

Comparisons with the National accounts

One objective of the NHIES is to provide data for improvements of the national accounts, notably for a benchmark estimates of final consumption expenditure by households (FCH).

In the national accounts, FCH include all expenditure, in cash and in kind, by residents on goods and services for the purpose of consumption minus sales of any existing such goods. At present, there is no independent estimate of FCH in Namibia's national accounts. Instead, FCH is calculated as a residual as displayed in the table below. The values refer to 1994, the year that is relevant for a comparison with the NHIES.

Household final consumption expenditure
according to the national accounts (N$ million)

A	Gross domestic product	10393.8
B	Imports	6158.3
C	Total resources	16552.1
D	Final consumption expenditure by government	3274.0
E	Final consumption exp. by non-profit institutions	372.0
F	Gross fixed capital formation	2243.8
G	Changes in inventories	269.6
H	Exports of goods and services	5651.3
I	Total uses exclusive of household consumption	11810.8
J	**Household final consumption expenditure (C–I)**	**4741.3**

The weakness of the national accounts method of estimating FCH is of course that errors and omission in all other items in the above table are reflected in FCH.

The total FCH according to the NHIES is N$ million 3 129 for the period November 1993 - October 1994. Thus, there is a substantial difference compared with the national accounts estimate.

It has not been possible to fully reconcile the NHIES with the national accounts. Based on the NHIES estimate, the following FCH can be derived. This estimate will still leave a discrepancy of about N$ million 500 compared to the national accounts estimate presented above.

A	Final consumption expenditure, NHIES	3129
B	Add 13 percent	407
C	Institutional households	34
D	Probable underestimation of certain items in the NHIES	485
E	Deduct vehicles for gross fixed capital formation	-70
F	Direct purchases abroad by Namibian residents	238
G	Final consumption expenditure after adjustments	4223

B   The upper limit of the 95 percent confidence interval is about 10.5 percent above the point estimate above. Another 2.5 percent can be added to take into account that the time period in the NHIES and the national accounts is slightly different. Prices increased between November-December 1993 and the same months in 1994.

C   Institutional households were not included in the NHIES. They comprise about 90 000 individuals. The consumption figure are based on average consumption by low income households in the NHIES.

D   This is based on a comparison with detailed import statistics. Items that seem to be underestimated in the NHIES are: alcoholic beverages, tobacco, clothing and footwear, furniture and major household appliances, medicine, leisure equipment and appliances, goods for personal care and miscellaneous personal effects.

E   Part of vehicles bought by households with a business are recorded as gross fixed capital formation in the national accounts.

F   Direct purchases abroad by Namibian residents are not included in the NHIES. The estimate, used in the national accounts, comes from the balance of payments (Bank of Namibia).

9.6 Confidence intervals

The statistics compiled from the NHIES can deviate from the target parameters (section 5.2) for basically two reasons:

a) the estimators used are biased i.e. on the average the statistics compiled from the estimator will deviate from the target parameter independently of the sample size

b) the random variation of the estimator i.e. the estimates will differ depending on which households have been randomly selected.

All sources of errors discussed in earlier sections of this chapter will influence to a larger or smaller extent both of these basic error components. However, the random variation of the estimator depends mainly on the sample design and the sample size. The random variation is quantified by a 95 % confidence interval as explained in section 5.5.

In the following table we present in the third column the 95 % confidence interval for selected NHIES statistics. The corresponding NHIES estimates are presented in the second column. In the fourth column the size of half the confidence interval is presented in percent of the estimate in column 2. The figures in the fourth columns we call "confidence coefficients".

Confidence intervals are presented for three groups of parameters - national, rural/urban and regional parameters. For the regional parameters, only those regions are presented which have the minimum and maximum values concerning confidence coefficients.

PARAMETER	ESTIMATE (E)	CONFIDENCE INTERVAL (CINT)	CONFIDENCE COEFFICIENT (100*0.5*CINT/E)
National parameters			
Number of private households in Namibia	244 827	230 043 - 259 611	6.0
Private household population in Namibia	1 389 017	1 288 701 - 1 489 333	7.2
Average household size in Namibia	5.7	5.5 - 5.9	3.5
Total annual private household consumption in Namibia, Mill N$	3 129	2 808 - 3 450	10.3
Average annual private household consumption in Namibia, N$	12 783	11 538 - 14 028	9.7
Rural parameters			
Number of private households in rural areas	161 962	147 905 - 176 019	8.7
Private household population in rural areas	987 691	891 233 - 1 084 149	9.8
Average household size in rural areas	6.1	5.8 - 6.4	4.9
Total annual private household consumption in rural areas, Mill N$	1 231	1 076 - 1 386	12.6
Average annual private household consumption in rural areas, N$	7 601	6 842 - 8 360	10.0
Urban parameters			
Number of private households in urban areas	82 864	76 809 - 88 919	7.3
Private household population in urban areas	401 325	366 565 - 436 085	8.7
Average household size in urban areas	4.8	4.5 - 5.1	6.3
Total annual private household consumption in urban areas, Mill N$	1 898	1 614 - 2 182	15.0
Average annual private household consumption in urban areas, N$	22 912	19 679 - 26 145	14.1
Regional parameters			
Number of private households in Karas	11 545	10 934 - 12 156	5.3
Number of private households in Oshana	24 198	16 131 - 32 265	33.3
PARAMETER	ESTIMATE	CONFIDENCE	CONFIDENCE

	(E)	INTERVAL (CINT)	COEFFICIENT (100*0.5*CINT/E)
Average household size in Khomas	4.7	4.4 - 5.0	6.4
Average household size in Omaheke	5.1	3.9 - 6.3	23.5
Total annual private household consumption in Khomas, Mill N$	1 164	922 - 1406	20.8
Total annual private household consumption in Karas, Mill N$	181	89 - 273	50.8
Annual private consumption per capita in Caprivi, N$	1 012	868 - 1 156	14
Annual private consumption per capita in Karas, N$	3354	1 185 - 5 523	65

From the table we can see that the confidence coefficient for national estimates is 10 % or less, for rural and urban estimates 15 % or less while the confidence coefficient of regional estimates varies a lot. For the regional estimates the confidence coefficients in the table vary from 5.3 to 65 %.

Chapter 10. SPECIAL OBSERVATIONS

10.1 Rural/urban classification

In the master sample frame, PSUs were classified in urban and rural strata using the classification adopted during the time of the 1991 Census. This was easily done since the PSUs were formed using the enumeration areas of the Census. There were 68 urban and small urban areas at the time of the Census. In 1992, the Ministry of Local Government and Housing proclaimed 15 municipalities and 12 towns. These 27 areas are considered as urban when compiling the NHIES statistics. The remaining 41 out of the 68 "old" urban areas are now considered as villages and fall under the rural areas as does the rest of the country.
The disaggregation of statistics from NHIES for urban and rural areas has been done according to this new classification decided by the Government.
In the data files the former classification is under the variable UR while the new classification is under NUR (See chapter 8 and annex 3 of this report, section 0.5).
The following table shows the classification of the PSUs in the master sample. 6 of the former urban PSUs are considered as rural according to the new classification.

PSUs in the master sample by rural/urban classification.
Namibia excluding Walvis Bay.

Old classification	New classification		Total
	Urban	Rural	
Urban	75	6	81
Rural	0	111	111
Total	75	117	192

No PSUs were defined for Walvis Bay and the sampling procedure in Walvis Bay is explained in section 4.6. Altogether 144 households were selected and 131 responded. Out of these, the Topnaar households were treated as rural and the rest as urban.

10.2 The effect of weighing

In section 5.3 - 4 of this report it has been explained why all estimators to be used when compiling the NHIES tables should use weights (raising factors). The reason is that the NHIES sample is far from "self-weighing". The weight used should be the inverse of the inclusion probability for each household. In the NHIES sample, this weight varies between 6.08 and 245.3. This means - expressed in a popular way - that the minimum number of households which a household in the NHIES sample represents is about 6 while the maximum number is about 245. The NHIES sample would be self-weighing only if each household in the sample represented the same number of households of the Namibian population.

In this section we will compare some estimates compiled with and without weights. The purpose is to illustrate the effect of weighing in the estimation procedure.

Comparison of weighed and unweighed estimates of the population of Namibia distributed by age and sex.

Age Group	Weighed estimates						Unweighed estimates					
	Female		Male		Total		Female		Male		Total	
	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
0-4	109 312	15	111 738	17	221 050	16	1 949	15	1 990	17	3 939	16
5-9	98 302	13	98 358	15	196 660	14	1 754	13	1 740	15	3 494	14
10-14	93 105	13	82 538	13	175 643	13	1 614	12	1 480	12	3 094	12
15-19	84 105	11	78 781	12	162 886	12	1 472	11	1 424	12	2 896	12
20-24	72 099	10	59 795	9	131 893	9	1 308	10	1 078	9	2 386	10
25-29	56 488	8	45 913	7	102 401	7	1 031	8	847	7	1 878	8
30-34	48 242	7	36 264	6	84 506	6	905	7	680	6	1 585	6
35-39	35 449	5	28 152	4	63 601	5	653	5	529	4	1 182	5
40-44	27 944	4	25 466	4	53 410	4	516	4	460	4	976	4
45-49	19 987	3	19 388	3	39 374	3	375	3	351	3	726	3
50-54	18 851	3	16 380	2	35 231	3	336	3	302	3	638	3
55-59	12 623	2	11 751	2	24 375	2	230	2	230	2	460	2
60-64	15 771	2	11 552	2	27 323	2	277	2	209	2	486	2
65+	36 299	5	27 874	4	64 173	5	656	5	484	4	1 140	5
Non resp:	2 986	0	3 505	1	6 491	0	48	0	56	0	104	0
Namibia	731 562	100	657 454	100	1389 017	100	13 124	100	11 860	100	24 984	100

It is very evident that the *totals* will be quite different if weighing is used or not. If no weights are used, the totals will only picture the size of the sample in the different age groups and not the size of the Namibian population.

However, there are only occasional differences among the weighed and unweighed estimates of the *percentages* shown in this table in spite of the big differences in weights between the households of the NHIES sample. The reason should be that there is a low correlation between the studied age-sex distribution on one hand and the distribution of the weights on the other hand.
Comparison of weighed and unweighed estimates of the total and average

annual private household consumption by region and rural/urban areas.

Region	Weighed estimates						Unweighed estimates					
	No of house-holds	Averg house-hold size	No of persons	Total consumption million N$		Average hhold consump	No of house-holds	Averg house-hold size	No of persons	Total consump. million N$		Average hhold consump
	%		%	Sum	%	N$	%		%	Sum	%	N$
Caprivi	6.9	5.4	6.6	92	3.0	5 479	5.3	5.4	5.0	1	2.3	5 844
Erongo	6.8	4.5	5.4	250	8.0	15 087	8.4	4.5	6.7	6	10.3	16 398
Hardap	5.1	4.3	3.9	168	5.4	13 484	5.0	4.5	4.0	3	5.3	14 072
Karas	4.7	4.7	3.9	181	5.8	15 722	5.2	4.8	4.3	3	5.8	14 949
Khomas	13.9	4.7	11.6	1 164	37.2	34 152	13.5	4.9	11.6	20	35.2	34 932
Kunene	4.2	5.7	4.2	81	2.6	7 882	5.2	5.7	5.2	1	3.4	8 636
Ohangwena	10.4	7.5	13.7	156	5.0	6 111	9.5	7.4	12.4	2	4.4	6 239
Okavango	8.3	6.1	9.0	153	4.9	7 537	7.1	6.2	7.7	2	4.2	8 020
Omaheke	3.7	5.1	3.4	118	3.8	12 936	4.8	5.3	4.5	2	4.8	13 163
Omusati	8.9	7.0	11.0	169	5.4	7 746	10.2	7.1	12.8	3	5.8	7 618
Oshana	9.9	6.7	11.6	216	6.9	8 928	9.5	6.4	10.7	3	6.5	9 179
Oshikoto	7.7	6.2	8.4	139	4.4	7 407	7.8	6.1	8.4	2	4.8	8 186
Otjozondjupa	9.3	4.4	7.2	236	7.6	10 374	8.4	4.6	6.8	4	7.2	11 413
NAMIBIA	100.0	5.7	100.0	3 129	100.0	12 783	100.0	5.7	100.0	58	100	13 382
Rural	66.2	6.1	71.1	1 231	39.3	7 601	61.1	6.2	66.4	21	35.7	7 822
Urban	33.8	4.8	28.9	1 898	60.7	22 912	38.9	4.9	33.6	37	64.3	22 102

All the columns in this table are computed as ratios except the two columns about the total consumption (sum). There are great differences between these two columns because the first one relates to the population while the second one relates to the NHIES sample.

In the other columns dealing with percentages and averages, there are certain evident differences. The percentage values for number of households, number of persons, distribution of total consumption (%) and the average household consumption differ clearly between the weighed and unweighed estimates. The differences in the estimates of average household size are less clear.

For example, the average household consumption of Namibia according to the unweighed estimate is N$ 13 382, while the weighed estimate is N$ 12 783. This may be explained by the fact that there is a certain amount of oversampling in urban areas in the NHIES sample (see section 4.3) and the unweighed estimate of the average household consumption reflects this since the average household consumption in urban areas is higher. This effect of the oversampling in urban areas is eliminated by means of the weighing procedure and the weighed estimate becomes lower.

Chapter 11. RESOURCES EMPLOYED

11.1   Introduction

Like the 1991 Population and Housing Census, the Namibia Household Income and Expenditure Survey, was a huge operation and it will not be easy to specify all resources used. This chapter will provide estimated figures for the major costs of the survey.

Firstly, the following costs are estimated:

- salaries, subsistance and travel allowances *for temporary staff*,
- transport,
-materials and supplies,
-other services and expenses

Secondly, costs in the form of salaries and subsistence and travel allowances *for permanent personnel* engaged in the survey operations are estimated.

The cost estimates are based on rough calculations because there is no administrative system available at CSO for a systematic follow-up of different cost components of a survey.

The report does not include expenses incurred by non-governmental agencies e.g SIDA and UNDP for the remuneration of consultants involved in the NHIES.

The costs are disaggregated in costs for the pilot survey and for the main survey. And the costs for the main survey are disaggregated in costs belonging to the financial year 1993/94 and the financial year 1994/95 respectively. There are also separate calculations of leave and bonus payments to temporary staff.

11.2 The pilot survey

Remuneration	**TOTAL (N$)**
12 Enumerators	
10 weeks = 5 x 10 =	
50days	
N$ 8.02/hour x 8 hours/day = N$ 64.16 /day	
12 x 50 x 64.16	38 496
Total Remuneration	**38 496**

	TOTAL (N$)
Subsistennce and Travel	
6 enumerators	
42 days	
N$ 70.00/day	
6 x 42 x 70	17 640
6 enumerators	
21 days	
N$ 70.00/day	
6 x 21 x 70	8 820
Total Subsistence and Travel Allowance	**26 460**
Transport	
(Monthly charge N$ 918.00)	
6 (4x4) vehicles	
10 weeks = 2.5 months	
6 x 918 x 2.5	13 770
(Kilometer charges @ N$ 1.25/km)	
6 (4x4) vehicles	
10 weeks = 2.5 months	
1 800 km/month for each vehicle	
6 x 1800 x 2.5 x 1.25	33 750
Total for Transport	**47 520**
TOTAL FOR THE PILOT SURVEY	**112 476**

11.3 The main survey -   TOTAL (N$)
September 1993 - March 1994

Remuneration
55 enumerators
Oct. 1993 - Mar. 1994 22.5*6= 135 days
8 hours/day @ N$ 8.02/hour = N$ 64.16
55 x 135 x 64.16   476 388

9 supervisors
Sept.1993- Mar.1994   12+135 days=147 days
8 hours/day @ N$ 9.52/hour = N$ 76.16
9 x 147 x 76.16   100 760

5 editors/coders
Dec.1993 - Mar.1994 22.5*4=90 days
8 hours/day @ N$ 8.02/hour = N$ 64.16
5 x 90 x 64.16   28 872

5 data-entry clerks
Dec.1993 - Mar.1994 22.5*4=90 days
8 hours/day @ N$ 8.02/hour = N$ 64.16
5 x 90 x 64.16   28 872
Total Remuneration   **634 892**

Travel and Subsistence
30 enumerators
182 days
N$ 70.00/day
30 x 182 x 70   382 200
Total Travel and Subsistence   **382 200**

Transport	**TOTAL (N$)**
Purchase of 20 bicycles	Not included
Purchase of 5 Ford Bantams	Not included
Monthly charges: 8 Sedan cars @ N$504.00/month each 6 months 8 x 504 x 6	24 192
11 4x4s @ N$ 918.00/month each 6 months 11 x 918 x 6	60 588
2 Bakkies @ N$ 523.00/month 6 months 2 x 523 x 6	6 276
Kilometre charges: 8 Sedan cars @ N$ 0.53/km 2 000 km/month 6 month 8 x 0.53 x 2 000 x 6	50 880
11 4x4s @ N$ 1.25/km 2 000 km/month 6 months 11 x 2 000 x 1.25 x 6	165 000
2 Bakkies @ N$ 0.71/km 2000 km/month 6 months 2 x 0.71 x 2000 x 6	17 040
Total for Transport	**323 976**
Materials and Supplies (equipment to enumerators etc.)	**4 000**
Other Services and Expenses (printing etc)	**20 000**
TOTAL MAIN SURVEY - FINANCIAL YEAR 1993/94	**1 365 068**

11.4 Main Survey-
April 1994 - March 1995

Remuneration
55 enumerators
Apr.-Nov.1994 22.5*7+12=169.5 days
8 hours/day @ N$ 8.02/hour = N$ 64.16
55 x 169.5 x 64.16   598 132

9 supervisors
Apr.-Nov.1994 22.5*7+12=169.5 days
8 hours/day @ N$ 9.52/hour = N$ 76.16
9 x 169.5 x 76.16   116 182

5 editors/coders
Apr.1994 -Mar. 1995 22.5*12=270 days
8 hours/day @ N$ 8.02/hour = N$ 64.16
5 x 270 x 64.16   86 616

5 data-entry clerks
Apr.1994 -Mar. 1995 22.5*12=270 days
8 hours/day @ N$ 8.02/hour = N$ 64.16
5 x 270 x 64.16   86 616
Total Remuneration   **887 546**

Travel and Subsistence
30 enumerators
215 days
N$ 70.00/day
30 x 215 x 70   451 500
Total Travel and Subsistence   **451 500**

Transport

Monthly charges:
8 Sedan cars @ N$504.00/month each
8 months
8 x 504 x 8   32 256

11 4x4s @ N$ 918.00/month each
8 months
11 x 918 x 8   80 784

2 Bakkies @ N$ 523.00/month
8 months
2 x 523 x 8   8 368

Kilometre charges:
8 Sedan cars @ N$ 0.53/km
2 000 km/month
8 month
8 x 0.53 x 2 000 x 8   67 840

11 4x4s @ N$ 1.25/km
2 000 km/month
8 months
11 x 2 000 x 1.25 x 8   220 000

2 Bakkies @ N$ 0.71/km
2 000 km/month
8 months
2 x 0.71 x 2 000 x 8   22 720
Total for Transport   **431 968**

Materials and Supplies   **4 000**

Other Services and Expenses   **20 000**

TOTAL MAIN SURVEY-
FINANCIAL YEAR 1994/95   **1 795 014**

11.5 Leave entitlement and bonus payment for temporary staff

Leave payment
Pilot Survey
12 x 2.5 x 2.5 x 64.16   4 812

Main survey- Oct. 1993 - March 1994
55 x 6 x 2.5 x 64.16   52 932
9 x 6 x 2.5 x 76.16   10 282
10 x 4 x 2.5x 64.16   6 416
Main Survey- April 1994 - Mar. 1995
55 x 8 x 2.5 x 64.16   70 576
9 x 8 x 2.5 x 76.16   13 709
10 x 12 x 2.5 x 64.16   19 248

Total amount for leave   **177 975**

Bonus Payment
36 enumerators
13 months @ N$ 80.20
36 x 13 x 80.20   37 534

2 enumerators
19 months @ N$ 80.20
2 x 19 x 80.20   3 048

7 supervisors
19 months @N$ 95.20
7 x 19 x 95.20   12 662

2 supervisors
12 months @ N$ 95.20
2 x 12 x 95.20   2 285

Total amount for bonus   **55 529**
TOTAL   **233 504**

11.6 Total costs for the NHIES

Costs for temporary staff, transport, materials and supplies, other services and expenses

Pilot Survey	**112 476**
Main Survey- Sept.93 - March 94	**1 365 068**
Main Survey - April 94 - Mar.95	**1 795 014**
Leave and bonus payment	**233 504**
TOTAL	**3 506 062**

If the salaries for the permanent staff, involved in the survey, are taken into consideration, the following costs should be added to the above total for the survey:

Officer's Position	Subsistence allowance	Salary & Bonus[**]	Total
2 Statisticians	N$ 15 200[*]	N$ 267 014	N$ 282 214
1 Statistician Assist.	N$ 7 600[*]	N$ 117 908	N$ 125 508
3 Senior Stats. Clerks	N$ 45 600[*]	N$ 99 099	N$ 144 699
1 Stats.Clerk	N$ 15 200[*]	N$ 33 033	N$ 48 233
Total	**N$ 83 600**	**N$ 517 054**	**N$ 600 654**

[*] The rate per day is N$ 190. The number of days was roughly 40 per person for the statisticians and statistician assistant, 80 for the senior statistical clerks and 20 for the statistical clerk.
[**] 1.5 years per staff member

TOTAL COSTS FOR THE NHIES	**N$ 4 106 716**