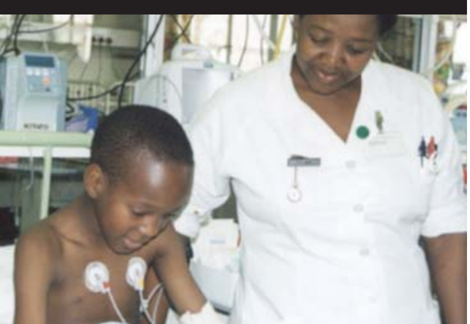
HEALTH
KwaZulu-Natal

COOPERAZIONE
ITALIANA

Evaluation of Antiretroviral Therapy Against HIV / Aids in KwaZulu-Natal South Africa



Dr Venanzio Vella, Italian Cooperation,
Dr Thiloshini Govender and Mr Dlamini Scelo DOH,
Dr Myra Taylor. Prof Indres Moodley, Ms Verona David and
Prof Champaklal C Jinabhai, University of KZN

Evaluation of the antiretroviral therapy in KwaZulu-Natal, South Africa

Dr Venanzio Vella, Italian Cooperation

Dr Thiloshini Govender and Mr Scelo Dlamini DOH/KZN

*Dr Myra Taylor, Prof Indres Moodley, Ms Verona David School of Family & Public Health and Prof Champaklal C. Jinabhai. School of Family & Public Health, Department of Community Health,
University of KZN*

October 2008

Front Cover provided by
Mrs Kathryn Potgieter and Ms Thobeka Shandu, KZN DOH

Editorial

HIV and AIDS constitutes one of the major diseases that ravages our societies since the first known cases in the early 1980s. Sub-Saharan Africa bears the brunt of this pandemic with scientists still in search of a cure. Current measures are still based on prevention particularly the ABC strategy – abstinence, being faithful or condomise. Current attempts have been made to reduce morbidity related to the disease thus prolonging people's lives and providing people living with HIV and AIDS with more productive years.

In April 2004 the Department of Health in KwaZulu-Natal rolled out the Antiretroviral (ARV) Programme (now known as the Comprehensive Care, Management and Treatment of HIV and AIDS – CCMT – plan). Eight hospitals were accredited to provide treatment and by the end of the year more hospitals had been accredited to provide the service. Currently the Province has more than 80 Public Sector Service Points providing the service to more than 140 000 people, with more Service Points planned for the future. These Service Points include hospitals, community health centres, non-governmental organizations and prisons. In order to achieve equitable access to treatment, the Department of Health has ensured that at least each sub-district has a Service Point that caters for its community.

As the number of patients that are part of the programme increases, a need exists to reflect back and draw lessons from experiences since 2004. These lessons will help guide us into the future and to indicate whether we have been able to achieve what we set out to do so that we could adjust and correct the gaps identified.

The purpose of this report is to provide information on lessons learnt from 32 Service Points that started the highly active antiretroviral therapy in KwaZulu-Natal. It is hoped that the results will inform programme and patient management in the Province. In addition, there is hope that the report will guide the programme into achieving the National Strategic Plan of 2007 – 2011.

This report, therefore, needs to be read with an intention of drawing practical solutions on the ever increasing challenges in the implementation of the CCMT plan. It should be used as a discussion document at all levels and should not be let to gather dust as it adds value to the programme in terms of its direction. The Department acknowledges the valued support and contribution from the Italian Cooperation in conducting this evaluation.

Dr. Y. Mbele
Acting Head of Department
KwaZulu-Natal Department of Health

Table of content

Tables

Figures

Acronyms

ARV	Antiretrovirals
ART	Antiretroviral therapy
CCMT	Comprehensive HIV & AIDS care management treatment
CD4	CD4 T-lymphocytes
CHWs	Community Health Workers
CI	Confidence Intervals
DOH	Department of Health
EQ-5	The Euro Quol is a measure of quality of life
FT	Full-time
GPS	Global Position System
HIV	Human Immunodeficiency Virus
HRQL	Health Related Quality of Life
KZN	KwaZulu-Natal
LY	Life Years
MD	Medical Doctor
M&E	Monitoring & Evaluation
MIS	Management Information System
NSP	National Strategic Plan
PSA	Probabilistic sensitivity analysis
PT	Part-time
PYO	Person-Years of Observation
QALY	Quality Adjusted Life Years
SPN	Senior professional nurse
VCT	Voluntary Counselling & Testing
UNAIDS	United Nations Programme Against AIDS
WHO	World Health Organization

Acknowledgement

The evaluation was a concerted effort between the Department of Health (DOH) of KwaZulu-Natal (KZN), the Italian Cooperation and the University of KZN.

The evaluation was supported financially and technically by the Italian Cooperation.

The data collection involved several survey teams. A first team interviewed the managers, the doctors and the nurses, a second team did the task analysis and the exit interview, a third team conducted an expenditure review in each delivery site for FY05/06, a fourth team extracted a representative sample of the records from the archives and entered them into an ACCESS database, a fifth team re-visited the delivery sites and the laboratory of Albert Luthuli Hospital in 2007 to fill the missing information.

The managers, the doctors and the nurses at the delivery sites were critical to provide the teams with the information on the management aspects of the programme and to access the medical records.

The staff at the provincial laboratory of Albert Luthuli Hospital provided access to the database of the laboratory results of the records which had missing information on viral load and CD4.

The patients interviewed and the patient records provided critical information to improve the ART programme which if implemented should improve ART retention.

The authors are grateful for the statistical advice provided by Prof Maurizio Vichi from the Faculty of Statistics of the University of Rome “La Sapienza” who was contracted by the Istituto Superiore di Sanità (Italian National Institute of Health).

Executive Summary

Introduction

The treatment programme to deliver antiretroviral therapy (ART) against HIV in the public services of KwaZulu-Natal (KZN) started in March 2004. Potential ART sites were assessed against minimum standards of infrastructure, data processing capability and personnel, and those ones meeting the criteria were accredited and began treating patients. A strengthening plan was implemented to ensure that the sites falling below minimum standards were following the guidelines. The short time frame available for planning caused some critical areas, such as information systems being overlooked. It is against this backdrop that it was decided to evaluate 32 sites which started rolling out ART in 2004.

The report is structured into two parts. **Part 1** gives an overview of what has happened in the 32 delivery sites between March 2004 and September 2006 while **Part 2** provides estimates of cost-effectiveness of several delivery options. Seventy six percent of the patients were still on ART after two years, but the probability of remaining on ART and its cost-effectiveness varied according to different staff and workload profiles. These findings provide practical guidelines on how to improve ART effectiveness by taking into account the absorption capacity of the treatment sites.

Objectives

To evaluate the delivery of ART in KZN. This included the description of the patients and the delivery sites, the identification of implementation problems and their potential solution, and the estimation of the cost-effectiveness of available delivery options.

Methodology

The evaluation was a concerted effort between the KZN Department of Health (DOH) of KwaZulu-Natal (KZN), the Italian Cooperation and the University of KZN. The evaluation, which was financially and technically supported by the Italian Cooperation, was conducted between May and September 2006. It consisted of a cross-sectional survey and a retrospective analysis of a representative sample of records. The cross-sectional survey collected information on the characteristics of 32 delivery sites and included interviews with managers, staff and patients, and a task analysis. The retrospective analysis was carried out on a representative sample of 2835 records selected from the archives of the 32 delivery sites.

The analysis consists of two parts. SPSS version 15 was used to provide the frequency distribution, cross-tabulations, Cox regression and Kaplan Meier curves described in **Part 1**. These statistical methods were used to identify the probability of remaining on ART if the patients attended delivery sites which were characterized by different profiles of personnel and annual intakes of new patients (delivery options).

Part 2 of the evaluation provides the costs per delivery site and per person year of observation (PYO) for fiscal year 2005/06 (FY05/06) and the cost-effectiveness of the different delivery options identified in Part 1. The registers of the delivery sites provided the basis for estimating the costs due to personnel, drugs, laboratory tests, supplies, equipment and utilities. The total costs were divided by the estimated number of PYOs for FY05/06 to estimate the cost per PYO. The cost per PYO was also estimated for the sites which were identified in Part 1 as having different delivery options. A second order Monte Carlo simulation was carried out to take into account the effect of uncertainty on the cost-effectiveness of these delivery options. TreeAge Pro software was used for the Monte Carlo simulation.

Results

One major management problem was the lack of a defaulters tracing system. Because the information was recorded on paper forms and there was no electronic patient register, the delivery sites did not update the number of patients who defaulted. The defaulters were usually contacted over the phone and very few were visited at home. Only one quarter of the delivery sites used CHWs to trace defaulters, making the system too static to maintain an effective coverage.

Patient perception

There was a high level of satisfaction among the patients who were interviewed. About 90% of the patients who were interviewed were satisfied with the time dedicated by the staff. The patients did not have problems in understanding what was said during the consultation and they found that they had enough privacy. More than 60% of those who were interviewed rated as very good or excellent the availability of drugs, the attitude and the availability of the staff, and the cleanliness of the delivery site.

Effectiveness

Effectiveness was measured as the proportion of patients remaining on treatment. According to the retrospective analysis of the representative sample of 2835 records, the patients still of ART declined to 85% by the end of the first 6 months. The proportion of patients still on ART was reduced to 80% and 76% respectively by the end of the first and second year of treatment.

A Cox regression model singled out the most significant risk factors for treatment discontinuation. Beginning treatment with CD4<100 cells/µl, not having a telephone contact number and being male were the most significant individual risk factors for discontinuation. The delivery sites which were significantly more at risk for loosing patients were located in unsafe areas[1], had one part time (PT) Senior Professional Nurse (SPN) and one PT medical doctor (MD), and had an intake of more than 200 new patients per MD per year. The patients who attended delivery sites with the following staff and workload profiles had very different probabilities of remaining on treatment by the end of the second year:

(a) Delivery sites with part time (PT) senior professional nurse (SPN) and PT doctor (MD) and with less than 200 new patients per doctor per year had a retention rate of 85% over a period of two years;
(b) The above mentioned delivery sites reduced their retention rate to 50% when the workload increased to 200 or more new patients per doctor per year;
(c) Delivery sites with FT SPN and FT MD and with less than 200 new patients per doctor per year had a retention rate of 88% over a period of two years; and
(d) The above mentioned delivery sites reduced their retention rates to 70% when the workload increased to 200 or more new patients per doctor per year.

The above findings provide clear guidelines on the absorption capacity of the delivery sites. As far as the annual intake of new patients per MD is less than 200, there is no significant difference in retention rates between delivery sites that are staffed with PT senior staff and delivery sites that are staffed with FT senior staff. The added value of having FT senior staff is when the workload reaches 200 or more new patients per MD per year. Such level of workload is expected to impact negatively on the retention of patients compared with the lighter workload of less than 200 new patients per MD per year. However the presence of FT senior staff will at least significantly attenuate the negative impact compared with the presence of PT senior staff.

Costs

The economic evaluation estimated the costs for FY05/06. The average annual cost per site was R2.8 million with a 95% Confidence Interval of R2.4 - R3.2 million. Forty-seven per cent of the total costs were spent on ART and non ART drugs, and 28% was spent on the personnel. The average cost per person year observation[2] (PYO), was R6846 (+/-R96) with a range of a minimum R6070 and a maximum of R7530 per PYO. The annual cost per PYO was then estimated for the four staffing and workload profiles listed in the previous section under (a), (b), (c) and (d).

Cost-effectiveness

A Markov model second order Monte Carlo simulation was carried out to assess the influence of uncertainty of all parameters. The objective was to predict the costs to retain 1000 patients for a period of 10 years if the previously mentioned staff and workload profiles (a) through (d) were implemented. After 1,000 simulations, option (c), characterized by delivery sites with FT SPN and FT MD with less than 200 new patients per MD per year, was the most cost-effective. With this option, 728 of the initial 1000 patients would still be

---

1 The delivery site was considered located in an unsafe area if the manager replied affirmatively to the question "Are there problems with security in the area, which may sometimes prevent patients from coming to the health facility ?"

2 The cost born by the average delivery site to follow up a patient on ART for a whole year does not include the costs of other health services (e.g. hospitalization).

on ART after 10 years, at a cumulative cost of R57.9 million[3] at 2006 value. The incremental cost-effectiveness ratio for the most cost-effective option was slightly less than R85,000, which is to say that it would cost about R85,000 over a period of 10 years to retain each extra patient.

Discussion

The evaluation has provided a comprehensive picture of the delivery of ART in KZN. This included the characteristics of the delivery sites and the patients which significantly influenced ART retention, and the estimation of the cost-effectiveness of four staff and workload profiles. The measure of effectiveness was the proportion of patients who were still on ART after two years. Most patients died or discontinued ART during the first six months and by the end of the first two years only 76% were still on ART.

The analysis identified four typologies of delivery sites which, independently from the other risk factors, significantly influenced ART retention. About half of the patients began ART with CD4<100 cells/μl which was one major risk for discontinuing treatment. The patients were also at higher risk of discontinuing ART if they were males, did not have a telephone contact number and they attended a delivery site which was located in an unsafe area. Notwithstanding these individual risk factors, the probabilities of remaining on treatment increased if the delivery site had FT senior staff and a workload of less than 200 new patients per MD per year. The simulation confirmed the cost-effectiveness of maintaining workload under the 200 new patients per MD per year.

Conclusions

The analysis has identified specific staff and workload profiles that were critical to maintain an efficient delivery system. Not exceeding 200 new patients per MD per year would be the safest option because it coincides with optimal absorption capacity. If the annual intake exceeds 200 new patients per doctor per year, having FT senior staff would make a significant difference. In any case, if a delivery site exceeds 200 new patients per MD per year its manager should prepare contingency plans to scale up tracing strategies.

Full coverage is not sustainable if there is a mismatch between supply and demand. The staff and workload profiles identified in this analysis can be used to improve the sustainability of ART expansion. Planners can use the results of this evaluation to rank the delivery sites according to their staff and workload profile to estimate the risk for default and thus quantify the gaps in human and other resources required to sustain coveragetargets.

3 The costs are at the value of 2006, without adjustment for future inflation and without discounting for future values

Recommendations

The results of the evaluation suggest that ART delivery can improve through the following actions:

(a) ***Adopt an electronic register, which could be programmed to routinely produce the names of the patients who do not come as scheduled.*** The expansion to other delivery sites would require intensive training and supervision, frequent trouble shooting and substantial workload. To ensure smooth implementation of the electronic register, a mobile team should visit each unit on a rotating basis. This will ensure that the system works, staff is well trained and supervised and information is updated and transmitted to the DOH.

(b) ***Monitor the patients who are more at risk for default especially in the first semester of ART***. The electronic register should provide routine management reports to staff, including a report ranking the patients according to the presence of the risk factors for discontinuation identified by the evaluation. The patients with a high risk profile should be given extra counselling and should be put under surveillance to be traced as soon as they default.

(c) ***Increase the tracing effectiveness***. At registration, the staff should ensure the recording of the telephone contact numbers and the address. The tracing team should take the Global Position System (GPS) coordinates of the households of the defaulters;

(d) ***Conduct an audit of the ART delivery sites***. The audit should provide for each delivery site an update on the number of patients on ART, the staff and the workload profile, and the prevalence of the risk factors identified in this evaluation;

(e) ***Rank the delivery sites*** according to the staff and workload profiles identified by the audit;

(f) ***Estimate the number of staff by type*** to be recruited to reach the best delivery option which has been identified by the evaluation;

(g) ***Set realistic coverage targets*** according to the staff and workload profiles that can be afforded;

(h) ***Update the status of the programme by analysing the data collected by the itinerant team***. At regular intervals, the previously mentioned itinerant team will forward the updated files to the Monitoring & Evaluation officer of the ART programme for the analysis. The objective is to monitor and verify that the management strategies are indeed implemented and are improving the ART retention;

(i) ***Produce a plan of action to integrate the ART and TB programmes***. Although there is a high co-infection rate between TB and HIV, the ART and TB programmes run in parallel. The integration should take into account the staff and workload profiles matching the absorption capacity according to the methods used by the evaluation; and

(j) ***Assess the feasibility of further integrating ART within Primary Health Care (PHC) services***. It is only after solving the problems of integrating the ART and the TB programmes that further integration within the mainstream PHC services could be planned, implemented, monitored and evaluated

Background

The human immunodeficiency virus (HIV) destroys the immune system causing the Acquired Immune Deficiency Syndrome (AIDS). The main transmission routes of HIV include unprotected sex, exchange of contaminated blood products and sharing of injection equipment. Vertical transmission from mother to child occurs during pregnancy, delivery and breastfeeding. HIV infects and destroys the CD4 T-lymphocytes (CD4), causing the insurgence of opportunistic infections leading to death.

According to UNAIDS, Sub-Saharan Africa is the region which is the most affected by HIV/AIDS and Southern Africa bears the brunt of the pandemic. In 2007, UNAIDS estimated that 33.2 million people were living with HIV/AIDS and 2.1 million people were expected to have died of HIV/AIDS. Sixty eight percent of the global HIV burden is estimated to be in Sub-Saharan Africa, with Southern Africa accounting for about one third of all people living with HIV/AIDS worldwide. Within the region, South Africa contributes a sizeable HIV burden with a national HIV seroprevalence of 29% among women attending antenatal care sites in 2006 (DOH, 2007). The lowest and highest HIV seroprevalence rates were found respectively in Western Cape (15%) and KwaZulu-Natal (39%).

The antiretroviral therapy (ART) reduces AIDS mortality and improves quality of life. After Zidovudine was licensed in 1987, several other antiretroviral drugs were discovered and they are presently used in combination in the so called highly active antiretroviral therapy. These drugs suppress viral replication (Wood *et al.* 2000), restoring immune function and reducing opportunistic infections (Ledergerber *et al.* 1999) which in turn improve health conditions. The impact of ART on morbidity and mortality has been extensively documented (Palella *et al.* 1998) (Hogg *et al.* 1998) (Moore *et al.* 1999) (Hogg *et al.* 2000) (Jordan *et al.* 2002) (Egger *et al.* 2002). However ART does not eliminate HIV, which starts replicating again if treatment is interrupted. As in any chronic condition, patients tend to stop treatment as they improve, which besides having consequences for the patient, increases the probability of developing HIV resistance to ART. It is therefore critical that any ART programme is evaluated before problems become ingrained.

DOH ART programme

The 2000 – 2005 National Strategic Plan (NSP) included key interventions to bring the country close to the realization of breaking the AIDS chain. In 2003, the South African Cabinet approved the National Operational Plan for Comprehensive HIV and AIDS Care, Management and Treatment (CCMT). The main purpose of the plan was to provide interventions to reduce the morbidity and mortality due to HIV/AIDS and to bring all partners on board in the fight against HIV/AIDS. The accreditation was coordinated by the National Department of Health through the implementation of the Operational Plan of 2003 and the National ART Guidelines of 2004. Potential ART sites were assessed against minimum standards of infrastructure, data processing capability and personnel. A strengthening plan was implemented to ensure that the sites falling below minimum standards were following the guidelines. The KwaZulu-Natal (KZN) Department of Health (DOH) started rolling out ART in March 2004.

ART in KZN as in other provinces of South Africa has involved the regimens outlined in Table 1. Regimen 1a was to be dispensed to all adult patients excluding females who still intended to bear children as well as male shift workers for whom regimen 1b was provided.

Regimen 2 was the second line drug in case of failure of regimen 1a and 1b. There were two paediatric regimens till three years of age and two other regimens for older children.

Table 1: ART regimens for adults and children

Adult patient regimen			
Regimen 1°	**Regimen 1b**	**Regimen 2**	**Special Regimen**
• Lamivudine (3TC) • Stavudine (d4T) • Efivarenz (EFV)	• Lamivudine (3TC) • Stavudine (d4T) • Nevirapine (NVP)	• Didanosine (ddI) • Zidovudine (AZT) • Lopinavir/otonavir (LPV/r)	• Any combination of treatment as a replacement of drugs from 1a or b
Paediatric patient regimen			
Regimen 1		**Regimen 2**	
<3yrs	**>3yrs**	**<3yrs**	**>3yrs**
• Lamivudine (3TC) • Stavudine (d4T) • Kaletra	• Lamivudine (3TC) • Stavudine (d4T) • Efivarenz (EFV)	• Didanosine (ddI) • Zidovudine (AZT) • Efivarenz (EFV) Nevirapine (NVP)	• Didanosine (ddI) • Zidovudine (AZT) • Kaletra

The urgency of the situation resulted in a very short time frame between the policy decision to provide treatment and the roll out. As a consequence, the planning process was fast tracked and some critical areas such as information systems were overlooked. It is against this backdrop that the KZN DOH decided to evaluate the 32 sites that started rolling out in 2004.

Objectives

The objectives were to describe the ART programme, identify management problems and propose solutions. Evaluation was not to be meant to judge but to document the status of the programme in order to improve it. The Italian Cooperation has provided technical assistance and funds to the KZN DOH to achieve these objectives.

Rationale

Effective management requires that problems be identified and corrected before they become ingrained. Identifying the factors behind the variation in performance across the sites delivering ART helps to find out what can work in local circumstances, at what costs and with what effectiveness. Even in the case of an effective therapy like ART, the results depend on the amount of resources available and how they are used. It was therefore necessary to check how ART has been delivered since 2004, how resources have been used and what impact has been produced. It was therefore decided to take a cross-sectional survey on the characteristics of the delivery sites and a representative sample of records from the archives of the sites.

Structure of the report

This report is structured into two parts. **Part 1** deals with the description of the programme and the factors influencing the probability of remaining on treatment in the first two years. **Part 2** is related to the economic evaluation and the estimation of the cost-effectiveness of delivering ART through several staffing and workload strategies. The reason for separating the report into two parts is related to the relatively demanding statistical methodology used in Part 2 which might not be of interest for the average reader. Both parts are divided into introduction, methodology, results, discussion, conclusions and recommendations.

Part 1

Overview of the antiretroviral therapy in KwaZulu-Natal

1.1 Introduction

The complexity of ART requires a strong monitoring system on the status of the ART programme. Although a management information system was part of the accreditation process, the ART delivery sites did not have an electronic data entry of the patients' information. Furthermore, no information was available on the absorption capacity and the efficiency of the delivery system. This situation did not allow to monitor the progress of the ART programme to identify problems and propose solutions, and to measure its impact.

The only monitoring system has been based on weekly indicators, which have been described in Issue 10 of the DOH Epidemiology Bulletin[4]. At the end of each week the delivery sites faxed to the KZN DOH a Weekly Form containing the number of patients who were screened for CD4, were eligible for ART (CD4<200 cells/μl), begun ART, experienced side effects, discontinued and were known to have died during the week.

Figure 1 provides a snapshot of the reliability problems of the weekly monitoring system. The three lines in Figure 1 show the number of new patients who were tested for the first time for CD4, had an initial level of CD4<200 cells/μl and begun treatment between March 2004 and May 2006. According to the monitoring system, the first line should have represented the number of new patients who were tested for the first time for a CD4 count. However, if this were the case, the gradient of the first line should have maintained a relatively constant gap with the second and third line. Instead, its exponential increase suggests that the first line of Figure 1 is likely to represent the first as well as the repeat CD4 tests carried out on new and old patients under treatment. As the ART programme became established, an increasing number of patients on ART were re-tested every six month. The reporting of the cumulative number of tests is the most likely cause of the exponential increase of the monthly number of CD4 tests and the increasing gap between the first and the second line. Another problem in Figure 1 is the irregularity of the monthly numbers represented by the frequent drops in the three lines, which worsen in the months of December.

Figure 2 shows another reliability problem affecting the number of interruptions and deaths reported by the weekly system. The two lines in Figure 2 show that the reported number of interruptions and deaths had similar trends between March 2004 and January 2005. However the number of interruptions and deaths started to diverge after February 2005, and after May 2005 interruptions dropped down to almost zero. These different methods of reporting do not allow to compare rates of defaulting and deaths across months and the two outcomes should be aggregated into a common indicator of patients who discontinued for whatever reason. Because these reporting problems are likely to bias the estimation of survival, treatment retention should be used as an alternative proxy of effectiveness.

[4] For further information consult the site of the KwaZulu-Natal DOH http://www.kznhealth.gov.za/

Figure 1 Number screened, eligible, began ART, KZN weekly reporting

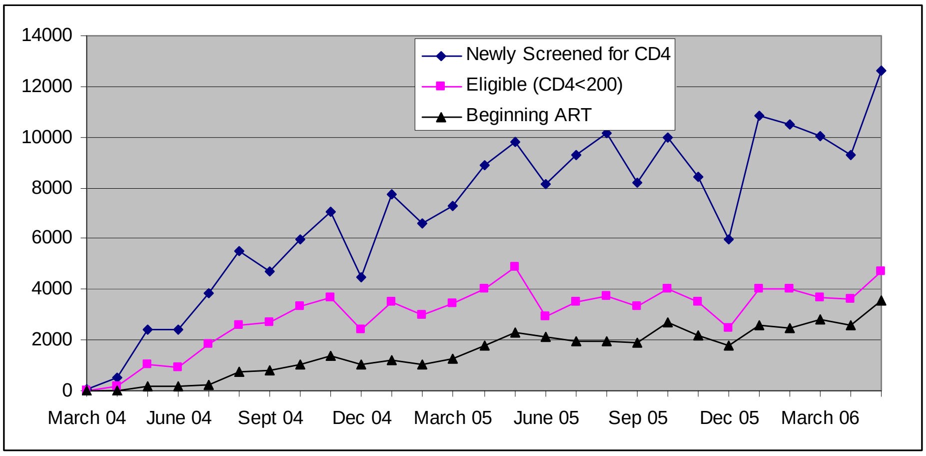
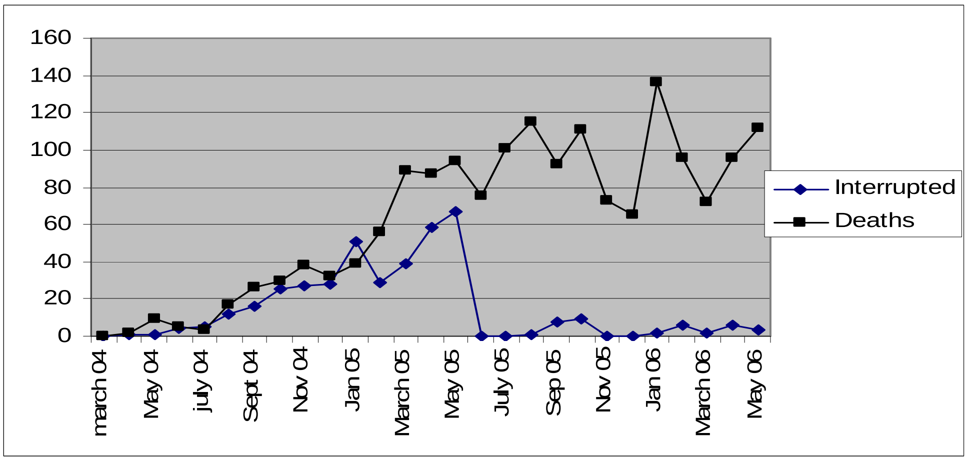


Figure 2 Number of deaths and defaulters, KZN weekly reporting



Individual records

While the previously mentioned weekly monitoring system provided aggregated numbers, no electronic system was set up to record the information from individual patients. Besides the provision of the aggregate numbers in Figure 1 and Figure 2, no in-depth analysis was conducted on the characteristics of the individual patients who died or defaulted. The information on each patient has been recorded on several paper forms, beginning with the registration and ending up with the follow up visits or the exit from treatment. Figure 3 shows the link between the different stages of the patients' flow and the individual forms which have been used by the staff to capture the patients' information. Table 2 provides the type of information which has been captured. Because there was never a systematic attempt to enter all this information into electronic format, no analysis was ever done by the ART programme on the individual patients' records.

Figure 3 Patient pathway for antiretroviral rollout in KZN

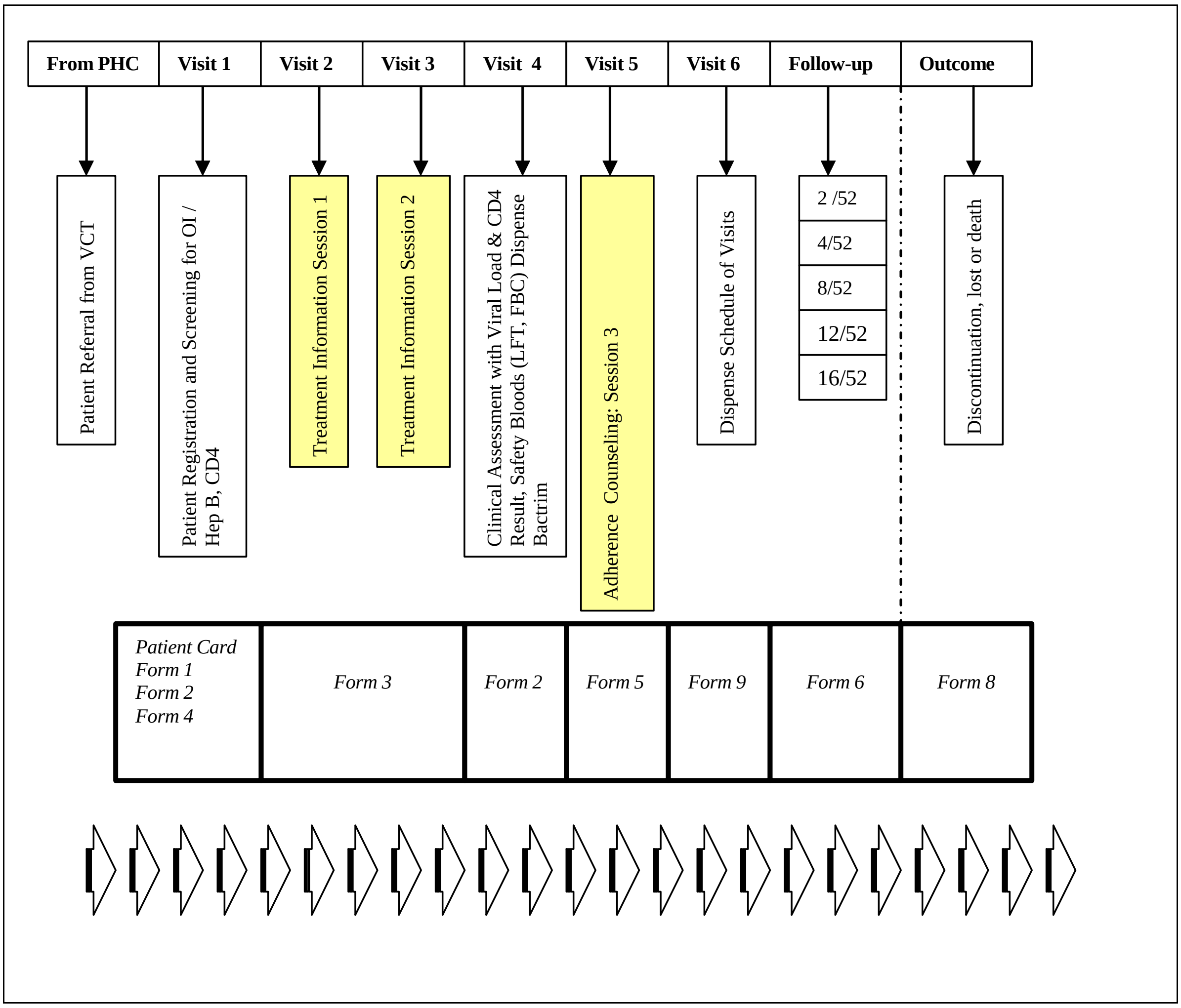


Table 2: Adult and Paediatric patient forms, KZN ART Programme

Adults	Paediatrics
Form 1: Adult Patient Registration	Form 1: Paediatric Patient Registration
Form 2: Adult Baseline Clinical Examination	Form 2: Paediatric Baseline Clinical Examination
Form 3: Adult Patient Counselling form	Form3: Paediatric Patient Counselling form
Form 4: Adult Baseline Laboratory Results	Form 4: Paediatric Baseline Laboratory Results
Form 5: Adult initiation/change of treatment	Form 5: Paediatric Prescription form
Form 6: Adult Patient follow-up	Form 6: Paediatric patient follow-up
Form 7: Adult Patient transfer form	Form 7: Paediatric inter-hospital transfer form
Form 8: Adult Patient Exit form	Form 8: Paediatric Patient Exit form
Form 9: Checklist	Form 9: Paediatric Checklist
Adult visit summary	Paediatric visit summary

ART delivery sites

Another information gap was the lack of a routine update on the staff, infrastructure and other characteristics of the delivery sites. A management information system should have theoretically included a routine update on the numbers and types of staff, the status of the infrastructure, the regularity of supplies, the economic costs and other factors influencing efficiency and effectiveness of delivering ART. Although it is frequently taken for granted that the information system provides such information, this is hardly the case. The only way to obtain such information was to visit the sites and collect it through a survey.

Justification for the evaluation

The absence of an electronic management information system has resulted in a number of opportunities being missed and a lack of documentation on progress. Processing all the individual records would have required a long time and the ART programme was at risk of proceeding in the dark with many assumptions and insufficient data to prove them. The only way to solve this deadlock was to conduct an evaluation.

The proposal for the study was drafted by the Italian Cooperation together with the Epidemiology Unit of the KZN DOH. The proposal was discussed with the ART programme management and it was approved by the Head of the Department of Health. The evaluation had to measure what happened during the period 2004-2006 in terms of ART retention to provide decision makers with different scenarios for expansion. Ethical approval was obtained from the Biomedical Research Ethics Committee of the University of KZN in May 2006.

1.2 Methodology

The design combined a cross-sectional survey on the delivery site and the data entry of a representative sample of records for a retrospective analysis. The delivery sites in Pietermaritzburg provided the opportunity to test the methodology and to finalize the survey instruments. The 32 delivery sites that started rolling out in 2004 were contacted to introduce the study proposal and to get the permission to visit them. Several teams visited the 32 delivery sites a first time between June and September 2006 and a second time in 2007 to fill missing information and to verify inconsistencies. The missing data on the CD4 and viral load of the sampled records were also filled thanks to the collaboration of the provincial laboratory at Inkosi Albert Luthuli Central Hospital.

Delivery sites' characteristics

The University of KwaZulu-Natal was contracted to collect the information on the delivery sites. Each manager was interviewed on several aspects related to the infrastructure, crowding conditions, number and type of part time and full time staff, links with the community health workers and volunteers in the catchment area, data recording, regularity of drug supplies and access to laboratories and equipment, and other variables which could shed light on the status of the ART delivery.

A separate interview was carried out with all the doctors and nurses staffing each delivery site. The information included length of service, training and supervision, record keeping, compliance with the treatment guidelines, information provided to the new patients, strategies to trace the defaulters, working conditions, major problems and suggested solutions.

A task analysis was conducted to observe the interactions between patients and staff. Having obtained informed consent from the staff and the patients, the survey team observed the interaction between patients and staff while they were registered, visited, counseled and monitored during follow up visits. This allowed to check how the expected tasks were performed, the time taken, the recording of the information on the relative paper forms, the information provided by the staff and the questions asked by the patients. The same patients were interviewed when they left the delivery sites to assess their perception on the quality of service and to measure their quality of life through the EuroQuol described in Annex I.

Sample of records

A representative sample of the patients' records was selected from the archives of the 32 delivery sites. The sampling was carried out with systematic random sampling with probability proportional to the size of the delivery site. In this way, the probability of selecting a record in any given delivery site varied proportionally with the number of records at the archive of the site. The sample size was estimated to capture the proportion of patients who discontinued ART in the first two years of treatment with an error of +/- 2%. The weekly indicators suggested that such proportion could have been around 17%. If simple random sampling had been possible, 1355[5] records would have been required to be 95% certain to capture such estimate within an error of +/-2%.

However, the fact that the survey was based on 32 delivery sites could have created some clustering effect. Patients attending the same site are likely to share common characteristics (e.g. geographic area, socioeconomic status) leading to similar subjects having similar outcomes (Chuang JH *et al* 2002). Therefore "… low variance due to differences between clusters (which is relatively large) is mixed with variation between subjects within clusters (which is relatively small because of the positive correlation), resulting in unnecessarily large standard deviations, larger *p* values, wider confidence intervals…." (Localio *et al* 2001).

Thus, to obtain the same precision of a simple random sampling, the stratified sample had to be adjustment to take into account potential clustering. In theory, the sample of 1355 had to be multiplied by a design effect (Deff) term given by the following formula: Deff = 1+(m-1)ρ. The m in the formula is the number of records per site and ρ is the intra-cluster correlation coefficient, representing the ratio between cluster variability and total variability. In practice these terms were unknown and the adjustment had to be based on the experience from previous surveys. The World Health Organization which has promoted the 30 cluster methodology for the estimation of immunization coverage and for Primary Health Care Review has applied a design effect of two (Lameshow *et al.* 1985) in this type of survey. Therefore, the sample size was doubled to 2,710 and topped up to 2,800 to take into account non sampling errors, which are deviations from the true values that are not a function of the chosen sample.

The sample corresponded to about 10% of the universe of patients which had started ART in the 32 clinics. According to the weekly reporting, about 28,000 patients had begun treatment between March 2004 and May 2006, thus the sampling interval selected for the systematic random sampling was 10 (28000/2800). In each delivery site, the first record was selected by drawing a random number between 1 and 10, while the following records were selected with an interval of 10 till exhausting all the records. The selection was carried out through the help of the staff of the archives of the delivery site and the information was entered into a Microsoft® Office Access 2003 database. A total of 2835 records related to the patients who started ART between March 2004 and the time of the survey were sampled, representing a universe of 28,350 patients.

The provincial laboratory at Inkosi Albert Luthuli Central Hospital provided the results of the CD4 and viral load of the sampled patients in case this was missing. Although the paper forms should have recorded the results of these tests, there were many missing data of the results of the laboratory tests. The missing values were traced in the database of the provincial laboratory.

Analysis

The first three sections of the results provide a descriptive analysis of the delivery sites and the patients. The first section is based on the interviews with the managers and the staff. The second section is based on the task analysis and the patient exit interview. The third section is based on the retrospective analysis of the representative sample.

The fourth section deals with the factors influencing the probability of remaining on ART. It was not possible to use survival as an outcome measure because the recording system grossly underestimated mortality. Only 20% of those patients who were lost to follow up were confirmed dead while the fate for the remaining 80% who discontinued ART was unknown. According to retrospective studies half of the defaulters are likely to have died (Dalal *et al*; Rosen et al; Bisson *et al*). Therefore, because survival would have been biased it was decided to use treatment retention as the best available alternative for effectiveness.

A Cox regression model was used to identify the risk factors that were significantly associated with ART retention. Cox regression was used to identify covariates which were significantly associated with the probability of remaining on ART. Staffing and workload profiles, which at parity of other factors were significantly associated with a higher probability of remaining on ART, were used as proxies for the delivery options used by the ART delivery sites.

The probabilities of remaining on ART given the above mentioned delivery options were measured through Kaplan Meier curves. The patients were categorized according to the delivery profiles of the sites they attended to. The relative retention rates estimated through Kaplan Meier curves provided the transition probabilities which were used in the cost-effectiveness analysis described in Part 2. Therefore, the methodology of this Part 1 should be considered as preparatory to the methodology used in Part 2.

1.3 Results

The results are divided into the following sections:

1.3.1 Characteristics of the delivery sites. This involved the infrastructure, staff composition, access to the site, problems experienced by managers and staff and suggested solutions;

1.3.2 Task analysis. The enumerators recorded the time that was taken by the staff to perform the different tasks related to registration, counselling, baseline examination, treatment regimens and follow up visits.

1.3.3 Exit interview. After the patients left the delivery sites, exit interviews were carried out to know their perception about the quality of the service;

1.3.4 Retrospective analysis on the sample of 2835 patients who began ART in 2004; and

1.3.5 Implications for the expansion of the delivery of ART.

1.3.1 Characteristics of the delivery sites

The 32 delivery sites were scattered across the province (Figure 4). According to the Geographic Information System (GIS), each delivery site served an average of slightly less than 200,000 people living within an average distance of 17 kilometres (kms) from the delivery site. Table 3 shows the delivery sites according to their catchment population, average travelling distance and accessibility index. The index was estimated through a GIS methodology (Annex II) that took into account the average travelling distance to the nearest road, difficulty of terrain (e.g. hills) and average distance in kms to reach the delivery site.

Figure 4 Location of the 32 ART delivery sites in KZN

Table 3 The 32 ART Delivery sites covered by the evaluation, GIS profile

Delivery site Name	Population within catchment area of the delivery site in 2007	Average Distance Travelled in kms by the catchment population	Accessibility Index
ADDINGTON	59984	3	1
BENEDICTINE	143395	16	15
BETHESDA	132697	27	24
CHARLES JOHNSON	224094	18	27
CHRIST THE KING	193851	25	32
CHURCH OF SCOTLAND	213096	21	30
DUNDEE	147857	23	22
E.G. USHER	59169	21	8
EDENDALE	281715	14	26
ESTCOURT	179552	24	28
G. J. CROOKES	259102	21	37
GREYS	61816	12	5
HLABISA	272594	25	46
KING EDWARD	103629	3	2
LADYSMITH	258399	22	38
MADADENI	349882	14	32
MAHATMA GHANDI	144005	5	5
MANGUZI	52339	15	5
MOSVOLD	133491	23	20
MSELENI	76061	23	12
MURCHISON	289804	24	46
NEWCASTLE	67280	16	7
NGWELEZANA	246756	15	25
NKONJENI	162111	17	18
NORTHDALE	193412	15	19
PORT SHEPSTONE	100273	16	11
PRINCE MSHIYENI	416321	8	21
R.K. KHAN	499817	7	25
ST APOLLINARIS	215017	26	38
ST MARY’S	454988	13	39
STANGER	72141	10	5
VRYHEID	152713	31	32
AVERAGE	**194292**	**17**	**22**

Managers' interview

The managers provided the information on the ART sites. Ten percent of managers mentioned that the presence of security problems did not allow easy access to the sites. Active tracing of defaulters was rare, although nine out of 10 sites had transport, and only 28% of the sites were involving the community health workers (CHWs) in their area. One major problem in preventing discontinuation of ART was the absence of an effective monitoring system to alert the staff when patients were not coming as scheduled. As a consequence, only one fourth of the managers knew how many patients had defaulted in the previous month.

The delivery sites were in good conditions but were overcrowded. Most delivery sites were open daily between 7 am and 4 p.m, had a regular supply of electricity and water, and were attached to a hospital. According to 60% of the managers, it was common for the patients to wait outside the site because of overcrowding. The delivery sites had a median of nine rooms. The median number of clinical staff included one enrolled nursing assistant, one professional nurse, less than one senior professional nurse, one chief professional nurse, one doctor and five counsellors. Most delivery sites received the results of the CD4 count before the patients' next appointment. Only three out of ten delivery sites received regular outside supervision.

Figure 5 Managers' opinion about the ART delivery sites

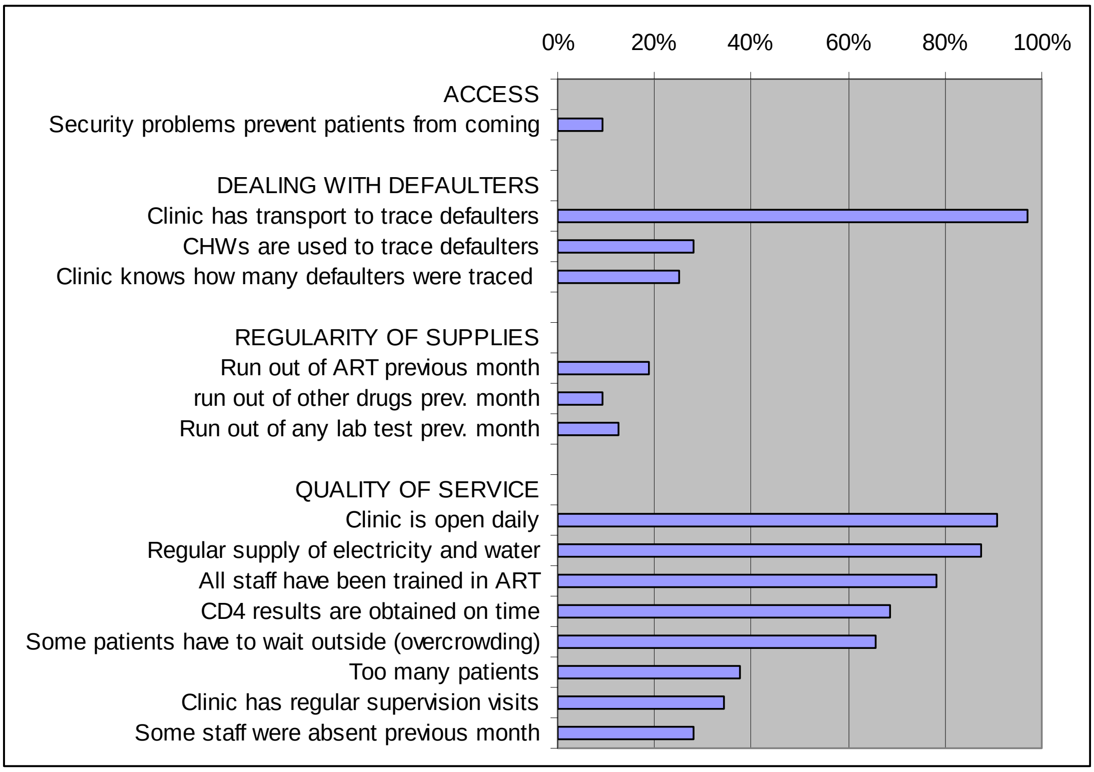


Table 4 Median number of rooms per delivery site

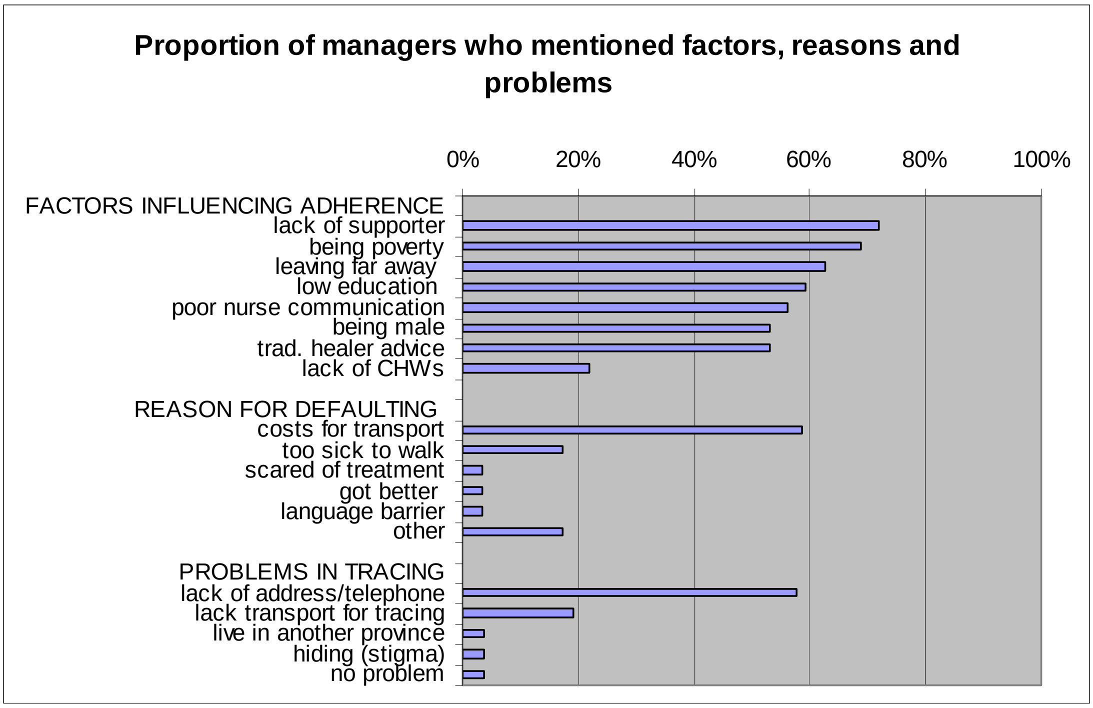
Type of Room	Median Number
Waiting	1
Counselling	2
Consultation	2
Data capture	1
Kitchen	1
Ablution	1
Other	1
Total	9

Table 5 Median number of full time staff per delivery site

	Median Number
Enrolled Nurse Assistant	1
Professional Nurse	1
Senior Prof. Nurse	0
Chief Prof Nurse	1
Counsellors	5
Doctors	1

The managers provided their opinion on the factors influencing adherence, default and tracing. Risk factors for poor adherence were mentioned in the following order: lack of patient supporters, socio-economic problems, living far away from the delivery site, low level of education, nurse scarce effectiveness in communicating the importance of adherence to ART, being a male patient and receiving conflicting advice from the traditional healers. The major reasons for defaulting were the costs of transport, the presence of side effects and severe clinical conditions not allowing patients to walk. If patients missed their appointment, the staff telephoned the family but home visits were carried out only by a minority of delivery sites. The main problems in tracing the defaulters were lack of a telephone contact number and unclear address.

Figure 6 Managers' opinion on adherence, default and tracing *



*the sum does not add to 1 (100%) because more than one reply was given

Staff interviews

All the nurses and the doctors who were present during the day of the survey were interviewed to ascertain problems in delivering ART. About 60% of the staff had been based at the delivery sites for more than one year, two thirds had received a first training on ART. One fourth had received refresher training on ART and a first training on the integration with treatment for other sexually transmitted infections and tuberculosis (TB). About 20% of the staff received training on how to fill the ART forms to record the information on the individual patients during registration, baseline examination, counselling, follow up visit and exit.

A few questions were asked about the knowledge gained through training and the information provided to the new patients. Most staff knew that the eligibility for ART was the presence of a CD4<200 cells/μl, while only a few gave incomplete or not relevant answers (Figure 7). In terms of information provided to the new patients (Figure 8), about one fourth emphasized the importance of positive living, one fifth told patients what HIV was and how to prevent its transmission, and 15% communicated the importance of adherence and the possibility of side effects. The rest of the information given to the patients dealt with the type and duration of treatment, the need to return monthly and to bring the supporter, the importance of nutrition, the presence of opportunistic infections and the need to bring their sexual contacts for counselling and testing.

Figure 7 Staff knowledge

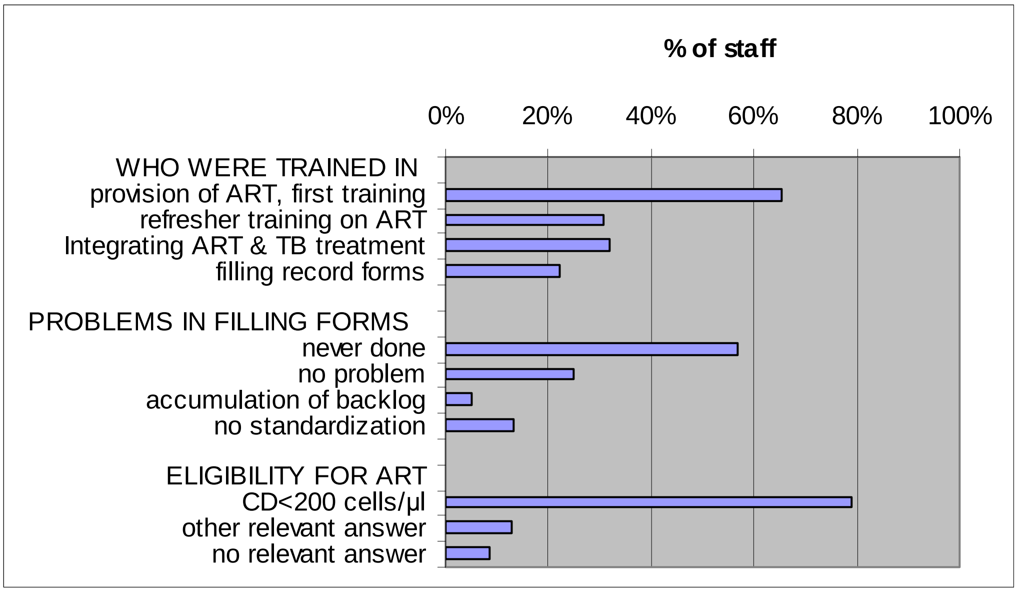
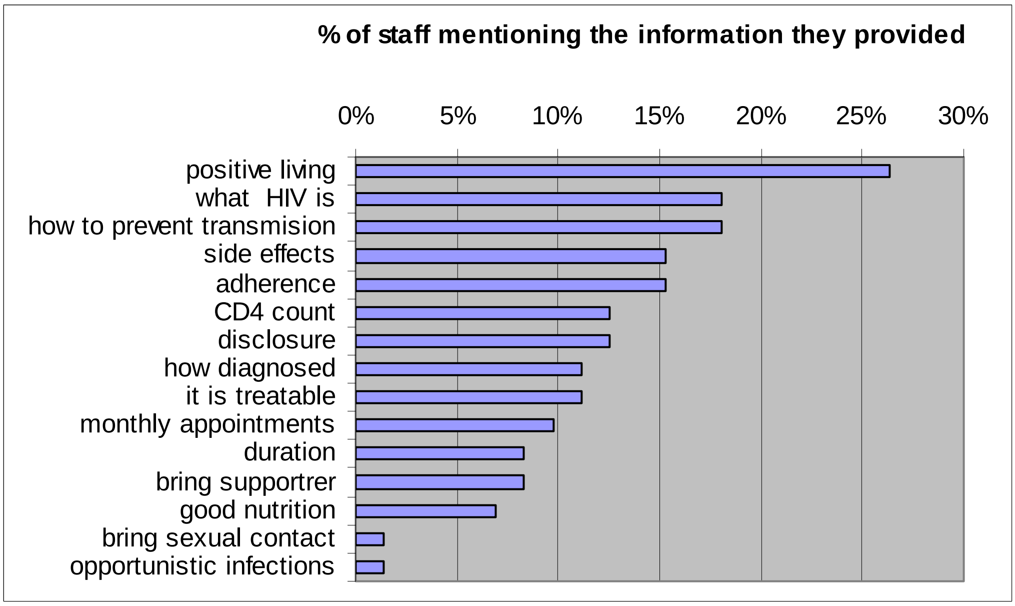
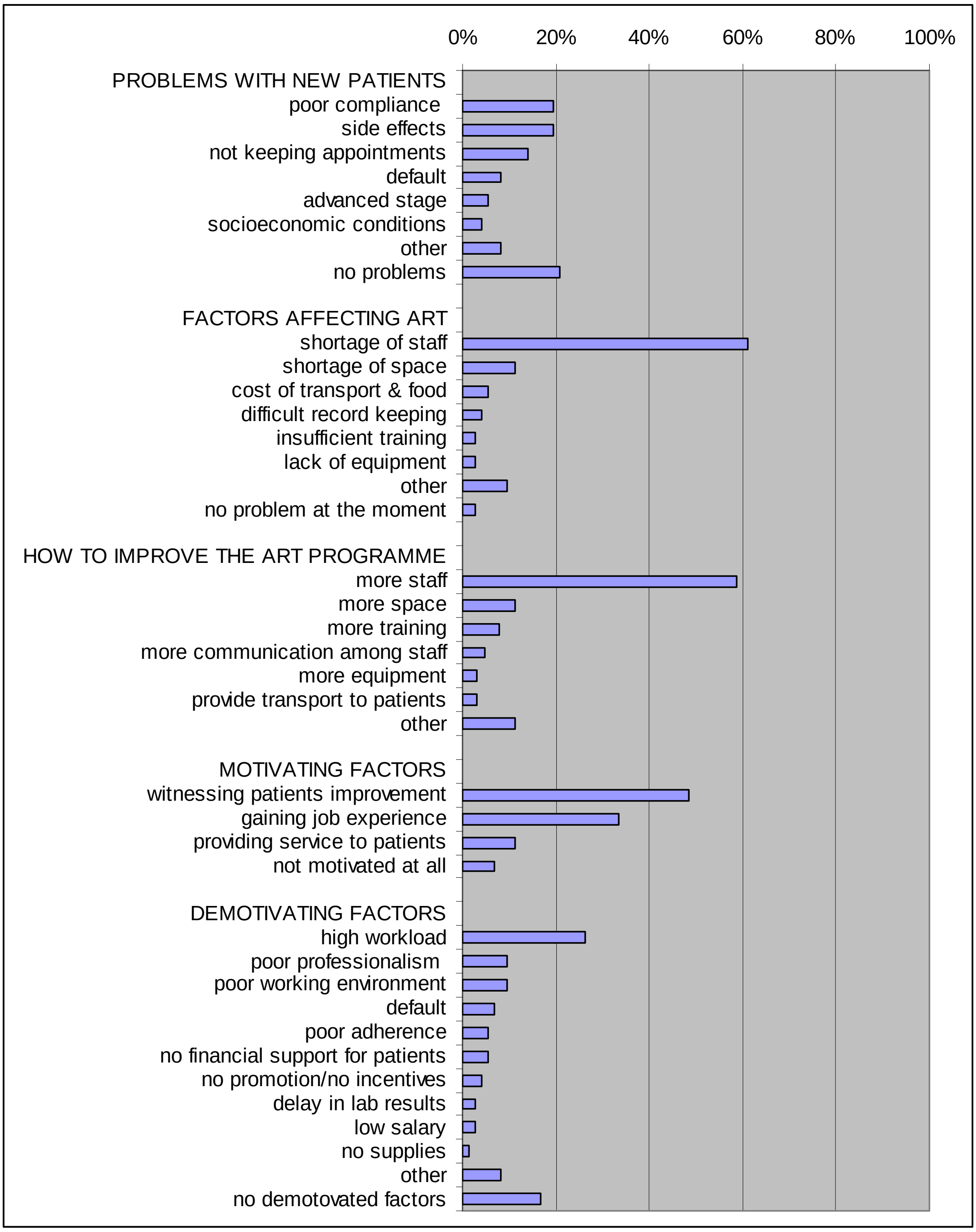


Figure 8 Proportion of staff providing information to new HIV patients



Staff provided their opinion on the problems affecting the delivery of ART (Figure 9). The major problems among new patients were the low level of compliance, the presence of side effects, the failure of keeping the appointments, the default, the advanced stage with which the patients started ART and the poor socioeconomic conditions. Most nurses mentioned that shortage of staff and space were significantly affecting the delivery of ART. Recruitment of more staff, upgrading the infrastructure to increase the waiting area, more training and improved communication among staff were the main suggestions on how to improve the ART programme. The most common motivating factors for the staff included the improvement in the clinical conditions of the patients and professional growth through job experience. The major demotivating factors were the high workload, poor professionalism and poor working environment- The high workload was caused by too many patients and too few staff. Poor professional behaviour meant that a few staff were not doing a proper job or were not proactive in helping patients. Poor working environment included moody behaviour and negative attitude by some staff, lack of cleanliness of the site and high level of stress. Low salaries, lack of incentives and no career promotions were also mentioned as demotivating factors.

Figure 9 Problems affecting the staff

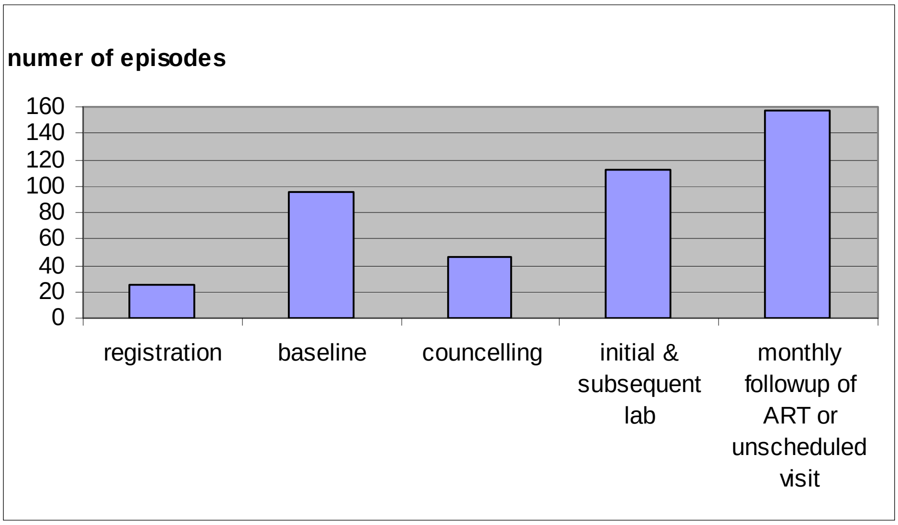


1.3.2 Task Analysis

Informed consent was obtained from the staff and the patients attending the delivery sites to observe the daily activities. Overall, the survey team observed 361 patients who were selected at random from the list of patients attending the delivery site on the day of the survey. These patients underwent registration, counselling, baseline clinical examination, treatment initiation and follow up visits. These steps have been described in the introduction and have been depicted in Figure 3 and Table 2.

The objective of the task analysis was to observe what happened during these visits. This included the estimation of the time taken for each visit, the completeness of the expected tasks (e.g. weighing the patient), the compliance with record keeping (e.g., recording the weight), the interaction between patient and health personnel (e.g. explanations given, questions asked) and the other factors that could be observed. Some patients underwent more than one type of visit/episode during the same day (e.g. registration and baseline examination), while others had one episode only (follow up visit). The 361 patients contributed 437 episodes in total between registration, baseline, counselling, lab tests and follow up visits.

Figure 10 Type and number of episodes which were observed



The longest and the shortest time were respectively dedicated to the clinical examination and the baseline laboratory test (Table 6). The median time taken to conclude each episode of the ART processes hid a high variation. For example, registration took a median of 13 minutes with a range of 4-30 minutes, baseline laboratory tests took a median 6 minutes with a range of 5-36 minutes, the range of baseline clinical examinations and follow up visits took respectively 5-64 and 5-80 minutes.

Counselling was conducted in groups with a median of 13 patients and lasted more than one hour per session. The size of the group ranged between 3 and 39 patients. The shortest median time was 67 minutes for drug readiness and the longest one was 90 minutes for home environment. Also in the case of counselling there was a high variation ranging between less than 10 minutes to a maximum of two hours per session.

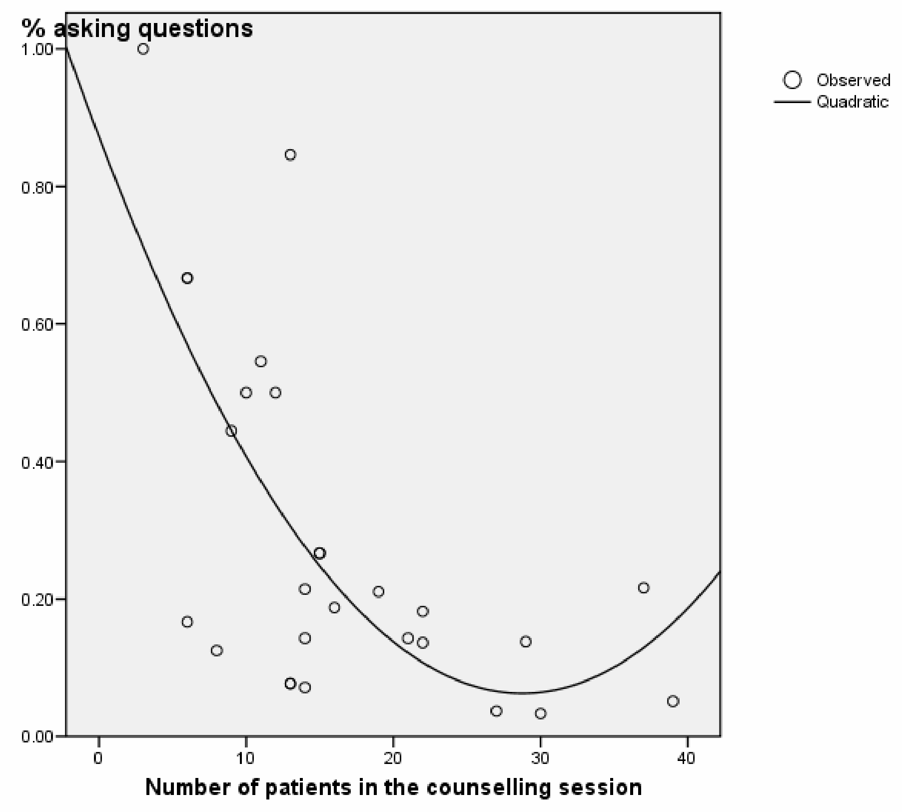
Participation declined with the number of patients attending the counselling session. Each point in Fig 11, representing a counselling session, is plotted according to the number of participants (X-axis) and the proportion of patients who asked questions (Y-axis). The regression curve, which explains 46% of the variance, shows that as the group size increased the proportion of participants asking questions declined. The highest participation was in groups with less than ten patients and the lowest was in groups with more than 25 patients.

Table 6 Task Analysis

Visit	Number of episodes which were observed	Average Time	Median Time	Range in minutes
Registration	24 episodes observed	13 m	11 m	4 - 30 m
Counselling *				
a)positive living	24 sessions observed	91	80	17-180
b) drug readiness	27 sessions observed	75	67	8-165
c) drug adherence	26 sessions observed	70	65	5-180
d) nutrition	13 sessions observed	89	68	17-180
e) disclosure	29 sessions observed	88	80	8-180
f) supporter	19 sessions observed	88	80	8-180
g) home environment	5 sessions observed	98	90	66-132
Initial lab test	109 episodes observed	7	6	5-36
Baseline examination & Treatment initiation	91 episodes observed	15	12	5-64
Follow-up visits	154 episodes observed	11	8	5-80

* The sessions had a median of 13 patients with a range between 3 and 39 patients.

Figure 11 Proportion asking questions and size of the counselling group



Completion of recording

During the task analysis, the observers noted how frequently the staff recorded the information on the relative forms which were described in the introduction (see Table 2). Figure 12 shows that more than 80% of the staff filled the various sections of the registration form. The baseline clinical examination form was recorded for less than half of the patients. It was observed that only for about one third of the patients, the staff recorded if the patients had been previously hospitalised and if they had been previously treated for TB. Symptoms and results of the physical examination were scarcely recorded on the relative sections of the form. Recording was low for the weight, height, temperature, World Health Organization (WHO) stages and WHO performance, lab requests and other key information collected during the baseline clinical examination.

Figure 12 Task analysis, compliance with recording

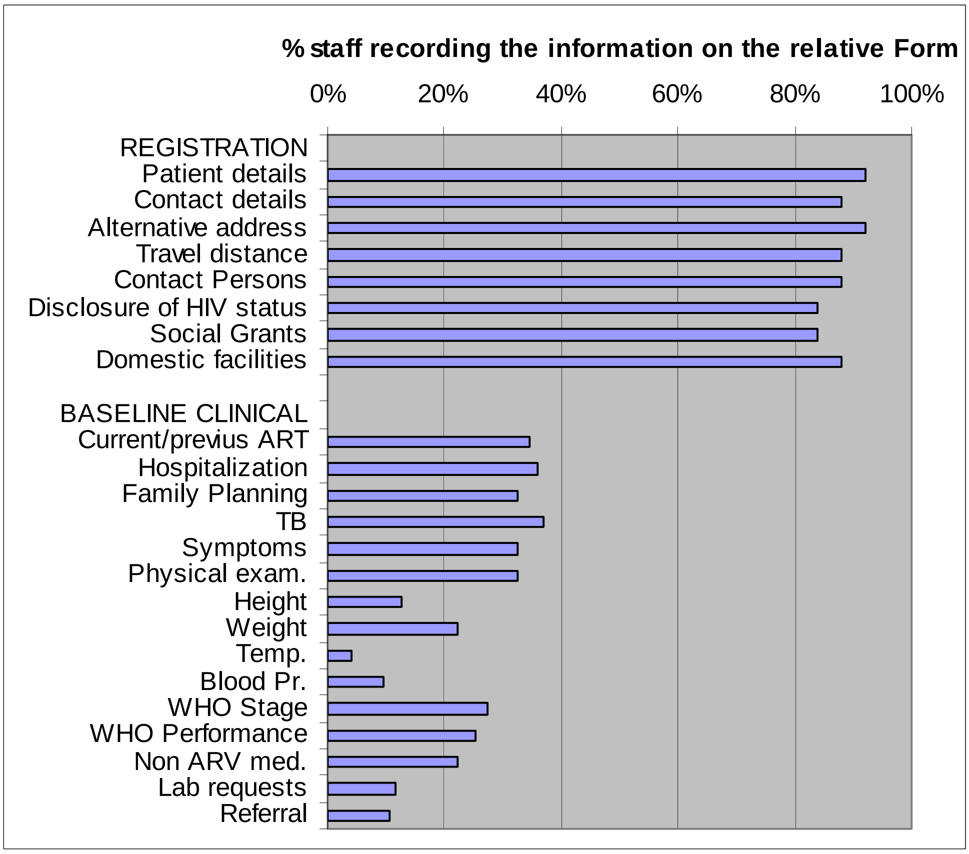
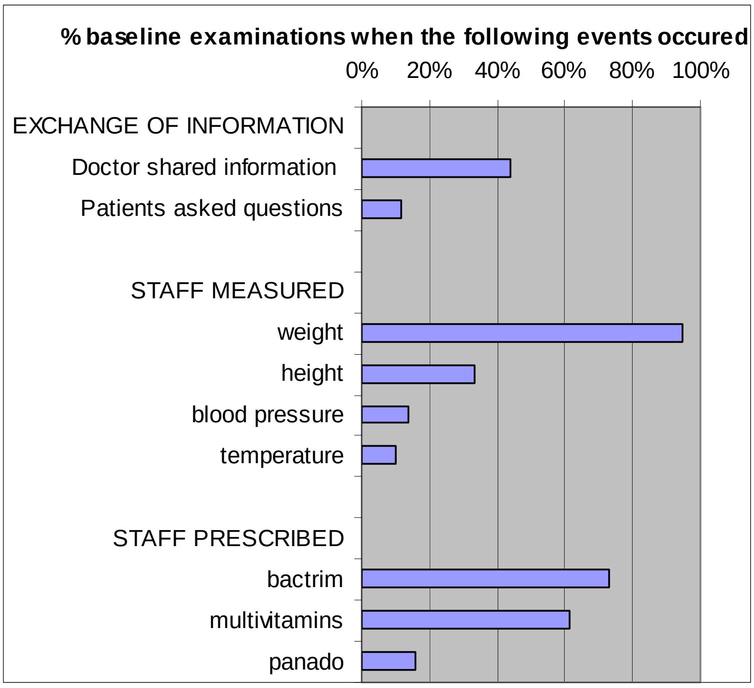


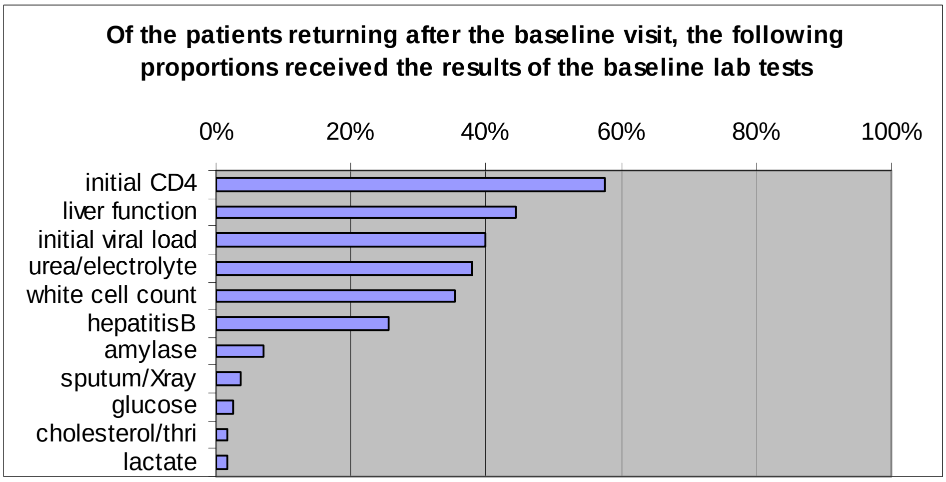
Figure 13 shows the completeness of the most important tasks performed by the staff during the baseline clinical examination. Less than half of the staff shared the findings of the clinical examination and very few patients asked questions. Most patients were weighed but height, blood pressure and temperature were measured for a minority of cases. Full Blood Count (FBC), urea, liver functions, hepatitis B test, Pap smear, sputum and chest X-Ray were requested for about 10% of the patients. It was observed that the most common prescriptions at baseline were Bactrim and Multivitamin, while Fluconazole was prescribed for a minority of cases.

Figure 13 Task analysis, events observed during the baseline examination



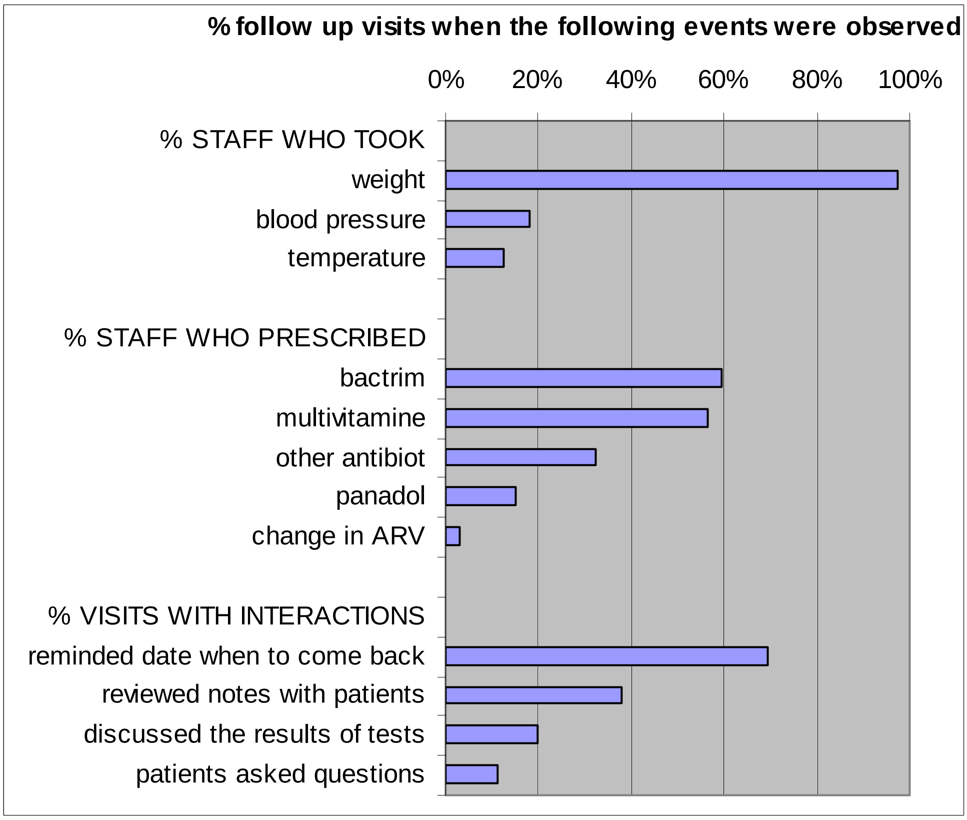
The most frequent baseline laboratory tests were the CD4 value (Figure 14). During the task analysis of the patients who were returning to know the results of the baseline laboratory tests, 60% received the results of the initial CD4, about 40% received the results of the liver function tests, the initial viral load and the values for urea/electrolytes and the white blood cell count. About one fourth of the patients received the results of the hepatitis B test, while very few results were related to the remaining baseline tests. These results show that not all the patients receive the results when they come back after the baseline laboratory visit and that some laboratory tests are carried out on a small proportion of patients.

Figure 14 Task analysis, baseline laboratory tests



The completion of recording declined further for the patients who were observed during the follow-up visits (Figure 15). After the regimen assignment, the patients returned for their monthly check up and to get their drug replenishment. Weight was regularly measured while less than 20% of the patients had their blood pressure and temperature measured. Bactrim and multivitamins were prescribed in more than half of the follow up visits which were observed. In about 70% of the follow up visits the patients were reminded to come back by the expected date. Only a minority of the patients asked questions.

Figure 15 Observation of follow up visits

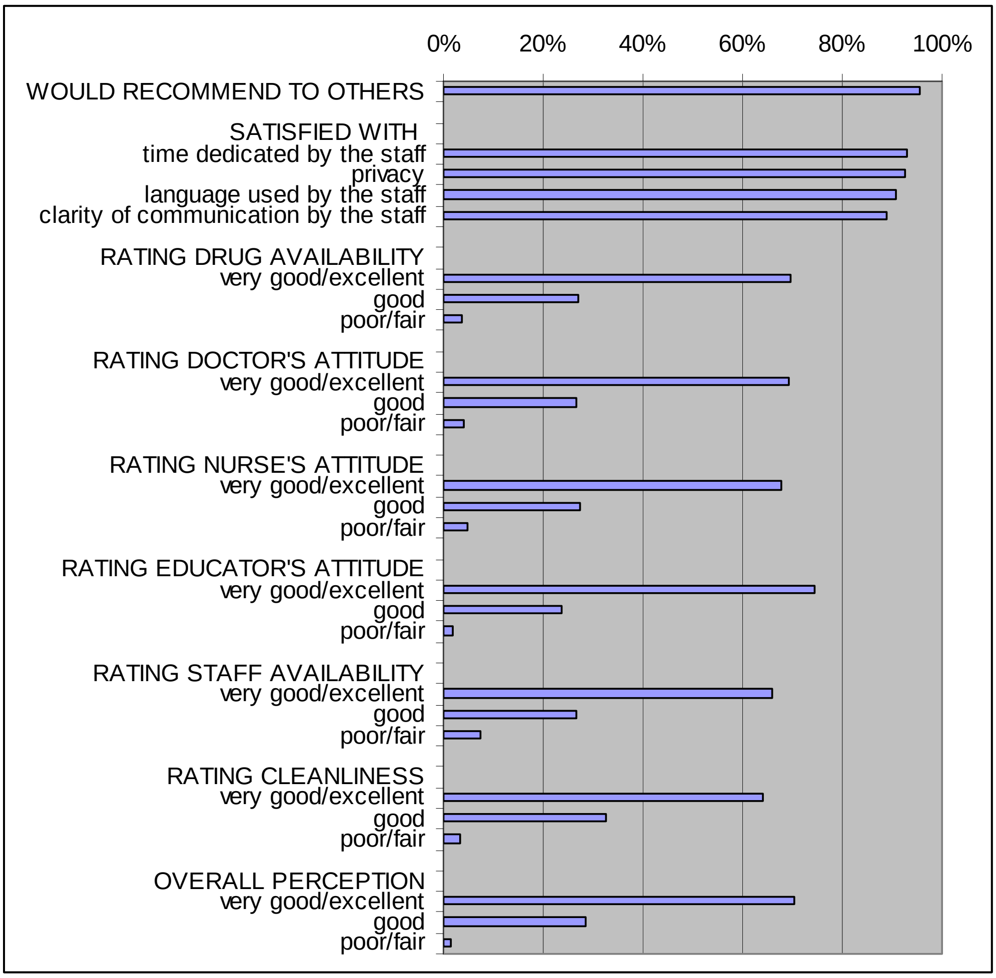


1.3.3 Exit interview

The patients who were observed during the task analysis were interviewed outside the ART delivery site. After obtaining informed consent and after making sure that there was enough privacy, 361 patients were interviewed. In the case of children (16%) the guardian was interviewed. About two thirds were females, more than half lived in formal dwellings, and the majority had not completed secondary education. This sub-sample of patients who were interviewed did not differ in mean age, prevalence of women and other characteristics from the 2835 records which were sampled for the retrospective analysis.

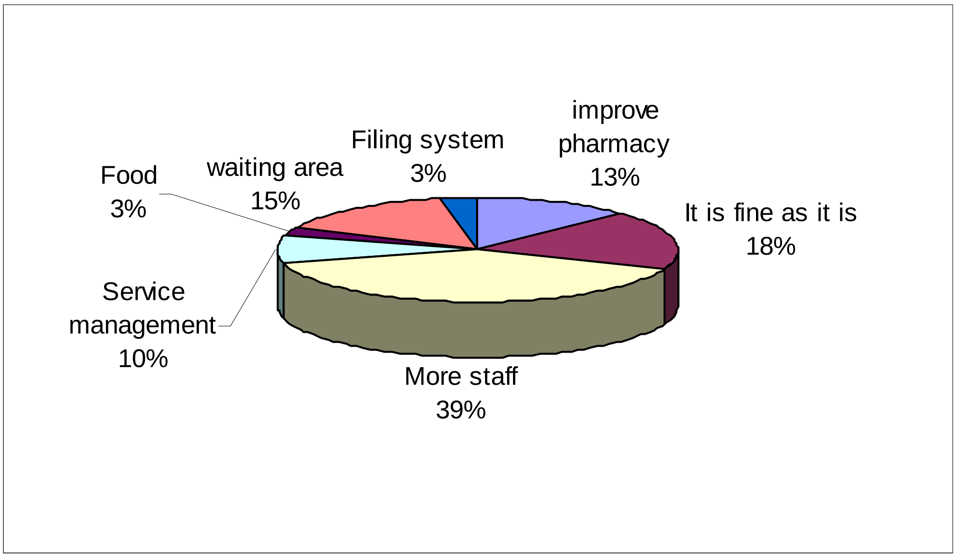
Patients were very satisfied with the service (Figure 16). Almost all the patients who were interviewed would have recommended the service to other patients in need of ART. About 90% of the interviewees were satisfied with the time dedicated by the staff, the privacy offered by the site, the language used by the staff and the clarity of what the staff said. More than 60% rated very good or excellent the drug availability, the attitude and the availability of the staff, and the cleanliness of the delivery site. The overall perception was very good or excellent for about 70% of the interviewees.

Figure 16 Satisfaction with the service



To verify that the high level of satisfaction was genuine other questions were asked about how services could have been improved (Figure 17). About one every five interviewees found that the existing services did not need any improvement. The rest suggested that the delivery sites needed more staff, more space and more chairs in the waiting area, speedier services at the pharmacy with the organization of different queues according to the severity of the clinical status. A minority of patients suggested to provide food while patients were waiting and to give food rations to be taken home. The rest of the suggestions dealt with service management including the provision of longer opening hours, punctuality of staff, more respect for patients, confidentiality, integration of services to reduce too many appointments and the provision of ambulances to transport patients at home.

Figure 17 Client's suggestions on how to improve ART



Some patients had contacted other services before attending the delivery site where they were interviewed. About 15% had contacted a traditional healer because he/she was located nearby their home and 22% had moved from a previous delivery site. The reason for changing delivery sites included access to paediatric regimens, better quality of services and lower risk of being seen by neighbours.

Most interviewees knew how HIV was transmitted. Almost everybody (93%) mentioned sexual intercourse, 30% added[6] the sharing of blades and needles, 21% added donating blood, 11% added mother to child and 4% added injections as potential transmission routes of HIV. The fact that one fifth mentioned that donating blood is risky is probably related to the fear that non-sterile equipment could be used to draw blood.

Perceived quality of life

Quality of life was measured through the EuroQuol standardized questionnaire, which is described in Annex I. The patients were asked about limitations of mobility, self care, usual activities, pain/discomfort, anxiety and depression. The degrees of functionality included "no problem" "some problem" and "severe problem". Most patients were self-sufficient, about half were suffering from moderate and extreme pain and discomfort, 10% were suffering from anxiety and depression, and 16% had impaired mobility. The above mentioned limitations were transformed into health related quality of life (HRQL). The HRQL is a standardized score varying between 0 and 1, with 1 defining perfect health and decreasing values defining declining quality of life. Figure 19 shows that for about half of the patients who were interviewed HRQL was between 0.9 and 1, 30% had an HRQL of 0.8-0.9, 10% had an HRQL of 0.7-0.8 and the rest had lower HRQL values.

[6] Because the same interviewee could add more than one cause of transmission the sum of the percentages of all the causes does not add to 100%

Figure 18 Quality of life of ART patients attending the delivery sites

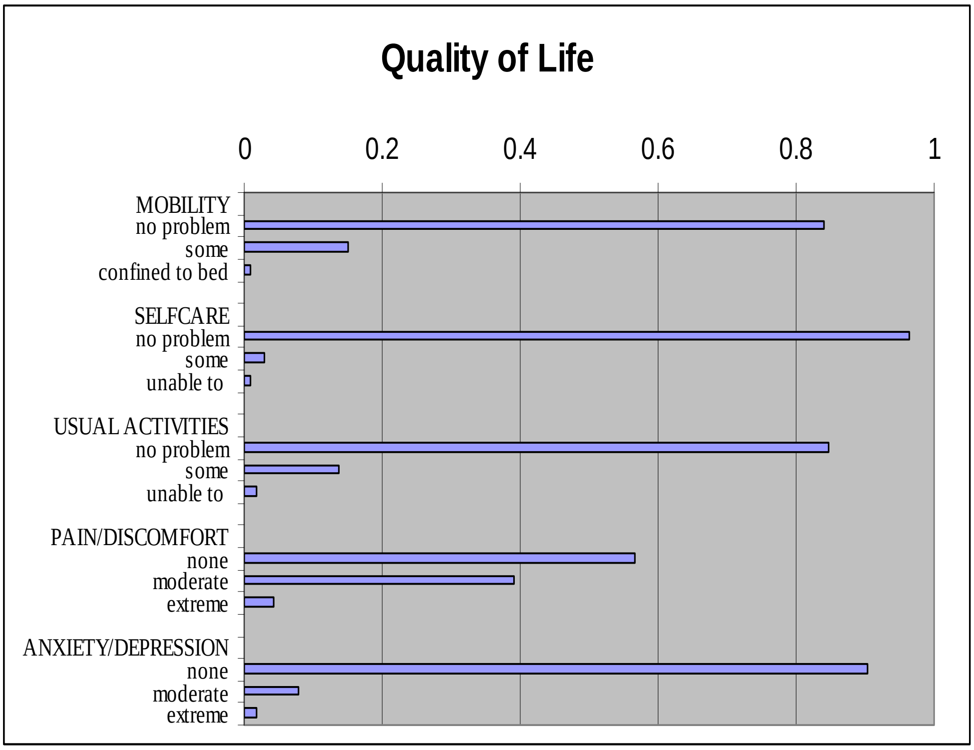
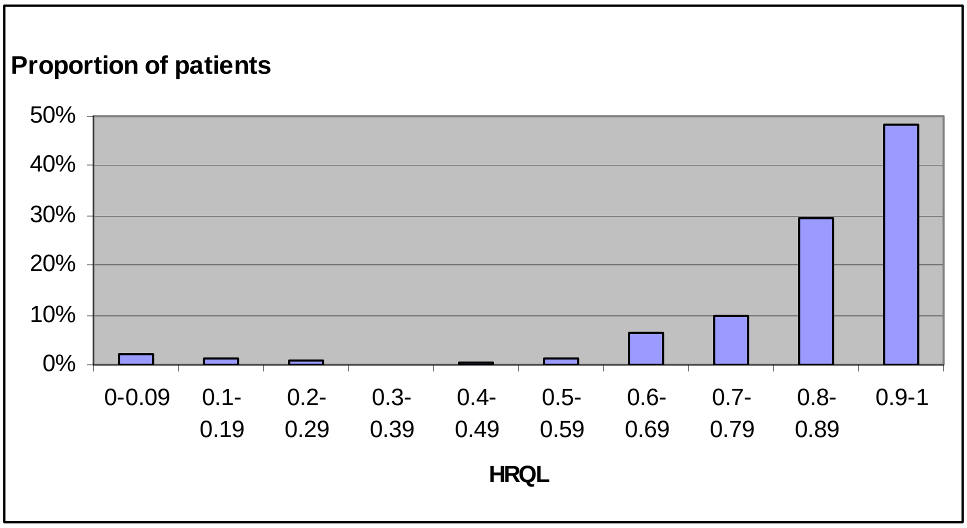


Figure 19 Health Related Quality of Life



1.3.4 Retrospective Analysis

1.3.4.1 Registration

This section deals with the retrospective analysis of the 2835 records which were representative of the universe of patients who started ART between March 2004 and September 2006. Each record was composed of information related to the stages of registration, counselling, baseline examination and laboratory tests, treatment initiation, follow up visits and exit. The age distribution was bimodal because of the presence of children and adults and with the known concentration of adult patients in the most productive age groups. The typical patient was female, single, Zulu speaking, unemployed, receiving grants and not having completed secondary education (Figure 20). The households where patients lived were usually provided with running water and electricity, half of them used electric stoves for cooking and one third had flushing toilet. About half of the patients were taking more than one hour to reach the delivery sites, and most of them were using taxis (Figure 21). Slightly less than half of the children were orphans and they were equally divided between loss of the father, mother or both parents (Figure 22).

Figure 20 Registration form, characteristics of patients

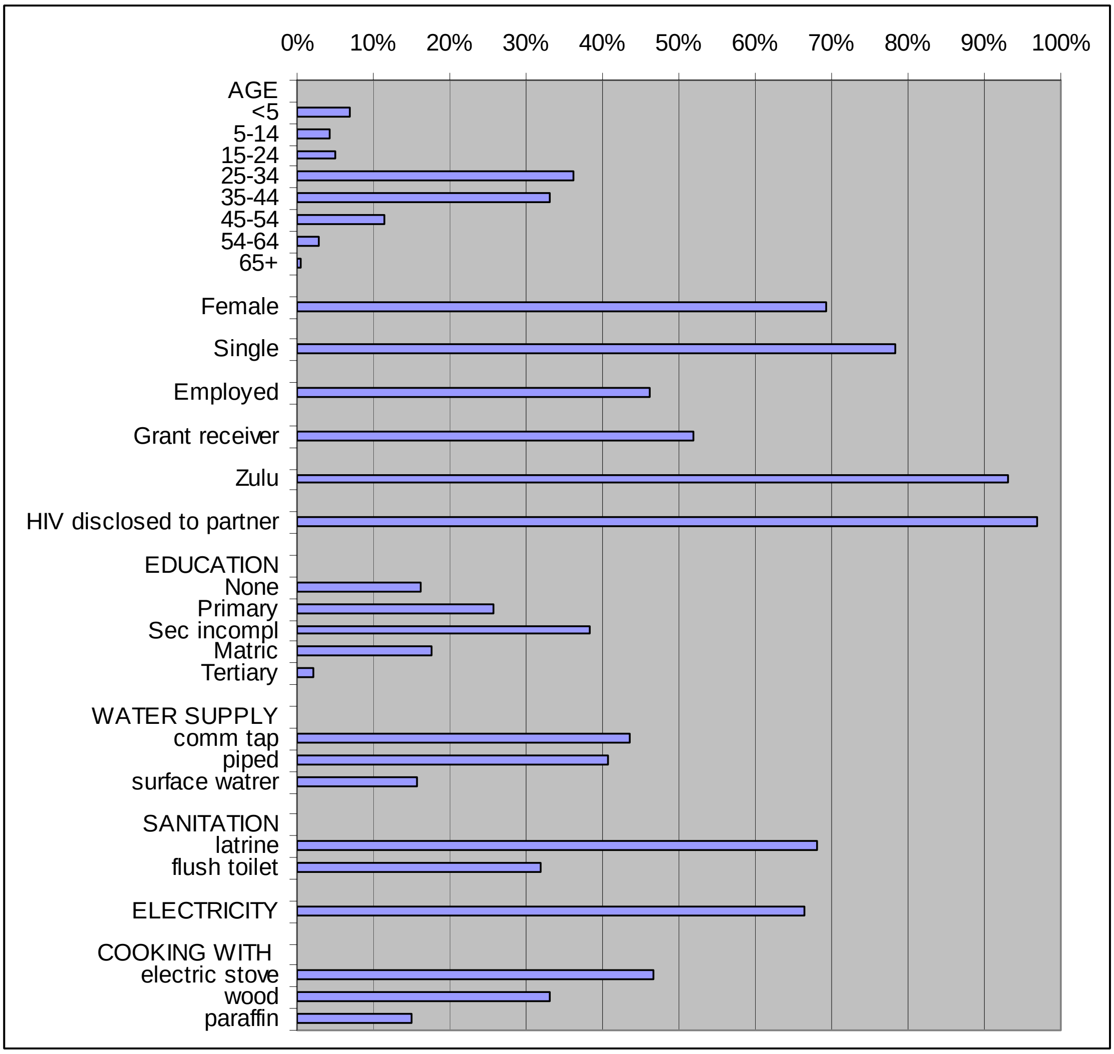


Figure 21 Registration form, access to delivery sites

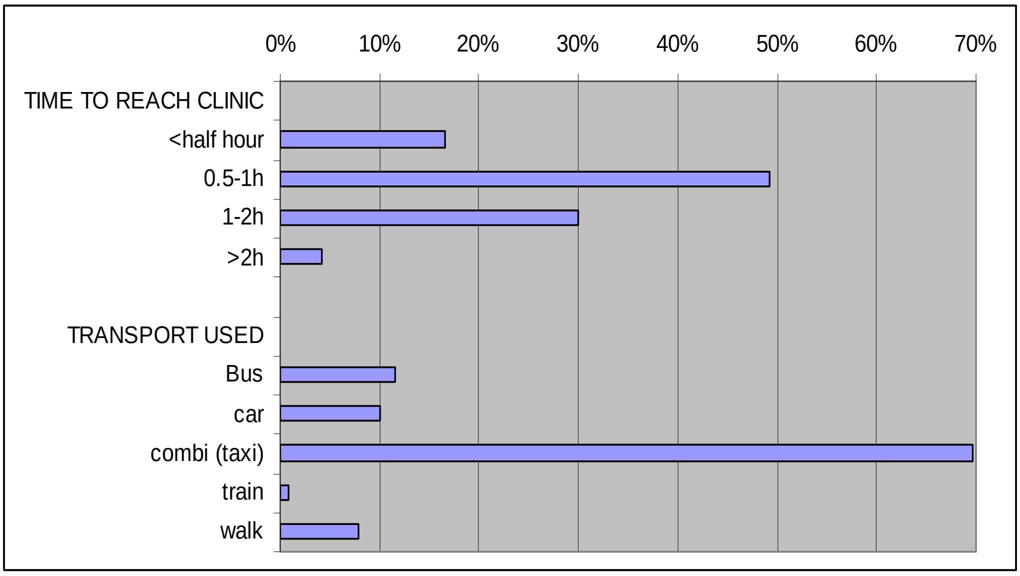
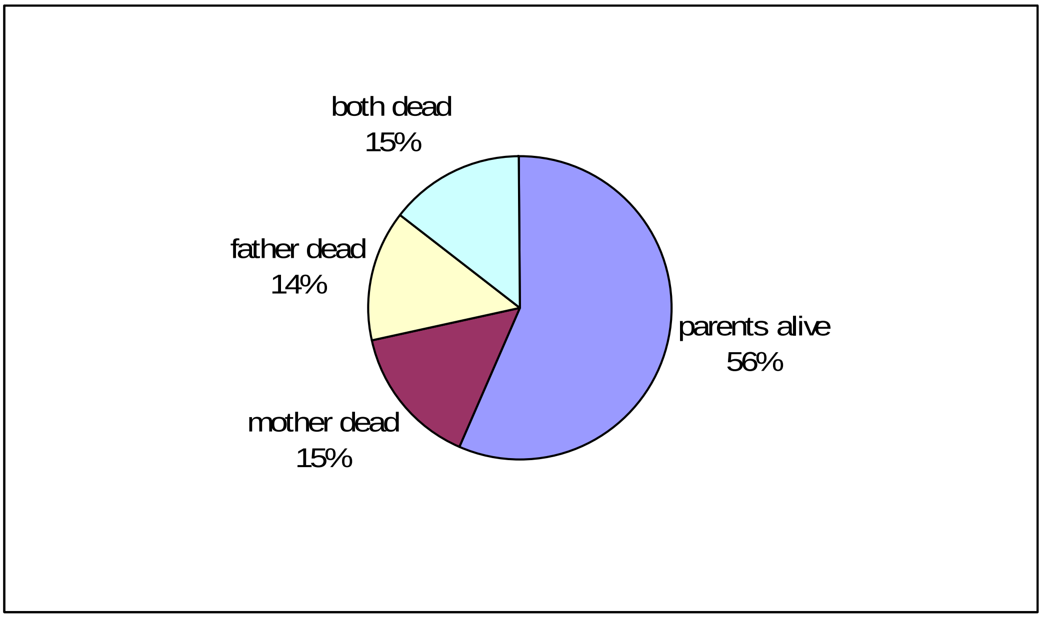
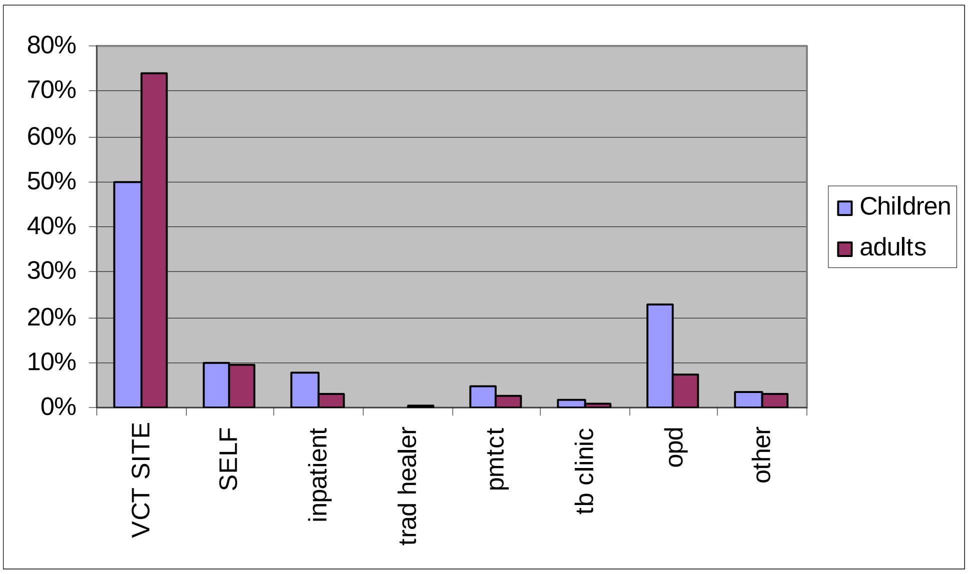


Figure 22 Registration form, prevalence of orphans



Most patients were referred to the ART delivery sites through Voluntary Counselling & Testing (VCT) sites. Figure 23 shows that half of the children and two thirds of the adults were referred through VCT. One out of ten patients was self-referred and more paediatric than adult patients were referred from hospitals and outpatient delivery sites.

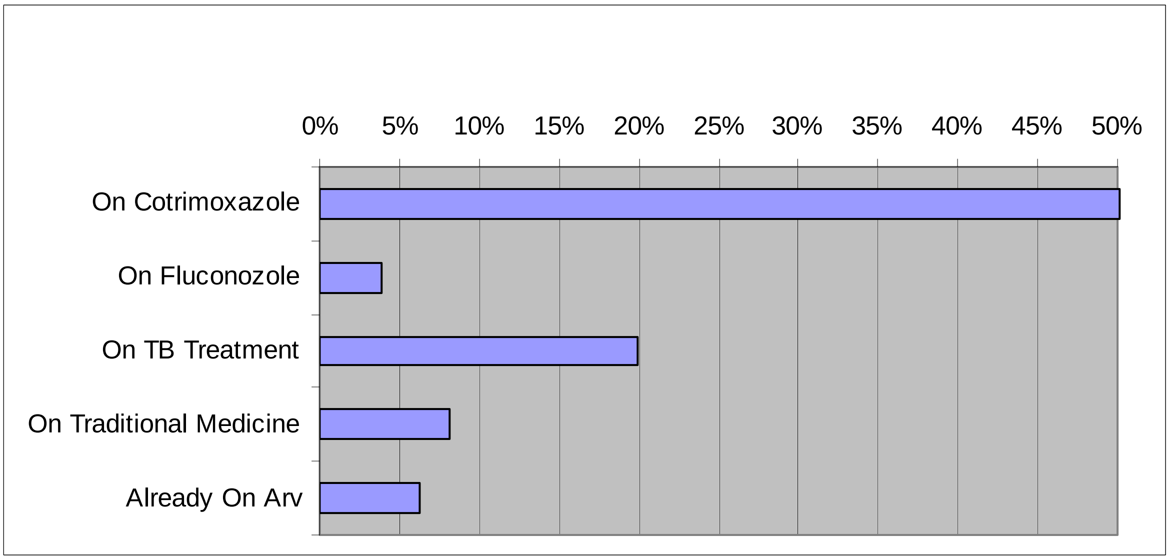
Figure 23 Registration form, type of referral to the ART delivery site



1.3.4.2 Baseline Clinical Examination

The most common previous medical problem was TB. While slightly more than one third of the patients had suffered from TB in the previous year, about 1% had suffered from liver, renal and mental disorders. Fifteen percent had been hospitalised in the previous year. At the baseline visit, half of the patients were on cotrimoxazole, 20% were on TB treatment, 8% were taking traditional medicine, 4% were on fluconazole and 6% were already on antiretroviral drugs (Figure 24).

Figure 24 Baseline examination form, medication at baseline



At baseline, the most common symptom among adults was the loss of weight and the most common sign was the enlargement of lymphnodes. Figure 25 shows that at the baseline visit the symptoms which were more reported by the patients were weight loss (50%), fatigue (42%), headache (34%) and cough (33%). Three out of ten patients reported night sweats, rash and thrush, while about two out of ten patients reported fever, vaginal/penile discharge, diarrhoea, dizziness, visual change abdominal pain and shortness of breath. Figure 26, reporting the most common findings at the physical examination, shows that about 8% had abnormal lymphnodes, 6% had skin problems and 4% had abnormal findings affecting lungs, ear nose throat, abdomen, and other systems.

Figure 25 Baseline examination form, symptoms reported by the patients

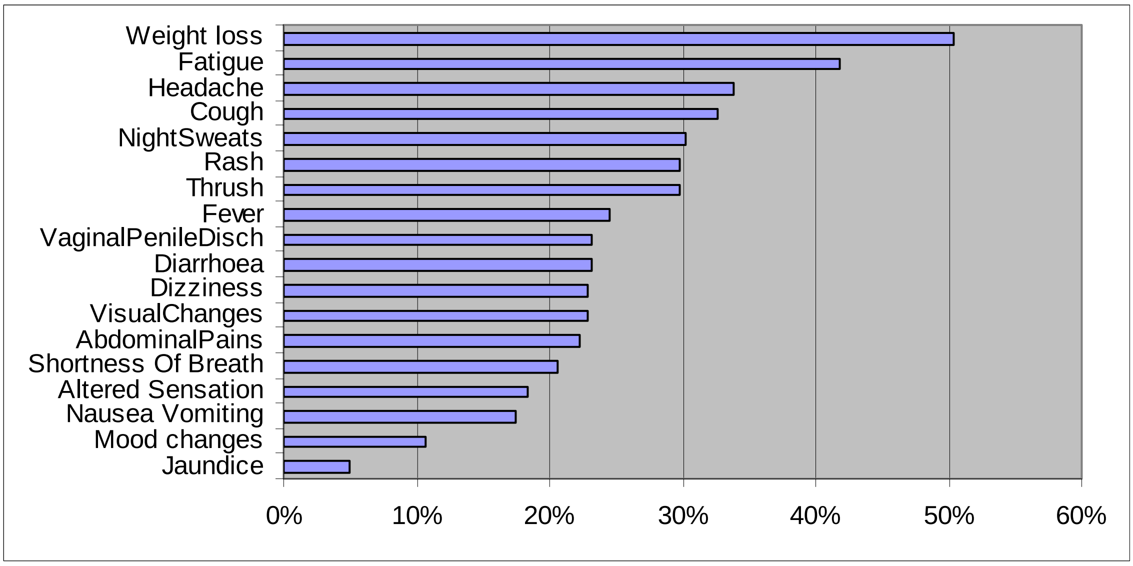
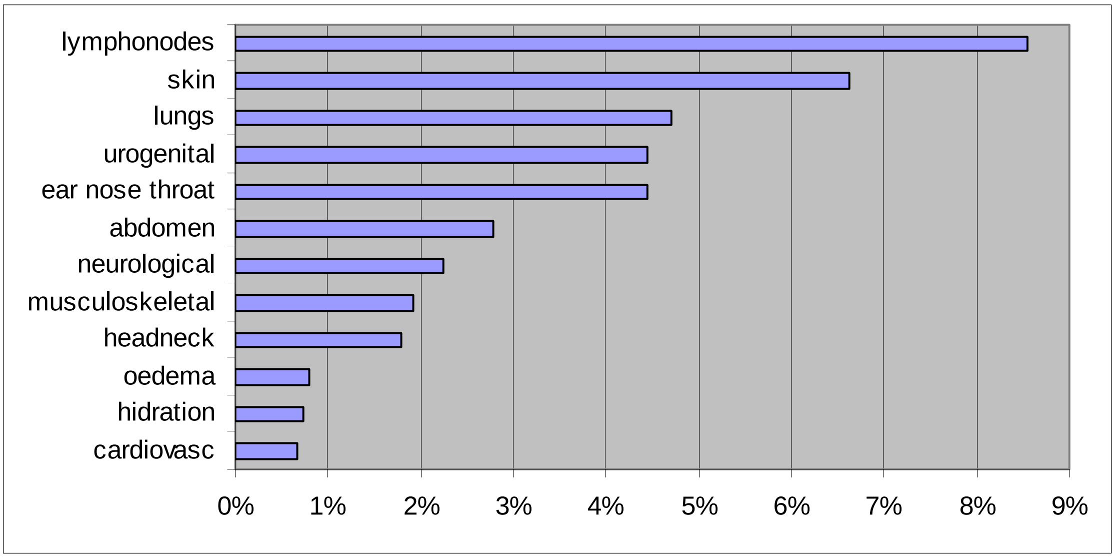
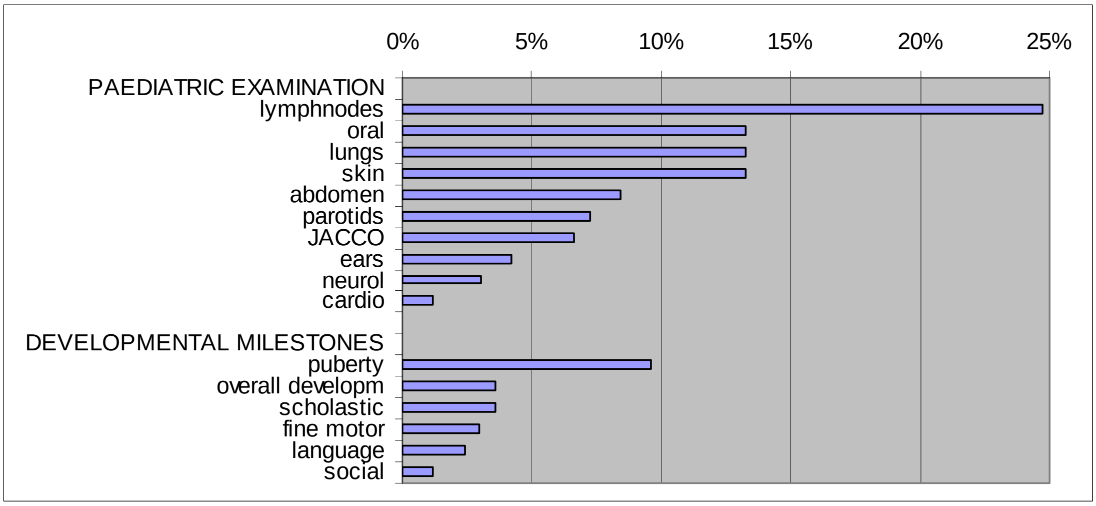


Figure 26 Baseline examination form, clinical signs among adults



The baseline paediatric examination revealed a higher prevalence of abnormal findings compared to the adults. One fourth had abnormal lymphnodes, 13% had problems affecting the oral cavity, the lungs and the skin (Figure 27). Less common were other signs related to the abdomen, parotids and other organs. About 10% had a delay in pubertal developmental milestones and less than 5% had other developmental delays.

Figure 27 Baseline examination form, clinical signs among children



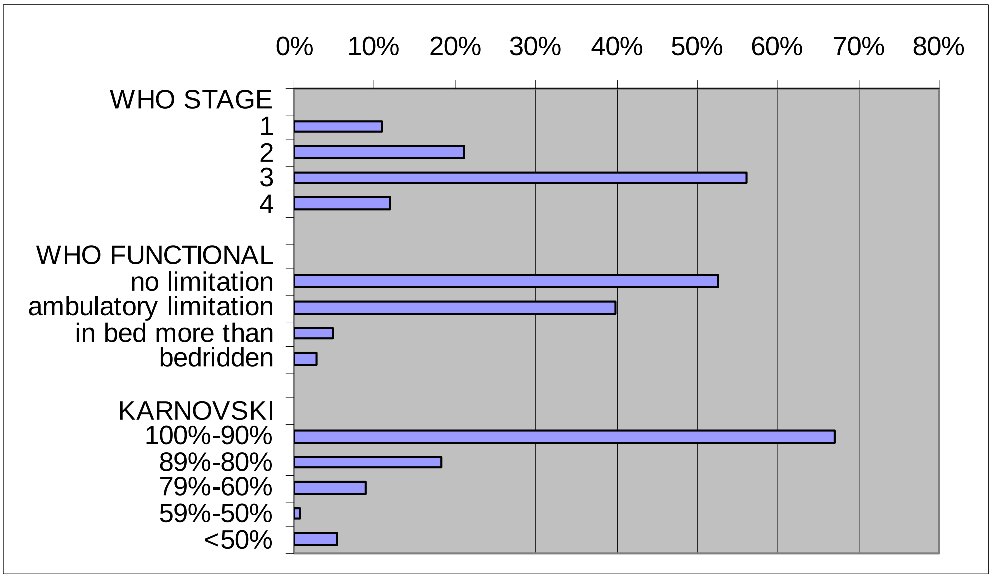
The clusters of symptoms listed in Table 9 were used to determined the WHO stage. The symptoms varied from the less severe (Stage 1) to the most severe ones (Stage 4) and as each patient has symptoms belonging to several groups, the highest group prevailed. For example, a patient with asymptomatic HIV or generalized lymphadenopathy was assigned to Stage 1, while another one with lymphadenopathy and recurrent upper respiratory tract infections was assigned to Stage 2. More than half of the patients were in WHO stage III (Figure 28).

The WHO Functional status and Karnofsky were used to measure the prevalence of disabilities. According to the WHO functional status about 40% had ambulatory problems, 5% were in bed most of the time and 3% were bed ridden. The Karnofsky is a disability scale ranging from 0% to 100%, with optimal functions (100%-90%), minor symptoms (89%-80%), normal functions with some efforts (79%-70%), inability to carry on normal activities but ability for self caring (69%-60%), inability for self care (59%-50%) and disabled for values under 50%. The median Karnovski was 84 with a prevalence of about 6% under 50% Karnofsky.

Table 7 Clusters of clinical signs defining WHO stages

WHO Stage	Symptoms
1	Asymptomatic HIV
	Persistent generalized lymphoadenopathy
2	Herpes Zoster (within last five years)
	Minor Mucocutaneous Manifestations
	Recurrent Upper Resp. Infe.
	Weight Loss <=10% body weight
3	Severe bacterial infections (within the last 5 years)
	Oral candidiasis (Thrush)
	Unexplained chronic diarrhoea for more than one month
	Unexplained prolonged intermittent or constant fever for more than one month
	Oral hairy leukoplakia
	Pulmonary TB (within last year)
	Weight loss of 10% body weight or more
4	Candidiasis (oesophageal, bronchial, trachea, lungs)
	Extrapulmonary cryptococcosis
	Cryptosporidiasis with diarrhoea for more than one month
	Herpes simplex
	HIV encephalopathy
	HIV wasting syndrome
	Kaposi's Sarcoma
	Lymphoma
	Atypical disseminated mycobacteriosis
	Mycosis (i.e. Histplasmosis, Coccidiomicosis)
	Extrapulmonary TB
	Pneumocystis pneumonia
	Progressive multifocal leucoencephalopathy
	Septicemia
	Toxoplasmosis Central Nervous System

Figure 28 WHO stage, WHO Functional status, Karnofski at baseline



Anthropometric indicators

The recording of anthropometric measurements was incomplete and unreliable and therefore the results need to be viewed with caution. While weight was available for 70% of the sample and had a smooth distribution (Figure 29), height was available for only 36% and its measurement was less accurate. The height measurements were more concentrated around 30 cms and one meter (Figure 30), because of the tendency to round up or down the height without taking a precise measurement. This caused unreliability of the body mass index, which is estimated by dividing the weight in kgs by the height in meters squared. After taking out the outliers' height data, the BMI followed a bell shaped curve with similar mean and median (Figure 31). Most patients with valid measures (Figure 32) were normally nourished or overweight (BMI 18.5-19.9) while 21% were undernourished (BMI <18.5) and 10% were obese (BMI>30). For the patients with valid WHO stages and valid BMI, the prevalence of under-nutrition (BMI<18.5) increased from nil to 30% between WHO stage I and stage IV (Figure 33).

Figure 29 Weight distribution at baseline

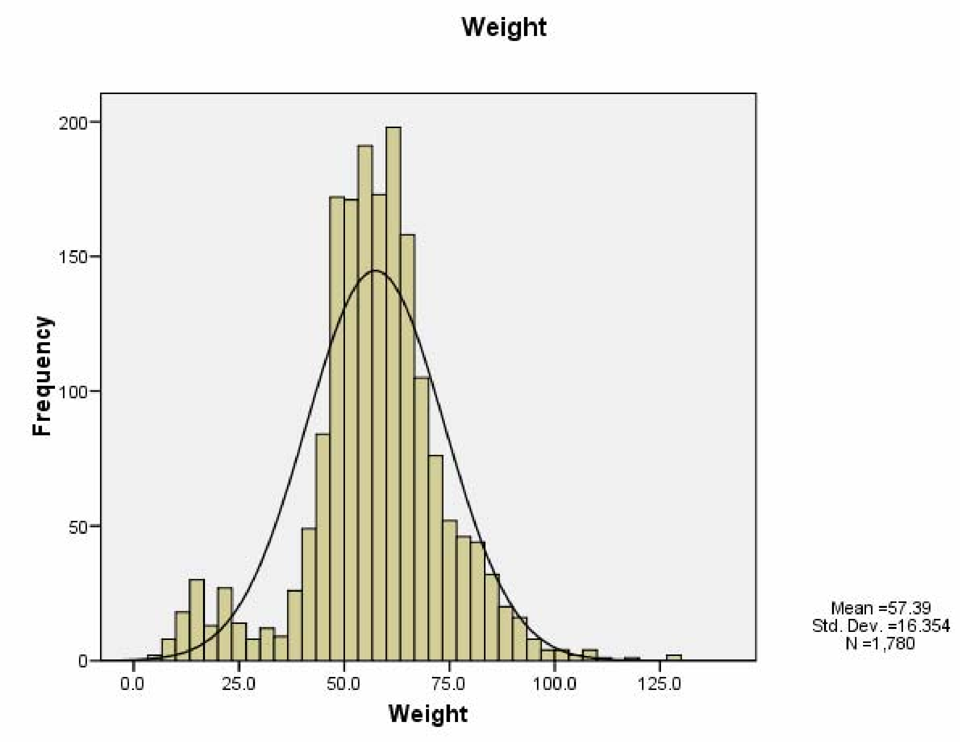


Figure 30 Height distribution at baseline

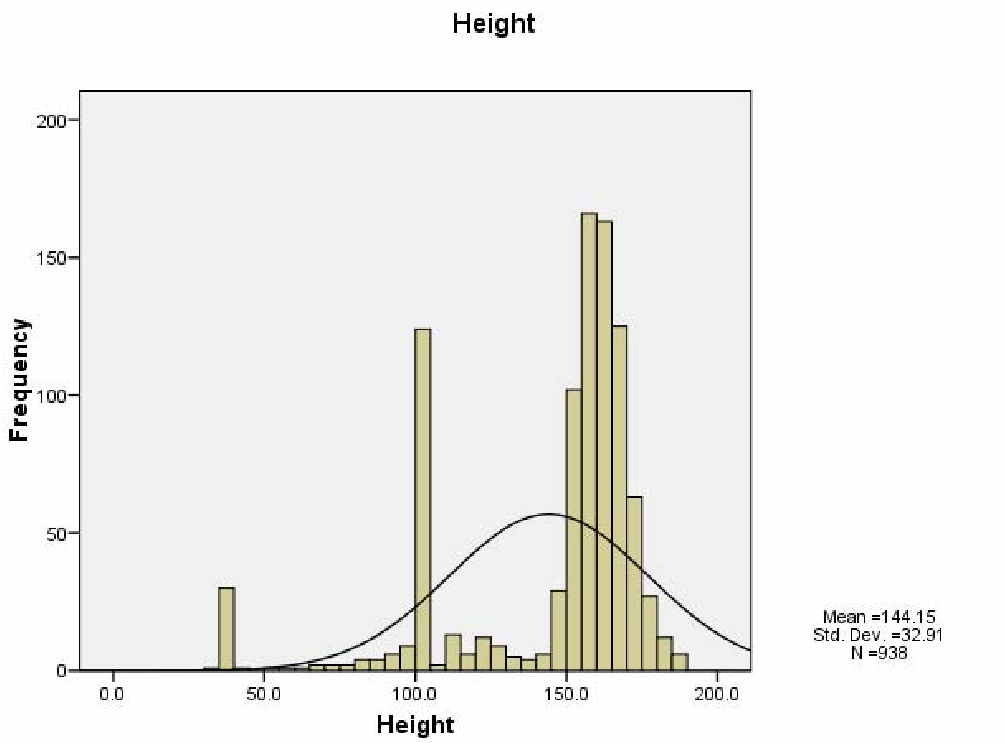


Figure 31 BMI distribution at baseline

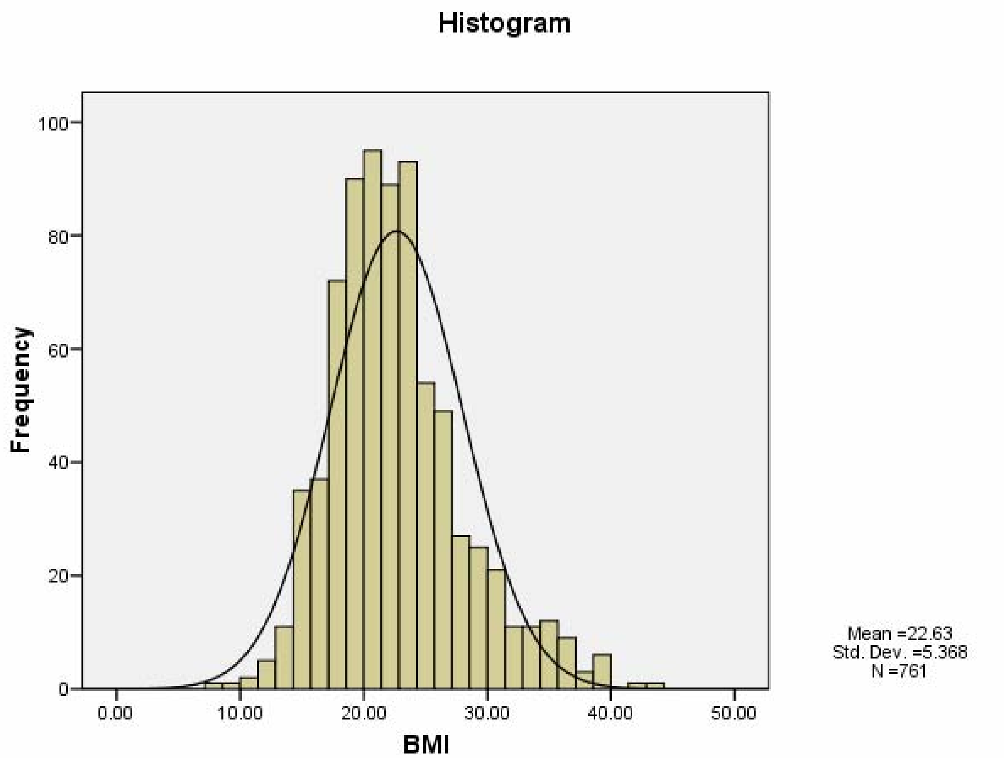


Figure 32 Undernutrition (BMI<18.5) and obesity (BMI>=30) at baseline

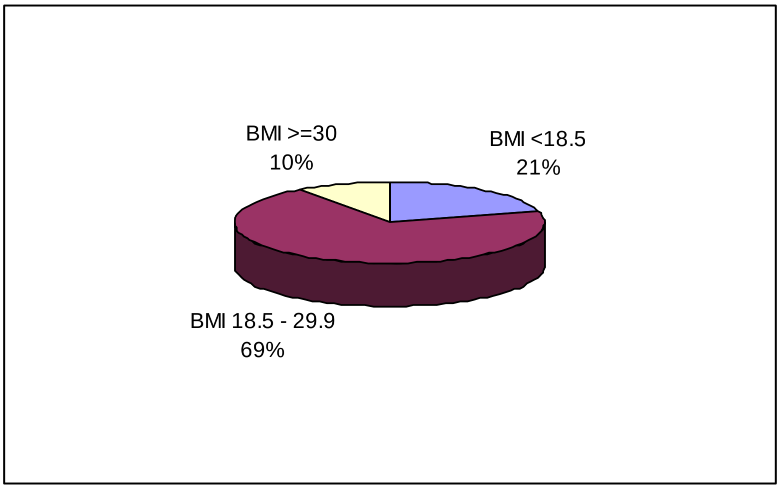
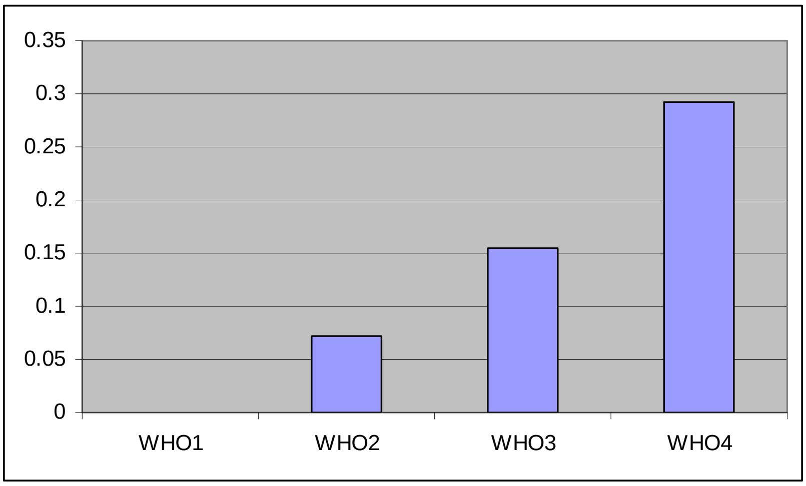


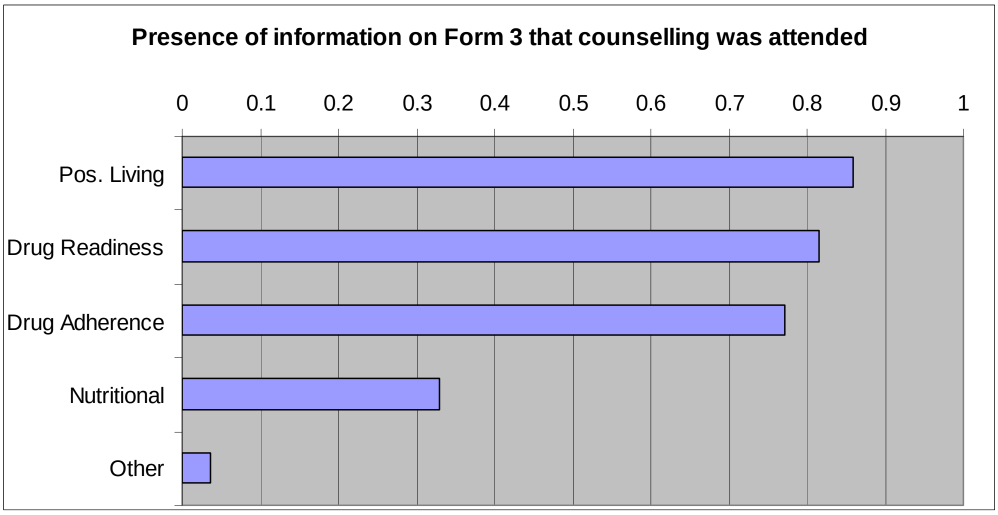
Figure 33 Undernutrition (BMI<18.5) by WHO Stage at baseline



1.3.4.3 Counseling

Figure 34 shows that 85% of the ART patients were recorded as having attended positive living while the recording declined for other types of counselling. As recording was given a low priority, many forms were left blank and it is likely that the proportions in Figure 34 are under-estimates of the real coverage of the different types of counselling. Nevertheless, the decline after the first sessions (positive living) suggests that the attendance was slightly lower for drug readiness and drug adherence and much lower for nutrition counselling. The counselling form had incomplete information on people to be contacted (70%) direction on how to reach the address of the patient (60%), and cell phone or the home telephone number (30%). Because of the incomplete information, a high proportion of defaulters could not be traced.

Figure 34 Counselling form, incomplete recording of counselling attended



1.3.4.4 Baseline Laboratory

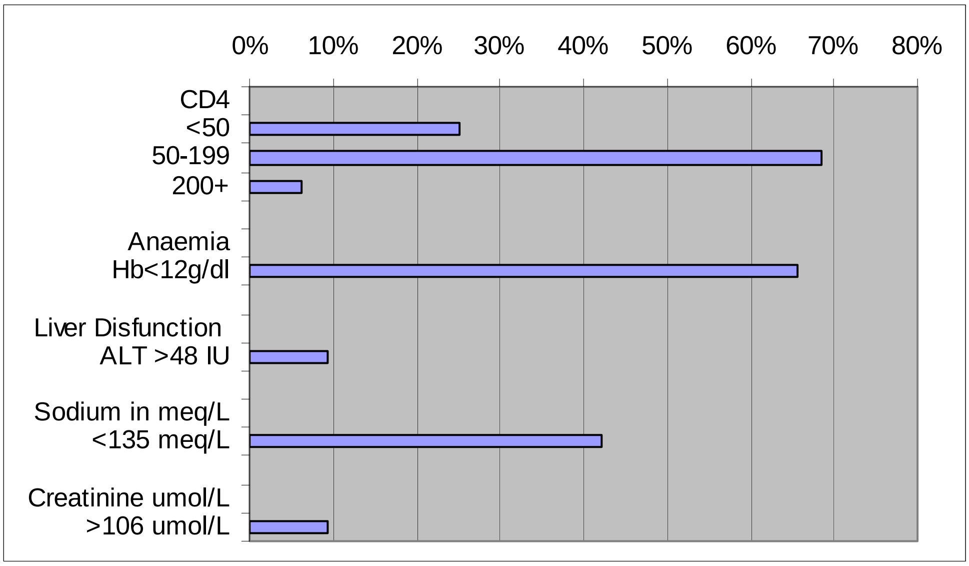
Table 8 and Figure 35 shows the baseline laboratory values for the patients having such baseline information. The median CD4 value was slightly above 100 cells/μl and the median viral load was above 40,000 copies/ml. About two thirds were anaemic, 10% had high level of transaminasis and creatinine, and 40% had low values of sodium (Figure 35). The baseline CD4 and viral load values were re-checked by searching the database of the provincial laboratories. The other values need to be viewed with caution due to their under-recording on the laboratory Forms.

Table 8 Baseline laboratory form, laboratory values

	CD4 (a)	**Viral Load (b)**	**WBC (c)**	**Haemo globin (d)**	**ALT(e)**	**AST(f)**	**Creatinine (g)**
Sample	2451	2269	1276	1469	1364	115	1291
Mean	139	231,854	5.1	11.1	30.9	38.2	78.3
SE	3	18,446	0.1	0.1	1.2	2.4	1.4
Median	116	40,333	4.8	11.2	21.0	32.0	75

(a) cells/μl (b) copies/ml (c) White Blood Cells 10^9/L (d) g/dl (e) Alanine transaminase IU/L (f) Aspartate transaminasi IU/L (g) umol/L

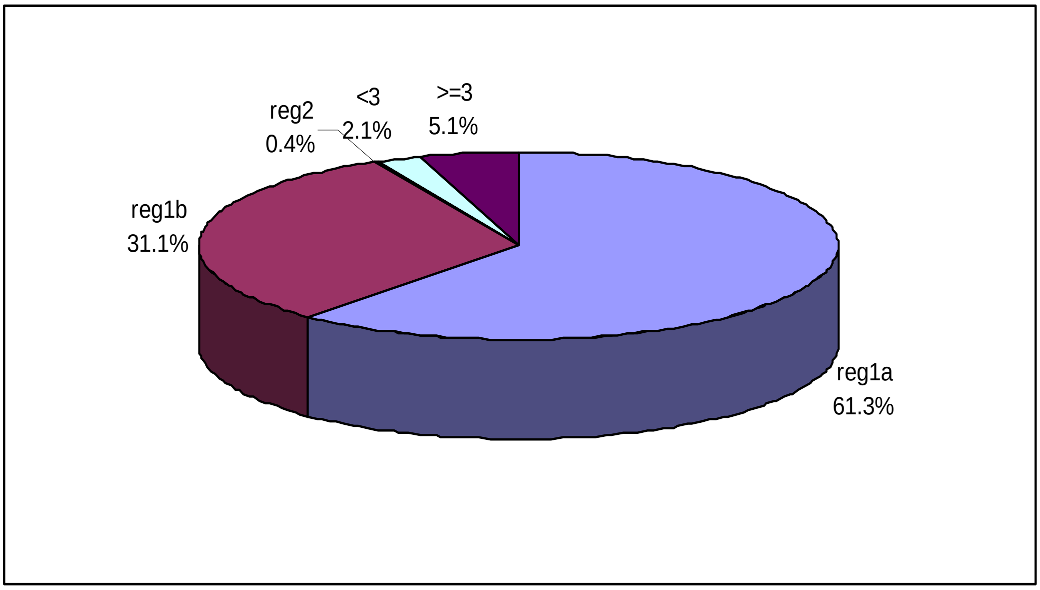
Figure 35 Prevalence of baseline laboratory values



1.3.4.5 Initial Regimens

Figure 36 shows the regimens at the beginning of treatment. About two thirds were assigned to Regimen 1a, about one third was assigned to Regimen 1b, while 7% were assigned to paediatric regimens and less than 1% was assigned to Regimen 2. The very low assignment to regimen 2 was due to the fact that only 6% were already using ART and a few of them had developed resistance to Regimen 1a or 1b.

Figure 36 Regimen assignment



1.3.4.6 *ART Retention*

The first semester after treatment assignment was the most critical for remaining on ART. Figure 37 shows that the proportion remaining on ART declined to 85% by the first semester, 80% by the second semester and 76% by the end of the second year. In comparison, the ART programme in Khayelitsha near Cape Town, reported that 83% of the patients were still on ART after two years (Cleary *et al.* 2006). This higher performance is probably due to the fact that Khayelitsha covers a relatively limited geographic area with a privately managed well-funded programme.

Figure 37 Probability of remaining under ART in the first two years

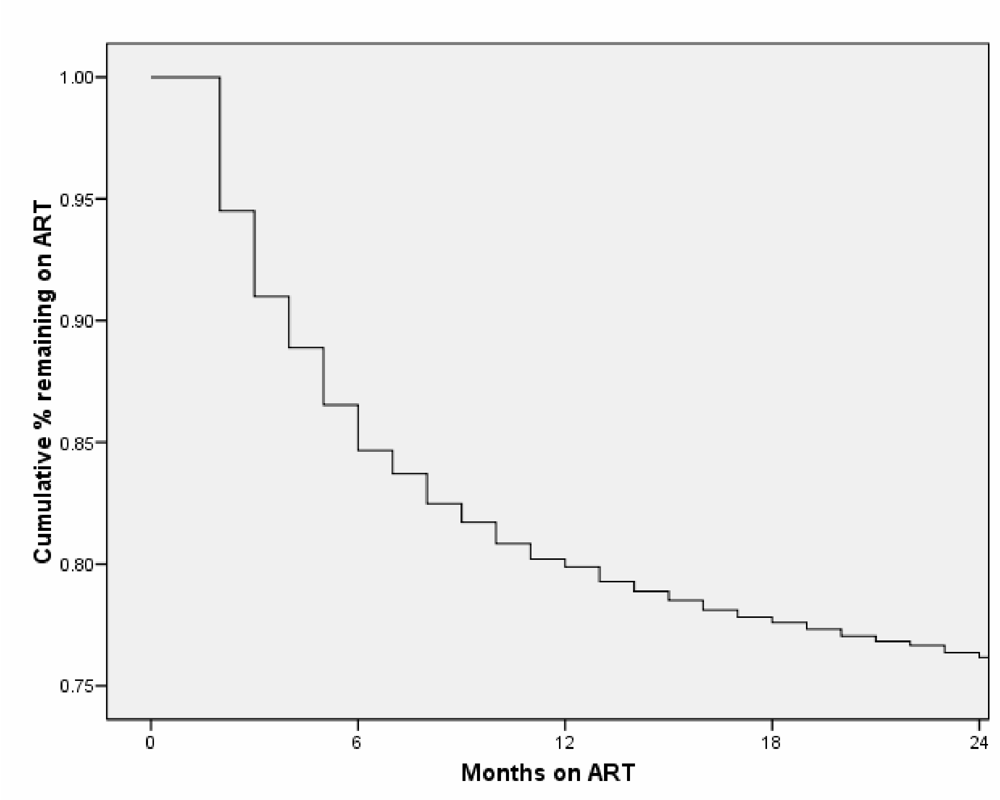
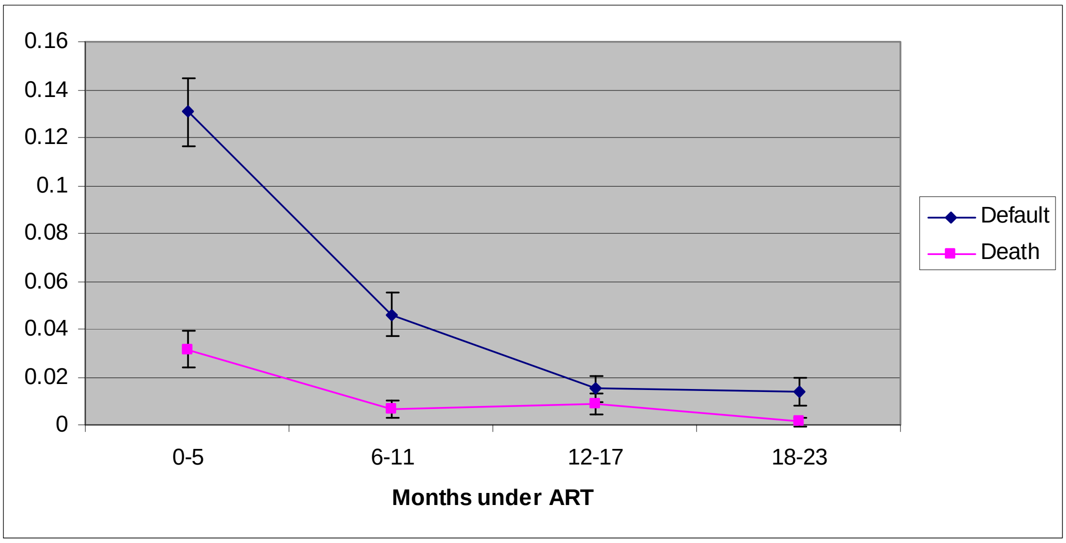


Figure 38 shows that in the first six months 3% were certified as dead and 13% were recorded as having defaulted, but their fate was unknown. This is slightly better than the average mortality of 6% and default rate of 12% estimated to occur in the first semester of ART in developing countries (Braitstein *et al* 2006). The differentiation between default and death should be viewed with caution because most patients who defaulted and whose fate was unknown have probably died. Nonetheless, the differentiation is useful to show that intensive tracing during the first semester could have reduced the discontinuation significantly. Using the person year of observation (PYO) the default rate in the first two years was 25% with a 95% confidence interval of +/- 1.9% while the recorded death rate per PYO was 5.8% +/-1%.

Figure 38 Default and death rates by six-month intervals



Even if under-reported, the above mortality rates are higher compared with other cohort studies carried out in South Africa. A community based ART programme near Cape Town recorded a mortality of 2.5 per PYO during the first year of treatment (Lawn *et al*.2005). Another private-sector programme recorded a crude mortality rate of 3.5% during a median follow up period of 1.8 years (Nachega *et a1*.2006). These studies had superior recording systems that successfully captured the fate of most patients who defaulted. Therefore, it is likely that the real difference in mortality between the ART programme in KZN and these pilots would have been higher if the fate of the majority of patients who were lost to follow up had been known.

Figure 39 shows the Kaplan Meier curves for the patients who started ART at different levels of CD4 cells/µl. The higher and lower lines in Figure 39 represent respectively the proportion of patients who started ART with CD4 100-199 cell/µl and CD4<100 cells/µl and who remained under treatment by the end of each month. The two lines diverge significantly, indicating the persistent lower chance of remaining on treatment for patients starting at CD4 <100 cells/µl.

Starting at low levels of CD4 values had long-term consequences even when patients remained on treatment. Figure 40 shows the average increase in CD4 values by semester for the group which started ART with CD4<100 cells/µl and with CD4 100-199 cells/µl. Although improving, the first group was significantly lagging behind at every semester compared with the second group.

Figure 39 Proportion remaining on ART by initial CD4 values

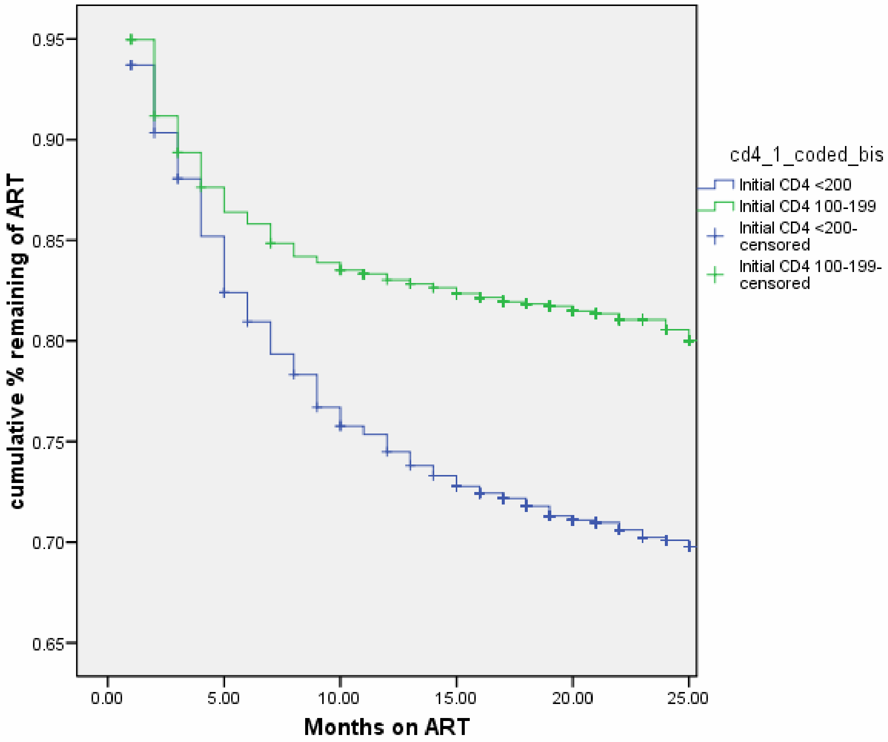
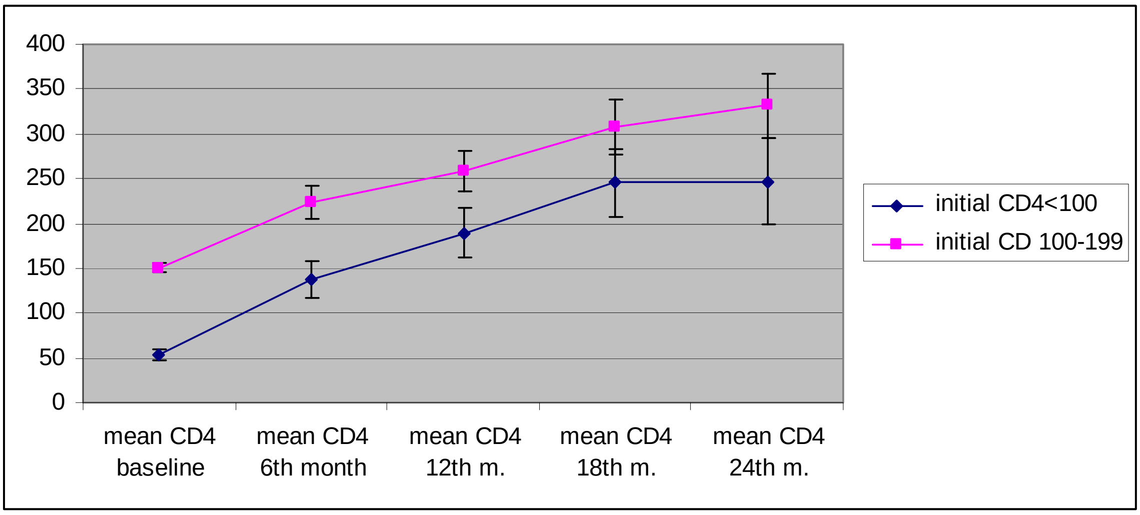


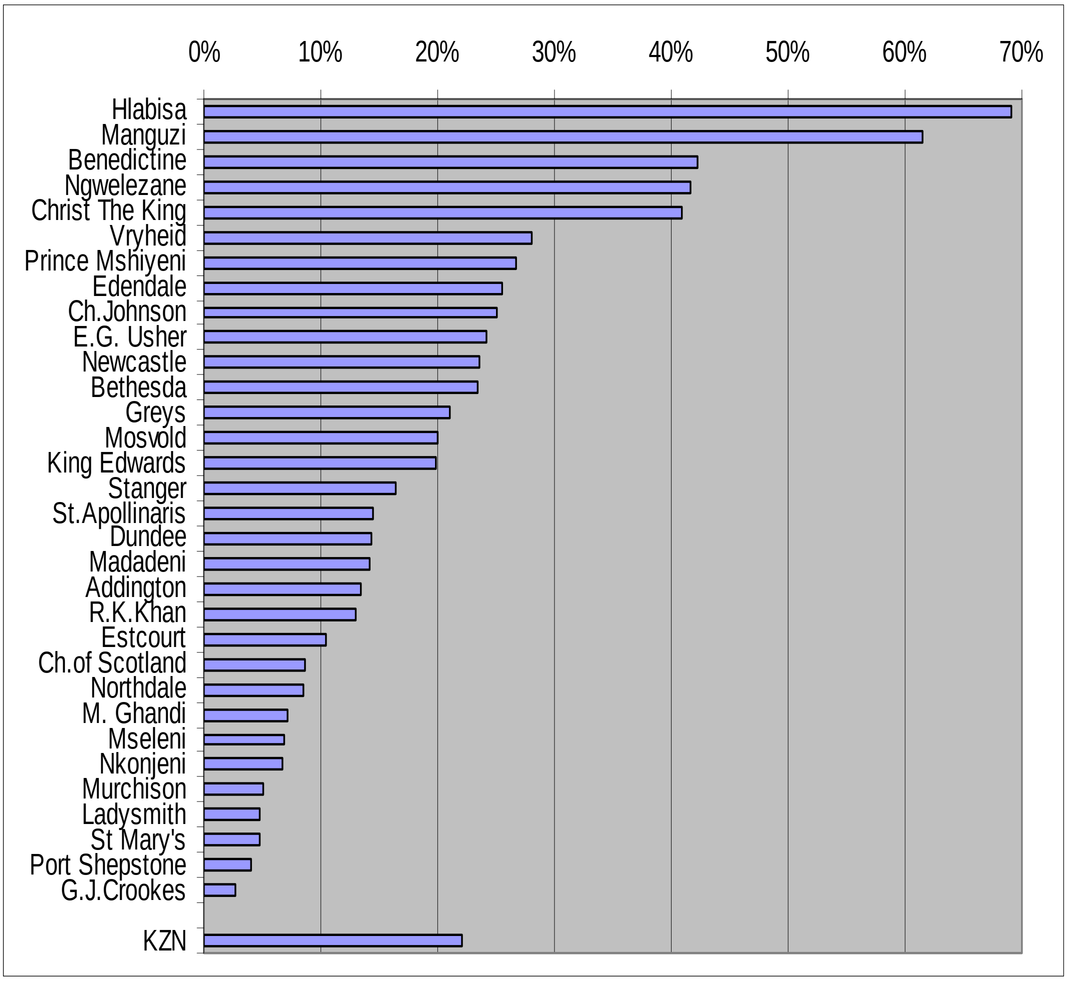
Figure 40 Change in mean CD4 values, according to baseline CD4 values



Variation across delivery sites

Figure 41 shows the proportion of patients who discontinued ART during the first two years across the delivery sites. There was a wide variation between the worst situation of Hlabisa, where the majority of patients were lost within the first two years, and G.J.Crookes, which lost very few patients. The average for KZN masks a high inter-site variation, which is the result of different levels of efficiency across sites. The characteristics behind this variation have been analysed to shed light on the performance of the ART programme.

Figure 41 Proportion discontinuing (death & default) by delivery site



Factors favouring retention

Figure 42 summarizes the relationship between the characteristics of the patient and the delivery site, and the probability of remaining on ART in the first two years. The figure provides in a snapshot the results of several cross-tabulations, with each line representing the proportion of patients with certain characteristics who remained on ART. For example 80% of females versus 73% of males were still on ART at the end of the second year and so on for the other variables.

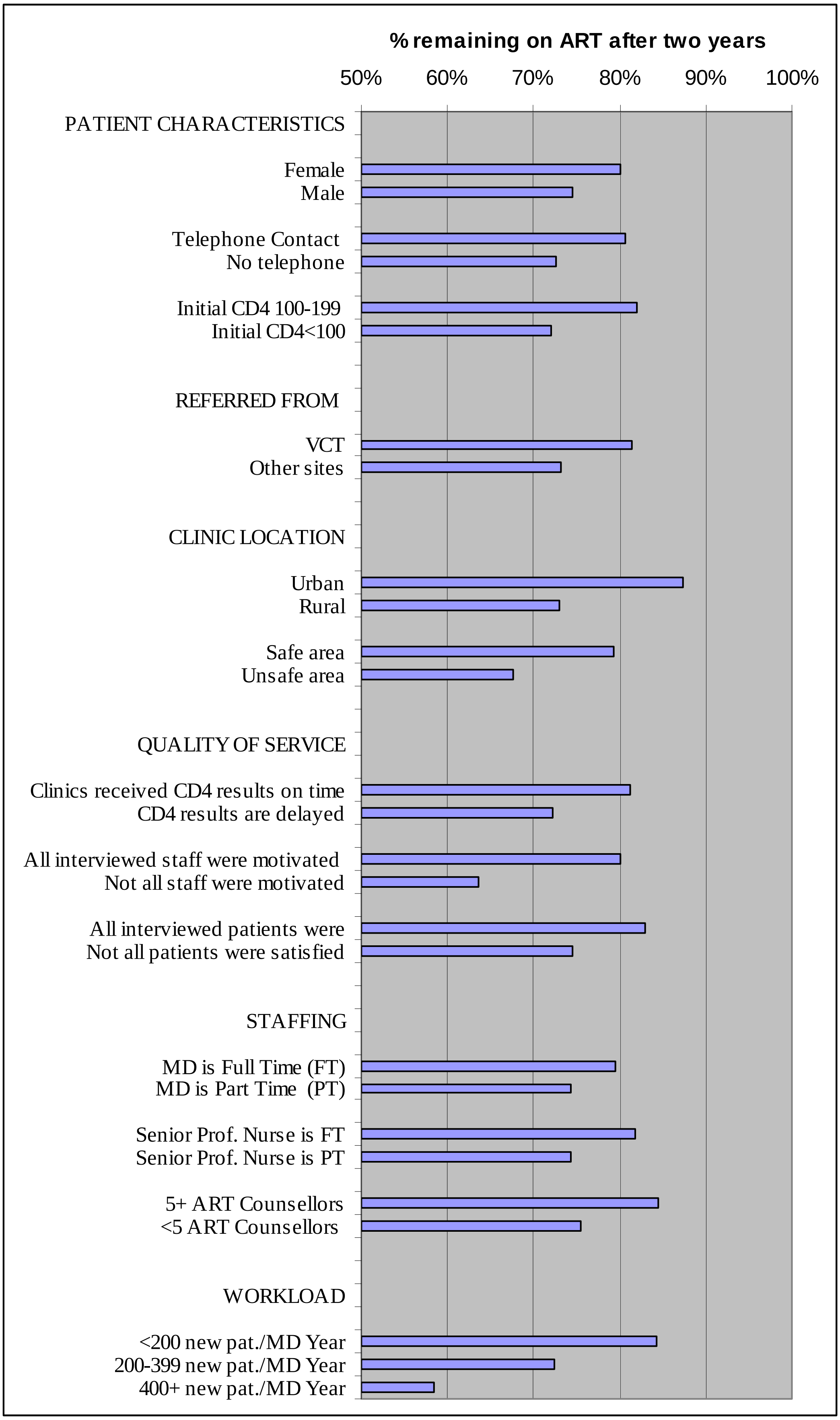
The variables are related to the patient's characteristics, the referral system, the location, the quality of service and the staffing pattern. For each category, the longer the bar the higher the prevalence of remaining on ART for the patients having that category. Favourable patients' characteristics for remaining on ART included being female, having a contact phone number and having started ART with CD4 100-199 cells/μl. Having being referred from Voluntary Counselling and Testing (VCT) instead of a hospital or other services was associated with a higher chance of remaining on ART.

Certain characteristics, which were proxies of access and quality of services, were associated with a higher retention rate. The proxies of access which were associated with ART retention were location of the delivery site in urban and in safe areas[7]. The proxies of quality, which were associated with retention, included the confirmation by the manager that the CD4 results were always received before the patients' next visit and that no staff had been absent during the previous month. There was a higher retention rate in the delivery sites where no staff who was interviewed mentioned that he/she was "not motivated at all". Delivery sites where the majority of the patients who were interviewed declared that they were fully satisfied with the service had also a significantly higher retention. Retention was higher in the delivery sites where all staff mentioned that the quality of information provided to the patients was critical for compliance.

Delivery sites with certain staffing pattern and workload had also a significant higher retention rate. In this context, it has to be kept in mind that the average delivery site had a core staff of one cleaner, one data clerk, one assistant nurse, and one registered nurse. There were also part-time nutritionist/dieticians, pharmacists and social workers. But the only categories of staff which were significantly related to a higher retention rate were the counsellors, the senior professional nurse (SPN) and the medical doctors (MD). Another factor which was related to retention was the workload which was measured as the annual intake of new patients per MD per year. It has to be considered that the SPN and the MD could be full time in the site or part time if they were seconded from the local hospital to spend part of the time in the hospital and part of the time in the ART delivery site. For each delivery site, the workload was estimated by dividing the average number of new patients enrolled in the first two years since the site started rolling out ART by the number of doctors employed in the delivery site during the same period. Although the workload was doctor based, it was a proxy of the overall staff workload.

[7] Security problems were considered present if the managers replied affirmatively to the question "Are there problems with security in the area, which may sometimes prevent patients from coming to the health facility ?"

Figure 42 Proportion with given characteristics remaining on ART

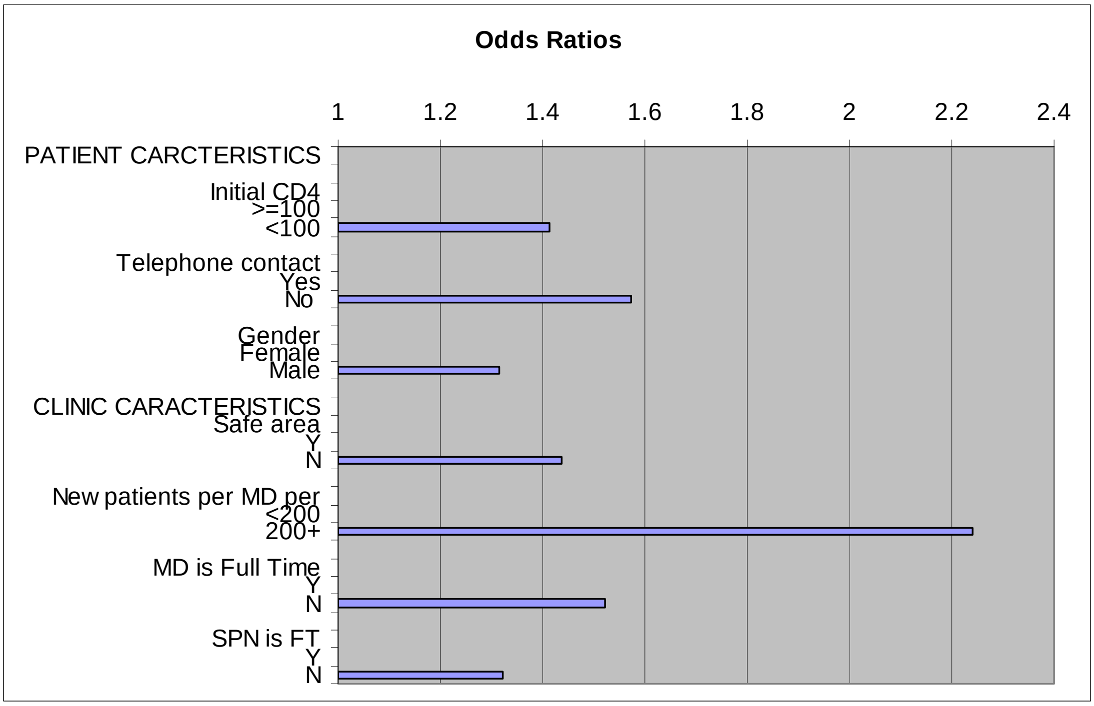


Multivariate analysis

Although the individual variables in Figure 42 were statistically significantly associated ($p<0.05$) with ART retention some of these associations could have been spurious ones. For example, patients with contact telephone numbers might have been socio-economically better off, which might per se have lead to higher retention rates. This potential spurious association can be excluded by disaggregating the patients into different socio-economic groups. This will allow to checking if the association between telephone contact number and retention persists in each socioeconomic group. However, controlling for too many variables (e.g. Figure 42) would have led to such a disaggregation that very few patients would have remained per group. With each extra variable, the sample size of each table would have decreased, reducing the possibility of drawing firm conclusions. For this reason, Cox regression was used to exclude spurious associations.

A Cox regression model was run to check the net effect of each variable on the probabilities of leaving ART within the first two years. The Cox regression model identified the most critical variables, which per se made a significant difference in retaining patients. All the variable were included in an initial model and taken out one by one in backward fashion if their coefficient was not statistically significant ($p<0.05$). The variables in the last model (Annex III) were all highly significant ($p<0.01$) and are represented as odds ratios in Figure 43.

Figure 43 Odds Ratios for ART discontinuation



The odds ratios measure the risk for leaving ART, if certain characteristics are present vs. a baseline. For example Figure 43 shows that patients attending delivery sites with 200 or more new patients per MD per year had more than twice the risk of leaving ART compared to their counterparts attending sites with <200 new patients per MD per year. Other significant factors that independently from the others increased the risk of leaving ART included presence of FT SPN and FT MD and location of the delivery site in a safe area. Being males, not having a telephone contact number and having an initial level of CD4<100 cells/µl increased the risk of discontinuing ART.

Delivery options

The results of the analysis were further transformed into staffing and workload profiles which could be proxies of delivery options/strategies. The odds ratios are helpful in visualizing the risk of discontinuing ART when certain factors are present. Nonetheless, there is a need to transform this information into strategic options that can be used to improve the ART programme. For this reason, it was decided to focus on the factors that could be changed through management strategies. Therefore, the last variables in Figure 43 were used to create the following staffing and workload options for the delivery sites:

(a) Part time (PT) doctor (MD) and PT senior professional nurse (SPN) with less than 200 new patients per doctor per year;
(b) Same as above but with 200 or more new patients per doctor per year;
(c) FT MD and FT SPN with less than 200 new patients per doctor per year; and
(d) Same as above but with 200 or more new patients per doctor per year.

Each patient was categorized according to the above-mentioned profiles/delivery options to check the probability of discontinuing ART which was associated with each profile. Figures 44 and 45 represent the Kaplan Meier curves of the above-mentioned delivery options. The proportion of patients discontinuing ART because of death, default or any other reason is on the Y-axis and the duration of treatment is on the X-axis. Figure 44 shows that with an annual intake of less than 200 new patients per MD per year, the proportion still on ART at the end of the second year is expected to be 88% and 85% respectively with PT (first line) and FT (second line) SPN and MD. In other words, as far as the annual intake of new patients per MD is less than 200, there is no significant difference in retention rates between delivery sites that are staffed with FT senior staff and delivery sites that are staffed with PT senior staff.

The added value of having FT senior staff is when the workload reaches 200 or more new patients per MD per year. Such level of workload is expected to impact negatively on the retention of patients compared with the lighter workload of less than 200 new patients per year. However the presence of FT senior staff will at least significantly attenuate the negative impact compared with the presence of PT senior staff. The delivery sites which have the higher annual workload of 200 or more new patients per MD and are staffed with FT MD and FT SPN are expected to have 70% of patients still on ART at the end of the second year. For the same higher workload, this proportion is expected to decline to 50% in the delivery sites which are staffed with PT SPN and PT MD. The cost-effectiveness of these staffing and workload profiles is discussed in Part 2 of the report.

Figure 44 Workload < 200 new patients per year & ART retention

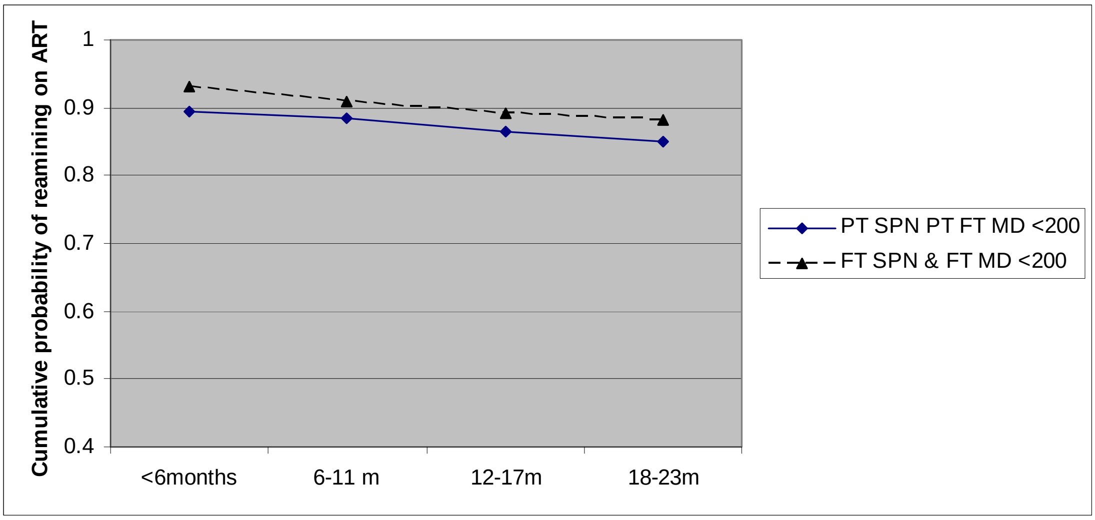
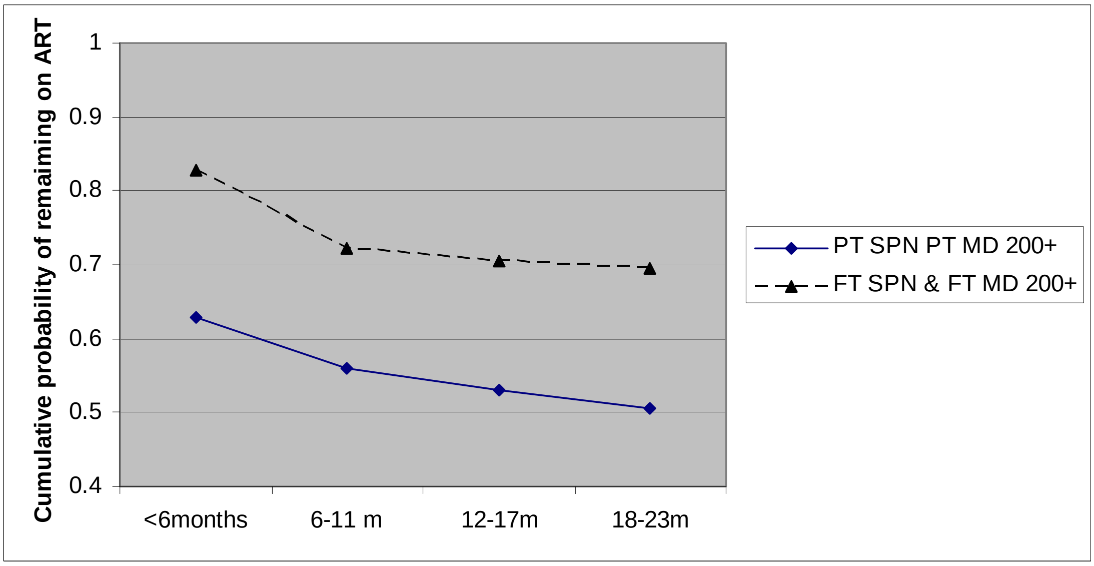


Figure 45 Workload 200+ new patients per year & ART retention



1.3.5 Implications for the expansion of the ART programme

This section focuses on the manpower implications of the findings related to the staffing and workload characterizing the above mentioned delivery options (a) through (d). Although there is pressure to move from the static units to other less facility-base delivery systems, it is critical to see the implications of staff profile and workload on the present static sites. Even if other delivery systems were implemented, they would still have to rely on the support from the static sites. Thus, the staffing and workload options previously mentioned under (a) through (d) will continue to play a primary role and they can be used to estimate human resource needs.

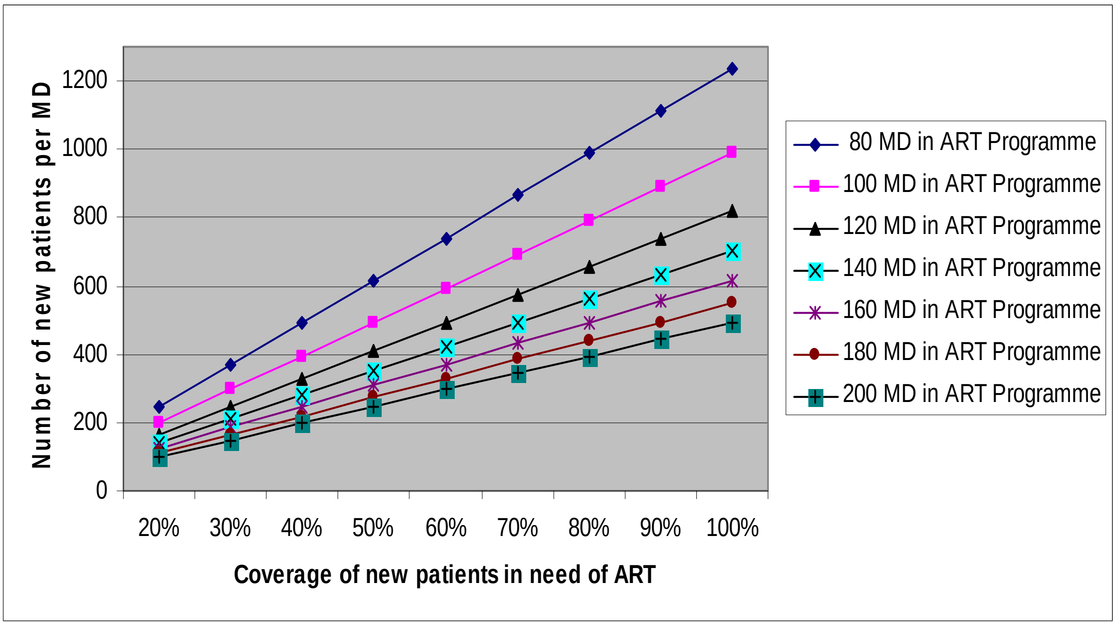
The results of the analysis were used to assess the implications on the staff requirements for the fixed sites. The staffing scenarios are based on each ART site having the following full time (FT) core staff: one enrolled nursing assistance (ENA), one professional nurse (PN), five counselors, one data capturer and one cleaner The part time (PT) core staff includes: one pharmacist, one social worker and one nutritionist/dietitian. The delivery options (a) – (d) will have in common the above mentioned core staff but will differ in respectively FT or PT SPN and MD, and on the annual intake of new patients per MD. PT staff is seconded by the nearby hospital to work in the attached ART sites for an average of 20 hours per week. The number of new patients per doctor per year has been selected as a proxy of the overall absorption capacity of the delivery site on the basis of the retention rates obtained from the analysis. For most sites, which have only one MD, the cut off of <200 and 200 or more new patients can be considered as number of new patients per delivery site per year. However, in case a delivery site had two MDs, the annual cut off points of new patients per site will double to <400 and 400 or more.

Expansion of ART

Planning methods try to strike a balance between what is desirable and what is achievable. This means matching supply with demand, resources with absorption capacity and efficiency with coverage. The 2003 model of the Actuarial Society of South Africa (ASSA03) predicts about 150,000 new AIDS sick patients in 2008. The demographic model gives the most reliable predictions according to the most reliable assumptions and to the knowledge on the status of the epidemic. Of these 150,000 new patients in need of ART, about one third (Lawn *et al* 2005) is likely to die before accessing ART, bringing the demand for the above mentioned period to about 100,000 new patients.

An effective coverage of 100,000 new patients depends on the absorption capacity of the health services. The lines in Figures 46 show the relationship between the number of MDs which are available to the ART programme, the coverage targets (X-axis) and the workload in terms of new patients per MD per year (Y-axis). For example, the 80 ART Unit operating at the moment can count on an average of 80 MDs and relative core staff. Therefore coverage of just 50%, would lead to a workload of about 600 new patients per MD in 2008. This is obtained by drawing a vertical line from the axis X (50%) up to the last line (80 MDs) and checking the corresponding value on the Y-axis (600). As the coverage increases, the increasing number of new patients overloads the 80 MDs, exceeding 1200 new patients per MD at 100% coverage. In other words even if it were possible to achieve full coverage with a staff of 80 MDs, this will be at the cost of an unsustainable burden for the staff.

Figure 46 Relationship between coverage and workload, KZN 2008

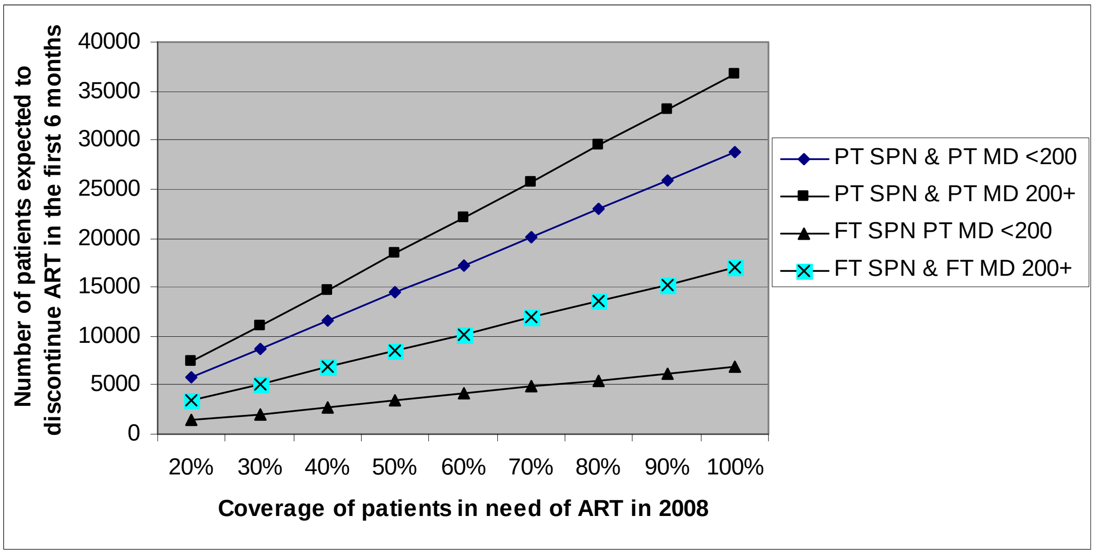


Sustainable coverage targets depend on the relationship between efficiency of the delivery system, supply and demand. If full coverage is the goal, adequate planning should identify how this will be sustained without running the risk of excessive default rates leading to drug resistance. The four lines in Figure 47 represent the predictions of the number of defaulters (Y-axis) expected by level of coverage (X-axis) if delivery options (a)-(d) were implemented in 2008. For example at 50% coverage the number of defaulters expected in the first semester varies between about 3,000 to about 18,000 depending on the delivery option.

The use of Figures 46 and 47 will help the DOH to adjust coverage targets according to the absorption capacity of the system and to plan the human resource needed to achieve certain targets. The new patients per MDs were used in this analysis as a proxy of the general staff workloads. Therefore, Figures 46 and 47 will be used to predict the rate of increase of the MDs and SPN required for the four delivery options but also the rate of increase for the staff.

This allows to estimate how excessive the workload is according to the target and what is the personnel gap to be filled to decrease the workload and reduce the number of defaulters. As previously mentioned, if the ART programme can count on 80 MDs, a target of 50% is expected to lead to 600 new patients per MD which is associated with 3000-18000 defaulters according to the delivery option. To reduce such workload to the more favourable level of less than 200 new patients per MD, the present staff profile will have to be increased by three times to maintain a sustainable coverage of 50%. If the gap cannot be covered through recruitment, the alternative is to lower the coverage targets or to face the consequences of the increasing number of defaulters shown in Figure 47 In other words planners should not limit themselves to setting targets but be more specific on how to achieve and then maintain them.

Figure 47 Coverage and expected discontinuations by delivery option



Achieving full coverage without hampering the efficiency of the delivery sites would require an unrealistic high number of staff. If full coverage were to be achieved without exceeding the cut off of 200 new patients per site, the staff will have to increase by six times to a sustainable full coverage. Because this is unlikely to be possible, there are two options: (a) keep full coverage as a target, whatever the capacity of the system, but expect an unsustainable workload and an increasing number of defaulters; or (b) adopt targets, according to the capacity of the system to sustain a realistic workload, with a higher chance of keeping the coverage once this is reached. Because option (b) is more sustainable than option (a), any increase in coverage must be realistically set against presently available staff and sustainable levels of workload.

Another clarification to be made is the role of the static units versus a less facility based delivery system. Whatever alternative system is selected, this will have to still rely on the support from the static units. Because of the high default rate characterizing the first six months, it is likely that any new patient will have to be catered for by high-level staff at least during the first semester. It is after the patient gets stabilized that less resource intensive systems will be able to take an effective role, integrating themselves into the existing delivery system. There is a need to investigate different delivery models in which static and outreach systems are able to build on each other strengths. A balanced approach should always take into account strengths as well as weaknesses that any alternative strategy is bound to have, limiting the tendency to privilege only one or the other system.

1.4 Discussion

The evaluation has provided the first comprehensive picture of the status of the ART programme in the delivery sites which began distributing ART in 2004. The literature has provided the experience of well-funded projects in South Africa (Cleary et al 2006) but the information on more routine services has been lacking. For the first time, the ART programme of KZN has been described in terms of management characteristics, coverage and effectiveness in retaining patients. Although a high proportion of the patients who were interviewed expressed a high level of satisfaction, there were some problems. The interview with the managers, the staff and the patients confirmed an excessive number of patients. This resulted in overcrowded waiting areas, shortage of space and high workload.

The task analysis enabled the observation of the interaction between patients and staff during the registration, baseline clinical examination, counselling and follow up visits. The counselling was conducted in groups of 3-39 patients and the participation, in terms of patients asking questions, was higher when the size of the group was less than 10 patients. While the recording of the registration form was high, less than half of the staff recorded the baseline clinical examination form and very few filled the follow up visit form.

The representative sample of 2835 records provided the information on how many patients left ART. Only 20% of those who discontinued were confirmed dead while the fate of the rest was unknown. Considering the high mortality experienced by patients discontinuing ART, it is likely that most of the patients who defaulted died. The limited information on record precluded a survival analysis; hence, the proportion of patients who remained on ART in the first two years was used to estimate the effectiveness of the delivery system.

The 76% retention rate within the first two years put the performance of the ART programme above average for Sub-Saharan Africa. Rosen *et al.* did a systematic review on ART programme in Sub-Saharan Africa and found that the average programme in Africa retained 60% of the patients in the first two years. The higher retention rate of the ART programme in KZN reflects the more advanced status of the delivery system of KZN compared to other African countries.

The implications of the findings are discussed according to the following sections: integration with other programmes, management information system, tracing of defaulters, infrastructure, quality of services, human resources, problems and potential solutions.

Integration

The ART programme is run vertically with little integration with other programmes. The lack of integration with the TB programme is counterproductive because TB and HIV have a high co-infection rate and thus should be dealt together, possibly by the same programme. Nonetheless, there is a need to carefully study what such integration implies in terms of logistics and workload. The feasibility of further integrating ART within the mainstream PHC activities should be even more carefully evaluated. This includes management models to assess the consequences that integrating too many services will have on the efficiency and cost-effectiveness of a complex therapy like ART.

Information system

The management information system is not functional given the low priority that staff put on information gathering. Paper modules/forms which were introduced since 2004 have been described in the introduction. They were designed to capture the information during the scheduled visits related to registration, baseline clinical examination, follow up visits and exit from the programme due to interruption, default, transfer and death. However, most staff did not see the need for recording even the most important information on the laboratory results. It was easier to put the laboratory slips directly into the patient folder where they could be retrieved when needed, rather then spending time to write the results on a form.

This shows the differences in information priorities between front line staff and managers. The lack of enthusiasm that the former category has for filling forms is akin to the low interest that most people have for filling questionnaires. This contrasts sharply with the needs of managers who require data recording and processing to produce aggregate numbers.

The situation does not bode well for the new forms that are expected to replace the old ones. Whatever the design of the new forms, recording will continue to be felt as a bureaucratic exercise by front line staff, especially if no staff is dedicated full time to the data gathering. One solution is to introduce an itinerant team which will visit each delivery site on a rotating basis. This team would train and supervise the staff at each site, to ensure that data are entered and used to trace defaulters and to facilitate the data flow between the sites and the central level. This support and feedback will reinforce the importance placed on information and the need for consistent and accurate recording and reporting.

Defaulters Tracing

Most units did not have an effective system to update the number of the patients who defaulted and needed to be traced. The key patients' data should be entered into a database (electronic register), which should be programmed to flag the patients who are at high risk of defaulting according to the presence of the risk factors which were identified by the evaluation (e.g. male, CD4<100 cells/μl). The electronic register will be programmed to assign the dates for the scheduled visits and to produce a daily list of patients who did not come as scheduled.

Most ART sites did not actively trace the defaulters. Although almost all the delivery sites had access to transport, the defaulters were usually contacted by phone, with very few ART sites conducting household visits. Telephone contact numbers and addresses were frequently missing, thus limiting the chances of locating the defaulters. Although about half of the delivery sites had CHWs in their area, only one quarter of the delivery sites involved the CHWs in tracing the defaulters. Each site should have a dedicated tracing team that besides conducting household tracing will involve the CHWs in taking care of the patients who are more at risk for default according to the presence of risk factors.

Infrastructure

One of the most significant risk factors for ART discontinuation was the mismatch between supply and demand. Sites were overcrowded because of limited waiting areas and insufficient number of chairs, with the patients frequently waiting outside. The solution would be an upgrade of the infrastructure of the existing delivery sites to add more space and the planning of new delivery sites with larger waiting areas and more consultation rooms.

Quality of service

Although the proxies for the quality of service were measured at the time of the survey, they nonetheless reflected the variation in quality across delivery sites and helped to explain variation in ART retention. Delivery sites had a lower retention rate if (a) managers confirmed that the results of the CD4 were not ready by the time the patients were coming back after the baseline visit, (b) there was at least one person among the staff who was interviewed who declared a complete lack of motivation to do his/her job, and (c) at least one patient, among those who were interviewed, found that the ART service was not "fully satisfactory".

Human resources

The further expansion of the ART programme depends on matching supply with demand. The staffing profile and the workload of each delivery site had a significant relationship with the retention rate. The management model used in this analysis has shown the potential consequence of aiming at full coverage without planning on how to maintain it. Planners should use this model to forecast what is likely to happen when available resources are overwhelmed.

Problems

The Cox regression allowed one to identify key individual risk factors that remained statistically significant after controlling for other variables. Patients without a recorded telephone contact number were 1.6 times more at risk of leaving ART during the first two years. The problem could be related to carelessness in recording or to the patients' refusal to provide a contact number because of stigma.

Slightly less than half of the patients began ART with CD4<100 cells/µl which was a major risk factor for discontinuation and recovery of immune functions. These patients had 1.4 times higher risk of discontinuing ART, probably because of the higher mortality that has been reported among patients starting ART with such low CD4 values (Nachega *et al* 2006). Even when remaining on ART, the average increase in CD4 cells/µl was significantly lower than the gains experienced by the group who started ART between 100-199 CD4 cells/µl.

Although many researchers emphasize that ART should begin between 200-350 CD4 cells/µl, many patients do not even begin with a CD4 of 100-199 cells/µl. While this debate is important at a policy level, there is also a need to focus on how to ensure that patients are tested early enough to start ART when their CD4 values have not declined under 100 cells/µl. The solution depends on the difficultly in tackling the many factors impeding access, including the stigma surrounding HIV and AIDS.

Males defaulted significantly more than females. The odds ratios of 1.3 obtained from the Cox regression translate in about 30% higher discontinuation among males compared to females, which is similar to the Kenyan experience (Ochieng *et a*l 2008). Also in this case, behaviour is likely to play a causal role, and as in any behavioural factor, policy directions on what to do are not very specific. More documentation is needed on the effectiveness of specific strategies aimed at changing longstanding cultural values which are at the basic roots of behavior.

Potential solutions

The delivery options explain the variation in retention across delivery sites and allow one to distill lessons from what has worked. The message is that some delivery sites have been working within their own absorption capacity and others have exceeded it with subsequent decline in their effectiveness. This allows one to focus on the staffing profile and levels of workload that had a high potential of maintaining patients on ART and that could be implemented. The four delivery options can make a difference at parity of other risk factors and they provide a practical way of increasing retention.

The estimates of how many patients would leave ART according to the delivery options should raise the awareness of the challenges lying ahead. Most delivery sites are experiencing a high workload of more than 200 new patients per doctor per year at present coverage levels. Reducing the workload to more sustainable levels would require opening new centres, recruiting more core and senior staff, implementing more training and supervision, improving the management information system and integrating alternative delivery systems.

Coverage targets should be established after strategies are set up on how to recruit sufficient staff to reduce the annual workload under 200 new patients per delivery site. This is better than applying a theoretical coverage without the necessary assumptions on how to maintain it. The criteria proposed in this analysis are derived from a multivariate model, which was based on what happened in the first two years of the programme. It thus provides a better perspective of the relationship between efficiency and effectiveness, and of the risks incurred if certain targets are too optimistic or certain conditions are not met.

The cost-effectiveness of ART can vary according to local conditions, coverage targets, staff availability and mixes of critical inputs influencing quality and efficiency of services. Resources are unlikely to be sufficient to achieve full coverage by implementing the best delivery option. Nonetheless it is better to use previous experience to predict what will be the maximum workload that the staff is able to sustain. It will be up to the DOH to strike a balance between desirable coverage and feasibility. Each target needs to be matched with the available staff to estimate the maximum number of patients that the system can afford according to the predictions outlined in the model used by this analysis.
An area requiring strengthening is the tracing during the riskier period of the first semester after the initiation of ART. Most of the patients discontinued ART during the first six months, suggesting that increasing tracing efforts during the first semester has the highest potential for increasing ART retention. The introduction of an electronic patient register, the registration of the telephone contract number and address, the increased use of household visits and the recording of the GPS coordinates, and the involvement of CHWs should increase the effectiveness of tracing.

The previously mentioned electronic register will require an intensive supervision which can be achieved through an itinerant team. Information systems which rely solely on the contribution of front line staff are bound to fail without a strong back up support system. There are several reasons why most management information systems (MIS) do not work:

(a) They are based on the assumption that health staff has the same enthusiasm and understanding for the data gathering as those who designed the MIS;
(b) There is more emphasis on the informatics aspects than on the analysis of the data and the use of the analytical results; and
(c) There are no financial or logistic reasons to collect the information.

Issue (a) and (b) cannot be solved without first tackling (c). Whatever the informatics architecture, any MIS can only gather information if the staff has compelling financial or logistic reasons to fill the information. An example of a financial reason is the need for hospitals in many developed countries to provide the discharge diagnoses to the central level to get reimbursed. Because the hospitals in KZN are paid on the basis of the number of beds, their managers do not have any financial reason to produce statistics on the diagnoses of discharge. It is for this reason that any attempt to process the diagnoses of discharge in KZN has not been successful and the only available statistics of the causes of discharge was from one survey (Epidemiology Bulletin 6, 2004).

An example of a logistic reason for the MIS is the update of the stock inventories (e.g. pharmacy) which is done because the staff needs to get replenishments. The clear link between information gathering and the daily activities compel the staff compelled to fill the information. However, in the case of recording the laboratory tests on a patient's form, nurses and doctors will not be affected if they simply put the laboratory slips into the patient's folders without entering the results on the patient's form. Any MIS which is based only on theoretical guidelines or methodological motivations is hard to implement. It is only when non compliance with recording criteria results in an impediment to routine functioning that staff will be penalized for not filling the records.

An incentive which could make the MIS work is the intensive support coming from an itinerant team visiting the delivery sites on a regular basis. This team will train and supervise the staff to enter the most critical information into the electronic patient register. Having visited the last delivery site, the team would revisit the first delivery site and would check the information which is entered into the electronic register. At each visit, the rotating team will coach the nurses, discussing the problems and showing the advantages of entering the information. This would lead to better understanding on why and how the use of the patient register helps to trace the defaulters. At the end of each round the rotating team will check the data and will forward the cleaned files to the provincial level. Here, the Monitoring & Evaluation officer will update key monthly indicators, analysing the reasons behind the weakness and strengths of the ART programme.

1.5 Conclusions

This evaluation has identified several conditions to be met to improve the effectiveness of ART in routine settings in KZN. The pharmaceutical companies focus on estimating the efficacy of drugs through trials but they do not quantify the factors that were present in these privileged conditions. Evidence based medicine (EBM) use the results from trials to suggest ‘what can work’ but it is less specific about ‘how’ the efficacy could be transferred in routine settings (effectiveness). This analysis has identified the factors that were significantly associated with patients’ retention in the ART programme in KZN and that can be used to improve the delivery of ART.

It is important to recognize that the services are only part of the equation of ART effectiveness. Stigma, lack of security, distance from the delivery site and cost of transport impaired access to the services. Socio-economic conditions, gender, initial low levels of CD4 and availability of telephone contacts were critical in increasing ART retention. The cost-effectiveness of ART depended on the interaction between many factors that were not under the control of the health services. Health services have no influence on socio-economic factors and security, and have limited impact on stigma and behaviour, which depend on the slow changes in cultural values. Therefore, it is important to keep in mind that the ART programme like any other health delivery system can deal with only part of the problem, which is to provide access to an effective treatment.

The ART programme needs to prioritise the attention to the risk factors which were found significantly associated with treatment retention and which are within the scope and mandate of the programme. These include improving the infrastructure and the quality of service, and keeping the staffing profile and the workload at acceptable levels. ART retention increased in those delivery sites which: (a) were more accessible, (b) recorded the patient telephone contact number, (c) had FT SPN and FT MD, and (e) recruited less than 200 new patients per MD per year. As discussed in Part 2, the above mentioned staff and workload profiles were critical in enhancing efficiency and cost-effectiveness, and therefore should be taken into account when planning the expansion of ART.

1.6 Recommendations

Implement an electronic register. A database containing the individual information of the patients and their schedule of appointments is needed. The register should be programmed to produce a daily list of patients who did not come as scheduled.

Monitor the patients at greater risk for default, especially during the first six months of treatment. The patients who are ranked at higher risk for default according to the presence of the risk factors identified in this analysis should receive enhanced counselling. Monitoring these patients more carefully during the first semester should help decrease the number of patients who are lost to follow up.

Enhance tracing. The staff should ensure that all the patients provide a telephone contact number and an address at the time of registration. If these are not available, the tracing team should accompany the patient home and take the GPS coordinates for future tracing. A tracing team should visit the defaulters after a first attempt to contact them by phone and should assign them to the local CHWs to reduce the risk for future default.

Conduct an audit on the staff and workload profile of the delivery sites. The ART programme does not have updated figures of its staff and it lacks the profile of the sites which were not covered by this evaluation. The audit will update the information on the categories and number of staff and on the status of the infrastructure.

Rank the delivery sites according to delivery profiles. The audit will provide the information to rank the delivery sites according to the delivery options used in this analysis.

Identify strategies to increase efficiency. The information on the personnel and the workload will allow checking if the system is overstretched and to quantify the personnel gaps. The frequency of other factors (e.g. unsafe areas, delivery sites with < 3 consultancy rooms, overcrowded waiting areas) will provide the basis to quantify other gaps and to come up with a plan on how to fill them.

Establish a rotating MIS team. This team should visit each unit to provide training and supervision, ensure the updating of the patient register, identify inconsistent data and correct them, and provide cleaned electronic files to the monitoring & evaluation (M&E) officer at the central level.

Monitor the programme indicators. The M&E officer will monitor the trends in the prevalence of the risk factors (e.g. initial CD4<100) to confirm that the strategies to reduce them are working.

Produce a plan of action for the integration of TB and ART. Integrating vertical programmes can be associated with a decline in efficiency because of an increase in management complexity. Management models based on risk analysis should help predict the human resource needs and the logistic implications for an effective integration. Once an integration plan is implemented and evaluated, further integration with other PHC services should be taken into account.

1.7 Bibliography

AIDS Epidemic Update report. UNAIDS 2007.

Antenatal HIV and Syphilis seroprevalence survey 2006. National Department of Health, South Africa. 2007.

Badri M, Bekker LG, Orrell C, Pitt J, Cilliers F, Wood R. Initiating highly active antiretroviral therapy in sub-Saharan Africa: an assessment of the revised World Health Organization scaling-up guidelines. AIDS 18: 1159–1168; 2004.

Badri M, Maartens G, Mandalia S, Bekker LG, Penrod JR, Platt RW, Wood R, Beck EJ. Cost-effectiveness of highly active antiretroviral therapy in South Africa. PLoS Med 3(1): e4; 2006.

Bisson GP, Gaolathe T, Gross R, Rollins C, Bellamy S, Mogorosi M, Avalos A, Friedman H, Dickinson D, Frank I, Ndwapi N. Overestimates of survival after ART: implications for global scale-up efforts. PLoS ONE 3(3): e1725; 2008.

Braitstein P, Brinkhof MW, Dabis F, Schechter M, Boulle A, Miotti P, Wood R, Laurent C, Sprinz E, Seyler C, Bangsberg DR, Balestre E, Sterne JA, May M, Egger M; Antiretroviral Therapy in Lower Income Countries (ART-LINC) Collaboration; ART Cohort Collaboration (ART-CC) groups. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. Lancet 367: 817-824; 2006.

Chuang JH, Hripcsak G, Heitjan DF.Design and analysis of controlled trials in naturally clustered environments: implications for medical informatics. J Am Med Inform Assoc. 9: 230-238, 2002.

Cleary SM, McIntyre D, Boulle AM. The cost-effectiveness of antiretroviral treatment in Khayelitsha, South Africa--a primary data analysis. Cost Eff Resour Alloc.4:.20; 2006

Corey DM, Kim HW, Salazar R, Illescas R, Villena J, Gutierrez L, Sanchez J, Tabet SR. Brief report: effectiveness of combination antiretroviral therapy on survival and opportunistic infections in a developing world setting: an observational cohort study. J Acquir Immune Defic Syndr 44: 451-455; 2007

Dalal RP, Macphail C, Mqhayi M, Wing J, Feldman C, Chersich MF, Venter WD. Characteristics and outcomes of adult patients lost to follow-up at an antiretroviral treatment clinic in Johannesburg, South Africa. J Acquir Immune Defic Syndr 47: 101-107; 2008

Egger M, May M, Chene G, Phillips AN, Ledergerber B, Dabis F, et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. Lancet 360: 119–129; 2002.

Epidemiology Bulletin, Issue 6. An analysis of the hospital admissions in KwaZulu-Natal between 1998 and 2002. KZN DOH March 2004.

Epidemiologic Bulletin, Issue 10. Weekly Monitoring System of the Antiretroviral therapy for HIV/AIDS in KwaZulu-Natal. KZN DOH June 2005.

HIV & AIDS and STI Strategic Plan for South Africa 2007-2011. National Department of Health South Africa. 2007.

Hogg RS, Heath KV, Yip B, Craib KJ, O'Shaughnessy MV, Schechter MT & Montaner JS. Improved survival among HIV-infected individuals following initiation of antiretroviral therapy. JAMA 279: 450–454; 1998.

Jelsma1 J and Ferguson G. The determinants of self-reported health-related quality of life in a culturally and socially diverse South African community. Bull World Health Organization 82: 206-212; 2004.

Jelsma J et al. Preferences of urban Zimbabweans for health and life lived at different ages. Bull World Health Organization 80: 204-209; 2002.

Jelsma J et al. The reliability and validity of the Xhosa version of the EQ-5D.
Disability and rehabilitation 26: 103–108; 2004.

Jordan R, Gold L, Cummins C, Hyde C. Systematic review and meta-analysis of evidence for increasing numbers of drugs in antiretroviral combination therapy. BMJ 324: 757; 2002.

Krabbe P, Weijnen T: Guidelines for analysing and reporting EQ-5D outcomes. In The Measurement and Valuation of Health Status using EQ-5D: A European Perspective. Edited by: Brooks R, Rabin R and de Charro F. Dordrecht, Kluwer Academic Publishers. 2003.

Initial steps to developing the World Health Organization's Quality of Life Instrument (WHOQOL) module for international assessment in HIV/AIDS. The WHOQOL HIV Group World Health Organization, Geneva, Switzerland. AIDS Care 15: 347-357; 2003.

Interim WHO Clinical Staging of HIV/AIDS and HIV/AIDS case definitions for surveillance. Africa Region. WHO/HIV/2005.02

Lawn SD, Myer L, Orrell C, Bekker LG, Wood R. Early mortality among adults accessing a community-based antiretroviral service in South Africa: implications for programme design. AIDS 19: 2141-2148; 2005.

Lemeshow S, Robinson D. Surveys to measure programme coverage and impact: a review of the methodology used by the expanded programme on immunization. World Health Stat Q 38: 65-75; 1985.

Ledergerber B, Egger M, Erard V, Weber R, Hirschel B, Furrer H, Battegay M, Vernazza P, Bernasconi E, Opravil M, Kaufmann D, Sudre P, Francioli P, Telenti A. AIDS-related opportunistic illnesses occurring after initiation of potent antiretroviral therapy: the Swiss HIV Cohort Study. JAMA 282: 2220-2226; 1999.

Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. Ann Intern Med 135:112–23; 2001.

Moore RD, Chaisson RE. Natural history of HIV infection in the era of combination antiretroviral therapy. AIDS 13: 1933–1942; 1999.

Nachega JB, Hislop M, Dowdy DW, Lo M, Omer SB, Regensberg L, Chaisson RE, Maartens G. Adherence to highly active antiretroviral therapy assessed by pharmacy claims predicts survival in HIV-infected South African adults. J Acquir Immune Defic Syndr 43: 78-84; 2006.

Nathan G, Nattrass N & Raubenheimer C. The cost of HIV prevention and treatment interventions in South Africa. Centre for Social Science Research, UCT 2003.

National Antiretroviral Treatment Guideline. National Department of Health South Africa. 2004.

Ochieng V., Ochieng D., Sidle J., Holdsworth M., Siika A., Yiannoutsos C., Nyandiko W., Owiti M., Kimaiyo S., Braitstein P. Gender and loss-to follow-up (LTFU) from a large HIV treatment program in Western Kenya. XVII International AIDS Conference, Mexico City. Abstract TUAB0202. 2008

Operational plan for comprehensive HIV and AIDS care, management and treatment for South Africa. National Department of Health South Africa. November 2003

Palella FJ, Jr., Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, Aschman DJ, Holmberg SD. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators [see comments]. N Engl J Med 338:853–860;1998.

Rassool GH: Health professionals applaud the WHO 3 by 5 initiative to bring treatment to AIDS sufferers. J Adv Nurs 46: 569; 2004

Rosen S, Fox MP, Gill CJ. Patient retention in antiretroviral therapy programs in sub-Saharan Africa: a systematic review. PLoS Med 4(10): e298; 2007

Taylor K, DeYoung P: WHO's 3-by-5 target. Lancet 362: 918; 2003

Tawfik Y, Kinoti S, Blain GC. Introducing Antiretroviral Therapy (ART) on a large scale: hope and caution. AED November 2002.

Wood E, Yip B, Hogg RS, Sherlock CH, Jahnke N, Harrigan RP, O'Shaughnessy MV,Montaner JS. Full suppression of viral load is needed to achieve an optimal CD4 cell count response among patients on triple drug antiretroviral therapy. AIDS. 14: 1955-1960; 2000.

Part 2

Economic evaluation of the antiretroviral therapy in KwaZulu-Natal

2.1 Introduction

Even in the case of an effective therapy, the results depend on the resources available and how they are used. In developed countries, ART has transformed HIV/AIDS in a chronic disease because there are sufficient human and financial resources to effectively follow up each patient. Although South Africa has fewer resources compared to more advanced countries, these resources are higher than in less developed ones and if they are used efficiently, they could significantly improve the survival of ART patients.

One limitation of available research on ART is the scarce knowledge about the cost-effectiveness of different delivery options. While international agencies like WHO and UNAIDS advocate ambitious coverage targets, services are struggling to cope with the existing coverage. The objective of this analysis was to reply to questions such as “how are existing ART sites coping ?”, “are there different delivery strategies being put in place, at what cost and with what effect ?”. Answering these questions is instrumental to tackle the issue of the absorption capacity of the delivery system and to prevent delivery failure when resources are overstretched.

As described in Part 1, the 32 government sites which started providing ART in 2004 were evaluated. This included a cross-sectional survey to collect information on the delivery sites and a retrospective analysis of a representative sample of 2835 records of patients. After two years, about 76% of patients were still on ART and several risk factors for leaving treatment were identified. The second major finding was that at parity of other risk factors, the probabilities of leaving ART were significantly influenced by the delivery site having the following profiles:

(a) Part time (PT) doctor (MD) and PT senior professional nurse (SPN) with less than 200 new patients per doctor per year;
(b) Same staff as above but with 200 or more new patients per doctor per year;
(c) Full Time (FT) MD and FT SPN with less than 200 new patients per doctor per year; and
(d) Same staff as above but with 200 or more new patients per doctor per year.

Part 2 deals with the economic evaluation of the ART sites which were described in Part 1. This includes the estimation of the running costs of the 32 delivery sites, which were surveyed, and the cost effectiveness of the above-mentioned delivery options (a) through (d). To ensure that the estimates were robust, uncertainty was taken into account through Monte Carlo simulation which was run to predict the cost-effectiveness of following up 1000 patients over a 10 year period.

2.2 Methodology

This is a cost-effectiveness analysis of the ART programme from the provider's perspective and it is limited to the 32 ART sites which started providing treatment in 2004. The analysis was based on the data collected from the 32 delivery sites and from the representative sample of 2835 patients' records, as described in Part 1. This section deals with the data collection related to the estimation of the costs and with the structure of the simulation.

Cost estimates

A survey team collected the expenditures sustained by the 32 ART delivery sites in the financial year 2005/06 (FY05/06)[8]. These did not include the cost of other health services such as the cost of hospitalisation, and the personal costs born by the patient to reach the delivery sites. Data were collected according to commonly used methodology (Annex V). The registers of the delivery sites were used to get the number and type of staff employed by the site and the number of hours worked per week, the consumption of ART and non-ART drugs, the laboratory tests carried out and the supplies which were consumed between the 1st April 2005 and 31st March 2006.

All the units of consumption (e.g. personnel hours) were multiplied by the unit cost provided by the Department of Health of KZN (Annex VI). For example, the personnel cost was estimated by multiplying the number and type of full time staff by the relative gross salaries. The cost for the part time staff was estimated by multiplying the number of hours worked in FY05/06 by each type of staff by the cost per hour paid by the DOH for each specific category of staff. The ART and non-ART drugs and the laboratory investigations were multiplied by the unit cost. The annual cost of the equipment was obtained by multiplying the cost of each type of equipment found in the delivery site by its replacement value (unit costs/ lifetime in years). The costs of the utilities (water, electricity, telephone and waste disposal) were estimated by multiplying the total cost of the utilities of the hospital, to which the delivery site was attached, by a weighing factor (squared meters of the ART unit/total square meters of the hospital). The costs were expressed in R 2006.

The estimates of the person-years of observation (PYOs) for FY05/06 were extrapolated from the representative sample of 2835 records. Because patients entered and left the ART programme at different points in time, some patients contributed less than 12 months in FY05/06. Those in the sample who begun ART before 1/4/05 and continued after 31/3/06 contributed 12 months each. Those ones who begun ART after 1/4/05 and remained on treatment contributed the number of months separating the date of initiation and the end of the FY (31/3/06). In case of interruption, default or death, the patients contributed the number of months up to the date of exit. The sum of the total number of months contributed by the sample in FY05/06 was divided by 12 to get the number of PYOs. These were inflated by the sampling fraction to extrapolate the number of PYOs contributed by the universe of patients in FY05/06.

The cost per PYO was obtained by dividing the total costs for FY05/06 by the total number of PYOs for FY05/06. The average cost per PYO was estimated for the total sample of the 32 delivery sites and for the cluster of delivery sites having the delivery options (a) through

[8] The period covered is 1/4/05-31/3/06

(d) which were mentioned in the introduction. This allowed one to estimate the cost to follow up an average patient for one year according to the four delivery options.

Probabilities of remaining on ART

The probabilities of remaining on ART according to the delivery options (a) through (d) were already described in Part 1. This provided the transition probabilities used in the six-month cycles of the Markov model which was used in this economic evaluation to estimate the cost-effectiveness of the above mentioned delivery options. The analysis on the retention rate described in Part 1, covered the first two years of treatment, providing the probabilities for the first four cycles of the duration of six months each. Because no information was available after the 2^{nd} year, the probabilities of the fourth cycle were applied to the cycles 5-20. A total of 20 cycles of six-month duration were used to simulate a period of 10 years and the 95% confidence intervals of the transition probabilities provided the degree of uncertainty.

Probabilistic sensitivity analysis

The main objective of health policy and planning is to simulate the long term effects of the strategic directions taken on the basis of the past experience. The objective of this analysis was to project what could be the long term costs and retention rate of the different delivery options which were mentioned in the introduction. However, one important aspect to factor in the projections is the uncertainty of the estimates on the costs and retention rates of each delivery option, which were mentioned in the previous paragraph. Probabilistic sensitivity analysis (PSA) *(*Drummond M et al 2001) *(*Briggs et al. 1998) (Sonnenberg et al 1993*)* is commonly used to take into account for the uncertainty of the estimates.

PSA was used to assess the uncertainty through Monte Carlo simulations (Annex VII). These techniques have been widely used in economic evaluation of ART (Cleary et al 2006) (Hurwitz et al 2001) (Vijayaraghavan 2007) (Braithwaite et al 2007) (*Ghosh 2007) (Corzillius et al 2004)* to predict the cost-effectiveness of several scenarios. In our case, the scenarios of the four delivery options (a) through (b) were evaluated in terms of their cost to cover 1000 patients for 10 years and their effectiveness in retaining the highest number of patients.

The simulations were based on a series of six-month cycles. Each cycle represented a period of six months in which each patients were consuming resources and could remain on ART according to the probabilities associated with the delivery option (see previous paragraphs) and the length of treatment (e.g. higher probabilities of discontinuation in the first cycle). This process was repeated for 20 cycles of six months each according to the time horizon to be simulated (10 years). Without entering into the technicalities which are described in Annex VII, the simulations can be simplified as follows: (a) one thousands virtual patients were submitted one at a time through each delivery option; (b) each patient went through each cycle, where all the costs and the probabilities of remaining on ART were decided through a random process, within the variation of costs and probabilities of each cycle predicted by the given delivery option; (c) if the virtual patient remained on ART the next cycle followed, otherwise it dropped out and another virtual patient entered the simulation.

During this process, each virtual patient experienced the range of costs and probabilities of remaining on ART typical of each delivery option. Whenever a patient dropped out or arrived at the end of the 20th cycle, a new virtual patient went through another simulation where different random values were drawn. Each of the 1000 simulations per delivery option produced a different estimate of the cost-effectiveness of the four delivery options. At the end of 1000 runs, the model calculated the average cost and ART retention for a given delivery option based on the averages of the total patient runs. Acceptability curves were produced to indicate the proportion of times a delivery option was considered more cost-effective than the others according to the willingness to pay (Fenwick *et al* 2001).

2.3 Results

The results describe the cost for FY05/06 and the cost-effectiveness of the delivery options. The costs are presented as (i) average cost per delivery site and cost per PYO for FY05/06, and as (ii) cost per PYO for each delivery option (a) through (d) which were described in the introduction. The transition probabilities of remaining on ART according to the four delivery options are derived from the model that was presented in Part 1. The simulations on the cost effectiveness of the different delivery option are presented in the last section.

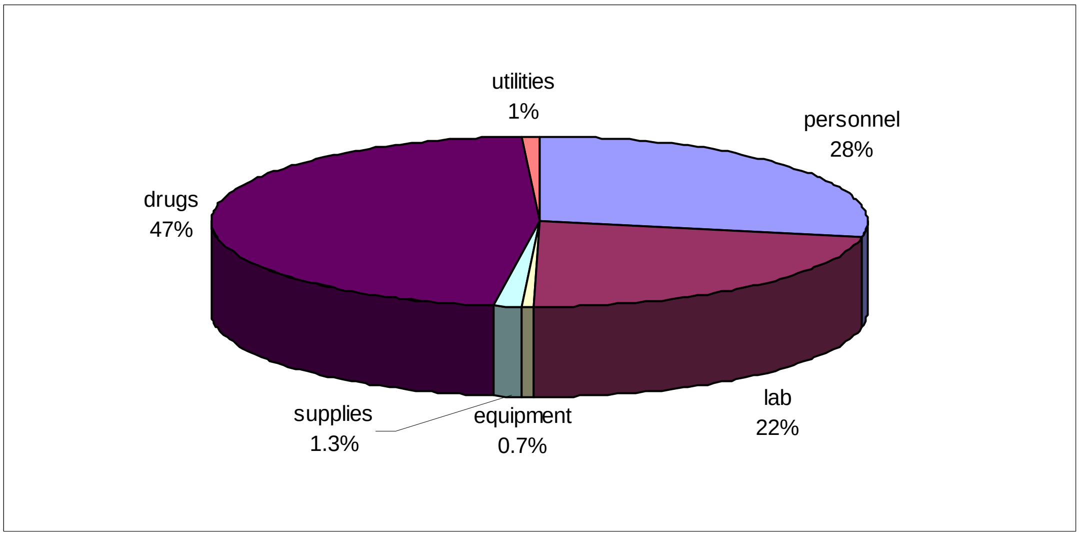
Expenditures for FY05/06

The highest share of the costs was due to ART and non-ART drugs followed by personnel, laboratory tests, supplies, utilities and equipment. Table 9 shows that the total average cost per delivery site was around R 2.8 million with a standard error of about R 0.2 million. The range varied between a minimum of slightly less than R 0.8 million and a maximum of R 5.2 million. Drugs, personnel and lab tests accounted respectively for 47%, 28% and 22%, while equipment, supplies and utilities accounted for another 3% (Figure 48).

Table 9 Average cost per ART delivery site for FY05/06

	Personnel	Lab	Equipment	Supplies	Drugs	Utilities	Total
Mean	R 792 367	R 622 462	R 18 010	R 37 083	R 1 321 331	R 25 626	R 2 816 879
SE	R 59 818	R 46 939	R 1 372	R 2 950	R 101 278	R 1 938	R 214 296
Min	R 215 435	R 169 225	R 4 856	R 10 205	R 355 770	R 7 058	R 762 550
Max	R 1 456 827	R 1 155 884	R 32 727	R 68 084	R 2 452 084	R 47 136	R 5 212 741

Figure 48 Expenditure categories



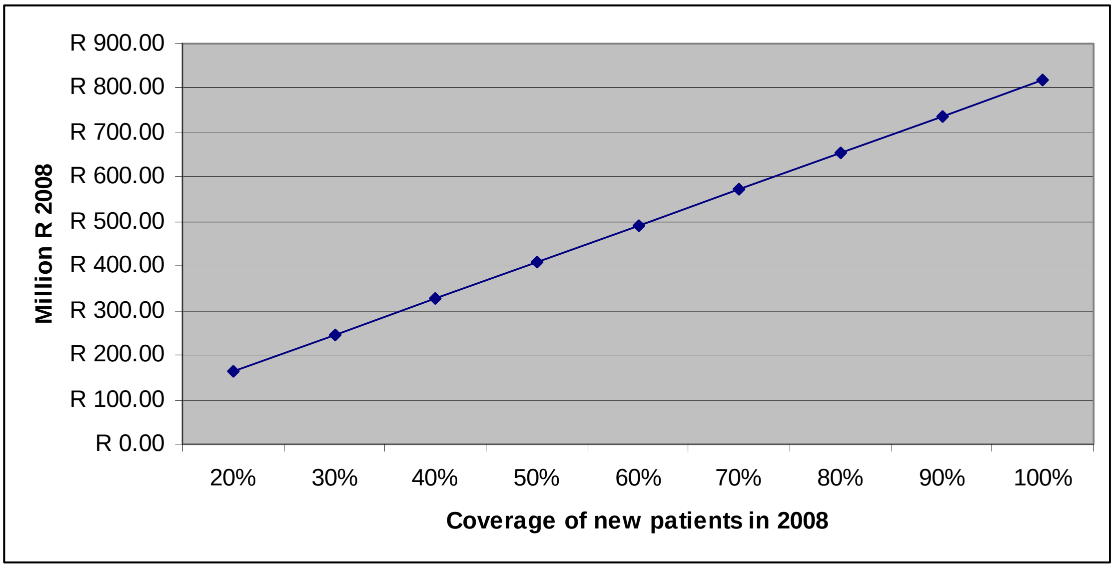
The total costs divided by the total number of PYOs provided the cost per PYO in Table 10. The average cost per PYO, which is the cost which was spent to follow up an average patient for one year, was slightly less than R 6800 with a 95% CI of about +/-R 100. Personnel, lab tests and drugs contributed respectively an average of about R 1900, R 1500 and R 3200, while the use of equipment, the supplies and the utilities contributed for slightly less R 200.

Table 10 Average costs per PYO, KZN, FY05/06

	Personnel PYO	Lab PYO	Equipment PYO	Supplies PYO	Drugs PYO	Utilities PYO	Total PYO
Mean	R 1,927	R 1,514	R 44	R 90	R 3,208	R 62	R 6,846
SE	R 11	R 8	R 1	R 2	R 25	R 1	R 48
Min	R 1,717	R 1,336	R 38	R 36	R 2,891	R 51	R 6,070
Max	R 2,070	R 1,608	R 49	R 104	R 3,633	R 67	R 7,530

The above estimates provided the basis for estimating the running costs of the static units according to coverage targets. As already discussed in Part 1, about 100,000 new patients are expected to require ART in 2008 in KZN. The line in Figure 49 provides the estimates of how many R million (Y-axis) would be required for each level of coverage (X-axis) in 2008. The estimation was obtained by multiplying the average cost per PYO, adjusted for inflation by applying the SA Consumer Price Index, by the proportion of the expected number of new patients in need for ART, according to what discussed in Part 1. Even if it were possible to ensure full coverage through the existing delivery sites, their running costs would easily consume the lion’s share of the budget for all activities related to HIV/AIDS, which in 2008, was in the order of R1 billion (Ndlovu N). Only 20% would remain for upgrading/maintaining existing and building new units, outreach services, prevention and other activities.

Figure 49 Estimated costs to deliver ART by coverage, KZN 2008



Delivery options

The results of the analysis described in Part 1 allowed one to identify four delivery options that at parity of other risk factors significantly influenced the ART retention. The four delivery options of the ART delivery sites were characterized by a common core staff, part time (PT) or full time (FT) senior level staff and different levels of workload. The core staff included one cleaner, one data clerk, five lay counselors, one assistant professional nurse, one professional nurse, one dietitian, one social worker and one pharmacist. The senior staff included one senior professional nurse (SPN) and one medical doctor (MD) who were attached to the nearby hospital and were assigned to the ART delivery site as part time (PT) or full time (FT). The workload differentiating the strategies was the annual intake of less than 200 new patients per doctor or 200 and more new patients per doctor. Thus, the delivery options had the same core staff mentioned above but they differed in terms of senior staff and workload as follows:

(a) PT MD and PT SPN with less than 200 new patients per doctor per year;
(b) Same staff as above but with 200+ new patients per MD per year;
(c) FT MD or FT SPN with less than 200 new patients per doctor per year; and
(d) Same as above but with 200+ new patients per MD per year.

As described in Part 1, the Kaplan Meier curves of the patients who attended the delivery sites with the above mentioned profiles provided the transition probabilities for each six-month cycle[9]. Because the data were related to the first two years only, the transition probabilities of the 4th cycles were applied to the cycles 5 through 20. This should not be far from what happens in reality because the losses of patients which were very high in the first semester stabilized after the 2nd cycle. Furthermore, uncertainty around the estimates was taken into account by the simulation.

Table 11 Probability (P) of leaving ART and standard errors (SE) by cycle

Delivery option in terms of Staff & workload	Cycle 1**		Cycle 2		Cycle 3		Cycle 4+***	
	P	SE	P	SE	P	SE	P	SE
*PT MD & PT SPN <200 New Patients per MD per Year	0.107	(0.030)	0.010	(0.010)	0.022	(0.016)	0.019	(0.018)
PT MD & PT SPN >=200 New Patients per MD per Year	0.373	(0.042)	0.069	(0.022)	0.037	(0.018)	0.030	(0.021)
FT MD & FT SPN <200 New Patients per MD per Year	0.070	(0.012)	0.022	(0.007)	0.019	(0.007)	0.012	(0.006)
FT MD & FT SPN >=200 New Patients per MD per Year	0.172	(0.034)	0.107	(0.028)	0.018	(0.012)	0.011	(0.011)

* PT Part time FT Full Time ** Each cycle has a duration of six month *** It is assumed that the a patient survivng the 4th cycle will continue to experience the same probabilities of the 4th cycle for the cycles following the 4th one

[9] More information on the Markov cycle is available in Annex VII

Cost per cycle

The 32 ART sites were categorized according to the staffing profiles and the workload related to the four delivery options (a) through (d). The total cost (FY05/06) and the total PYOs of the cluster of sites with the staffing and workload profile (a) provided the basis to estimate the cost per PYO of the relative option (a) and so on for the other options. Being the cost per PYO equivalent to the cost to follow up a patient for one year, the cost for each cycle of six months was equivalent to half the cost per PYO for the Markov state "remained on ART". As it was assumed that on average those patients who left ART did so in the middle of the cycle, the cost of the Markov state "discontinued ART" was equivalent to one quarter of the cost per PYO (Table 12).

Table 12 Costs per cycle for staying or leaving ART, R2006

Options in terns of staff & workload	Cycle remained on ART Average	(SE)	Cycle discontinued ART Average	(SE)
*PT MD & PT SPN <200 New Patients per MD per Year	3240	(79)	810	(20)
PT MD & PT SPN >=200 New Patients per MD per Year	3331	(26)	833	(7)
FT MD & FT SPN <200 New Patients per MD per Year	3508	(29)	877	(7)
FT MD & FT SPN >=200 New Patients per MD per Year	3659	(85)	915	(21)

Cost effectiveness

The first cost-effectiveness estimates were carried out with a deterministic approach using the average values for probabilities and costs. Table 13 provides the costs and effectiveness of following up a cohort of 1000 patients for ten years. The delivery options defined by staffing and workload (columns a-b) are listed in order of increasing costs[10] to follow up the cohort of 1000 patients for 10 years at R2006 value (column c). The incremental costs (column d) provide the difference between the costs of one strategy versus the previous less expensive one (e.g. 49,585,842-31,550,663=18,035,179). Column (e) provides the number of patients still on ART at the end of the 10th year while column (f) provides the incremental number of patients that would be retained by one option versus the previous less expensive one (e.g. 629-336=293). The incremental cost-effectiveness ratio (ICER) in column (g) is obtained by diving the incremental cost by the incremental effectiveness (e.g. 18,035,179/293 = R61,554). The ICER is the extra cost to retain one extra patient if a more cost-effective strategy is adopted versus a less cost-effective one.

It is customary to estimate the cost-effectiveness of one strategy against the previous one, starting with the least costly one. The ICER compares the extra cost required to retain an extra patient if a more cost-effective strategy were adopted. Within a period of 10 years it would cost an extra R61,554 (R2006 value) to retain an extra patient if the delivery option in the second row of Table 13 were to replace the option in the first row. Replacing the option

[10] The costs are at R2006 values and were neither adjusted for inflation nor discounted for the 10 year-period

in the second row with the option in the third row would be more costly but would retain a lower number of patients (602 vs. 629). A strategy which costs more and is less effective than another one is considered "dominated" and is excluded. For this reason, the option in the fourth row is compared with the previous most cost-effective option which is in the second row (not with the third one) and its ICER is R84,392.

Table 13 Costs to follow 1000 patients for the first 10 years of ART, R2006

a	b	C	D	e	f	G
Senior Staff	**Workload**	**Cost in R2006**	**Incremental Cost**	**Effectiveness**	**Increm. Effectiv.**	**Incremental Cost-Effectiveness (ICER)**
SPN & MD PT	200+	31,550,663		336		
FT SPN PT MD	<200	49,585,842	18,035,179	629	293	61,554
SPN & MD FT	200+	50,101,604	515,762	602	-27	Dominated
SPN & MD FT	<200	57,940,682	8,354,840	728	99	84,392

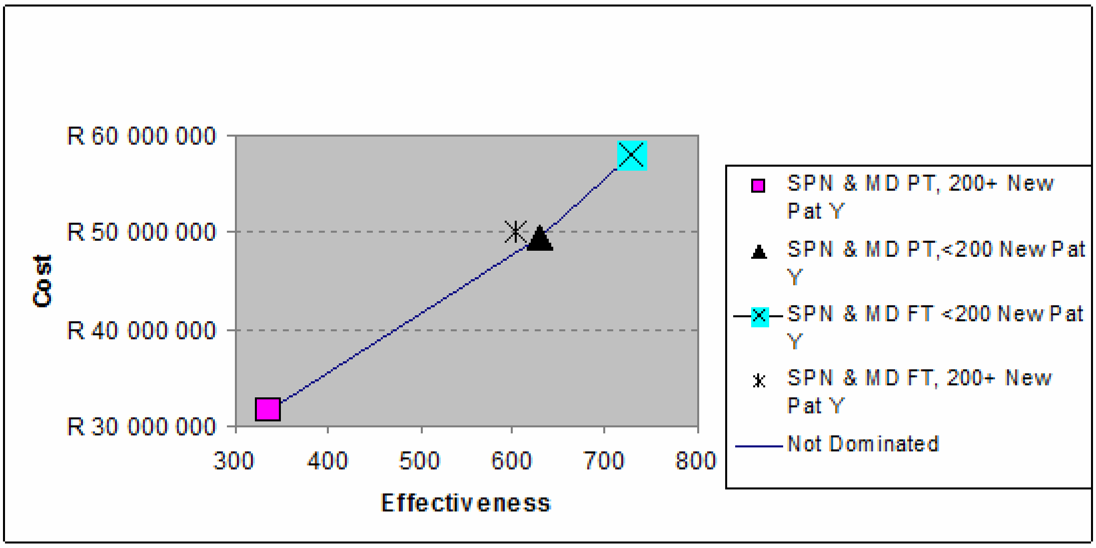
SPN senior professional nurse, MD doctor, PT part time, FT full time ** New patients per MD per year

Table 13 provides an idea of how much is likely to be obtained from the invested resources if alternative options were adopted. The option in the second row of Table 13 would cost 18 million more than the option in the first row but would retain 293 patients. This is tantamount to say that if the DOH and society were willing to adopt option two versus option one they will have to accept to spend 18 million more to save an extra 293 patients. This is equivalent to R61,554 per each extra retained patient if a more expensive option is adopted versus a less expensive one.

Figure 50 shows the same information of Table 13 with the interventions plotted according to their costs and effectiveness. The X-axis shows the effectiveness in terms of the number of patients expected to remain on ART at the end of the 10th years. The Y-axis represents the cost in R million at 2006 value that would be required to follow up the hypothetical cohort of 1000 patients for 10 years.

The optimal interventions are linked by a line, which is also known as cost-effective frontier. The option "SPN & MD FT 200+" at the left of the frontier is dominated because it costs more and retains less patients than the option on its right "SPN & MD PT <200". On the other hand, FT SPN & FT MD is the most cost-effective option when the annual intake of new patients is less than 200.

Fig 50 **Cost-effectiveness to follow up 1000 patients for 10 year**



Monte Carlo simulations

The results in Table 13 are based on a deterministic cohort approach and they do not take into account the uncertainty of the predictions. In real life conditions, the costs and the effects of ART vary around the average, limiting the extrapolation of the results based only on average values. The variation of the transition probabilities and the costs is expressed as standard errors around the average in Tables 11 and 12.

The effect of uncertainty was taken into account through probabilistic sensitivity analysis (PSA) also known as Monte Carlo simulation. All the costs and the transition probabilities were allowed to vary around the 95% confidence intervals (CI) through a stochastic process of 1000 simulations (see methodology). At each simulation, a virtual patient went through a strategy and according to the random process it dropped out at any cycle or continued up to the 20th cycle. This resulted in 1000 different cost-estimates for each option with a dispersion of values which are presented in Table 14 and Figure 51.

While in Table 13 the cost-effectiveness and the ICER were presented as single values, Table 14 presents the 95% confidence interval of the values obtained from the 1000 simulations. For example, if the virtual 1000 patients were followed up for 10 years, the 10 year cost of the first strategy could vary between R31 and R32.3 million at R2006 value. Similarly the number of patients retained on ART at the 10th year could vary between 330 and 348 and so on.

The value added of the simulation has been to show that the predictions are robust. After 1000 iterations, the average cost-effectiveness of SPN & MD FT <200 is higher than the other options. Even considering the uncertainty around the estimates, the 95% CI of the incremental effectiveness of the 4th option is significantly higher then the previously not dominated one. Over a period of 10 years, the 4th intervention would retain 719-738 of the 1000 initial patients, which is significantly higher than the 623-643 patients retained by the second best option. The ICER of the 4th option over the 2nd one would vary between R75,180 and R98,810 per extra retained patient.

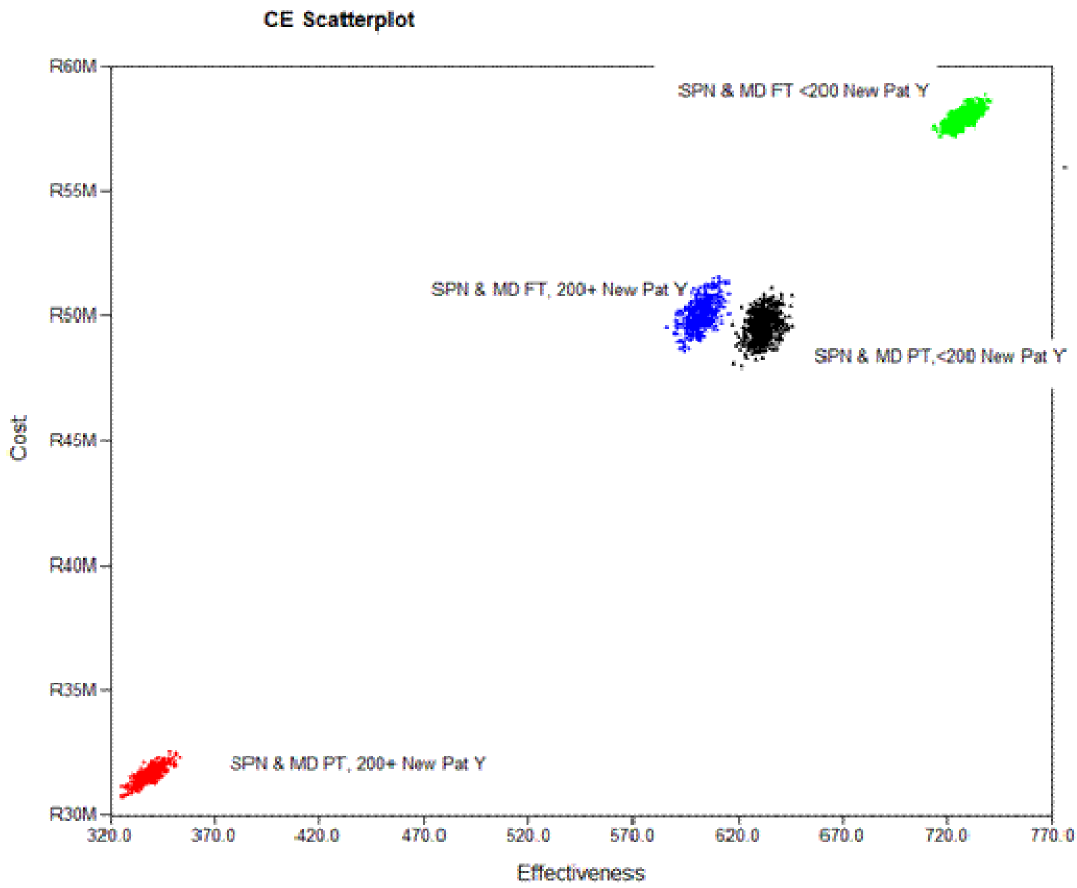
Table 14 Results of the 1000 simulations, time period 10 years

Senior Staff	Work load	Cost in R million (95% CI)	Effective Ness** (95% CI)	Incr.*** Cost in 2006R million (95% CI)	Incr. Effective Ness (95% CI)	Incr. Cost-effeciveness ratio (ICER) in R2006 (95% CI)
SPN & MD PT	200+	R 31.6 (R31.0 to R32.2)	339 (330 to 348)	Baseline	Baseline	
SPN & MD PT	<200	R 49.6 (R48.6 to R50.7)	633 (623 to 643)	R 18.0 (R16.9 to R19.2)	294 (284 to303)	R 61,354 (R57,926 to R64,838)
SPN & MD FT	200+	R 50.1 (R49.1 to R51.2)	603 (593 to 613)	R 0.5 (-R0.9 to R1.9)	-29 (-38 to-21)	Dominated
SPN & MD FT	<200	R 57.9 (R57.4 to R58.6)	728 (719 to 738)	R 8.3 (R7.2 to R9.5)	96 (89-103)	R 87,075 (R75,180 to R98,810)

* R Million at 2006 value to follow up1000 Patients for 10 years **patients remaining on ART after 10 years.
*** The incremental cost is the difference between the costs of one intervention vs the previous less costly one.

Figure 51 provides a graphic representation of the variation of the 1000 simulations presented in Table 14. The Y-axis represents the cost in R million at 2006 value required to follow up 1000 patients for 10 years. The X-axes represents the number of patients remaining on ART at the end of 10 years. Compared with Figure 50, in which there was only one "cost-effectiveness point" per delivery option, Figure 51 provides four clouds of points, corresponding to the cost-effectiveness estimates of the 1000 simulations. The spread shows that the clouds of cost-effectiveness points related to the options are significantly distant from each others. The option SPN & MD FT 200+ is still “dominated” by the clouds immediately to the right (SPN & MD PT <200). The cloud of points of the most cost-effective option at the upper right hand corner of Figure 51 is well separated from the other options.

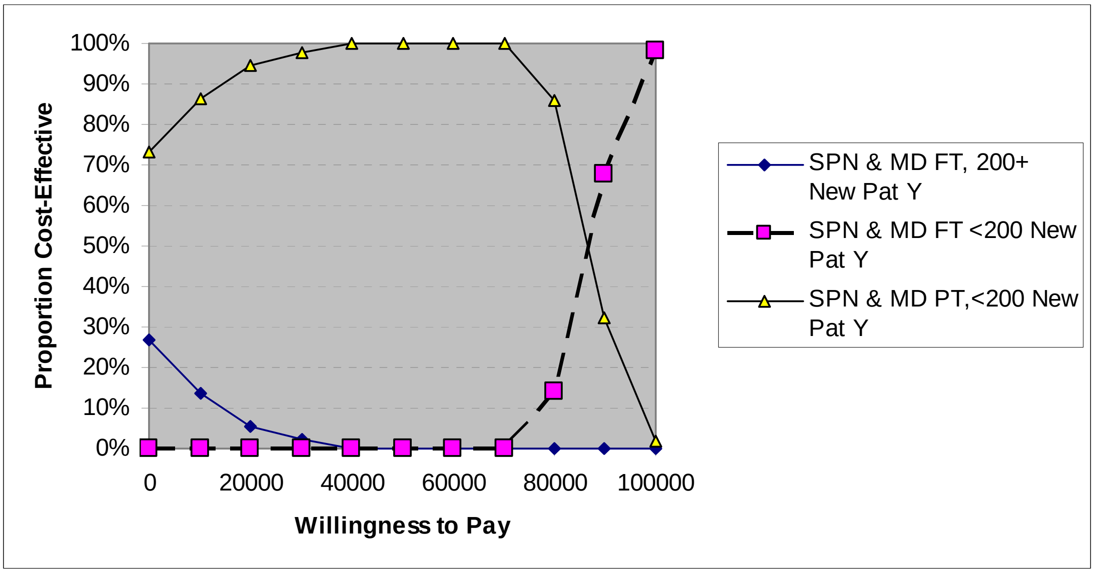
Figure 51 Plots of the 1000s simulations



The acceptability curves in Figure 52 represent the proportion of the 1000 simulations when an option was the most cost-effective according to the willingness to pay. The X-axis shows how much society would be willing to pay (in R 2006) to retain an extra patient on ART over a 10-year period. The Y-axis measures the proportion of times a strategy was most cost-effective if society were willing to pay the given amount on the X-axis. If health priorities were so many that very few resources could be spared to retain an extra patient, the willingness to pay would be near zero. In this case, SPN & MD PT <200 would be cost-effective for about 70% of the time and SPN & MD PT 200+ would be selected for about 30% of the rime.

As more resources are available and society is willing to spend more to retain an extra patient, the values on the X-axis move towards the right. In this case, the proportion of times SPN & MD PT <200 would be selected would increase and would become the sole intervention when willingness to pay is between R40,000 and R70,000. With willingness to pay above R70,000 "SPN & MD FT <200" gradually replaces SPN & MD PT <200 and would be selected 100% of the time as the most cost-effective option when the willingness to pay reaches 100,000.

Figure 52 Acceptability curves of cost-effectiveness



2.4 Discussion

Simulations are widely used in medical decision analysis and economic evaluation. Far from being academic exercises, simulations are part and parcel of the progress of health planning and they cannot be ignored. Probabilistic models are critical to decide how the resources could be better spent to obtain the highest return in health gains. To make an example, the National Institute of Clinical Excellence[11] (NICE) in the United Kingdom requires that evaluations of interventions include simulations. Therefore, simulations are common practice in modern health planning.

To the best of the authors' knowledge this evaluation has unique characteristics. It has identified management options and it has estimated their cost-effectiveness, on the basis of the experience of a publicly funded ART programme. The analysis of the 32 delivery sites, which started rolling out in 2004, has provided the basis to identify significant risk factors for discontinuing ART. Delivery options based on staff profiles and workload were found to have a significant effect at parity of the other risk factors. During the 1000 simulations, the probabilities and costs were allowed to vary within their 95% confidence intervals (CI) taking uncertainty into account.

The advantage of this analysis was to create awareness about the uncertainty surrounding the estimates of cost-effectiveness. Interventions, which have proven their efficacy in trials, might fail to become effective in operational conditions. Trials have a high concentration of human and other resources, ensuring a high level of compliance and other factors that are critical for drug efficacy. When the same drugs are delivered through routine health services the potential efficacy declines.

Therefore, any conclusion on cost-effectiveness should take into account the variation characterizing health interventions in real life conditions. The variation found in the 32 delivery sites confirms that routine services have a variable degree of effectiveness. This analysis has captured the factors causing such variation and has transformed them into potential delivery options.

The simulations have confirmed that the predictions of ART retention associated with the delivery options were robust. The clouds of points in Figure 52 show that although the cost-effectiveness could vary, the ranking of the interventions did not change in terms of cost-effectiveness. The best option was: FT SPN & FT MD <200 followed by PT SPN & PT MD <200; while FT SPN & FT MD 200+ was dominated.

[11] NICE is the British body dealing with the evidence base medicine

Limitations

Only the delivery options related to the static units were taken into account in this evaluation. The evaluation has shed light on the variation in treatment retention across the static units. Other strategies were not taken into account because the ART programme of KZN was overwhelmingly based on the static units and there were no data on other potential strategies. This leaves open the question on how to integrate less resource intensive outreach activities.

The costs of hospitalization were not taken into account. This might have under-estimated the cost-effectiveness of the delivery options with the highest retention rate, because those remaining on ART are likely to require less hospitalization. On the other hand, the cost saving due to lower need for hospitalization is uncertain because unless full coverage is achieved, there will still be an increasing number of AIDS patients not covered by ART who will continue to overload the hospitals.

The analysis did not consider the costs to build extra delivery sites. The scenario in Figure 49 was based on the extrapolation of the costs per PYO of the existing units. This was done more to create awareness that full coverage with ART would almost exhaust the present budget for the whole HIV/AIDS programme and not much will be left for other lines of expenditures (e.g. prevention). The objective was not to provide a comprehensive budget for all type of activities but to estimate the potential cost of covering with ART all the patients in need.

The overall costs were derived from the records of the delivery sites, which might have been affected by under-recording of the personnel, the drugs and the other consumables. Although this might have affected the exact estimation of the costs in a few delivery sites, it is unlikely that this might have affected the ranking of the delivery options in terns of cost-effectiveness. This would have required a systematic under or over-reporting in all the units which had the same staffing profile and workload at the basis of one or the other option.

The measure of effectiveness has been limited to the ART retention because it was the most reliable outcome that was available. Most simulations in the literature take into account survival to estimate the cost per year of life (YL) and per quality adjusted life years (QALYs). The simulation could have estimated these more comprehensive measures of outcome by using assumptions and by taking secondary data from the literature. However, this would not have made much of a difference because a strategy that is cost-effective in retaining patients is also likely to be cost-effective in terms of YL and QALYs.

It could be argued that ART retention tells only part of the story and that a complex treatment like ART depends on compliance. However, with the exception of a few well-funded pilot projects it is unlikely that government services will ever be able to afford the luxury of measuring compliance in a reliable manner on a great number of delivery sites. It is therefore inevitable that any evaluation of publicly-funded health services covering wide areas will have to rely on the best available information, which in this case was ART retention.

Implications

The implications of the findings should be interpreted with a certain degree of flexibility and open-mindedness. The economists and the health professionals tend to view the health problems from their own perspective. This lack of consideration from the others' point of view is frequently affected by ideologically rigid positions and a certain degree of dogmatism. Saying that a certain strategy is more cost-effective does not mean that it has to be implemented blindly across the whole health care delivery system. By the same token, doctors might refuse to put any economic perspective into account because of their individualistic approach to solve problems affecting individual patients. The results do not suggest that not exceeding 200 new patients per doctor per year is the panacea because it would require an implicit rationing that is hardly feasible. Nonetheless, the fact that the delivery options were based on the 32 delivery sites suggests that these profiles and workload levels are already in place and are behind the variation in retention rate.

The policy question is how to expand the most cost-effective delivery option to achieve the highest possible coverage. The economic evaluation has shown the high potential of ART retention with delivery options which do not exceed 200 new patients per MD per year. However, this cut off would only allow a limited coverage of the new patients needing ART. A higher coverage can only be achieved by: (a) exceeding the cut off point of 200 new patients per year by several times, facing the consequences of an increasing number of defaulters; (b) expanding the number of delivery sites distributing ART; and (c) integrating less resource intensive options.

Less intensive alternative delivery systems are frequently proposed on the basis of successful pilots. These have the potential of increasing access to ART with the proviso that management models take into account the risks involved of a rushed expansion. ART is a complex therapy, which is associated with side effects needing high skilled staff. Increasing coverage by delivering ART through alternatives options could run the risk of a higher number of defaulters if the staff are not skilled enough to cope with ART complexity. Thus, any simplification of the delivery system needs to be carefully assessed in terms of ART retention.

There has been more emphasis on increasing coverage than in planning how to maintain it, with the consequence that coverage has become an end in itself. It is no use to aim at full coverage if this is likely to overstretch capacity, producing tens of thousands of defaulters with the risk of spreading drug resistance. This might be paid at higher cost later on, as the TB programme has found out with the insurgence of extremely drug resistant TB. Because ART effectiveness requires a very high level of adherence, a lower but feasible coverage would deny ART to everybody but at least would have a higher chance of serving a feasible number of patients effectively. An increased level of coverage going beyond the absorption capacity of the delivery system would still not cover everybody and would run the risk of providing a poor service to those who are under treatment.

The results of this evaluation should be used to understand the reasons behind the variability in the cost-effectiveness of delivering ART. The results can be used to forecast realistic coverage targets which can be achieved without impairing the effectiveness of ART. This means that the coverage targets need to be based on the average number of new patients per delivery site and per staff. This will have to be accompanied by an estimation of the expected number of defaulters according to the predictions of this analysis and the planning of relevant actions to trace them. This should provide the policy makers with sufficient information on choices and potential consequences, which are at the basis of informed decisions.

2.5 Conclusions

This analysis has put cost-effectiveness into the context of what has happened in the ART programme so that the DOH can make informed decisions on how to expand coverage. Any policy debate about full coverage of ART should take into account the limitations of the health system in maintaining the minimum level of efficiency to ensure effectiveness. The experience of the first two years of the ART programme has provided some answers about the capacity of the ART delivery sites to effectively retain patients.

Taking informed decisions includes the selection of policy options which take into account operational conditions. Any coverage target should be accompanied by estimates on the expected number of patients per staff. Knowing that full coverage will result in more than 1,200 new patients per MD in 2008 should help to put coverage in context. Even if the DOH were to change the present policy of relying mainly on static units, the implementation of other delivery systems needs to be planned carefully on the basis of the experience accumulated in a few pilots.

Expanding alternative delivery systems is hampered by the scarce documentation on their effectiveness and logistic implications. Besides a few research projects, most systems are described in generic terms without any quantification on the inputs, outputs and outcome. Knowing that a certain pilot project has increased coverage to 80% is not very helpful if there are no details on the size of the geographic area and the population that has been covered, the human and the other resources used per patient and the retention rate. What is still missing is a thorough quantitative analysis of the problems and risks, cost and effectiveness, and the logistics and absorption capacity of these pilots. The same methodology of this analysis could be used to clarify these issues and to identify the factors that allowed successful pilots to deliver ART effectively. This is critical to quantify the resources required to provide alternative services to achieve and maintain an equitable access province wide.

Whatever alternative delivery systems are proposed, they will still have to rely on the static units, especially during the first six months, which is the most critical period for default. The alternative delivery systems might be more suitable for patients which have been on ART for sometime and therefore can be followed up through a less sophisticated system. The first few months are the most difficult and they require very experienced staff that can ensure compliance and can effectively control the side effects. It is after the patients have been stabilized that less specialized delivery systems could be involved in taking care of the maintenance of the patients already on ART. This means that whatever alternative systems are added, the brunt of the new patients approaching ART every year will continue to fall on the expertise of the high level staff that are based at the static units. This means that whatever alternative systems will be added, the results of this analysis will still be relevant to estimate the human resource needs of the static units.

2.6 Recommendations

Most of the recommendations were already listed in Part 1, to which the following ones are added as they are more related to the results of the economic evaluation.

Estimate the resource gaps to reduce the workload to less than 200 new patients per unit per year. This can be done by estimating the present ratio between the number of new patients per unit recruited in 2008 and the existing staff to estimate the present workload and the staff to be recruited to reduce it to more sustainable levels.

Set coverage targets according to capacity. The gaps estimated in the previous step will be at the basis of realistic coverage targets and the estimation of the budget required to achieve it.

Evaluate alternative delivery systems. Alternative delivery systems that are being implemented across the province are variegated and cover relatively small areas. These scattered alternatives should be evaluated to quantify human and other resources per ART patient that make these pilots work. Simulations should be carried out to take uncertainty into account as it was done in this analysis. This will provide planners with the information to estimate feasibility and cost effectiveness of alternatives.

7 Bibliography

Beck, RJ, Pauker SG. The Markov process in medical prognosis. Medical Decision Making 3: 419-458; 1983.

Briggs A, Sculpher M. An Introduction to Markov Modeling for Economic Evaluation. Pharmacoeconomics 13:397-409; 1998.

Braithwaite RS, Shechter S, Chang CC, Schaefer A, Roberts MS. Estimating the rate of accumulating drug resistance mutations in the HIV genome. Value Health. May- 10:204-213; 2007.

Corzillius M, Mühlberger N, Sroczynski G, Jaeger H, Wasem J, Siebert U. Cost effectiveness analysis of routine use of genotypic antiretroviral resistance testing after failure of antiretroviral treatment for HIV. Antivir Ther. 9:27-36; 2004.

Drummond M, McGuire A. Economic evaluation of health care, merging theory with practice. Oxford University Press. 2001.

Fenwick E, Claxton K Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ 10: 779-787; 2001.

Ghosh P, Vaida F. Random changepoint modelling of HIV immunologic responses. Stat Med. 26:2074-2087; 2007.

Hurwitz SJ, Asif G, Schinazi RF. Development of a population simulation model for HIV monotherapy virological outcomes using lamivudine. Antivir Chem Chemother 18:329-341; 2007.

Luce BR, Manning WG, Siegel JE, Lipscomb J. Estimating Costs in Cost-Effectiveness Analysis. In Cost-effectiveness in health and medicine Edited by: Gold MR, Siegel JE, Russell LB, Weinstein MC. New York, Oxford University Press; 1996.

Nhlanhla N. Budget allocations for HIV and AIDS in 2005/6 provincial social sector budgets: Implications for improved spending. Budget Brief No. 156 IDASA Budget Information Service. August 2005.

Roberts, MS, Sonnenberg FA. Decision modeling techniques. G.B. Chapman, F.A. Sonnenberg, eds., Decision Making in Health Care: Theory, Psychology, and Applications. Cambridge University Press, 2000.

Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Medical Decision Making. 13: 322-38; 1993.

Progress on Global Access to HIV Antiretroviral Therapy: A Report on "3by5" and Beyond. WHO and UNAIDS. 2006

Tree Age Software Inc. 2008.

Vijayaraghavan A, Efrusy MB, Mazonson PD, Ebrahim O, Sanne IM, Santas CC. Cost-effectiveness of alternative strategies for initiating and monitoring highly active antiretroviral therapy in the developing world. J Acquir Immune Defic Syndr. 46:91-100; 2007.

Annexes

Annex I	-	EuroQol measure of quality of life
Annex II	-	GIS Accessibility Index
Annex III	-	Cox Regression Model
Annex IV	-	Kaplan Meier curves, probability of remaining on ART
Annex V	-	Costing
Annex VI	–	Unit Cost FY05/06
Annex VII	-	Monte Carlo simulation

Annex I - EuroQol measure of quality of life

Dr Aamina Peer, Italian Cooperation

Background

The World Health Organisation (WHO) defines quality of life as an individual's perception of their position in life in context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychological state, personal beliefs, social relationships and their relationship to salient features of their environment. (WHO/MSD/MER/02.3, 2003 and WHOQOL-HIV Group, 2003). There is a growing recognition internationally of the importance of understanding the impact of healthcare interventions in patients' lives rather than just their bodies. This is particularly important for patients with chronic, disabling or life threatening diseases who live without the expectation of cure and have conditions that are likely to have an impact on their physical, psychological, and social well-being (WHO/MSD/MER/02.3, 2003 and WHOQOL-HIV Group, 2003). Quality of life studies are being conducted to assess the effects health interventions have on these aspects of patients' lives. The benefits of quality of life studies can be seen in medical practices, improving doctor-patient relationships, in assessing the effectiveness and the relative merits of different treatments, in heath service evaluations, in research and in policy making (WHO/MSD/MER/02.3, 2003 and WHOQOL-HIV Group, 2003)

Health related quality of life (HRQL) instruments are the tools used to determine the quality of life of patients that undergo health interventions (Meunning 2002). There are many different HRQL instruments that have been developed to describe quality of life in diverse populations (Delate *et al.* 2001, and Gielen *et al*. 2001, Badia *et al*. 2000) These include the Health Utility Index (Furlong *et al*. 2001), SF36 (Han *et al*. 2002) and the Nottingham Profile (WHOQOL-HIV Group. 2003). There are also HIV specific instruments such as Medical Outcomes Study – HIV (MOS-HIV) and Medical Quality of Life-HIV (MQOL-HIV) (Albert 1999, Badia X et al. 2000), and WHOQOL-HIV (World Health Organisation Quality of Life – HIV) (WHOQOL-HIV Group, 2003).

EuroQuol

The European Quality of Life (EuroQol) measures the Health Related Quality of Life (HRQL) through questions assessing mobility, self-care, usual activities, anxiety/depression and pain/discomfort (Brooks 1996). For each question, the interviewee selects one option of no limitation, medium and severe limitation, which is associated with a score. The total score, obtained by adding the individual scores range from 0 to 100, with 0 equal near death experience and 100 equal to the best possible health.

The European Quality of Life (EuroQol) Group has developed the EQ-5D. The EQ- 5D provides a measure of overall HRQL based on five descriptive questions with three levels of answers and a rating scale. Utilities valued between 0 (death) and 1 (full health) for the different combinations of possible answers in the descriptive part have been established for different population groups. This HRQL describes health states in terms of five dimensions (mobility, self-care, usual activities, anxiety/ depression and pain/discomfort) (Badia et al. 1999) The EuroQol has been used in Zimbabwe to measure the quality of life of patients affected by HIV/AIDS (Rabin R et al. 2001, Badia 2000, Delate et al 2001, Jelsma et al. 2003), and other medical conditions (Rabin et al. 2001).

There are a few studies that have used the EQ-5D in South Africa (Cleary 2003, Jelsma et al. 2005, Hughes et al. 2004). The EuroQuol has been used in a Xhosa speaking population living in a sub-urban area of Cape Town, where it was found that patients living with HIV/AIDS had a significant lower EuroQuol than the rest of the community (Hughes et al 2004). The Zulu and English EQ-5D version which have been used in South Africa have been produced by adhering to strict official translation protocols and have been recognised by the EuroQol Group. It is a self-administered instrument and takes less than 10 minutes to complete. The limitation is that a South African utilities value set for health states is not available. However, the UK population value set has been used in studies in South Africa (Cleary 2003) and the Zulu version of the EQ-5D was considered appropriate to determine the HRQL for patients on the KwaZulu-Natal ARV rollout programme.

Aim

The aim of this study was to measure the HRQL of patients attending the ART sites in KwaZulu-Natal.

Method

Subjects

All participants in this study are registered in the KZN ARV rollout programme through one of the ARV sites. The criterions for inclusion on the KZN ARV rollout programme are those in the National Guidelines for ARV treatment in South Africa. The inclusion criteria are all patients eligible for ARV therapy and all patients on ARV therapy.

Instrument

Both the Zulu and the English version of the EQ-5D were used. The health states were described using the five EQ-5D domains: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The measurement of health status was done using the visual analogue scale of the EQ-5D. Although the instrument can be self-administered, the interviewer gave assistance to participants on how to complete the questionnaires.

Procedure

Patient recruitment on the KZN ARV rollout programmes began in March 2004. New patients are constantly being recruited and old patients are coming in for follow-up visits. A survey was planned to visit each clinic and to select the ARV patients attending the clinic during the day of the survey to fill the EQ-5D questionnaire.

Ethical consideration

Ethical approval was obtained from the Biomedical Research Ethics Committee of the University of KZN and Informed consent was obtained from the participants, who were made aware that lack of participation would not have any negative consequences. The individual results were kept confidential and were not shared with anybody, including the staff of the clinic. The interviews were conducted in private with only the participants being present.

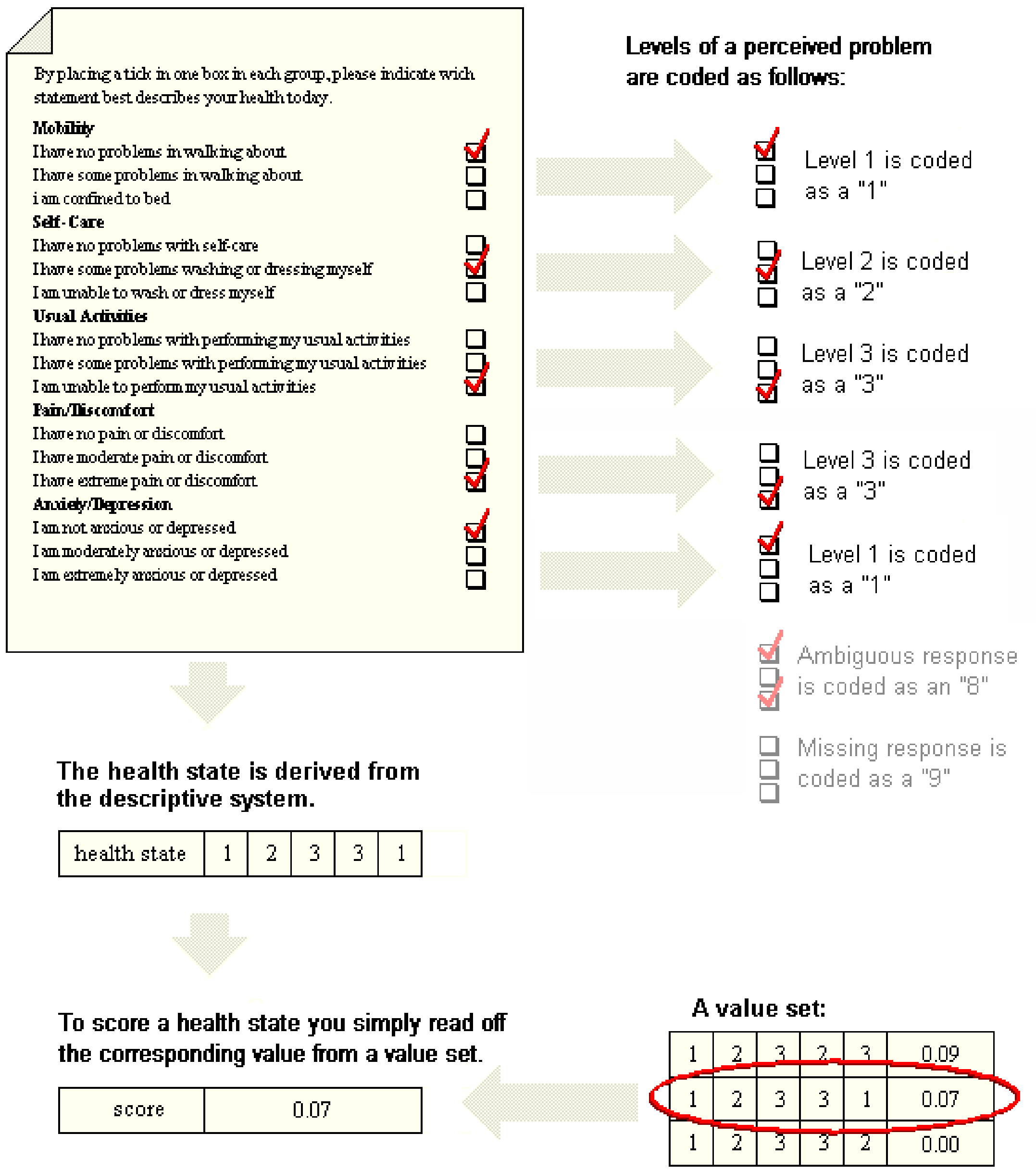
HIV specific HRQL instruments

The questionnaire consisted of 35 questions, which assessed eleven dimensions of health. All items were measured on a two, three five or six response options. The Medical Outcomes Study – HIV (MOS HIV) subscales are scored as summated rating scale and transformed on a 0-100 scale where the higher score indicates better health. In addition to this, a Physical Health and Mental Health score is also generated. The MOS HIV benefits are that it has shown to be more sensitive to change in short-term 3-month studies than MQOL HIV, it has been used internationally, it is self administered in less than 10 minutes, it has adequate internal consistency, and it has well documented psychometrics data.

The Medical Quality of Life-HIV (MQOL HIV) consists of 40 questions, which assess ten dimensions of health. This includes Mental Health and Physical Health amongst other dimensions. All answers are measured on seven or eight response options. The subscale scores are summed on a 4-28 point scale where higher scores represent better health. The MQOL HIV benefits are the same as the MOS HIV. However, this questionnaire has been shown to have adequate reliability and to be valid and responsive to changes in perceived quality HRQL changes and symptoms severity over 5.5 months in symptomatic HIV infected patients. It has the same limitations as the MOS HIV but the difference is that it is new and there are not many studies available.

EQ-5D questionnaires

The respondent was asked to indicate his/her health state by ticking (or placing a cross) in the box against the most appropriate statement in each of the 5 dimensions. This decision resulted in a one-digit number expressing the level selected for that dimension. The digits for five dimensions can be combined in a five-digit number describing the respondent’s health state. It should be noted that the numerals 1-3 have no arithmetic properties and should not be used as a cardinal score.



Scoring EQ-5D health states

This note describes the process by which responses to EQ-5D were converted into a single index format (referred to as EQ-5D$_{index}$). The following factors have to be taken into account when transforming EQ-5D responses into this single index format.

(a) The weights were generated using Time Trade-Off (TTO) in a nationally representative sample of the general population of the UK in 1993 (TTO is a method for assessing preferences for a given health state, in which the respondent is asked how much time he or she would be willing to trade from a given lifespan in the health state, to have the remaining lifespan in perfect health. For example, a respondent might have a 40 year life expectancy in a given health state, and might be willing to trade 10 years in order to have a 30 year life expectancy in perfect health)

(b) TTO generates weights that are widely accepted for the purposes of quality-adjustment in computing QALYs for use in economic evaluation

(c) The TTO weights listed here may or may not be those that represent the current social preferences of the UK general population. They are more than 10 years old. (Although, these health state values are not representative of the South African population, it will be appropriate to use because changes in the health state for the different stages of the disease will be observed.)

(d) The TTO weights listed here may or may not be representative of the social preferences of the general population of other countries (now or in the past)

Values for the 243 EQ-5D health states were calculated using a regression model based on UK national survey data (reported elsewhere). The following worked example indicates how these coefficients were used to compute the estimated values for each state.

<u>Calculating EQ-5D state scores - a worked example</u>

EQ-5D dimension	**Level 2 "some problem"**	**Level 3 "extreme problem"**
Mobility	0.069	0.314
Self-care	0.104	0.214
Usual activity	0.036	0.094
Pain / discomfort	0.123	0.386
Anxiety / depression	0.071	0.236
	Constant = 0.081	N3 = 0.269

The arithmetic needed to recover the estimated value for any health state from this table of decrements is given by the following example:

Taking health state 1 1 2 2 3

Full health (1 1 1 1 1) = 1.0

Constant term (for any dysfunctional state)(subtract 0.081)

Mobility.. level 1	subtract 0
Self-care.. level 1	subtract 0
Usual activity.. level 2	subtract 0.036
Pain / discomfort.. level 2	subtract 0.123
Anxiety / depression.. level 3	subtract 0.236

Level 3 occurs within at least 1 dimension hence subtract N3 parameter 0.269
The estimated value for state 1 1 2 2 3 (EQ-5D$_{index}$) is given by

1.0 - 0.081 - 0.036 - 0.123 - 0.236 - 0.269 = .255

Bibliography

Albert W Wu. MOS-HIV Health Survey, John Hopkins University. 1999

Badia X, Podzamczer D, Casado A, López-Lavid C, García M. Evaluating changes in health status in HIV-infected patients: Medical Outcomes Study-HIV and Multidimensional Quality of Life-HIV quality of life questionnaires. Spanish MOS-HIV and MQOL-HIV Validation Group. AIDS. 14:1439-47; 2000.

Badia X, Podzamczer D, Garcia M, López-Lavid C C, Consiglio E. A randomized study comparing instruments for measuring health-related quality of life in HIV-infected patients. AIDS. 13:1727-1735; 1999.

Benatar SR, Solomon R. Health Reform and crisis of HIV AIDS in South Africa. The New England Journal of Medicine: Health Policy Report, 351: 81-92; 2004.

Brooks R. EuroQol: the current state of play. Health Policy; 37: 53-72; 1996.

Clark PA, O'Brien K. Fighting AIDS in Sub-Saharan Africa is a public-private partnership a viable paradigm? Med Sci Monit 9:28-39; 2003.

Cleary, S, Boulle A, McIntyre D, Coetzee D. Cost-effectiveness of antiretroviral treatment for HIV-positive adults in a South African township. Health Economics Unit, University of Cape Town. 2003

Delate T, Coons S J. The use of health related quality of life measures in a sample of persons infected with HIV, Clinical Infectious Diseases, 32:e47-e52; 2001.

Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health utilities index (HUI) system for assessing health-related quality of life in clinical studies. Ann Med 33: 375-384; 2001.

Gielen AC, McDonnell KA, Wu AW, O'Campo P, Faden R. Quality of life among women living with HIV: the importance violence, social support, and self care behaviors. Soc Sci Med. 52:315-22; 2001.

Han C, Pulling CC, Telke SE, Huppler Hullsiek K; Terry Beirn Assessing the utility of five domains in SF-12 Health Status Questionnaire in an AIDS clinical trial. AIDS. 16:431-439; 2002.

Hughes J, Jelsma J, Maclean E, Darder M, Tinise X. The health-related quality of life of people living with HIV/AIDS. Disabil Rehabil. 26:371-376; 2004.

Jelsma J, Maclean E, Hughes J, Tinise X, Darder M. An investigation into the health-related quality of life of individuals living with HIV who are receiving ART. AIDS Care 17:579-588; 2005.

Jelsma J, Hansen K, De Weerdt W, De Cock P, Kind P. How do Zimbabweans value health states? Popul Health Metr. 1:11; 2003.

Meunnig P. Designing and conducting cost-effectiveness analyses in medicine and health care, John Wiley & Sons, Inc 2002.

Rabin R, de Charro F. EQ-5D: a measure of health status from the Euroqol Group, Annals of Medicine, 33: 337-343; 2001.

Report on the global AIDS epidemic. UNAIDS 2004. http://www.unaids.org

WHO/MSD/MER/02.3. WHOLQOL-HIV instrument. Mental Health: Evidence and research department of mental health and substance dependence World Health Organisation Geneva. 2003.

WHOQOL (World Health Organization's Quality of Life Instrument) HIV Group. Initial steps to developing the World Health Organization's Quality of Life Instrument (WHOQOL) module for international assessment in HIV/AIDS. World Health Organization, Geneva, Switzerland. AIDS Care. 15:347-357; 2003.

WHOQOL-HIV Group. Introducing the WHOQOL instrument. World Health Organisation Geneva. 2003.

Annex II - GIS ACCESSIBILITY INDEX

Shannon Rushworth, GIS consultant, rushworth@sai.co.za

The model was developed by using a WEIGHTED pairwise comparison in IDRISI GIS software and then further a multi-criteria evaluation (MCE) in IDRISI.

WEIGHT is used to develop a set of relative weights for a group of Factors in a multi-criteria evaluation. The weights are developed by providing a series of pairwise comparisons of the relative importance of factors to the suitability of pixels for the activity being evaluated.

These pairwise comparisons are then analyzed to produce a set of weights that sum to 1. The factors and their resulting weights can be used as input for the MCE module for weighted linear combination or ordered weighted average. The pairwise comparison compared the relative importance of the variables as follows:
The eigenvector of weights is: Slopes were the least important variable and weighted relative to other variables as 0.0556
Roads were the most important variable and weighted relative to other variables as: 0.5996
Game reserves as barriers were weighted relative to other variables as: 0.2109
Major rivers as barriers were weighted relative to other variables as: 0.1339

Consistency ratio = 0.10
Consistency is acceptable.

The above pairwise comparison weightings were used in a spatially Explicit multi criteria evaluation, which is a decision support tool. The decision was a choice between alternatives (which area is more accessible than another is). The basis for a decision is known as a criterion. In a Multi-Criteria Evaluation, a set of criteria was combined to achieve a single composite of accessibility.
Criteria included proximity to roads, slope gradient, exclusion of Reserved lands, and so on. Through a Multi-Criteria Evaluation, these criteria images representing suitability were combined to form a single accessibility map from which illustrates areas of good, moderate, moderately poor and poor accessibility.

Annex III - Cox Regression Model

Omnibus Tests of Model Coefficients[a,b]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8402.247	214.759	7	.000	206.422	7	.000	206.422	7	.000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8608.669

b. Beginning Block Number 1. Method = Enter

Variables in the Equation

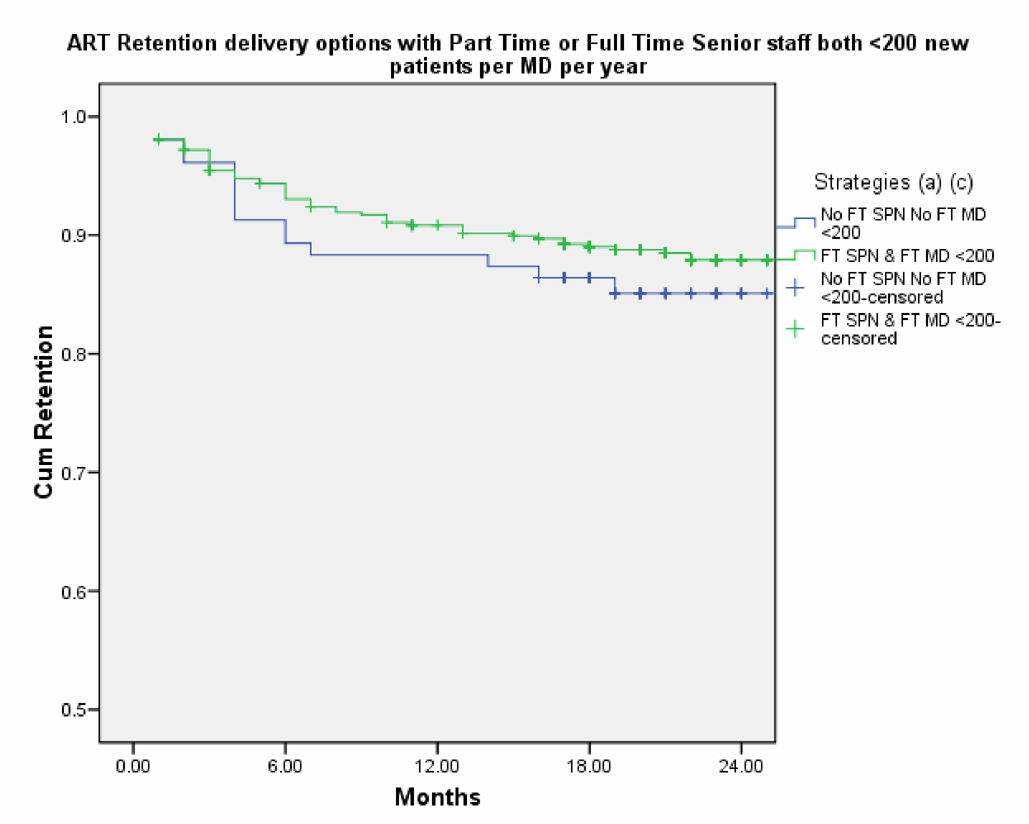
	B	SE	Wald	df	Sig.	Exp(B)
cd4_1_coded_bis	.346	.090	14.603	1	.000	1.413
tel_comb_2	.454	.086	27.762	1	.000	1.574
gender_comb_2	.274	.087	9.889	1	.002	1.315
security	.362	.114	10.134	1	.001	1.436
MD_FT_Y_N	.420	.095	19.529	1	.000	1.522
SPN_FT_Y_N	.280	.093	9.038	1	.003	1.323
newP_Y_MD_PSP_ CMO_LT200_MT200	.807	.087	86.820	1	.000	2.240

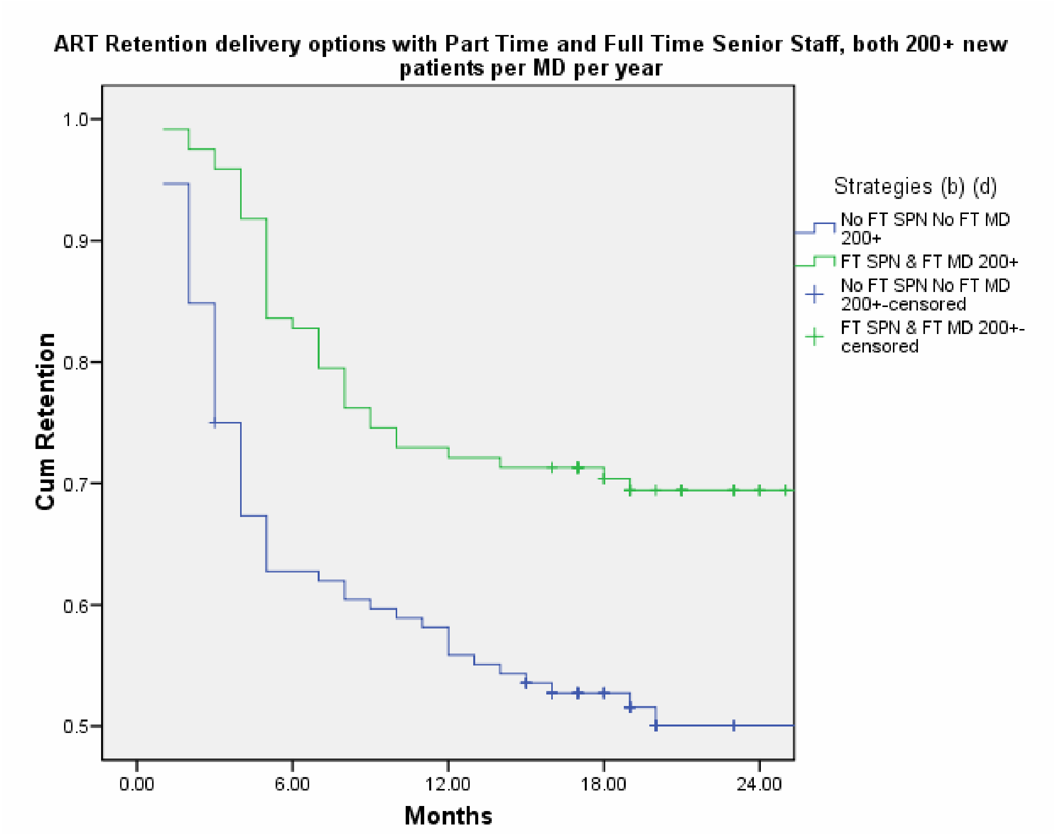
Covariate Means

	Mean
cd4_1_coded_bis	.531
tel_comb_2	.332
gender_comb_2	.383
security	.127
MD_FT_Y_N	.325
SPN_FT_Y_N	.518
newP_Y_MD_PSP_ CMO_LT200_MT200	.386

Cd4_1_coded_bis = initial CD4 <100µl vs. CD4 100-199 µl
Tel_comb_2 = no telephone contact number on registration form vs. presence of a telephone contact number
Gender_comb_2 = male vs. female
Security = delivery site where the manager replied that security impeded access vs. delivery site where there was no such problem
MD_FT_Y_N = delivery site with at least one full time doctor vs. delivery site with part time doctor
SPN_FT_Y_N = delivery site with at least one full time senior professional nurse vs. delivery site with part time senior professional nurse
newP_Y_MD_PSP_CMO_LT200_MT200 = less than 200 new patients per doctor per yeas vs. 200 or more new patients per doctor per year

Annex IV – Kaplan Meier curves, probability of remaining on ART





Annex V

Procedures for calculating unit costs of the antiretroviral therapy in KwaZulu-Natal

…………………………………

Compiled by:
Health Outcomes Research Unit
University of Kwa Zulu-Natal

………………………………..…………

This Annex described the data collection and cost estimation that was used for the 32 ART sites included in the economic evaluation for the period 1/4/05-31/3/06. The resources measured included:

Staff - health care professionals and support staff.
Drugs - medications and dietary supplements.
Lab and Radiology – tests and x-rays
Equipment - medical and non-medical, includes vehicles.
Supplies & materials - medical and non-medical materials e.g. food and sterilised supplies
Utilities - phones, water, gas and electricity.
Buildings and land, where ART programme takes place.

Data on both the consumption levels and the costs of each of these resources were used to calculate the costs incurred by each of the 32 delivery sites during the period related to the Financial Year 2005/06. The procedures for data collection for each resource are discussed below:

(A) Staff

The full staff lists were collected, with work location(s) job title, and salary grade from the programme manager (or a designated responsible person). This information was collected once only, at the beginning of the study.

The names of the staff were entered in the data forms, which was in the layout of a staff roster form. These were circulated to the programme manager, who verified the staff list.

Every month and for each member of staff, staff rosters were collected, stating the location and hours of work.

Every month, the following were identified: overtime hours worked, how many of these were paid, and the size of payment for these. The information was gathered from the accounts department and from the staff themselves. As a quick and easy method of ascertaining the importance of the costs of overtime, the programme manager was asked about the regularity of overtime hours worked, and overtime payments made.

Every month and for each member of staff, gross salaries (including tax, bonuses and allowances) were obtained from the accounts department (payroll).

(B) Drugs, medications & supplements

To estimate the costs of these, a bottom up approach was used.

Data forms were provided to the pharmacists allocated to the programme, for information on drug name and drug quantity given to the patients.

The unit prices of all the relevant drugs were located from: invoices and records in the pharmacy, stores or accounts departments.

The data forms for every month were filled.

(C) Equipment

The cost of each item of equipment was estimated as follows:

A full list of equipment being used by the programme was drawn (including equipment shared with other departments), stating its location, age, and expected length of service (separate medical and non-medical equipment). The age of equipment was available from lists held at the stores department (or buying office). The expected length of service was estimated, using informed guesses from either the users of that equipment, or government replacements rules.

The proportion of time; the equipment was used for the ARV programme was estimated by asking to the users of that equipment.

At the beginning of the study, the historic cost and replacement value of all items of equipment was obtained from the accounts departments, or from private suppliers. The scrap value (estimated market value of used equipment) was obtained from the hospital directors.

Whenever possible, it was specified (a) whether the equipment was imported, and (b) whether it was supplied by the government or direct from the manufacturer.

The accounts and maintenance departments was queried about the annual service, repair and maintenance costs of the equipment, and running costs, such as fuel.

(D) Supplies and materials

.
The sources of information for the quantity of supplies consumed came from the programme.

A Top-down method was used to estimate the costs of supplies and materials

First it was identified where the programme received their supplies from and if supplies came straight from a supplier.

The items were listed according to description, quantity of each item received by the programme every week.

The costs were sourced from invoices and records kept in the stores, the accounts department or suppliers.

(E) Utilities

These include telephone, water, gas and electricity.

The monthly cost of each utility was obtained from either payments made to utility companies by hospital accounts departments / Provincial Health Office, or from the expenditure statements. It was verified that payments on invoices added up to the value in expenditure statement, and reasons were asked for differences if they did not.

The total utility cost of the hospital to which the ART clinic was attached was weighed by the ratio between the square meters of the ART unit and the total size of the surface of the hospital.

(F) Laboratory and Radiology

All tests, x-rays, scans etc. undertaken for the programme for the financial year 2005/2006. (Over a twelve month period) was estimated.

Bibliograhy

1. Drummond MF, Jefferson TO (1996) "Guidelines for authors and peer reviewers of economic submissions to the *BMJ*." British Medical Journal 313(3 August): 275-283.

2. Drummond MF, O'Brien B, et al. (1997) Methods for the economic evaluation of health care programmes, Oxford University Press.

3. Shepard DS, Hodgkin D, Anthony YE. (2000) Analysis of hospital costs: a manual for managers. Geneva, WHO.

4. Creese A, Parker D (1994) Cost Analysis in Primary Health Care: A Training Manual for Programme Managers. Geneva, WHO.

5. Jefferson T, Demichelli V and Mugford M. (1996) Elementary Economic Evaluation in Health Care. London, BMJ Publishing Group.

6. Mugford M, Hutton G, Fox-Rushby J (1998). Methods for economic evaluation alongside a multicentre trial in developing countries: a case study from the WHO antenatal care trial. Perinatal and Paediatric Epidemiology 12, supplement 2.

7. Grannemann, TW, Brown RS, et al. (1986). “Estimating hospital costs: A multiple-output analysis.” Journal of Health Economics **5**: 107-127.

8. Hill DB (1976). “Identification of hospital cost determinants: A health planning perspective.” Inquiry **13**(March): 61-70.

9. Jeffers JR and Siebert CD (1974) “Measurement of hospital cost variation: Case mix, sevice intensity, and input productivity factors.” Health Services Research **Winter**: 293-307.

10. Lave JR, Lave LB, et al. (1972). “Hospital cost estimation controlling for case-mix.” Applied Economics **4**: 165-180.

11. Lave JR and Lainhardt S (1976). “The cost and length of a hospital say.” Inquiry **13**: 327-342.

12. Philips MA (1987) “Why do costings?” Health Policy and Planning **2** (3): 255-257.

Annex VI – Unit Cost FY05/06 -

Dr Aamina Peer, Italian Cooperation

Annual Costs for Human Resource (Source: Department of Health KZN)

Professional Nurse: level 6: R75411

Senior Professional Nurse: level 7: R93936

Chief Professional Nurse: level 8: R116658

Medical Officer: level 9: R139302

Senior Medical Officer: level 10: R173868

Principal Medical Officer: level 11: R271797 (all inclusive package)

Pharmacist: level 7: R93936

Senior Pharmacist: level 8: R116658

Chief Pharmacist: level 9: R139302

Administration Clerk: level 4: R51492

Senior Administration Clerk: level 5: R60915 and level 6: R75411

The formula to cost the post is notch plus 29.7% plus housing which is (4488 pa):

Counsellor is 50% of nurse salary (MTCT study and Cleary, 2003)

Referred to circular 83 of 2005 for overtime rates.

Example of calculation of Human Resource costs using secondary data
(Source: Department of Health KZN)

Calculations of Human Resource Costs

Medical Officer (level 9)	
Gross salary	139 302.00
29.7% incentives	53 856.00
Housing	41 372.69
Overtime	74 091.00
Total salary	308 621.69

Professional Nurse (level 6)	
Gross salary	75 411.00
29.7% incentives	22 397.07
Housing	41 372.69
Total salary	139 180.76

Total number of hrs = 48weeks * 5days * 8hrs/day = 1920
Patient contact = 75% * 1920 = 1440
Time worked in health facility with patient contact= 1440 hrs

Total clinical staff cost per hr (1 professional nurse and 1 doctor) = 308 621.69/1440 + 139 180.76/1440 =319.97/hr

Total clinical staff time per patient (nurse and doctor combined) = 18.53 minutes

Cost of clinical staff per ART patient visit = 18.53/60 * 310.97 =96.04

Cost of lay counsellor = 75 411/2 / 1440 = 26.18/hr

Time spent counselling = 30 minutes

Cost of lay counsellor per ART patient visit = 13.09

Price list for drugs (Source: Department of Health KZN)

ANTIRETROVIRAL PRICES: CONTRACT RT 71/2004 ENDING ON THE 31 AUGUST 2005

Drug	***Price***
Abacavir 20mg/ml solution	235.25
Didanosine 100mg tablets	72.25
Didanosine 25mg tablets	68.37
Didanosine 50mg tablets	65.37
Efavirenz 200mg capsules	297.72
Efavirenz 50mg capsules	24.81
Efavirenz 600mg (1a) tablets	204.48
Lamivudine 10mg/ml solution	24.48
Lamivudine 150mg tablets	41.57
Lopinavir/Ritonavir 133/33mg capsules	312.95
Lopinavir/Ritonavir 80/20mg/ml solution	62.59
Nevirapine 10mg/ml solution	199.5
Nevirapine 200mg (1b) tablets	41.61
Ritonavir 100mg capsules	73.02
Ritonavir 80mg/ml solution	62.59
Stavudine 1mg/ml solution	87.99
Stavudine 20mg capsules	22.23
Stavudine 30mg capsules	22.23
Stavudine 40mg capsules	24.16
Zidovudine 100mg capsules	113.21
Zidovudine 10mg/ml solution	25.59
Zidovudine 300mg tablets	77.52
Cotrimoxazole -56	5.75
HIV tests – capillary tubes and buffer soln	602.49
HIV tests – Rapid Determine Tests	537.66
Multivitamin Syrup –Paeds -150ml	20.82
Multivitamin Tabs 5-14 yrs -30s	60.50
Multivitamin Tabs 14yrs and older -30	8.20
Multivitamin Tabs- Centrum Select-60	106.67
Fluconazole Tabs -200mg	83.70
Fluconazole Suspension	118.02
Fluconazole Injection	106.67

__Monthly costs for ART Regimens in FY05-06__

	ADULT	Costs per month	
Adult Regimen 1a:	Efavirenz 600mg –	R 204.48	
	Lamivudine 150mg-	R 41.57	
	Stavudine 40mg -	R 24.16	
		Subtotal	R 270.21
adult Regimen 1b:	Stavudine 40mg	R 24.16	
	Lamivudine 150mg	R 41.57	
	Nevirapine 200mg	R 41.61	
		Subtotal	R 107.31
Adult Regimen 2:	Zidovudine 300mg	R 77.52	
	Didanosine 400mg	R 289.00	
	Ritonavir/Lopinavir	R 312.95	
	33mg/133mg	Subtotal	R 679.47
	PAEDIATRIC	Costs per month	
Paediatric Regimen 1 6-36 m	Lamivudine 10mg/ml solution	R 24.48	
	Lopinavir/Ritonavir 80/20mg/ml solution	R 62.59	
	Stavudine 20mg capsules	R 11.12	
		Subtotal	R 98.19
Paediatric Regimen 1 >= 3 y	Efavirenz 200mg capsules	R 148.86	
	Lamivudine 150mg tablets	R 20.79	
	Stavudine 20mg capsules	R 22.23	
	Subtotal		R 191.88
Paediatric Regimen 2 6-36 m	Didanosine 25mg tablets	R 68.37	
	Efavirenz 200mg capsules	R 99.24	
	Zidovudine 100mg capsules	R 22.23	
		Subtotal	R 189.84
	Didanosine 100mg tablets	R 72.25	
Paediatric Regimen 2 >=3 y	Lopinavir/Ritonavir 133/33mg capsules	R 208.63	
	Zidovudine 100mg capsules	R 126.80	
		Subtotal	R 407.68

Price list for laboratory and radiology services (Source: Department of Health)

Lab Unit Costs

Test	Unit Cost
CD4	53.53
Viral Load	200.39
FBC	17.58
Differential Count	21.46
LFT	45.18
Urea & Electrolytes	24.32
Hepatitis B Sur Antigen	20.57
Pregnancy Test	10.15
Chol/Trig	17.87
Glucose	15.81
RPR	9.62
HIV	20.57
Tb Microscopy	15.91
Lipids	29.16
Lactate	26.88
Uric acid	15.81
LDH	28.71
CK	28.71
Amylase	15.81
Crypto Antigen	92.58
Ca/Phos/Mg	17.78
Pap smear	69.82
Urine culture	32.34
Iron	78.05
CSF Culture	70.34
HIV DNA PCR	255.69

Radiology costs

CXR	74

ANNEX VII - Probabilistic Sensitivity Analysis

Simulations

Extrapolating the results of the analysis to a wider context is the essence of health planning but it is limited by uncertainty. The results of the analysis allowed to predict the costs and the probabilities of remaining on ART during the first two years according to the following staffing and workload profiles:

(a) Part Time (PT) doctor (MD) and PT senior professional nurse (SPN) with less than 200 new patients per doctor per year;
(b) Same staff as above but with 200 or more new patients per doctor per year;
(c) Full Time (FT) MD and FT SPN with less than 200 new patients per doctor per year; and
(d) Same staff as above but with 200 or more new patients per doctor per year.

However, as in any analytical model, these predictions have a certain degree of uncertainty. Both the probabilities of remaining on ART and the costs estimated for each of the delivery options (a) through (d) had standard errors around the average estimates. To check if these uncertainties could have limited the use of the results for policy .directions, a probabilistic sensitivity analysis also known as second order Monte Carlo simulations was carried out.

Simulations can be deterministic or probabilistic. An example of a simulation is the sequence of events associated with medical decisions. The decisions to do a test will produce certain results (positive test), leading to other decisions (to treat or not to treat). All the decisions are associated with the consumption of resources and the probabilities of obtaining certain outcomes. If the probabilities are fixed (e.g. 75% test sensitivity) the simulation is deterministic because no variation is taken into account. In this case the simulation behaves predictably because whatever the number of times the simulation is run, the results will be the same.

Deterministic simulations produce average cost estimates. For example, a simulation could be composed of a hypothetical cohort of thousands of patients passing through cycles lasting six months each. If the model is deterministic, the cohort is subject to the average transition probabilities (e.g. discontinuing ART) with no variation around the average. At each cycle some patients are lost and the remaining number of patients proceeds to the next cycle. After many cycles, simulating a certain number of years worth of treatment, the cohort is gradually reduced in size and those who remain provide the retention rate.

Contrary to the deterministic process, the probabilistic simulations, take into account the variation around the estimates through a random processes (Drummond M *et al* 2001). While the deterministic simulations take into account only the average probabilities (e.g. of leaving ART in the first six months), the Monte Carlo simulations selects among ranges of probabilities around the average estimates (e.g. 95% CI) on the basis of random numbers.

The simulations carried out in this analysis approach real life conditions where the outcome is uncertain and varies within certain confidence intervals. Virtual patients were subject one at a time to the full range of probabilities and costs, within their 95% confidence intervals (CI), predicted to occur according to a certain delivery option and a certain cycle. Taking into account the uncertainty of the predictions allows to draw firmer conclusions about the robustness of the policy recommendations. To understand the structure of the simulation the Markov cycles which were used are described below.

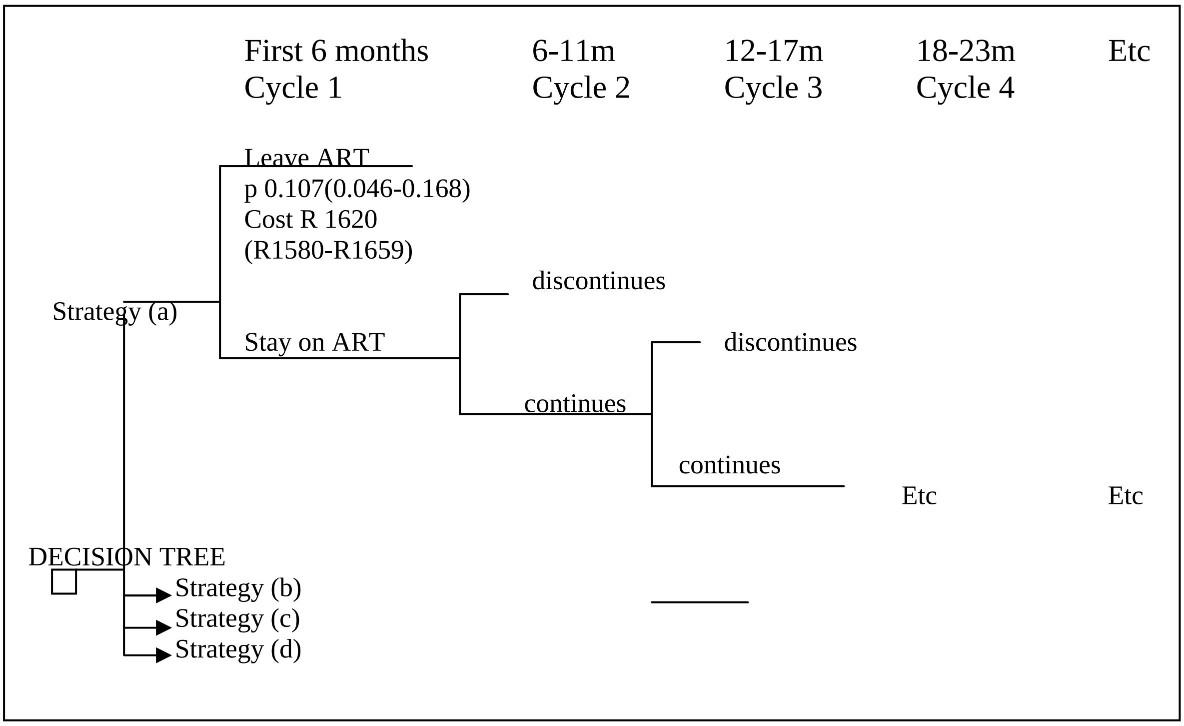
Markov cycles

The transition probabilities and the costs of the delivery options were structured in Markov cycles. These are used in medical decision analysis to track patients while they progress through time periods (Briggs *et al.* 1998). In this simulation, each virtual patient went through Markov cycles of six-month duration, during which, the patient experienced events, which are termed “Markov states”. The states were clearly defined and were mutually exclusive because the virtual patient could only be in one and not in the other state. Within each cycle, the patient could end up in two states: (i) “remained on ART” and (ii) “discontinued ART” because of death, default and any other reason. State (i) meant that the patients continued with the next cycle, while state (ii) meant that the patients stopped contributing to the model, which in Markov terminology is termed as “absorbed” state (Sonnenberg *et al* 1993). Because most Markov simulations are conducted to estimate survival, death is usually the “absorbing state” because it is irreversible and it cannot change into another state. Because this simulation was conducted to estimate ART retention, discontinuing ART for whatever reason was the “absorbed” state.

The transition probabilities and the costs of ending up in state (i) or state (ii) varied across the previously mentioned delivery options described in Part 1 and part 2 of the report. The transition probabilities were predicted by the Kaplan Meier curves of those patients who attended the delivery sites with delivery options (a) through (d) as described in Part 1of the report. The costs per PYO were estimated for those sites which had the staffing and workload profiles (a) through (d) as described in Part 2 of the report. Because each cycle had a six-month duration, the cost per each cycle was half the cost per PYO for the state “remained on ART”. Because it was assumed that on average those discontinuing ART were doing so in the middle of the cycle, the state “discontinued ART” was associated with 0.25 cost per PYO. The degree of uncertainty for both probabilities and costs was provided by the 95% confidence intervals around the average.

Distributions were specified for all transition probabilities and costs in part 2 of the report. For space reasons only the transition probabilities and costs (in R2006) of the state “discontinued ART” of the first cycle is provided in Figure 1, with the 95% CI in parenthesis. One subject at a time was submitted through the cycles according to a stochastic process. Once the first virtual patient entered the first Markov cycle of six months, the software[12] running the simulations selected one random number between 0 and 1. This random (stochastic) process decided the fate of the virtual patient according to the transition probabilities. For example a random number of 0.05 would have been within the 95% confidence intervals (e.g. 0.046 - 0.168) of the probabilities for the state "left ART". If this were the case, the virtual patient would have been "absorbed", thus stopping its contribution to the model in terms of ART retention. A second random number drawn within the 95% CI of the cost (e.g. R1580-R1659) decided the costs for the cycle. When one patient was "absorbed" the next virtual patient entered the simulation and the stochastic process continued through the drawing of random numbers.

Figure 1 Markov models



Uncertainty has been tackled through probabilistic sensitivity analysis (PSA). PSA has tested the uncertainty of the parameters (probabilities and costs) through second order simulations. In each second order simulation, all values of the probabilities and the costs were randomly selected within the 95% CI of the probabilities and costs distributions[13]. This was repeated 1000 times for each delivery option, capturing the uncertainty of the parameters. For each set of values, 1000 virtual patients went through the paths of the decision tree depicted in Figure 1 via a random walk. Each of the 1000 simulations per delivery option produced a different estimate of their cost-effectiveness in retaining patients on ART after 10 years. The spread of the estimates, which has been reported in Figure 5 of Part 2, provided a basis for judging their robustness.

[12] *The simulation was carried out through TreAge Software*

[13] The probbailities and costs had a triangular distribution around their 95% CI