MULTIVARIATE DATA REDUCTION:

A CASE STUDY ON LATENT DETERMINANTS OF SEXUAL HIV RISK AMONG FEMALE COMMERCIAL SEX WORKERS IN KISUMU.

AKINYI JOYCE

I56/60888/2010

A DISSERTATION SUBMITTED IN PARTIAL FULLFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTERS OF SCIENCE IN BIOMETRY OF THE UNIVERSITY OF NAIROBI

29TH NOVEMBER 2012

DECLARATION

DECLARATION BY THE CANDIDATE

This dissertation is my original work carried out at the University of Nairobi during 2011/2012 academic year and has not been presented for the award of any other degree in any university.

Signature.......... Date 29/11/2012

Joyce Akinyi

DECLARATION BY THE SUPERVISOR

This dissertation has been submitted for the partial fulfillment of the requirements of the degree of Master of Science in Biometry with my approval as the supervisor.

Dr. Nelson Owuor

School of Mathematics

University of Nairobi

for

Signature.......... Date 28/11/12

Ogutu Carolyne

DEDICATION

To my sons Malcolm, Martin and Marvin.

ABBREVIATIONS

age2:- Age of the respondent in years at enrolment

educ2:- The respondent's level of education

stay2:- Place of residence of the respondent

startearn2:- Age in years at which the respondent started earning money or goods from sex work

sexyrs2:- The total number of years the respondent had been involved in doing sex work

money2:- The respondent's alternative source of income

clients2:- The place from where the respondent got clients

price2:- The average cash price the respondent charged their clients

clientno2:- The average number of clients the respondent entertained per working day or night.

condomno2:- The average number of clients with whom the respondent used a condom per working day or night

alcohol2:- Whether or not the respondent uses alcohol on their working days or nights.

drugs2 :- Whether or not the respondent uses other drugs like heroine or bhang on their working days or nights

bfriend2:- Whether or not the respondent has a regular sex partner other than her clients

usecondom2:- Whether or not the respondent uses a condom with the regular sex partners

stisynd2:- Any syndromes of sexually transmitted infections at recruitment and during the study period

treatment2:- Nature of treatment sought for the STIs identified

noprg2:- Total number of times the respondent has been pregnant ageprg2:- Age in years of the respondent at their first pregnancy

birthno2:- Total number of times the respondent has given birth

contra2:- Use of any form of contraception by the respondent

care2:- Whether or not an HIV sero-positive respondent was on HIV care at enrolment.

ABSTRACT

A comprehensive understanding of the factors that determine the acquisition and transmission of HIV is fundamental for monitoring the epidemic and for providing treatment, care and support services to the infected and their families. The purpose of this paper was to identify the latent determinants of sexual HIV transmission and acquisition among female commercial sex workers in Kisumu town and its rural environs. Data was collected from January 2006 to December 2008 from 1,647 female commercial sex workers, using snowball sampling method. Exploratory factor analysis with principal component extraction method and varimax rotation was used. The analysis produced eight latent factors; sex work experience, general intelligence at the start of sex work, other intimate relationships outside sex work and their nature, substance abuse, sexually transmitted infections (STIs) and their management, environmental influence, financial status and possible daily exposure to the virus.

Table of Contents

1 INTRODUCTION

The introduction will consist of the background, statement of problem, objectives and the significance of the study.

1.1 BACKGROUND

Thirty years after AIDS was first reported, HIV continues to spread (WHO, 2011). Although 2010 estimates suggest that the annual number of people newly infected with HIV has declined 20% from the global epidemic peak in 1998, an estimated 2.7 million people acquired the virus in the year 2010 alone, (UNAIDS, 2011).

In sub-Saharan Africa, where most of the people newly infected with HIV live, an estimated 1.9 million people became infected in 2010.This was 27% fewer than the annual number of people newly infected from 1996 through 1998, when the incidence of HIV in sub-Saharan Africa peaked overall (UNAIDS, 2011).

In Kenya, national HIV prevalence decreased from about 14% in the mid-1990s to 6% in 2006, and has remained constant (NACC, 2006).

According to the 2007 Kenya Aids Indicator Survey, 100,000 new HIV infections are identified yearly in adults in Kenya and the distribution of HIV infections varies greatly across the country. Prevalence remains the highest in Nyanza at 15.3%, more than double the national prevalence estimate of 6 % (KAIS, 2007).

New surveillance data confirm that the epidemic disproportionately affects sex workers, men who have sex with men, transgender people, people who inject drugs, prisoners and migrants in both concentrated and generalized epidemics (WHO, 2011). One third of all new yearly infections in Kenya are attributed to these people, the Most At Risk Populations (MARPs) (KAIS, 2007).

Continuing evidence indicates that unprotected paid sex is a significant factor in the HIV epidemics in several sub-Saharan African countries (Gelmon L. et al, 2009). It has been postulated that it was a more significant factor in early HIV epidemics in sub-Saharan Africa. However, paid sex can remain an equally important factor in mature epidemics (Chen L. et al, 2007). There is increasing evidence that intensive HIV prevention programmes among Female Sex Workers (FSWs) can be highly effective in reducing the percentage of future infections (Moses, 2011).

There are an estimated 200,000 commercial sex workers (CSWs) in Kenya, 150,000 of whom are women. An estimated 14% of the people acquiring HIV infection in Kenya are linked to sex work (HIV infection among sex workers, their clients or their other sex partners) (Gelmon L. et al, 2009).

Considering the importance of transforming CSWs into agents of change by becoming safer sex practitioners, protecting their clients from Sexually Transmitted Infections (STIs)/HIV, educating their peers and seeking prompt treatment when infected, this paper sought to examine the underlying unobserved risk factors that make CSWs very vulnerable to being infected and infecting their clients with HIV. Factor analysis was considered an ideal instrument to identify the underlying risk factors among CSWs, and to determine whether specific risk factors tended to form patterns.

1.2 STATEMENT OF PROBLEM

Different studies have shown that age (Gray, et al., 2001), gender (Nicolosi, et al., 1994), sexually transmitted infections (Rottingen, et al., 2001), use of hormonal contraceptives (Wang, et al., 1999; Martin, et al., 1998) and certain socioeconomic, demographic and behavioral factors

(Johnson & Budlender, 2002) strongly influence the acquisition and transmission of HIV among female sex workers. However, the pattern of associations within a set of these risk factors needs to be investigated. The risk factors have widely been taken individually, without considering the possibility of certain latent underlying factors that could better help explain them, in order to be addressed effectively.

Factor analysis was therefore considered an ideal instrument to identify a small number of underlying risk factors and patterns in the acquisition and transmission of HIV among female CSWs and to see whether specific risk factors tended to form patterns.

1.3 HYPOTHESIS

The hypotheses of this study were:

1 The unobserved risk factors that determine the acquisition and transmission of HIV among female CSWs can be categorized into four areas; poverty, ignorance, sex work experience and personal attitude.

2 The observed risk factors in the acquisition and transmission of HIV among female CSWs are closely related to one another, and tend to form patterns.

1.4 OBJECTIVES

The objectives consisted of overall and specific objectives.

1.4.1 OVERALL OBJECTIVE

The overall objective of this study was to examine the interrelation of the risks of HIV acquisition and transmission in a population of female commercial sex workers in Kisumu, by using factor analysis, in order to detect the unobserved underlying causes.

1.4.2 SPECIFIC OBJECTIVES

The specific objectives of the study were;

1. To determine the unobserved risk factors in the acquisition and transmission of HIV among female CSWs in Kisumu.
2. To determine the relationships between the observed risk factors.
3. To demonstrate the application of factor analysis to this type of data.

1.5 SIGNIFICANCE OF THE STUDY

The study aimed to identify the unobserved risk factors in the acquisition and transmission of HIV among female CSWs in Kisumu. This is a marginalized group in society who are difficult to reach with HIV prevention, treatment and care since the extent to which HIV is affecting them has not yet been fully explored. The study findings would provide a clearer insight into these risk factors and help in terms of learning how to better monitor the epidemic and provide treatment, care and support services to the infected and their families.

This study also aimed at contributing to the existing knowledge of the application of factor analysis in data reduction.

2 LITERATURE REVIEW

Commercial sex work is defined in this paper as the regular exchange of sexual services for money or goods. This exchange involves a set of actors, including the sex worker, the client, and sometimes a third party (Overs, 2002).

Sex work remains an important contributor to the transmission dynamics of HIV within early, advanced and regressing epidemics in sub-Saharan Africa (Chen, et al., 2007). HIV prevalence among sex workers and their clients today is 10-20 times higher than among the general population (Scorgie, et al., 2011). With high rates of client change, the potential for onward transmission of HIV from an infected sex worker to other clients or partners may be more than 100 times greater than from other people living with HIV (WHO, 2008).

In East Africa 1/3 of female sex workers had HIV, though levels of up to 75% were documented in Kisumu, Kenya in the early 2000s (Morison, et al., 2001;Aklilu, et al., 2001).

Studies have shown several determinants of HIV transmission and acquisition among female CSWs. Dunkle found that the risk of HIV infection among female sex workers is determined by the total number of unprotected sex acts with an HIV-infected partner and the efficiency of HIV transmission, which were marked by higher number of clients, duration of sex work and inconsistent condom use (Dunkle, 2005). The CSWs place of residence and work influences their condom use practices and hence their chances of contracting and transmitting HIV (Ntumbanzondo, 2006).Sex workers with harmful alcohol use have a higher risk of HIV infection (Chisholm, 2004). Age, gender, sexually transmitted infections, use of hormonal contraceptives and certain socioeconomic, demographic and behavioral factors have also been proven to strongly influence the acquisition and transmission of HIV among female sex workers.

Factor analysis has been applied in several fields of study to show relationships and to uncover underlying patterns in risk factors. Exploratory factor analysis seeks to uncover the underlying structure of a relatively large set of risk factors and the researcher's à priori assumption is that any indicator may be associated with any factor (Hare, et al, 1998).

Schute A. E, et al used factor analysis to model the possible risks of hypertension in a black South African population. From 23 risks the factor analysis disclosed five factors that explained 56.2% of the variance in the male and 43.5% of the variance in the female group (Schutte, et al., 2003).

In a study published in the International Journal for the Advancement of Counseling, Kimemia M, et al found that five factors, with strong loadings on Emotional Support and Instrumental Support explained coping mechanisms among caregivers for family members living with HIV/AIDS in Kenya (Kimemia, et al., 2011).

3 METHODOLOGY

This section will consist of a description of the data and that of factor analysis.

3.1 DATA

Data from a cohort study carried out in Kisumu town and its rural environs from January 2006 to December 2008 was used. 1647 female commercial sex workers were recruited, with the inclusion criteria being that one must be a commercial sex worker and willing to participate in the study. The sample was taken using snowball sampling method. This is a non-probability sampling technique commonly used in populations that are difficult to access. The recruitment was done in bars with the help of bar tenders.

The participants were followed for a period of three years, during which they consented to be tested for HIV and STIs at three month time intervals. Treatment was provided when necessary.

The variables measured were; age in years of the respondent at enrolment, level of education, place of residence, age in years at which they first earned money from sex work, number of years as a CSW, any other source of income and its nature, the price they charged their clients, the average number of clients per working day or night, the number of clients with whom they used a condom on their last working day, where they got their clients from, use of alcohol, use of other hard drugs, their marital status, whether or not they had a steady boyfriend and if they used a condom with him, history of STIs and the manner of treatment, the total number of pregnancies, the age at first pregnancy , current use of contraceptives and whether or not they were on HIV care at the time of recruitment. The outcome variable was their HIV status, being negative or positive.

The level of education was measured in 6 levels; none, primary incomplete, primary complete, secondary incomplete, secondary complete and post secondary. Marital status was measured in 4 levels; married, separated, single and divorced. The place of residence was classified as either in Kisumu town or a village outside. Age of the respondent, age at first pregnancy, age at which they first earned money from sex work and the number of years they have done this work were all measured in calendar years. The price they charge their clients, the average number of clients per working day or night and the number of clients with whom they used a condom on their last working day were all in numerical units. The history of STIs, manner of treatment, steady boyfriend, use of condom with the steady boyfriend, use of contraceptives, any other source of income, use of alcohol, use of other drugs and being on HIV care all had two levels; yes or no. The nature of source of income varied among participants, and the place where they got their clients was at home or in a brothel.

The data was entered into Ms Excel and then imported to SPSS version 17 for analysis.

Factor analysis with principal component extraction method and varimax rotation was used to analyze.

3.2 FACTOR ANALYSIS

Factor analysis is a method for investigating whether a number of variables of interest

$$X_1, X_{2,} \ldots, X_p$$

are linearly related to a smaller number of unobservable factors

$$F_1, F_2 \ldots F_q.$$

Each observable variable is a linear function of independent factors and error terms.

It is all about studying the co-variance (or correlation) and is based on a statistical model. This analysis describes the covariance relationships between many variables in terms of a few underlying, unobservable random quantities called factors. If there is a group of highly correlated variables, which in turn are uncorrelated with other variables, these represent realizations of some underlying phenomena that is responsible for the observed correlations. Factor analysis is normally carried out with a view to reification: the investigator usually has a conceptual model of some underlying entity which cannot be measured directly. These latent, or hidden, variables are the factors in factor analysis.

The aim of factor analysis is that each of the p observed variables can be represented by means of $q < p$ mutually uncorrelated common factors. This will leave some uncorrelated residual specific to each of the observed variables, the uniqueness, which is not correlated with any of the remaining $p-1$ variables.

Note that the diagonal of a correlation matrix is 1.Only part of this 1 is due to the $q < p$ latent variables - this part is known as the communality.

It is possible to rotate the q axes of common factors to new orthogonal or oblique axes to make the factor solution fit with existing theoretical ideas regarding the model.

3.2.1 The factor analysis model

The orthogonal model underlying factor analysis can be described as follows:

$$x = \mu + \Gamma\Phi + \varepsilon \qquad \text{(Eq.1)}$$

Where x is a $1 \times p$ random vector. μ represents a vector of unknown constants (mean values), Γ is an unknown $p \, X \, q$ matrix of constants referred to as the loadings. Φ is a $q \times 1$ unobserved random vector referred to as the scores. ε is $1 \times p$ unobserved random error vector having mean 0 and a diagonal covariance Ψ referred to as the uniqueness or specific variance.

3.2.2 Model assumptions

The model of factor analysis is based on two assumptions concerning the relationships.

These are;

1) The error terms ε_i are independent of one another, and are such that

$$E(\varepsilon) = 0 \text{ and } Cov(\varepsilon) = \Psi I$$

2) The unobservable factors Φ_i are independent of one another and of the error terms, and are such that

$$E(\Phi) = 0 \text{ and } \sum\nolimits_{\Phi} = I$$

With these assumptions, $\operatorname{cov}(\Phi, \varepsilon) = 0$, if $\sum_{\Phi} = I$ then $\operatorname{cov}(x, \Phi) = \Gamma$. It is worth emphasizing that unlike many multivariate techniques, factor analysis is a statistical model for our observations, with the following distributional form:

$$x \sim Normal(\mu, \Gamma\Gamma^T + \Psi)$$

3.2.3 Modeling a vector of variables

A vector of variables $x = x_1, \ldots, x_p$ is modeled as follows in factor analysis:

$$x_1 = \mu_1 + \sum_{k=1}^{q} \lambda_{1k}\Phi_k + \varepsilon_1$$

$$x_2 = \mu_2 + \sum_{k=1}^{q} \lambda_{2k}\Phi_k + \varepsilon_2$$

.

.

.

$$x_p = \mu_p + \sum_{k=1}^{q} \lambda_{pk}\Phi_k + \varepsilon_p$$

We note that under the terms of this model

$$\mathrm{var}(x_j) = \lambda_{j1}^2 + \lambda_{j2}^2 + \ldots + \lambda_{jq}^2 + \mathrm{var}(\varepsilon_j) \quad \text{(Eq. 2)}$$

One potential problem with this model is that there can be more parameters than data. For example, note that the covariance matrix $\sum$ has $p(p+1)/2$ parameters, the factor model ($\Gamma\Gamma^T + \Psi$) has $qp - q(q-1)/2 + p$ parameters. One issue arises whereby a factor analysis model must be constrained in order to ensure identifiability. Clearly,

$$p(p+1)/2 \geq qp - q(q-1)/2 + p,$$

Or:

$$q \leq \frac{2p+1-\sqrt{8p-1}}{2} \quad \text{(Eq. 3)}$$

Where $q < p$, the right side of equation 2 indicates how much of $\mathrm{var}(x_j)$ is explained by the model, a concept referred to as the communality.

Consideration of the order of the model leads us to the degrees of freedom, calculated as follows;

$$df = \frac{p(p+1)}{2} - qp + \frac{q(q-1)}{2} - p = \frac{(p-q)^2 - (d+m)}{2} \quad \text{(Eq. 4)}$$

3.2.4 <u>Centered and standardized data</u>

In practice it is often much simpler to centre the data, so that we model:

$$x_j - \mu_j = \sum_{k=1}^{q} \lambda_k \Phi_k + \varepsilon_j ; j = 1, \ldots, p$$

or even to standardize the variables so that in effect we are modeling the correlation matrix rather than the covariance matrix.

$$\frac{x_j - \mu_j}{\sigma_{jj}} = \sum_{k=1}^{q} \lambda_k \Phi_k + \varepsilon_j ; j = 1, \ldots, p$$

Regardless of the data matrix used, factor analysis is essentially a model for $\sum$, the covariance matrix of x,

$$\sum = \Gamma\Gamma^T + \psi \qquad \text{(Eq. 5)}$$

3.2.5 Factor indeterminacy

Another problem with factor analysis is that it is a very indeterminate model. Specifically, it is unchanged if we replace Γ by $k\Gamma$ for any orthogonal matrix k. However, this can be turned to our advantage; with sensible choice of a suitable orthogonal matrix k we can achieve a rotation that may yield a more interpretable answer. Factor analysis therefore requires an additional stage, having fitted the model we may wish to consider rotation of the coefficients.

3.2.6 Strategy for factor analysis

To fit the model, we need to:

_ Estimate the number of common factors q

_ Estimate the factor loadings Γ

_ Estimate the specific variances ψ^2

_ On occasion, estimate the factor scores ϕ

3.2.7 Fitting methods for factor analysis

The two most commonly used fitting methods are the principal component method and the maximum likelihood method. The maximum likelihood (ML) method is an iterative method that is computationally more demanding and is prone to Heywood cases, nonconvergence, and multiple optimal solutions. The principal component method is computationally efficient and has similarities to principal component analysis.

3.2.8 **Principal component extraction**

We use the spectral decomposition to obtain one possible factoring of the covariance matrix $\sum$

$$\sum = E\Lambda E^T$$

Which can be expanded as:

$$\sum = \lambda_1 e_1 e_1^T + \lambda_2 e_2^T + \ldots + \lambda_p e_p^T$$

$$= (\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \ldots, \sqrt{\lambda_p}e_p)\begin{pmatrix} \sqrt{\lambda_1}e_1 \\ \sqrt{\lambda_2}e_2 \\ . \\ . \\ . \\ \sqrt{\lambda_p}e_p \end{pmatrix}$$

In practice we don't know $\sum$ and we use S (or we standardize the variables and use R).Spectral decomposition yields linear principal components as follows:

$$z_1 = e_{11}x_1 + e_{12}x_2 + \ldots + e_{1p}x_p; \mathrm{var}(z_1) = \lambda_1$$

$$z_2 = e_{21}x_2 + e_{22}x_2 + \ldots + e_{2p}x_p; \mathrm{var}(z_2) = \lambda_2$$

.

.

.

$$z_p = e_{p1}x_2 + e_{p2}x_2 + \ldots + e_{pp}x_p; \mathrm{var}(z_p) = \lambda_p$$

Which in matrix notation can be expressed as:

$$Z = EX \qquad \text{(Eq. 6)}$$

Where $Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_p \end{pmatrix}$, $X = \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_p \end{pmatrix}$ and $E = \begin{pmatrix} e_{11} & e_{12}\cdots & e_{1p} \\ e_{21} & e_{22}\cdots & e_{2p} \\ \vdots & \vdots & \vdots \\ e_{p1} & e_{p2}\cdots & e_{pp} \end{pmatrix}$

Multiplying both sides of (Eq. 6) by E^{-1} gives:

$$E^{-1}Z = X$$

For orthogonal matrices $E^{-1} = E^T$ so we can invert the transformation by using

$$X = E^T Z$$

This can be expanded as:

$$x_1 = e_{11}z_1 + e_{21}z_2 + \ldots + e_p z_p$$

$$x_2 = e_{12}z_2 + e_{22}z_2 + \ldots + e_p z_p$$

.

.

.

$$x_p = e_{1p}z_p + e_{2p}z_2 + \ldots + e_{pp}z_p$$

Which we could express as;

$$x_1 = (e_{11}\sqrt{\lambda_1})\frac{z_1}{\sqrt{\lambda_1}} + (e_{21}\sqrt{\lambda_2})\frac{z_2}{\sqrt{\lambda_2}} + \ldots + (e_{p1}\sqrt{\lambda_p})\frac{z_p}{\sqrt{\lambda_p}}$$

$$x_2 = (e_{12}\sqrt{\lambda_1})\frac{z_1}{\sqrt{\lambda_1}} + (e_{22}\sqrt{\lambda_2})\frac{z_2}{\sqrt{\lambda_2}} + \ldots + (e_{p2}\sqrt{\lambda_p})\frac{z_p}{\sqrt{\lambda_p}}$$

.

.

.

$$x_p = (e_{1p}\sqrt{\lambda_1})\frac{z_1}{\sqrt{\lambda_1}} + (e_{2p}\sqrt{\lambda_2})\frac{z_2}{\sqrt{\lambda_2}} + \ldots + (e_{pp}\sqrt{\lambda_p})\frac{z_p}{\sqrt{\lambda_p}}$$

and if we set $\gamma_{jk} = (e_{jk}\sqrt{\lambda_j})$ and $\phi_j = z_j / \sqrt{\lambda_j}$ we have a clear link with the factor analysis model given in Eq. 1. If we try writing this in matrix terminology, our loadings matrix Γ is the $p \times p$ matrix where the j[th] column is given by $\sqrt{\lambda_j} e_j$ we now have:

$$S = \Gamma\Gamma^T$$

which is getting us part of the way to our factor analysis model. Note that under the principal component solution, the estimated loadings do not alter as the number of factors is increased or decreased.

We don't actually want to use a decomposition with $q = p$ variables. We partition Λ into $\Lambda_1 = \lambda_1, \lambda_2, \ldots, \lambda_q$ and $\Lambda_2 = \lambda_{q+1}, \ldots, \lambda_p$ with the corresponding eigenvectors. As a consequence, we reduce the size of Γ matrix, i.e. to neglect the contribution of $\lambda_{q+1} e_{q+1} e_{q+1}^T + \ldots \lambda_p e_p e_p^T$.

So when considering our model for the data, we wish to partition our factors as follows:

$$x_1 = e_{11}z_1 + e_{21}z_2 + \ldots + e_{q1}z_q + e_{q+1,1}z_{q+1} + \ldots + e_{p1}z_p$$

$$x_2 = e_{12}z_2 + e_{22}z_2 + \ldots + e_{q2}z_q + e_{q+1,2}z_{q+1} + \ldots + e_{p2}z_p$$

.

.

.

$$x_p = e_{1p}z_p + e_{2p}z_2 + \ldots + e_{qp}z_q + e_{q+1,p}z_{q+1} + \ldots + e_{pp}z_p$$

And if we set

$$e_{q+1,j}z_{q+1} + \ldots + e_{pj}z_p = \zeta_j ; j = 1, \ldots, p$$

we can rewrite this as:

$$x_1 = e_{11}z_1 + e_{21}z_2 + ... + e_{q1}z_q + \zeta_1$$

$$x_2 = e_{12}z_1 + e_{22}z_2 + ... + e_{q2}z_q + \zeta_2$$

.

.

.

$$x_p = e_{1p}z_1 + e_{2p}z_2 + ... + e_{qp}z_q + \zeta_p$$

This can be expressed as:

$$x_1 = (e_{11}\sqrt{\lambda_1})\frac{z_1}{\sqrt{\lambda_1}} + (e_{21}\sqrt{\lambda_2})\frac{z_2}{\sqrt{\lambda_2}} + ... + (e_{q1}\sqrt{\lambda_q})\frac{z_q}{\sqrt{\lambda_q}} + \zeta_1$$

$$x_2 = (e_{12}\sqrt{\lambda_1})\frac{z_1}{\sqrt{\lambda_1}} + (e_{22}\sqrt{\lambda_2})\frac{z_2}{\sqrt{\lambda_2}} + ... + (e_{q2}\sqrt{\lambda_q})\frac{z_q}{\sqrt{\lambda_q}} + \zeta_2$$

.

.

.

$$x_p = (e_{1p}\sqrt{\lambda_1})\frac{z_1}{\sqrt{\lambda_1}} + (e_{2p}\sqrt{\lambda_2})\frac{z_2}{\sqrt{\lambda_2}} + ... + (e_{qp}\sqrt{\lambda_q})\frac{z_q}{\sqrt{\lambda_q}} + \zeta_p$$

Where $\gamma_{jk} = (e_{jk}\sqrt{\lambda_j})$ and $\phi_j = z_i / \sqrt{\lambda_i}$. We notice that $\mathrm{var}(\zeta) = \psi$. If we consider this in terms of the decomposition matrix we have:

$$\Sigma=(\sqrt{\lambda_1}e_1,\sqrt{\lambda_2}e_2,...,\sqrt{\lambda_q}e_q)\begin{pmatrix}\sqrt{\lambda_1}e_1\\ \sqrt{\lambda_2}e_2\\ \cdot\\ \cdot\\ \cdot\\ \sqrt{\lambda_q}e_q\end{pmatrix}+\begin{pmatrix}\psi_1 & 0... & 0\\ 0 & \psi_2... & 0\\ \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot\\ 0 & 0... & \psi_p\end{pmatrix}$$

Where $\psi_j = \mathrm{var}(\zeta_j)=\sigma_{jj}-\sum_{k=1}^{q}\gamma_{jk}^2 \, for k=1,2,...,q.$

Estimates of the specific variances are given by the diagonal elements of the matrix $\hat{\Sigma}-\hat{\Gamma}\hat{\Gamma}^T$,

i.e: $\hat{\psi}=\begin{pmatrix}\psi_1 & 0... & 0\\ 0 & \psi_2... & 0\\ \vdots & \ddots & \vdots\\ 0 & 0\cdots & \psi_p\end{pmatrix}$ with $\psi_j=\sigma_{jj}-\sum_{k=1}^{q}\gamma_{jk}^2$

When using the principal component solution of $\hat{\Sigma}$, it is specified in terms of eigenvalue-eigen vector pairs $(\hat{\lambda}_1,\hat{e}_1),(\hat{\lambda}_2,\hat{e}_2),...,(\hat{\lambda}_p,\hat{e}_p)$, where $\hat{\lambda}_1\geq\hat{\lambda}_2\geq...\geq\hat{\lambda}_p$. If we wish to find a $q<p$ solution of common factors, then the estimated factor loadings are given by:

$$\hat{\Gamma}=(\lambda_1^{\frac{1}{2}}e_1,\lambda_2^{\frac{1}{2}}e_2,...,\lambda_q^{\frac{1}{2}}e_q)$$

The factors ϕ have identity covariance matrix

$$\mathrm{var}(\phi)=\mathrm{var}(\sqrt{\Lambda_1}\Gamma_1^T(x-\mu))=I_q$$

and are uncorrelated with the residuals:

$$\mathrm{cov}(\phi,\zeta)=\mathrm{cov}(\sqrt{\Lambda_1}\Gamma_1^T(x-\mu),\Gamma_2\Gamma_2^T(x-\mu))=\sqrt{\Lambda_1}\Gamma_1^T\Sigma\Gamma_2\Gamma_2^T=0$$

However, it can be seen that each ζ_i contains the same z_i so they are not mutually unrelated. Hence the latent variables obtained using the principal component method do not explain all the correlation structure in our data X. The covariance matrix for the error is:

$$\text{var}(\zeta) = \Gamma_2 \Lambda_2 \Gamma_2^T$$

3.2.9 <u>Diagnostics for the factor model</u>

We can define a residual matrix as:

$$\varepsilon = S - (LL^T + \psi)$$

By construction, the diagonal elements of this residual matrix will be zero. A decision to retain a particular q factor model could be made depending on the size of the off-diagonal elements. Rather conveniently, there is an inequality which gives us:

$$[\varepsilon = \hat{\Sigma} - (LL^T + \psi)] \leq \hat{\lambda}_{q+1}^2 + \ldots + \hat{\lambda}_p^2$$

So it is possible to check the acceptability of fit in terms of a small sum of squares of neglected eigenvalues.

In a similar manner to that used in principal components, it is possible to use the eigenvalues to indicate the proportion of variance explained by any given factor. So instead of examining discarded components we could examine those we intend to retain. Bearing in mind that $trace(\Sigma) = \sigma_{11} + \sigma_{22} + \ldots + \sigma_{pp}$, we know that the amount of variation explained by the first factor

$$\gamma_{11}^2 + \gamma_{21}^2 + \ldots + \gamma_{p1}^2 = (\sqrt{\lambda_1} e_1)^T (\sqrt{\lambda_1} e_1) = \lambda_1$$

So we know that the j-th factor explains the following proportion of total sample variance:

$$\frac{\lambda_j}{trace(S)}$$

which reduces to $\frac{\lambda_j}{p}$ when using standardized variables (the correlation matrix).

The Kaiser criterion was developed in the context of factor analysis. This is implemented by default in a number of computer programs, basically we retain factors which are explaining more than the average amount of variance; if we are decomposing the correlation matrix we retain all factors where the corresponding eigenvalues are greater than one (Hewson, 2009).

3.2.10 Communalities

In the case of standardized variables, the communalities indicate the proportion of variance of a manifest variable explained by its relevant factor structure. They are estimated as:

$$\xi_{jk}^2 = \gamma_{j1}^2 + \gamma_{j2}^2 + \ldots + \gamma_{jq}^2$$

$$\text{var}(x_j) = \underbrace{\gamma_{j1}^2 + \gamma_{j2}^2 + \ldots + \gamma_{jq}^2}_{communality\_of\_x_j} + \underbrace{\psi_i}_{specificity\_of\_x_j}$$

For standardized variables, $\text{var}(x_j) = 1$, therefore: $\gamma_{i1}^2 + \gamma_{i2}^2 + \ldots + \gamma_{iq2} \leq 1$ and $-1 \leq \gamma_{jk} \leq 1$

These are extracted from the matrix of loadings by squaring all entries and summing by row. The communalities are the proportions of the variance of the original variables that can be attributed to the common factors. As such, they should be in the interval [0, 1]. However, factor analyses that use iterative fitting estimate the communality at each iteration. For some data, the estimate might equal (or exceed) 1 before the analysis has converged to a solution. This is known as a Heywood (or an ultra-Heywood) case and it implies that one or more unique factor has a nonpositive variance.

3.2.11 Principal Factor solution

The diagonal elements of our covariance matrix are given by $\sigma_{jj} = \xi_j^2 + \psi_j$, so having determined the number q of common factors needed, we can decompose the reduced covariance matrix. If we obtain some initial estimates of ψ we can re-estimate the remaining part of the decomposition $\Gamma\Gamma^T$.

$$\sigma_{jj} = \xi_j^2 + \psi_j$$

If we had some initial estimate of ψ, $\tilde{\psi}$ say, we could obtain a "reduced" covariance matrix

$$S = \begin{pmatrix} \tilde{\xi}_1^2 & s_{12} \cdots & s_{1p} \\ s_{21} & \tilde{\xi}_2^2 \cdots & s_{2p} \\ \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} \cdots & \tilde{\xi}_p^2 \end{pmatrix}$$

and carry out an eigen decomposition of this matrix, updating our estimates of the uniqueness and repeat until convergence.

So all we need is an initial estimate of $\tilde{\psi}$. Many programs conduct a multiple regression of each manifest variable on each other, and use $s_{jj}r_j^2$. We then conduct a principal component analysis on $S - \psi$ to find Γ. ψ can then be recalculated as the diagonal of $S - \Gamma\Gamma^T$ and we extract a further set of principal components. These latter steps are repeated until convergence.

3.2.12 Rotation

Rotation yields a more interpretable factor structure. We seek a rotation:

$$\hat{\Gamma}^{(R)} = \hat{\Gamma}T$$

such that we obtain easy-to-interpret factor loadings. Where possible some components would be large, others small. The most obvious way to do this is to carry out the exercise by eye, and to rotate the axes around the origin so that some factor loadings become small. It is also easy to suggest a two dimensional rotation matrix:

$$T = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$$

for rotation angle $\phi; -\pi \leq \phi \leq \pi$ .We need to find a suitable value for ϕ .

For orthogonal rotations, two objective criteria are most commonly used to determine the optimal rotation: the Varimax procedure and the Quartimax procedure.

The Varimax procedure looks for a rotation which maximizes the objective V:

$$V = \frac{1}{p^2} \sum_{j=1}^{q} (p \sum_{j=1}^{p} \left[\frac{\gamma_{jk}^2}{\xi_j^2} \right]^4 - \left[\sum_{j=1}^{p} \left[\frac{\gamma_{jk}^2}{\xi_j^2} \right] \right]^2)$$

Where $\xi_i^2 = \sum_{k=1}^{q} \gamma_{jk}^2$ is the communality for each of the j variables as before.

3.2.13 <u>Factor scoring</u>

There are occasions where we may wish to estimate values for ϕ_i for a given individual *i*.

These values are referred to as the scores, the process of estimating them, which has to be carried out after Γ and ψ have been estimated is therefore referred to as scoring.

Bartlett (1937, 1938) proposed a method based upon weighted least squares.

Once we have estimates

$$x_1 - \bar{x}_1 = \sum_{k=1}^{q} \gamma_{1k} \hat{\phi}_1 + \zeta_1$$

$$x_2 - \bar{x}_2 = \sum_{k=1}^{q} \gamma_{2k} \hat{\phi}_2 + \zeta_2$$

.

.

.

$$x_p - \bar{x}_p = \sum_{k=1}^{q} \gamma_{pk} \hat{\phi}_p + \zeta_p$$

we need to estimate ϕ_j for $j = 1, 2, ..., q$. However as $\text{var}(\zeta_j) = \psi_j$ are not equal he argued that weighted least squares was the most appropriate technique.

The weighted least squares estimates thus obtained are:

$$\hat{\phi}_i = (\Gamma^T \Psi^{-1} \Gamma) \Gamma^T \Psi (x_i - \bar{x})$$

Thomson (1951) is based on assuming that both ϕ and ζ are multivariate normal, thus a concatenation of the manifest (x) and latent (ϕ) variables $y^T = (\phi^T, x^T)$ will also be normal with dispersion matrix:

$$\text{var}(y) = \begin{pmatrix} I & \Gamma^T \\ \Gamma & \Gamma\Gamma^T + \psi \end{pmatrix}$$

The mean of ϕ is zero by definition, therefore:

$$E(z/x_0) = \Gamma^T (\Gamma\Gamma^T + \Psi)^{-1} (x_o - \mu)$$

which gives the estimate for the scores as:

$$z = \hat{\Gamma}^T (\hat{\Gamma}\hat{\Gamma}^T + \hat{\psi})^{-1} (x_i - \hat{\mu})$$

4 DATA ANALYSIS AND RESULTS

The data analysis and results section will describe the analysis of data and give the results.

4.1 Data analysis

The data was analyzed using SPSS version 17.

4.1.1 Checking for multicollinearity

The determinant of the R-matrix (the correlation matrix) should be > .00001; if it is less then review the matrix for variables that correlate very highly (R > .8). Consider eliminating one or more of the variables depending on the extent of the problem.

For this data Determinant of R = .009, showing no multicollinearity.

4.1.2 Sample adequacy

KMO and Bartlett's test of sphericity produces the Kaiser-Meyer-Olkin measure of sampling adequacy. The value of KMO should be > 0.5 if the sample is adequate.

For this data KMO = 0.608, indicating the sample was adequate.

Bartlett's measure tests the null hypothesis that the original correlation matrix is an identity matrix. For factor analysis to work we need some relationships between the variables. If the R-matrix were an identity matrix then all correlation coefficients would be zero. We therefore want a significance value less than 0.05.

We test: $\mathrm{H}_o : R = I$ vs $H_1 : R \neq I$

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.608
Bartlett's Test of Sphericity	Approx. Chi-Square	3775.476
	Df	210
	Sig.	.000

For this data, Bartlett's test was highly significant ($p < .001$), therefore factor analysis was appropriate.

4.1.3 Eigen values

There should be as many eigenvectors as there are variables. The eigenvalues associated with each factor represent the variance explained by that particular linear component and we also get a display of the eigenvalue in terms of the percentage of variance explained. We see from the Table 1 that factors 1to 21 explain 13.368% , 10.408%, 7.991%, 7.116%,6.085%, 5.735%, 5.292%, 5.111%, 4.423%, 4.285%, 4.079%, 3.931%, 3.788%, 3.601%, 3.424%, 3.311%, 2.656%, 2.222%, 1.898%, 0.758% and 0.519% of the total variance respectively.

The first few factors explain relatively large amounts of variance whereas subsequent factors explain only small amounts of variance. All factors with eigenvalues greater than 1are extracted. This leaves us with 8 factors.

The part of the table labeled Rotation Sums of Squared Loadings displays the eigenvalues of the factors after rotation. Before rotation, factor 1 accounted for considerably more variance than the remaining 7 (13.368% compared to 10.408%, 7.991%, 7.116%, 6.085%, 5.735%, 5.292%, and 5.111%). After rotation it accounts for only 12.398% of variance compared to 9.274%, 7.852%, 7.619%, 7.312%, 5.785%, 5.471%, and 5.396% of variance explained by factors 2 to7 respectively. In total all the 8 factors explain 61.106% of the variance.

4.1.4 Communalities

Principal component extraction method works on the assumption that all variance is common, therefore before extraction the communalities are all one. The communalities in the column labeled Extraction reflect the common variance in the data structure. From Table 2 in the appendix, we see that 61.3% of the variance associated with age is common, or shared, variance; and so is 50.3% of the variance associated with the age when they started doing sex work, 74.3% of the variance associated with the average number of clients per working day, 64.4% of the variance associated with their age at first pregnancy, 81.9% of the variance associated with the number of times they have been pregnant, 43.9% of the variance associated with their level of education,48.6% of the variance associated with their place of residence, and so on.

The communalities are also looked at in terms of the proportion of the variance explained by the underlying factors. After extraction some of the factors are discarded and so some of the information is lost. The amount of variance in each variable that can be explained by the retained factors is represented by the communalities after extraction.

4.1.5 Factor loadings

The component matrix before rotation contains the loadings of each variable onto each factor. One important decision to make is the number of factors to extract. Kaiser's criterion states that if there are less than 30 variables and communalities after extraction are greater than 0.7 or if the sample size exceeds 250 and the average communality exceeds 0.6 then retain all factors with eigenvalues above 1. If none of these apply, a scree plot can be used when the sample size is large (around 300 or more cases).

In this case, the sample size =810 and the average communality =0.611. All the eight factors whose eigenvalues exceed 1 are retained.

4.1.6 The Scree plot

This is a plot, in descending order of magnitude, of the eigenvalues of the correlation matrix. It helps to visualize the relative importance of the factors. A sharp drop in the plot signals that the subsequent factors can be ignored.

The curve begins to tail off after the 8^{th} factor, as seen in Fig. 1 at the appendix. This is a clear indication that 8 factors should be retained.

4.1.7 Factor rotation

The rotated factor matrix is a matrix of the factor loadings for each variable onto each factor. It contains the same information as the component matrix except that it is calculated after rotation.

Comparing this matrix with the unrotated solution, we see that most variables loaded highly onto the first factor. However, the rotation of the factor structure clarified things considerably. There were 8 factors and 4 variables loaded very highly on factor 1(Table 4)

4.2 INTERPRETATION OF RESULTS

The CSWs current age in years, the total number of pregnancies, total number of births and total number of years doing sex work loaded highly on factor 1. These variables had loadings of [illegible]7.5%, 89.6%, 88.3% and 52.3% respectively. They related to the period of time involved in sex and the nature of involvement. Therefore factor 1 was labeled 'sex work experience'.

Having a regular sex partner and condom use with the regular partner had high loadings of 94.2% and 95.2% respectively on factor 2. Factor 2 was labeled 'other intimate relationships, outside sex work and their nature'.

The use of alcohol and use of other hard drugs loaded highly on factor 3. Their loadings were 85.3% and 85.2% respectively. Factor 3 was labeled 'substance abuse'.

Place of residence of the CSWs, where they got their clients and the average number of clients with whom they used a condom per day loaded highly on factor 4. The loadings were -54.7%, -64.1% and 54.5% respectively. Factor 4 was labeled 'environmental influence'.

Age at first involvement in sex work, age at first pregnancy and level of education loaded highly on factor 5. The loadings were 56.7%, 73.1% and 59.1% respectively. Factor 5 was labeled 'general intelligence at the start of sex work'.

STIs and STI treatment loaded highly on factor 6.The loadings were 77.7% and 73.4% respectively. Factor 6 was labeled 'STI infections and their management'.

Having another source of income other than sex work had a loading of 66.5% on factor 7. It related to the CSWs' financial status and therefore Factor 7 was labeled 'financial status'.

The average number of clients per working day had a loading of 85.0% on factor 8. It related to the frequency of possible daily exposure to HIV by the CSWs through their clients. Factor 8 was labeled 'possible daily exposure to the virus'.

5 CONCLUSIONS AND RECOMMENDATIONS

The study found that there are 8 latent factors that can explain the risk of HIV acquisition and transmission among female sex workers in Kisumu. These are; sex work experience, general intelligence at the start of sex work, other intimate relationships outside sex work and their nature, substance abuse, sexually transmitted infections (STIs) and their management, environmental influence, financial status and possible daily exposure to the virus.

The observed risk factors were found to be closely related and tended to form patterns. Their current age in years, the total number of pregnancies they have had, total number of births and the number of years of involvement in sex work were very closely related. The use of alcohol and use of drugs were also closely related, and so were the variables in each group that loaded on the same factor.

It is recommended that subsequent research considers investigating the associations of the 8 underlying risk factors with the disease outcome.

6 REFERENCES

Chen, L., Prabhat J., Bridget S. & Sema K. S., 2007. Sexual risk factors for HIV infection in early and advanced HIV epidemics in sub-Saharan Africa. *PLoS ONE,* p. 2.

Chisholm, D., 2004. Reducing the global burden of hazardous alcohol use: a comparative cost-effectiveness analysis. *Journal of studies on Alcohol,* Volume 65, pp. 782-793.

Costello, A. B. & Osborne, J. W., 2005. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assesment, Research & Evaluation,* 10(7).

Dunkle, K. L., 2005. Risk factors for HIV infection among sex workers in Johannesburg South Africa. *International journal of STD and AIDs,* Volume 16, pp. 256-261.

Elmore-Meegan, M., Conroy, R. & Agala, C. B., 2004. Sex Workers in Kenya, Numbers of Clients and Associated Risks: An Exploratory Survey. *Reproductive Health Matters,* 12 (23): 50-57

Gray, R., Wawer, M. & Brookmeyer, R., 2001. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *The Lancet,* Volume 357, pp. 1149-1153.

Hewson, P. J., 2009. *Multivariate Statistics with R.*

Johnson, L. & Budlender, D., 2002. HIV risk factors: A Review of the Demographic, Socio-economic, Bio-medical and Behavioural Determinants of HIV prevalence in South Africa. *CARE Monograph,* Issue 8.

KAIS, 2007. *Kenya Aids Indicator Survey preliminary report.*

Kimemia, M., Asner, S., Kimberly, K. & Daire, A. P., 2011. An Exploratory Factor Analysis of the Brief COPE with a Sample of Kenyan Caregivers. *International Journal for the Advancement of Counselling,* 33(3), pp. 149-160.

Martin, H., Nyange, P. & Richardson, B., 1998. Hormonal contraception, sexually transmitted diseases, and the risk of heterosexual transmission of human immunodeficiency virus type 1. *Journal of Infectious Diseases,* Volume 178, pp. 1053-1059.

Morison L., Weiss, H. A. & Buve, A.,2001. Commercial sex and the spread of HIV in four cities in sub-Saharan Africa. *AIDS,* 15(Suppl 4):S61–9.

National AIDS Control Council, 2009b. Kenya National AIDS Strategic Plan 2009/10-20012/13: Delivering on universal access to services. *Accessed October 1, 2012.*

Nicolosi, A., Leite, M. & Mussico, M., 1994. The efficiency of male-to-female and female-to-male sexual tranmission of the Human Immunodeficiency Virus: a study of 730 stable couples. *Epidemiology,* 5(6), pp. 570-575.

Ntumbanzondo, M., 2006. Unprotected intercourse for extra money among commercial sex workers in Kinshasa, democratic Republic of Congo. *AIDS Care,* Volume 18, pp. 777-785.

Overs, C. (2002). Sex Workers Part of the Solution: An analysis of HIV prevention Programming to prevent HIV transmission during commercial sex in developing countries.

Rottingen, J., Cameron, D. & gamett, D., 2001. A systematic review of the epidemiological interactions between classic sexually transmitted diseases and HIV: how much is really known? *Sexually Transmitted Diseases,* 28(10), pp. 579-597.

Schutte, A. E., Rooyen, V., Huisman, H. W. & Kruger, H. S., 2003. Factor analysis of possible risks for hypertension in a black South African population. *Journal of Human Hypertension,* Volume 17, pp. 339-348.

UNAIDS, 2011. Epidemic updates and health sector progress towards Universal Access. *GLOBAL HIV/AIDS RESPONSE,* p. 11.

Wang, C., Reilly, M. & Kreiss, J., 1999. Risk of HIV infection in oral contraceptive pill users: a meta-analysis. *Journal of Acquired Immune Deficiency Syndrome,* 21(1), pp. 51-58.

WHO, 2008. *Prevention and treatment of HIV/AIDS and other STIs targeting sex workers and their partners in Côte d'Ivoire, Report.* [Online] Available at: http://www.who.int/hiv/pub/populations/CI-PS-en08.pdf [Accessed 21 November 2011].

WHO, Unicef & UNAIDS, 2011. Epidemic updates health sector progress towards Universal Access. *Progress Report.*

7 APPENDIX

Table 1: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.807	13.368	13.368	2.807	13.368	13.368	2.603	12.398	12.398
2	2.186	10.408	23.776	2.186	10.408	23.776	1.948	9.274	21.672
3	1.678	7.991	31.768	1.678	7.991	31.768	1.649	7.852	29.524
4	1.494	7.116	38.883	1.494	7.116	38.883	1.600	7.619	37.143
5	1.278	6.085	44.968	1.278	6.085	44.968	1.535	7.312	44.454
6	1.204	5.735	50.703	1.204	5.735	50.703	1.215	5.785	50.239
7	1.111	5.292	55.995	1.111	5.292	55.995	1.149	5.471	55.710
8	1.073	5.111	61.106	1.073	5.111	61.106	1.133	5.396	61.106
9	.929	4.423	65.528						
10	.900	4.285	69.813						
11	.857	4.079	73.892						
12	.826	3.931	77.823						
13	.795	3.788	81.611						
14	.756	3.601	85.212						
15	.719	3.424	88.636						
16	.695	3.311	91.947						
17	.558	2.656	94.603						
18	.467	2.222	96.825						
19	.399	1.898	98.723						
20	.159	.758	99.481						
21	.109	.519	100.000						

Table 2: Communalities

	Initial	Extraction
age2	1.000	.613
educ2	1.000	.439
stay2	1.000	.486
startearn2	1.000	.503
sexyrs2	1.000	.402
money2	1.000	.618
clients2	1.000	.502
price2	1.000	.430
clientno2	1.000	.743
condomno2	1.000	.376
alcohol2	1.000	.745
drugs2	1.000	.764
bfriend2	1.000	.916
usecondom2	1.000	.918
treatment2	1.000	.603
noprg2	1.000	.819
ageprg2	1.000	.644
birthno2	1.000	.801
contra2	1.000	.323
care2	1.000	.533
stisynd2	1.000	.655

Table 3: Component Matrix

	Component							
	1	2	3	4	5	6	7	8
noprg2	.819	.142	.092	-.124	-.292	-.120	-.038	.048
birthno2	.784	.161	.159	-.141	-.313	-.123	-.013	.038
age2	.755	.069	-.087	-.021	.111	.030	-.069	-.108
sexyrs2	.510	.033	.150	-.157	-.030	.063	.221	-.200
bfriend2	-.329	.762	.118	-.440	-.039	.087	-.012	-.097
usecondom 2	-.364	.694	.069	-.505	-.061	.110	-.067	-.154
startearn2	.372	.384	-.218	.306	.247	.064	-.032	.101
contra2	.221	.383	-.277	.101	-.083	-.138	-.057	.107
drugs2	-.096	.395	.577	.479	-.153	-.057	-.023	-.097
stay2	.088	.108	.529	.004	.340	-.021	.196	.178
condomno2	-.009	.231	-.469	.112	-.075	.193	-.210	.050
clients2	.152	-.189	.446	-.364	.325	.042	-.048	-.050
price2	-.113	.316	-.377	-.050	-.210	-.001	.357	.016
alcohol2	-.145	.391	.362	.600	-.262	-.094	-.013	-.042
ageprg2	.221	.215	-.051	.136	.587	.147	.194	-.353
educ2	.022	.321	-.250	.291	.380	.051	.166	.119
treatment2	.134	.050	.201	.091	.005	.720	-.107	.069
stisynd2	.043	-.113	.089	.005	-.205	.650	-.029	.409
clientno2	.085	-.034	-.076	-.041	-.189	.139	.818	.054
money2	.226	.317	-.062	-.163	.266	-.171	-.184	.549
care2	.249	-.036	-.206	.072	-.089	.249	-.231	-.547

Extraction Method: Principal Component Analysis.
8 components extracted.

Table 4: **Rotated Component Matrix**

	Component							
	1	2	3	4	5	6	7	8
noprg2	.896	-.061	.026	.040	-.063	.016	.076	.000
birthno2	.883	-.020	.069	-.007	-.095	.016	.076	.022
age2	.675	-.140	-.150	.005	.308	.032	-.085	-.110
sexyrs2	.523	.017	-.052	-.226	.098	.026	-.170	.186
usecondom 2	-.087	.952	.032	.046	-.025	-.009	.016	-.010
bfriend2	-.056	.942	.118	.029	.036	-.006	.092	.031
alcohol2	-.046	.024	.853	.117	.030	-.002	.003	-.004
drugs2	.014	.114	.852	-.133	.046	.034	-.014	-.062
clients2	.128	.072	-.187	-.641	-.013	.062	.019	-.175
stay2	.054	.040	.218	-.547	.208	.074	.289	.046
condomno2	-.029	.091	-.095	.545	.165	.144	-.029	-.106
contra2	.262	.099	.076	.401	.189	-.106	.176	-.012
ageprg2	.060	.058	-.042	-.202	.731	-.062	-.236	.029
educ2	-.110	-.004	.059	.181	.591	-.006	.181	.096
startearn2	.262	-.043	.107	.262	.567	.069	.146	-.073
stisynd2	-.007	-.070	-.044	.049	-.149	.777	.111	.091
treatment2	.063	.050	.077	-.081	.135	.734	-.152	-.063
money2	.213	.127	-.134	.068	.208	.013	.665	-.220
care2	.220	-.010	-.081	.188	.116	.066	-.624	-.190
clientno2	.083	-.058	-.048	-.051	.030	.070	-.013	.850
price2	-.027	.242	-.035	.389	.067	-.096	.035	.450

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

Rotation converged in 6 iterations.

Fig 1: The scree plot