IDENTIFYING MOTIVATIONAL FACTORS IN JOB PERFORMANCE AND HOW THEY RELATE TO ATTRITION AMONG VILLAGE HEALTH WORKERS IN SHURUGWI AND CHIRUMANZU DISTRICTS, ZIMBABWE

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS

IN THE FACULTY OF MEDICINE

BY

Epiphania Nyakasoka

Supervisors: Prof S. Rusakaniko

Mr. W. Tinago

DEPARTMENT OF COMMUNITY MEDICINE

FACULTY OF MEDICINE

August 2013

UNIVERSITY OF ZIMBABWE

Contents

Contents ii
Abstract iv
Acknowledgements v
CHAPTER 1 1
INTRODUCTION 1
1.1 Background 1
1.2 Critical Appraisal 5
1.3 Problem Statement 6
CHAPTER 2 8
LITERATURE REVIEW 8
2.1 Motivation in job performance and attrition in VHWs 8
2.2 Research Question 11
2.3 Justification 11
2.4 Study Objectives 11
2.4.1 Broad Objective 11
2.4.2 Specific Objectives 12
CHAPTER 3 13
METHODOLOGY 13
3.1 Secondary Data Analysis 13
3.1.1 Description of the study 13
3.1.2 Sample Size 13
3.1.3 Variables for secondary data analysis 13
3.2 Methodology of Analysis 16
3.2.1 Data extraction 16
3.2.2 Data Management 16
3.2.3 Statistical Analysis 17
3.3 Ethical Considerations 19
RESULTS 20
4.1 Descriptive Statistics 20
4.3 Factors associated with attrition 27
4.3.1 Goodness of fit 29
CHAPTER 5 30
DISCUSSION AND CONCLUSION 30

5.1 Discussion 30
5.2 Conclusion 31
BIBLIOGRAPHY 33
APPENDIX A 35
A1 Review of statistical theory 35
AI.1 Factor analysis 35
Types of Factor Analysis 35
A1.2 Logistic Regression 40

List of Figures

Figure 1: Scree plot of Eigen values 23

List of Tables

Table 1: Data dictionary on variables selected for secondary data analysis 14

Table 2: Demographic characteristics of the 338 VHW survey respondents, Chirumanzu and Shurugwi, 2012 20

Table 3: Unrotated factors 22

Table 4: Rotated factor loadings (pattern matrix) and unique variances 24

Table 5: Univariate and multivariable logistic regression analysis for variables independently associated with VHW attrition in the Zvitambo-SHINE project 28

Abstract

Background: Motivation is one of the key determinants of village health worker performance. Motivation is difficult to assess and can be measured by a series of questions which can be grouped into smaller sets that accurately define motivation, through factor analysis. Motivation predicts job related behaviours like attrition. Such behaviours are also difficult to measure; however, proxies can be used, like intent to quit. Logistic regression can then be used to determine how the motivation groups are related to attrition.

Method: The village health worker survey enrolled 338 village health workers working in Chirumanzu and Shurugwi. All the respondents were asked a series of questions that assess motivation when they came for their routine meetings. Factor analysis was used to reduce the number of items assessing motivation from 16. Logistic regression analysis was used to determine the association between motivation factors and attrition.

Results: The 16 items measuring motivation were reduced to 2, namely peer support and dissatisfaction, through factor analysis. Logistic regression analysis showed that peer support was 36% protective to attrition while the odds of attrition were 3.20 times when there is dissatisfaction.

Conclusion: Peer pressure and dissatisfaction are two major variables that define motivation and both are associated with attrition. Peer pressure is a protective variable while dissatisfaction is a risk factor.

Acknowledgements

I would like to express my deepest gratitude to my main supervisor, Prof S Rusakaniko for guiding me through this project. I would also like to thank my co-supervisor Mr. W Tinago for his steadfast and first-rate effort in mentoring me through this project. His advice and critical comments have been of great value to me. I would not have had the courage and strength to complete my thesis without you.

My sincere gratitude goes to my biostatistics lecturers, Mr V Chikwasha and Mr G Mandozana for the mentorship they rendered to me to complete my thesis.

I would also like to express my heartfelt appreciation to Prof J Hakim, Dr M Mbuya, R Ntozini, B Chasekwa , R Kambarami for permitting me to use the Zvitambo-SHINE data in this study and for their excellent effort in guiding me throughout this project.

Last but not least, a special thank you goes to my fellow biostatisticians who assisted me throughout the project.

CHAPTER 1

INTRODUCTION

1.1 Background

Globally, there is a significant shortage of health care workers. World Health Organisation (WHO) estimates that a total of 4 250 000 health workers are needed to fill the gap. WHO recommends 23 health workers per 10 000 people. There are 57 countries, most of them in Africa and Asia with fewer than 23 health workers per 10 000 people. Sub-Saharan Africa faces the greatest challenges. While it has 11% of the world's population and 24% of the global burden of disease, it has only 3% of the world's health workers (WHO, 2006). It is in this situation of the health workforce crisis that many countries are investing in national village health worker programmes, however, sustainability of these programmes is threatened by poor motivation and high rates of attrition.

A village health worker (VHW) is a community member, trained in some way to promote health or to carry out some health care services and having no formal professional or paraprofessional certificate, degree or tertiary education (Lewin et al., 2005). VHWs are considered as a third health service delivery work-force (Sein, 2006).

Evaluation of village health workers' performance in general, is the focus of much attention at this time, as many countries invest in them as a strategy for the achievement of the millennium development goals (Haines et al., 2007). The effectiveness of VHWs has been demonstrated in some studies for example, a VHW programme in India resulted in significant reduction of low birth weight, preterm births and neonatal sepsis (Bang, Baitule, Reddy, &Deshmukh, 2005).

Motivation can be defined as how much a person is willing to put forth and maintain an effort towards the achievement of organisational goals (Franco L.M et al, 2002). Motivation is driven by many elements that can be intrinsic or extrinsic. Intrinsic factors include such factors as an individual's work related goals, as well as his/her sense of selflessness, effectiveness and organisational commitment. Extrinsic factors include peer approval, the incentives provided and the expectation of future paid employment (Syed Moshfiqur Rahman; 2010). It has been suggested that motivation is one of the major determinants of health care service quality and retention of VHW (WHO 2006; Dieleman et al 2006). Motivation of VHWs has the possibility of affecting health services quality. It has been recognised that low health worker motivation can relentlessly weaken demand for health services and may lead to wastage or loss of the limited number of workers available (Dieleman et al, 2006). The presence of motivated staff is necessary for the delivery of superior healthcare; this is one of the most difficult inputs to measure and guarantee (Haines et al, 2007). Motivation is a complex concept that is not directly observable, but is critical to the retention and performance of health workers.

Job motivation itself is hard to observe, instead it is reflected in different behaviours such as intent to quit. For a job motivation measure to be valid and useful, it should be able to predict work related behaviours such as attrition.

One of the most fundamental parts of managing any VHW program is to measure attrition and to identify its determinants. However, reporting health outcomes and process indicators such as the number of VHW recruited and trained is often prioritized. It is imperative to accept that focus on health outcomes and indicators are not the only reasons why there is lack of research on attrition. A major drawback to research on attrition is that it is hard to define and report. On top of their jobs as VHW, they usually involve themselves in other activities, for example agricultural activities. In societies where agricultural activities are done,

agricultural economic activity changes throughout the year. During the period when agricultural activities are relaxed, the VHWs, among other people tend to look for casual jobs in urban areas. During the periods when there is high agricultural activity, there are an increased number of field labourers and more working hours are needed. The VHWs then get back to their villages. The sometimes informal nature of VHW work makes it difficult to measure attrition as workers may be working in the field during peak periods. Hence it may be difficult to ascertain whether the VHW is absent because they are working in the fields or because they have quit.

The Village Health Worker survey was designed to gain insight into VHW work capacity, performance and perceptions about the motivational characteristics of their work. Understanding these types of VHW traits can provide insight into their performance. Traits such as motivation often manifest as latent traits, that is, we cannot observe them directly but we can infer their existence by the properties of observed variables. They can only be indirectly and incompletely measured via single survey questions. Individual survey questions are often imperfect measures of such complex traits of interest and there is frequently a need to distil survey data down into relevant information about the population.

Factor analysis is a method for identifying latent traits from question-level survey data. It is useful in survey analysis whenever the phenomenon of interest is complex and not directly measurable via a single question. In such situations, it is necessary to ask a series of questions about the phenomenon and then appropriately combine the resulting responses into a single measure or factor. Such factors then become the observed measures of the unobservable or latent phenomenon (which can also be called a construct). Having done factor analysis to group the items that index motivation, logistic regression can be done to find out the factors that predict attrition.

Logistic regression is used to analyze relationships between a dichotomous dependent variable and metric or dichotomous independent variables. Logistic regression involves fitting an equation to the data. The equation is of the form: Logit (p) = $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots\ldots\ldots + \beta_i X_i$.The goal is to find the best set of coefficients so that cases that belong to a particular category will, when using the equation, have a very high calculated probability that they will be allocated to that category. The principle by which it does so is maximum likelihood estimation which maximises the probability of getting the observed results given the fitted regression coefficients (Agresti A)

Description of Original Study

The Village Health Worker survey had no protocol developed; however four research questions were put forward:

1. Baseline characteristics of VHWs. How do the socio-demographic characteristics of VHWs differ by treatment?

a) Demographic characteristics – age, gender.
b) Work experience – prior experience, duration of tenure (phase of VHW recruitment).

2. VHW knowledge: What is the proportion of correct responses scored by VHWs to (Wash, IYCF and overall knowledge)

a) Define correct responses for each question in the survey
b) Assign score =1 if only the health worker provides a correct response to the i^{th} question; = 0 otherwise

3. What are the VHW perceptions of their work context, in terms of motivational characteristics and supervision? By supervisor? By randomized treatment?

a) Descriptive: % responses in each Lickert scale by statement (e.g. 50% strongly agree * statement X)

b) Factor analytic solution

4. What are the reported time commitments of VHWs in the different areas of their work? By randomized treatment?

The study was a cross sectional survey carried by ZVITAMBO on all VHW working in Chirumanzu and Shurugwi districts in Zimbabwe. The main parameter that the study intended to assess was VHW capacity to deliver interventions for a Sanitation hygiene infant nutrition efficacy (SHINE) project that is being implemented in the two districts. The study area was divided into 212 clusters, each equivalent to the catchment area of 1 to 5 VHW. Clusters were randomised using a restricted randomization program. Treatments were randomly allocated to the clusters and a VHW would belong to the treatment randomly assigned to their cluster. A VHW would then receive training on the treatment they have been randomly allocated and would deliver that treatment in the SHINE project. The VHW were approached when they came for their routine training and data was collected through face to face interviews.

The Village Health Worker survey used a series of questions to measure motivation and we aim to identify motivational factors that influence their work and how these factors relate to attrition. Attrition could not be directly assessed in this study and will be measured by a proxy, intent to quit.

1.2 Critical Appraisal

The study had no protocol developed, thereby making it difficult for anyone who intends to use the data to understand the study's objectives, methodology and justification. The data needed, however, to be used to answer research questions that had been put forward by this organisation. The problem that prompted the study to be carried out could not be established due to unavailability of the study protocol. The study's research questions were relevant for

the purposes for which the data was collected, that is, to assess VHW capacity to deliver randomized interventions. The questions were basically meant to answer how knowledgeable the VHW are in terms WASH and infant feeding so that their capacity to deliver the WASH and feeding interventions could be assessed. Besides knowledge, motivation, satisfaction and commitment also can affect the performance of VHW in delivering the interventions; hence the research questions were appropriate for the assessment. Randomisation was used to allocate treatment groups. This produces comparable groups in terms of general baseline characteristics, such as age or gender, and other key factors. Randomisation also ensures that the study results are not weakened by bias, especially due to human choices. Through randomisation potential known and unknown confounders were distributed evenly across groups. The study design was appropriate for the research questions because it represents the population of VHW in Shurugwi and Chirumhanzu were the SHINE interventions are targeted. The data collection method used addresses potential sources of bias, especially non response bias and also allowed the interviewers to probe more so as to solicit more information. The data was being collected on Netbook computers programmed with the questionnaires, which was appropriate as it makes transportation and storage of the questionnaires easy. The data is still in the process of being analysed using stata.

1.3 Problem Statement

Village health workers are agents of change in addressing health related issues such as nutrition, vaccinations and sanitation, largely in rural populations. Despite their presence, there is continued health crisis especially in hard-to-reach, underserved populations and there is also an increase in the village health worker attrition rate. The health workforce crisis is issues around staff shortages, brain-drain, low work motivation and poor performance being enormous in sub-Saharan Africa. Motivating them can potentially result in high performance,

and retaining of the village health workers also improves. Motivation is one of the key factors that affect the job performance of village health workers and it usually results in attrition. Motivation itself is difficult to measure and an assessment can be done through measuring its construct. Construct refers to any complex psychological concept like motivation, fear and love. A construct's height, weight or depth cannot be measured because constructs are not concrete materials in the visible world. We cannot describe in inches or pounds how much there are or where it starts and ends (AlleyDog.com). To determine motivational factors, we might use a series of questions that asks for the person's reactions to certain situations. Motivation is reflected in different behaviours such as attrition. Attrition has threatened the reduction of maternal mortality ratio which currently stands at 960 per 100 000, (ZDHS, 2010/2011) an increase from 560 per 100 000, according to ZDHS of 2005/2006. Neonatal mortality rate stands at 34 per 100 live births (ZDHS 2010/2011). Attrition could not be measured measured, instead a proxy, intent to quit was used to assess attrition. There are a series of questions in the survey, each of which is supposed to address a dimension of the latent variable under investigation, which is motivation, and as such, there is need for a data reduction technique such as factor analysis. Factor analysis groups the survey questions in some conceptually meaningful way. The construct being measured, motivation uses multiple items which need to be reduced to a smaller set of summary variables that accurately define motivation. Often the items measuring motivation are highly correlated and as such they are effectively saying the same thing. It may be useful to simplify the original set of variables to a new set of fewer uncorrelated and more manageable variables. The identified factors should be useful in predicting intent to quit the job (the proxy for attrition).

CHAPTER 2

LITERATURE REVIEW

2.1 Motivation in job performance and attrition in VHWs

African health systems are not only experiencing one of the greatest staff shortages, but clinical staff is currently faced with weak institutional frameworks and distortive incentive structures, ineffective management practices and adverse work environments at systemic and organisational level, resulting in an overburdened health workforce with low levels of work motivation (Mathauer & Imhoff, 2006). It is believed that this underperformance has not only undermined the capacity of health-care organisations, but even threatens the achievement of the Millennium Development Goals (MDGs) to "reduce child mortality; improve maternal health; and combat HIV/AIDS, malaria and other diseases" (UN 2007; WHO 2006).

As a way of increasing access to treatment for the sick where health services are geographically and financially inaccessible, several African countries are currently investing in VHW to deliver treatment. The first volunteer and paid VHW programs emerged in 1950s. The Indian Health Service was the first federal government agency to employ and utilize VHWs as community health representatives, in1968. Uganda was one of the first African countries to take the VHW policy to scale through the Home Based Management of Fever strategy, which aimed to improve prompt and appropriate treatment of presumptive malaria using VHW. In Mozambique, the use of VHW dates back to 1978 where they were (and still are) known as *AgentesPolivalentesElementares* (APEs). In Mozambique as is in Uganda and other settings, motivation is a key constraint area to the impact of integrated community case management when implemented through VHW. Mutale W et al have cited motivation as one of the key factors in retention of VHW.

The value of the healthcare is intensified when one has great need for healthcare and no direct access to it. In Zimbabwe, healthcare access is centralized through hospitals and clinics however, the clinics and hospitals are not easily accessible for people living in remote parts of the country. For people living in remote areas, who constitute **65%** of the population, a visit to their home by a Village Health Worker may be their only contact with a healthcare professional. The nation ideally requires **17 000** village health workers, but only **5000** Village Health Workers are functional, which translates to **29%** of the required personnel. (projecthigherlife, 2013).

It is not easy to state the number of village health workers that are available throughout the world because many programmes in which they are involved exist as small scale projects (projecthigherlife, 2013).

It has been agreed by many authors that health worker motivation is a key factor that determines health sector performance, however the availability of resources and the competence of the workers are also necessary (Dolea C, Adams O, 2005). A study in Zambia showed that 50% of health workers who were not motivated confirmed that they were likely to leave their posts in search of greener pastures (Kruse GR et al,2009).

According to Pinder, 2008, measuring and conceptualizing motivation is problematic from the onset because motivation cannot be observed directly and must be inferred from behaviour. Surveys are used and researchers come up with a list of potential motivators and ask participants to respond, usually in the form of Lickert style agreements. In a study to understand the motivation and attrition of HIV/AIDS village health workers by the Johns Hopkins University, the investigation consisted of a series of key informant interviews, in-depth interviews focus group discussions, and survey questionnaires to determine factors for motivation. In order to ascertain rates of VHW attrition, a literature review was conducted

according to the PRISMA method for systematic reviews. This latter process involved a primary database search, review of all resulting abstracts and elimination of non-relevant literature (Anne M, 2010).

One of the key challenges of VHW programmes that have been identified is attrition. Lungiswa N et al reported that in the 1980s levels of attrition were between 3.2 and 77% in the 1980s. In current programmes, the problem of attrition is still there: a 43% attrition rate was noted in a VHW programme in the Plurinational State of Bolivia. A tuberculosis intervention program in South Africa lost 11 out of 12 VHW in less than a year and in Bangladesh, a program which was aimed at improving newborn care was implemented and lost lost 32 out of 43 VHW over a four-year period. This is evidence that attrition is a common challenge in various health settings. A study in Bangladesh found reasons cited by VHW for leaving their posts included lack of time to attend to their own children and other responsibilities, insufficient salary/ profit (Khan SH et al; 1988). Another study in Nigeria found that VHW stopped working because of low salaries, lack of opportunity for advancement, lack of credibility by villagers and poor supervision (Gray HH; 1988).

Factor analysis serves several related purposes. One of its primary functions is to help an investigator in determining how many latent variables (that is, variables that are not directly observed) underlie a set of items. Thus, in the case of the 16 motivation items, factor analysis could help the investigator determine whether one broad or several more specific constructs were needed to characterize the item set. Factor analysis also can provide a means of explaining variation among relatively many original variables (for example, 16 items) using relatively few newly created variables (that is, the factors). This amounts to condensing information so that variation can be accounted for by using a smaller number of variables. For example, instead of needing 16 scores to describe how respondents answered the items, it might be possible to compute fewer scores (perhaps even one) based on combining items. In

Zambia, a study to measure health worker motivation was carried out. Factor analysis was used to group outcome constructs (Mutale W, 2013). The tools and concepts of this study have been applied in African settings to measure health worker motivation.

2.2 Research Question

What are the motivational factors in village health worker performance of their duties and how do these factors relate to attrition?

2.3 Justification

The analysis will enable a multi-dimensional construct, motivation, to be examined and identified. As a result of this analysis, the main factors influencing job motivation among VHW will be described, the relationship between motivation and attrition will be examined and recommendations for improving motivation and reducing attrition can be made. Recommendations from this analysis will be used to improve VHW performance motivational drivers identified and assist the Ministry of Health to redirect retention strategies and introduce ideas to minimise VHW attrition. This could produce continuity of quality care and more efficient recruitment and retention effort. If set targets are met, it will help reduce illness, morbidity and mortality.

2.4 Study Objectives

2.4.1 Broad Objective

To identify motivation factors in job performance and how these factors relate to attrition among VHWs in Chirumanzu and Shurugwi.

2.4.2 Specific Objectives

- To describe the socio-demographic characteristics of VHWs.
- To assess inter-item consistency reliability of items directed to motivation.
- To identify village health worker motivational factors.
- To determine the association between motivational factors and attrition.

CHAPTER 3

METHODOLOGY

3.1 Secondary Data Analysis

3.1.1 Description of the study

In this study, secondary data analysis was carried out using data from Zvitambo Project. The study is a sub-study of a Sanitation Hygiene Infant Nutrition Efficacy (SHINE) project that is being implemented in the Chirumhanzu and Shurugwi districts. Each of the VHW in the study was approached when they came for their routine meetings and were informed about the study. The study consists of a total of 338 respondents. Data was collected from the VHW through face to face interviews.

3.1.2 Sample Size

A sample size of 372 village health workers was available for analysis, but according to literature, to answer our objectives, at least 20 respondents are required for each item measuring motivation in order to do factor analysis. The study has 16 items measuring motivation, so it was estimated that at least 320 respondents would be required for the analysis.

3.1.3 Variables for secondary data analysis

For this project, the main study outcome is attrition. Attrition was measured by a proxy, intent to quit, which was measured by 1 item. Respondents were asked whether they were planning to leave the job. The answers were measured by a binary response, yes or no. Motivation was measured with 16 items, example items are; “I feel like I enjoy my work as a VHW, I feel well supported by other VHW I work with.” The items were measured by a 6-point Lickert scale on a scale of 0 to 5, 0 = N/A, 1= never, 2= rarely, 3= sometimes, 4= often, 5= always. Lickert scales are technically qualitative data as the scores are not ratio intervals. Table 1 presents a description of the variables extracted for analysis and how they are coded.

Table 1: Data dictionary on variables selected for secondary data analysis

	variable name	variable description	variable format	variable code
Outcome	Attrition	Would you want to leave this job if it were possible/if you had other alternatives? 1=yes 2=no	numeric	c20
Study variables	Age		numeric	a2
	Age group	1= < 30 2= 31- 40 3= 41- 50 4= 51- 60 5= 61^{+}	numeric	age_gp
	Gender	1= male 2= female	numeric	a2
	Marital status	1= single 2= married 3= divorced/separated 4= widowed	numeric	a4
	Education status	Have you ever attended school? 1= yes 2= no	numeric	a6
	Education level	What is the highest level of school you attended? 1=primary school 2=some secondary school 3=completed secondary school	numeric	a7
	Duration of tenure	How long have you been working as a village health worker?	numeric	a13
	Religion	1=Traditional 2=Roman Catholic 3=Protestant 4=Pentecostal 5=Apostolic sect 6=Other Christian 7=Muslim 8=None 88=Other	numeric	a22

Variables

measuring motivation

1=never
2=rarely
3=sometimes
4=often
5=always

I feel like I enjoy my work as a village health worker?	numeric	b1
I feel that my work will improve people's lives?	numeric	b2
I feel like I am adequately prepared for my responsibilities.	numeric	b3
I feel like I am responsible for so many activities that it's not possible to do them well in the time I have.	numeric	b4
I feel very connected with other village health workers.	numeric	b5
I feel well supported by the other VHWs I work with.	numeric	b6
If I were sick, I can easily find someone to help with my VHW duties.	numeric	b7
I can find at least one VHW whose advice I really trust.	numeric	b8
There is someone who gives me information to help me a VHW work's situation.	numeric	b9
There is someone I work with who I can turn to for advice about handling problems with my family.	numeric	b10
If I need help visiting mothers, I know that I can always get help from the other village	numeric	b11

health workers.		
I have had thoughts of leaving this job.	numeric	b12
I feel that I am responsible for more work than other colleagues.	numeric	b13
I am proud to tell others that I am a village health worker.	numeric	b14
Unless I am rewarded for it, I see no reason to spend extra effort promoting community health.	numeric	b15
How hard I work is directly linked to how much I am paid.	numeric	b16

3.2 Methodology of Analysis

3.2.1 Data extraction

The village health worker survey dataset was captured and stored in excel and saved as a text (tab delimited) file. It was then converted to excel comma separated value (CSV) file and imported to STATA 10. The original data had 257 variables. For secondary data analysis, we extracted 25 variables as they were the ones appropriate for answering our research question. The data was kept as 2 different files, 1 file had socio-demographic characteristics while variables on motivation and intent to quit were on the other file. To get one complete dataset, the files were merged using village health worker identification numbers as the matching variables.

3.2.2 Data Management

The data had been cleaned and coded by Zvitambo staff but after merging, it emerged that some village health worker identification numbers had been repeated, some had been erroneously captured, for example 003 captured as oo3 and some clinics where village health

workers report to had not been captured. The errors were corrected by checking and verifying with the netbook computers which were used for data collection.

3.2.3 Statistical Analysis

Village health workers' characteristics were described using means and standard deviation for continuous, normal variables. Medians and inter-quartile range (IQR) was used to describe continuous non normal variables. Continuous variables were tested for normality using histograms. Categorical variables were described using frequencies and percentages.

Cronbach's alphas were computed to check for inter-item consistency of the 16 items measuring motivation.

Factor analysis was used for reducing the items into a smaller, more manageable set of underlying concepts that accurately define motivation. Factor analysis proceeded in the steps briefly described below:

Step 1: Computing correlation matrix

A polychoric correlation matrix was computed since the items are ordinal. The correlation matrix is used to identify correlations between variables. Bartlett test of Sphericity was computed to test the hypothesis that the correlation matrix is an identity matrix. Bartlett test should be significant (that is $p< 0.05$). The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was computed to examine the appropriateness of factor analysis. High KMO values, between 0.5 and 1.0 indicate that factor analysis is appropriate.

Step 2: Factor extraction

Principal components analysis was used to extract the factors. Principal component analysis was preferred as it reduces a set of variables down to a smaller number of factors. To

determine the number of factors, Kaiser's criterion and scree plots were used. With the Kaiser criterion, the determination of the number of factors was done by choosing factors with eigen values greater than 1 and with the scree plot, we looked for a notable drop with abrupt transition from vertical to horizontal and a clear 'elbow.' We then extracted factors above the 'elbow.'

Step 3: Factor rotation

We rotated the factors using Varimax rotation. This was done to make the factors easier to interpret. Varimax rotation was chosen because it uses orthogonal rotations which are quite easier to interpret.This method assumes that the factors are independent

Step 4: Decision

The final decision about the number of factors to choose was made by grouping those variables that formed large loadings for the same factor. Factor names were decided upon, basing on past research.

Motivational factors obtained from factor analysis and demographic variables were assessed for association with attrition, using logistic regression analysis. Covariates with P values of < 0.2 in the univariate analysis were included in multivariate logistic regression to determine factors independently associated with attrition. Backward elimination method was used to derive covariates independently associated with attrition. The Hosmer-Lemeshow test was used to assess goodness of fit of the model. The results of the logistic regression analyses are expressed as odds ratio (ORs) and their 95% confidence intervals (CI).

3.3 Ethical Considerations

Approval to carry out the analysis was sought from the organisation and from JREC. Data was not shared with unauthorised persons and all information contained in the data set was kept confidential. All ethical considerations were duly observed

CHAPTER 4

RESULTS

4.1 Descriptive Statistics

A total of 338 respondents participated in the study. The mean (SD) age was 43.9 (8.6) years and 248 (73.6%) were female. Most of the respondents 281 (83.6%) had attained at least secondary education and 102 (30.1%) were Roman Catholic. Respondent characteristics are presented in Table 2. The median (IQR) duration of tenure and previous work experience as VHW was 37 (25-132) months and 36 (12-72) months, respectively.

Table 2: Demographic characteristics of the 338 VHW survey respondents, Chirumanzu and Shurugwi, 2012

Variable	N=338
Age (years), mean(SD))	43.9 (8.6)
Gender: Female	248 (73.6)
Education level	
Primary	55 (16.4)
Secondary	281(83.6)
Marital status	
Single	18 (5.3)
Married	258 (76.3)
Divorced/Separated	17 (5.0)
Widowed	45 (13.3)
Duration of tenure(months), median (IQR)	37 (25-132)
Religion	
Traditional	12 (3.6)
Roman Catholic	102 (30.2)
Protestant	68 (20.1)
Pentecostal	51 (15.1)
Apostolic Sect	90 (26.6)
Other Christian	4 (1.2)
None	11 (3.3)

*Data are n (%) unless otherwise stated. A traditional religion is a non Christian group that believes in spirit mediums.

4.2 Reliability and Factor Analysis of items used to measure motivation

The inter-item reliability of the 16 item scale, measured by Cronbach's Alpha (α) had a reliability coefficient of 0.64, which is acceptable according to the criteria proposed by George and Mallary, 2003. The criteria from George and Mallary provide the following rules of thumb: $\alpha > 0.9$ – excellent; $\alpha > 0.8$ – good; $\alpha > 0.7$ – acceptable; $\alpha > 0.6$ – questionable but acceptable; $\alpha > 0.5$ – poor but acceptable; $\alpha < 0.5$ - unacceptable

In the first step of explaratory factor analysis, polychoric correlations of all the 16 items were calculated. From the polychoric correlation matrix (Appendix B), we looked at whether the correlations were greater than 0.3, (a requirement for principal component analysis factor extraction) and if any patterns existed. Variable b2-b5 were correlated with each other while variable b5-b11 were significantly correlated with each other and variable b13-b15 correlated significantly with each other, suggesting that these variables may be measuring three underlying factors.

Before conducting factor extraction, the Kaiser-Meyer-Olkin approach was used to determine the sufficiency of the sample size for the component. The KMO measure of sampling adequacy was 0.74, indicating that the degree of common variance among the 16 items is "mediocre" bordering on the "middling," but however, if factor analysis is conducted, the factors extracted will account for a fare amount of variance but not a substantial amount. The Bartlett's test of sphericity showed that the sample intercorrelation matrix did not come from a population in which the intercorrelation matrix is an identity matrix ($p<0.001$). It is thus concluded that factor analysis was permissible.

The principal component method was used for factor extraction. From the unrotated principal factors, 16 factors were extracted, however, using the Kaiser Criterion, Factor 1 and Factor 2 were retained as they had Eigen values equal or higher than 1 (Table 3). The first factor with

an eigenvalue of 2.36, explains 69.4% of the total variance. Factors 3 through 16 have eigenvalues less than 1 and therefore explain less variance than a single variable.

Table 3: Unrotated factors

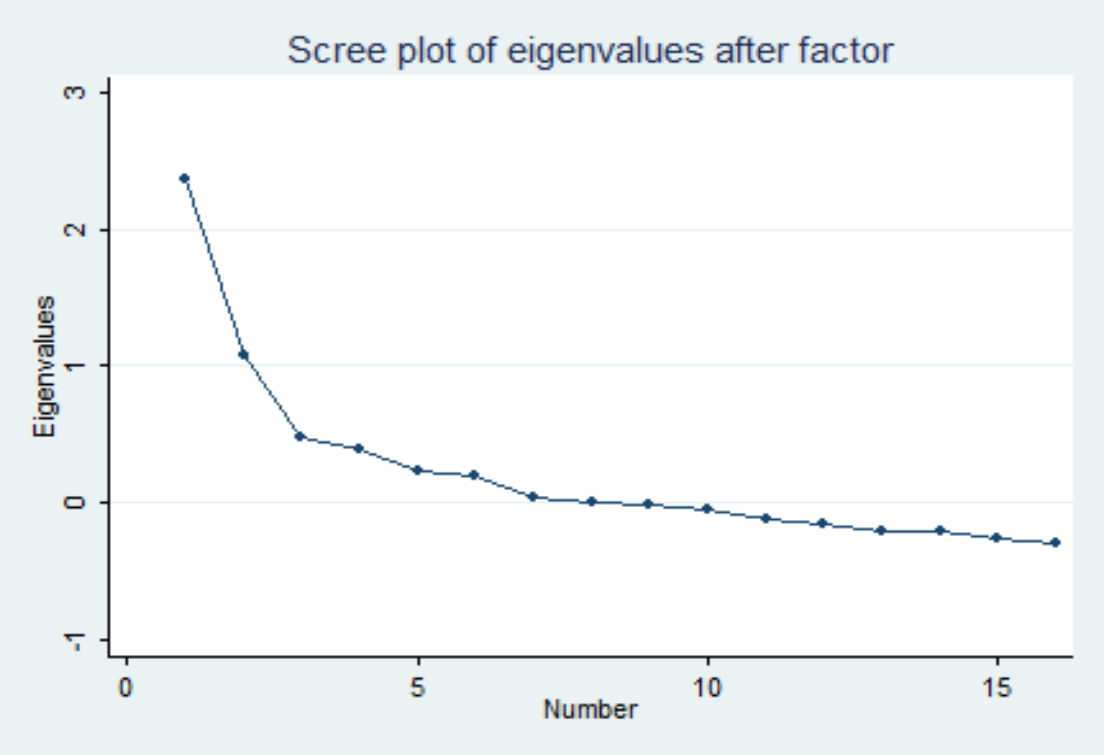
Factor	Eigen value	Difference	Proportion	Cumulative
Factor1	2.359	1.277	0.694	0.694
Factor2	1.081	0.601	0.318	1.012
Factor3	0.480	0.085	0.141	1.153
Factor4	0.394	0.162	0.116	1.269
Factor5	0.232	0.046	0.069	1.338
Factor6	0.187	0.159	0.055	1.392
Factor7	0.027	0.034	0.008	1.400
Factor8	-0.007	0.039	-0.002	1.398
Factor9	-0.015	0.077	-0.004	1.394
Factor10	-0.054	0.077	-0.016	1.378
Factor11	-0.131	0.037	-0.039	1.340
Factor12	-0.168	0.045	-0.050	1.290
Factor13	-0.212	0.004	-0.062	1.228
Factor14	-0.215	0.045	-0.063	1.165
Factor15	-0.260	0.038	-0.077	1.088
Factor16	-0.299	.	-0.088	1.00

This initial solution suggests that the final solution should extract not more than 2 factors since only 2 factors have Eigen values equal or higher than 1.

In Figure 1 below, we also confirm using the scree plot, the number of factors to be retained. The scree plot suggests that there are two factors above the point at which the graph forms an ‘elbow’ and trails off. The factors that are vertical and above the ‘elbow’ are the ones that account for a larger proportion of variance and are therefore more meaningful than the ones

that are horizontal and below the ‘elbow’, therefore based on the scree plot, two factors can be extracted.

Figure 1: Scree plot of Eigen values



Factor rotation

Factors were further rotated to clarify the factor pattern in order to better interpret the nature of the factors. Rotation was done using orthogonal varimax rotation method to get a clearer pattern and to maximise the loading of a variable on one factor while minimising the loading of the same variable on all the other factors. Factors that were loaded by a minimum of 3 variables were considered. Factor 1 was loaded by 5 variables while factor 2 was loaded by 4 variables; the rest of the factors had less than 3 variables loading on them. Factor 1 is mostly defined by variables b7-b11 (If I were sick, I can find someone to help with my VHW duties;

I can find at least one VHW whose advice I really trust; There is someone who gives me information to help me understand a VHW work's situation; There is someone I work with, who I can turn to for advice about handling problems with my family; If I need help visiting mothers, I know that I can always get help from the other VHWs respectively) and factor 2 is defined by variables b4, b12, b13 and b15 (I feel like I am responsible for so many activities that it's not possible to do them well in the time I have; I have had thoughts of leaving this job; I feel that I am responsible for more work than other colleagues; Unless I am rewarded for it in some way, I see no reason to spend extra effort promoting community health, respectively).

Table 4: Rotated factor loadings (pattern matrix) and unique variances

Variable	factor1	factor2	factor3	factor4	factor5	factor6	factor7	uniqueness
I feel like I enjoy my work as a village health worker				0.49				0.72
I feel that my work will improve people's lives.					0.41			0.76
I feel like I am adequately prepared for my responsibilities					0.38			0.73
I feel like I am responsible for so many activities that it's not possible to do them all		0.34						0.85

well in the time I have.				
I feel very connected with other village health workers		0.50		0.66
I feel well supported by the other VHWs I work with.		0.51		0.68
If I were sick, I can easily find someone to help with my VHW duties.	0.50			0.72
I can find at least one VHW whose advice I really trust.	0.64			0.54
There is someone who gives me information to help me understand a VHW work's situation	0.50		0.39	0.58
There is someone I work with who I can turn to for advice about handling	0.63			0.59

problems with my family.				
If I need help visiting mothers, I know that I can always get help from the other VHW.	0.63			0.57
I have had thoughts of leaving this job.		0.56		0.66
I feel that I am responsible for more work than other colleagues.		0.4472		0.7670
I am proud to tell others that I am a village health worker.				0.7484
Unless I am rewarded for it in some way, I see no reason to spend extra effort promoting community health.		0.3393		0.8207
How hard I work is directly linked to how			0.3127	0.8672

much I am paid.

(Blanks represent abs(loading) < 0.3)

The variables defining each of the two derived factors were tested for internal consistency reliability using Cronbach's alpha. Variables defining Factor 1 had a scale reliability coefficient of 0.73, which is acceptable while variables defining Factor 2 had a scale reliability coefficient of 0.50 which is poor but acceptable.

The derived factors were named according to the variables that define them. Factor 1 is peer support and Factor 2 is dissatisfaction.

4.3 Factors associated with attrition

Logistic Regression was done in order to determine if the factors derived from factor analysis and demographic factors predict attrition.

In the univariate logistic regression (Table 5), gender (OR = 0.47, 95%CI = 0.24 to 0.93; P = 0.03), and dissatisfaction (OR = 2.65, 95% CI = 1.68 to 4.19; p = 0.00), were statistically significant. For inclusion in the multivariable logistic regression analysis, variables with a p-value < 0.25 in the univariate analysis were considered. We implemented backward elimination method to remain with variables independently associated with attrition while adjusting for age, gender and marital status.

In the multivariable analysis, adjusted for age, gender education level and marital status, peer support (OR=0.64) and dissatisfaction (OR=3.20) independently predicted attrition. In particular, peer support was 36% protective of attrition while dissatisfaction results in a 3.2 increased odds of attrition

Table 5: Univariate and multivariable logistic regression analysis for variables independently associated with VHW attrition in the Zvitambo-SHINE project

Variable	Univariable OR (95% CI)	P-value	Multivariable OR (95% CI)	P-value
gender: male	1.0			
female	0.47 (0.24 to 0.93)	0.03	0.77 (0.35 to 1.72)	0.53
age	0.97 (0.93 to 1.00)	0.09	0.97 (0.92 to 1.02)	0.21
** marital status: single				
married	1.30 (0.29 to 5.88)	0.74		
divorced/separated	1.07 (0.13 to 8.56)	0.95		
widowed	0.37 (0.05 to 2.89)	0.34		
**Education status: Ever attended = 1				
Never attended =2	7.05* (0 to 274)*	1.00*		
Education level: primary school				
secondary school	1.99 (0.68 to 5.83)	0.21	1.22 (0.32 to 4.69)	0.77
tenure	0.99 (0.99 to 1.00)	0.10	1.00 (0.99 to 1.00)	0.37
peer support	0.78 (0.55 to 1.10)	0.16	0.64 (0.43 to 0.96)	0.03
dissatisfaction	2.65 (1.68 to 4.19)	0.00	3.18 (1.89 to 5.35)	0.00

*OR, CI and P-value derived from exact logistic regression analysis

**Variables not included in multivariate logistic regression.

4.3.1 Goodness of fit

The Hosmer-Lemeshow test was used to assess goodness of fit for the fitted model, a chi-square value 12.6 of with p=0.13 indicates that the fitted model is adequate.

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1 Discussion

The reliability of the items measuring motivation was determined by means of Cronbach's alpha coefficients and was found to be acceptable. Therefore, this indicates that these 16 items were useful in measuring motivation for job performance.

The findings from factor analysis in this study indicate that the main factors that influence motivation are peer support and dissatisfaction. The former is protective of attrition while the later increases the odds of attrition. The implementation of factor analysis resulted in the reduction of items that define motivation from 16 to 2. Peer support is generally when the VHW get assistance from fellow VHWs almost whenever they need it. Dissatisfaction is when the VHW show disgruntlement over their work.

Peer support comes in several forms. Several Non Governmental Organisation programs have successfully paired VHW so that they can work together and support each other. In the *Atencion Integral a la Ninez* (AIN) program in Honduras, VHW worked in groups of three. Working as a team allows VHW to divide their work and reduces the sense of isolation and complete responsibility for a particular geographic area. Teams provide mutual support and allow for local exchange of information (Marsh D, et al 1999). Holding monthly meetings might as well enhance peer support as it promotes bonding between the village health workers.

Dissatisfaction undermines the work performance of VHWs generally, and has the potential to compromise the health delivery system significantly. It is when VHW are dissatisfied that they may decide to quit, resulting in attrition.

Findings from Columbia, Mozambique, Nepal and Uganda show that peer support is an important motivation factor (Shetro 2000; Robinson and Larsen 1990; Taylor 2000), which is consistent with findings from this study.

There is limited research describing in detail the association between attrition and motivation factors. The results from logistic regression analysis in this study indicate that peer support is protective of attrition while dissatisfaction increases the odds of attrition. This study confirms earlier research that dissatisfaction amongst health workers is a major factor in healthcare staff leaving rural areas for the city or leaving the country altogether (Liese et al, 2003).

In this study almost all the selected demographic characteristics, that is, age, gender, marital status and education level predicted attrition, however the association was not statistically significant. There appeared to be no association between attrition and job tenure.

In the Solomon Islands, attrition was attributed to multiple causes, including lack of peer support, (Chevalier, 1993), which is in agreement with the results of this study.

Our study however had limitations despite these important findings. Firstly it could be possible that responses of VHWs were meant to comply with what they perceive to be an acceptable answer. Secondly the actual level of performance of the VHWs could not be assessed from the perspectives of either the community or the supervisors. The study also could not measure the levels of attrition.

5.2 Conclusion

Motivation can be defined by two factors, peer pressure and dissatisfaction both of which are associated with attrition. Peer support is protective of attrition; therefore there is need for the Ministry of Health to enhance peer support, particularly with regard to assistance they get from fellow colleagues when they are sick, when they have a sick relative to attend to, when

they need someone to confide in and advise about how to handle family problems and when they need information concerning their work's situation.

Dissatisfaction is a risk factor for attrition and must therefore be eliminated. There is need for the Ministry of Health to set clear performance goals and job descriptions. It is also important to reward work well done, as well as to give financial and non-financial incentives.

There is need for the Ministry of Health to address weaknesses identified and implement recommendations to motivate village health workers and consequently reduce attrition.

BIBLIOGRAPHY

Ajzen I (1991). The theory of planned behaviour. Organizational behaviour and decision processes. 50.178-211

Bjorklund C. (2001). Work motivation: Studies of its determinants and outcomes.

Chevalier C et al.(1993) Why do village health workers drop out? World health forum 14(3): 258-261

Dieleman M, Tooren J, Toure H & Martineau I. (2006). The match between motivation and performance management of health sector workers in Mali. *Human resources for Health*.

Dolea C, Adams O. (2005). Motivation of healthcare workers-review of theories and empirical evidence. Cah social Demogr Med. 135-161.

Furnham A, Jackson CJ, Miller T.Personality (1995). Learning style and work performance. Personality and individual differences.27.

George D & Mallary P (2003). SPSS for windows step by step. A simple guide and reference.4th ed. Bosto: Allyn & Boston.

Haines A, Sanders D, Lehmann U, Rowe AK, Lawn JE, Jan S. (2007). Achieving child survival goals: potential contribution of community health workers. *Lancet* ; 369

Kruse G.R, Chapula BT et al. (2009), Burnout and use of HIV services among health careworkers in Lusaka District. Zambia: A cross sectional study. Human Resources Health, 7: 55

Lewin S.A, Dick J, Pond P &Zwarenstein M. (2006). Lay health workers in primary and community health care. Cochrane Database Syst Rev.

Liese, Blanchet & Dussaul (2003). The human resource crisis in Health in Sub-Saharan Africa. Background paper for the World Bank.

Sein U.T. (2006). Health volunteers: Third workforce for Health-for-All movement. *Regional Health forum*.

Matheur I &Imhoff I. Health worker motivation in Africa (2006). The role of non-financial incentives and human resource management tools. *Human resources for Health*.

Mbindyo P, Gilson L et al (2009), Contextual influence on health worker motivation in district hospitals in Kenya. Implement Sci .

Mutale W (2012). Measuring health workers'motivation in rural health facilities. *Human resources for Health*.

Pinder C. (1998): Work motivation in organisationalbehaviour: Upper Saddle River, NJ: Prentice Hall.

UN (2007). Africa and the Millennium Development Goals 2007 update. United Nations.

WHO (2006). Working together for health. World health report 2006. Geneva. World Health Organisation.

Agresti A. Logistic Regression. Retrieved from http://lib.stat.cmu.edu/datasets/agrest.

Village Healthworkers Providing Healthcare in Rural Zimbabwe (2013, june 6). Retrieved from http://www.projecthigherlife.com/new.

Psychology Glosary. Retrieved from http://www.alledog.com/glossary/definition.

APPENDIX A

A1 Review of statistical theory

AI.1 Factor analysis

Types of Factor Analysis

There are two major classes of factor analysis: Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA). Broadly speaking EFA is heuristic. In EFA, the investigator has no expectations of the number or nature of the variables and as the title suggests, is exploratory in nature. That is, it allows the researcher to explore the main dimensions to *generate* a theory, or model from a relatively large set of latent constructs often represented by a set of items. Whereas, in CFA the researcher uses this approach to *test* a proposed theory (CFA is a form of structural equation modelling), or model and in contrast to EFA, has assumptions and expectations based on priori theory regarding the number of factors, and which factor theories or models best fit.

This discussion is limited to exploratory factor analysis.

Variable Reduction

Factor analysis reduces a large number of variables into a smaller set of variables (also referred to as factors). Factor analysis is a method for investigating whether a number of variables of interest Y1, Y2, : : :, Yl, are linearly related to a smaller number of unobservable factors F1, F2, : : :, Fk . It is used to explore the dimensionality of a measurement instrument by finding the smallest number of interpretable factors needed to explain the correlations among a set of variables – exploratory in the sense that it places no structure on the linear

relationships between the observed variables and the factors but only specifies the number of latent variables.

Assumptions

Factor analysis is designed for interval data, although it can also be used for ordinal data (e.g. scores assigned to Likert scales). The variables used in factor analysis should be linearly related to each other. Obviously the variables must also be at least moderately correlated to each other; otherwise the number of factors will be almost the same as the number of original variables, which means that carrying out a factor analysis would be pointless.

1. Calculate initial factor loadings.

This can be done in a number of different ways; the two most common methods are described very briefly below:

• Principal component method

As the name suggests, this method uses the method used to carry out a principal components analysis. However, the factors obtained will not actually be the principal components (although the loadings for the kth factor will be proportional to the coefficients of the kth principal component).

• Principal axis factoring

This is a method which tries to find the lowest number of factors which can account for the variability in the original variables that is associated with these factors (this is in contrast to the principal components method which looks for a set of factors which can account for the total variability in the original variables). These two methods will tend to give similar results if the variables are quite highly correlated and/or the number of original variables is quite high. Whichever method is used, the resulting factors at this stage will be uncorrelated.

What criteria will assist in determining factor extraction?

The aim of the data extraction is reduce a large number of items into factors. In order to produce scale unidimensionality, and simplify the factor solutions several criteria are available to researchers. However, given the choice and sometimes confusing nature of factor analysis, no single criteria should be assumed to determine factor extraction.Many extraction rules and approaches exist including: Kaiser‟s criteria (eigenvalue > 1 rule), the Scree test, the cumulative percent of variance extracted, and parallel analysis. It is suggested that multiple approaches be used in factor extraction.

Cumulative Percentage of Variance and Eigenvalue > 1 Rule

Cumulative percentage of variance (criterion) is another area of disagreement in the factor analysis approach, particularly in different disciplines, for example, the natural sciences, psychology, and the humanities. No fixed threshold exists, although certain percentages have been suggested, forexample, in the natural sciences, factors should be stopped when at least 95% of the variance is explained. In the humanities, the explained variance is commonly as low as 50-60%.

Scree Test

Interpreting Scree plots is subjective, requiring researcher judgement. Thus, disagreement over which factors should be retained is often open for debate. Although this disagreement and subjectiveness is reduced when sample sizes are large, *N: p* ratios are (>3:1) and communalities values are high. The Scree Test‟ was given its name by Cattell due to the Scree Test graphical presentation, which has visual similarities to the rock debris (Scree) at the foot of a mountain.

Inspecting and interpretation of a Scree plot involves two steps:

1. Draw a straight line through the smaller eigenvalues where a departure from this line occurs. This point highlights where the debris or break occurs. (If the Scree is messy, and difficult to interpret, additional manipulation of data and extraction should be undertaken).
2. The point above this debris or break (not including the break itself) indicates the number of factors to be retained.

Parallel Analysis

Parallel analysis is an under-used factor extraction technique5 and is often not reported in the literature. One possible reason for limited use is the analysis is not available in conventional statistical programs. In parallel analysis, actual eigenvalues are compared with random order eigenvalues. Factors are retained when actual eigenvalues surpass random ordered eigenvalues.

2. Factor rotation

Once the initial factor loadings have been calculated, the factors are rotated. This is done to find factors that are easier to interpret. If there are 'clusters' (groups) of variables — i.e. subgroups of variables that are strongly inter-related — then the rotation is done to try to make variables within a subgroup score as highly (positively or negatively) as possible on one particular factor while, at the same time, ensuring that the loadings for these variables on the remaining factors are as low as possible. In other words, the object of the rotation is to try to ensure that all variables have high loadings only on one factor. There are two types of rotation method, orthogonal and oblique rotation. In orthogonal rotation the rotated factors will remain uncorrelated whereas in oblique rotation the resulting factors will be correlated. There are a number of different methods of rotation of each type. The most common

orthogonal method is called varimax rotation; this is the method that many books will recommend.

Interpretation

Interpretation involves the researcher examining which variables are attributable to a factor, and giving that factor a name or theme. For example, a factor may have included five variables which all relate to pain perception; therefore the researcher would create a label of "pain perception" for that factor. Traditionally, three variables must load on a factor so it can be given a meaningful interpretation. The labelling of factors is a subjective, theoretical, and inductive process.The reason for thorough and systematic factor analyses is to isolate items with high loadings in the resultant pattern matrices.

A1.2 Logistic Regression

Logistic regression is used to analyze relationships between a dichotomous dependent variable and metric or dichotomous independent variables. Logistic regression combines the independent variables to estimate the probability that a particular event will occur, i.e. a subject will be a member of one of the groups defined by the dichotomous dependent variable. Logistic regression determines the impact of multiple independent variables presented simultaneously to predict membership of one or other of the two dependent variable categories. There are two main uses of logistic regression:

The first is the prediction of group membership. Since logistic regression calculates the probability of success over the probability of failure, the results of the analysis are in the form of an odds ratio. Logistic regression also provides knowledge of the relationships and strengths among the variables.

Principles of Logistic Regression

Logistic regression involves fitting an equation to the data. The equation is of the form: Logit (p) = $\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \ldots\ldots\ldots\ldots + \beta_iX_i$, in this case, the equation becomes; Logit(p) = β_0 + β1age+ β2educational qualification + β3manual dexterity+ β4motivation +β5previous experience+β6self discipline**.**

Categorical independent variables need to be coded because they do not have a continuous scale of measurement. Indicator (dummy) variables taking the values 0 and 1 to indicate the presence or absence of an attribute must be used to correctly represent the effects of such variables.

The logistic regression model can be used to express the relationship between a dichotomous outcome and an independent variable in 3 equivalent ways:

1. In terms of π, the probability of an event.
2. In terms of $\pi/1-\pi$, the odds of an event.
3. In terms of $\ln(\pi/1-\pi)$,the logit or log odds of an event.

The logistic regression model ln(p/1-p) is given the name because of the log transformation. For particular values of β_0, β_i and we have the explanatory variables, the probability of success may be greater than 1 or less than 0. The logit transformation is the most commonly used solution to prevent π_i from being outside the range of probabilities.

Variable selection

The process should begin with a careful univariate analysis of each variable, using likelihood ratio test or the Wald statistic. Upon completion, we select variables for the multivariate model. Any variable whose univariate test has a $p<0.25$ should be considered as a potential candidate, together with variables of known biological importance. Identify the significant interaction terms and include them in the model

Model fit and the likelihood function

The Maximum Likelihood (or ML) is used to find the function that will maximize our ability to predict the probability of Y based on what we know about X. Likelihood just means probability. It always means probability under a specified hypothesis. In logistic regression, two hypotheses are of interest: the null hypothesis, which is when all the coefficients in the regression equation take the value zero, and the alternate hypothesis that the model with predictors currently under consideration is accurate and differs significantly from the null of zero, i.e. gives significantly better than the chance or random prediction level of the null hypothesis. We then work out the likelihood of observing the data we actually did observe

under each of these hypotheses. The result is usually a very small number, and to make it easier to handle, the natural logarithm is used, producing log likelihood (LL). Probabilities are always less than one, so LL's are always negative. Log likelihood is the basis for tests of a logistic model[5].

The likelihood ratio test is based on –2LL ratio. It is a test of the significance of the difference between the likelihood ratio (–2LL) for the researcher's model with predictors (called model chi square) minus the likelihood ratio for baseline model[7].

Interpreting log odds and the odds ratio

The slope can be interpreted as the change in the average value of Y, from one unit of change in X.

Assumptions

- Logistic regression does not assume a linear relationship between the dependent and independent variables.
- The dependent variable must be a dichotomy (2 categories).
- The independent variables need not be interval, nor normally distributed, nor linearly related, nor of equal variance within each group.
- The categories (groups) must be mutually exclusive and exhaustive; (a case can only be in one group and every case must be a member of one of the groups).
- Group sizes of the dependent variable can be different.