UNIVERSITY OF SOUTHAMPTON
FACULTY OF SOCIAL AND HUMAN SCIENCES
School of Social Sciences

Measuring HIV Awareness and Knowledge: Analyses of Cross-sectional Surveys with a Focus on China

by

Olga Maslovskaya

Thesis for the degree of Doctor of Philosophy
March 2011

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

Doctor of Philosophy

MEASURING HIV AWARENESS AND KNOWLEDGE: ANALYSES OF CROSS-SECTIONAL SURVEYS WITH A FOCUS ON CHINA

by Olga Maslovskaya

HIV prevalence in China is currently less than one percent, but due to the large population this translates into a large number of people. The number of people living with HIV is growing and moving beyond high-risk groups to the general population. Ensuring adequate HIV awareness and knowledge is important for the successful prevention of HIV. This thesis investigates the evolution of HIV awareness and knowledge in China between 1997 and 2005. It also compares two methodological approaches to measuring HIV knowledge: a simple score approach and a latent variable approach. Three papers are presented and each addresses the main substantive issue using different methodologies. Various data sources and techniques used in the thesis provide each paper with its own perspective on the main substantive research question and unique insights into the main substantive and methodological issues.

The first paper examines the evolution of HIV awareness among women in China between 1997 and 2005. The aim of this paper is to compare the levels of HIV awareness at various points in time. A regression decomposition analysis technique is used in this paper in order to disentangle the two main components driving a change in HIV awareness: the change in a population structure and the change in effect sizes due to external factors such as political environment, interventions and programmes. The results show that HIV awareness increased over time in China. With time, lower awareness groups are catching up and gaps between groups with initially different awareness levels are narrowing. The results suggest that the main driver of the observed change in HIV awareness over time in China is the change in the environment such as in political commitment, interventions and campaigns.

The second and third paper both focus on the evolution of HIV knowledge among women in China between 1997 and 2005. The main aim of these papers is to assess whether China has succeeded in improving women's HIV knowledge over time, and if China is a relative success story in improving women's HIV knowledge when compared with other countries in the world with generalised (Kenya and Malawi) as well as with non-generalised (India and the Ukraine) HIV epidemics. The second paper uses a simple score approach to measuring HIV knowledge, whereas the third paper uses a latent variable approach. Partial proportional odds and multinomial logistic regression modelling techniques are employed for the analysis of patterns of HIV knowledge in China over time and in other countries included in the analyses. The main findings indicate that China has succeeded in improving women's HIV knowledge. HIV knowledge in China is comparable to HIV knowledge in other countries with non-generalised epidemics. The HIV knowledge in China has become more homogeneous over time across different groups. However, the gap between the groups still exists and, therefore, more efforts should be directed towards improvement of HIV knowledge among women in China as well as in other cultural and epidemiological contexts. The main methodological findings show that both simple score and latent variable approaches to measuring HIV knowledge are useful and provide unique insights into the topic of the evolution of HIV knowledge in China.

Contents

List of Figures

List of Tables

List of Boxes

Declaration of Authorship

I, Olga Maslovskaya, declare that the thesis entitled

'Measuring HIV Awareness and Knowledge: Analyses of Cross-sectional Surveys with a Focus on China'

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission

Signed:…………………………………………………………..

Date:…………………………………………………………….

Acknowledgements

This PhD in Social Statistics would not have been possible without 1+3 scholarship from the Economic and Social Research Council (ESRC) (scholarship number PTA-031-2006-00185).

First and foremost, I would like to express my sincere gratitude to my supervisors, Professor Peter W.F. Smith and Dr. Sabu S. Padmadas, for their attentive supervision of the research. I really appreciate that they always had time to answer my questions, and that they always read my drafts very quickly. I am grateful for their valuable comments, suggestions, ideas, help, guidance, and their moral support and encouragement throughout the entire process of my PhD. I learned a lot from my supervisors during the last three years and I am very grateful for all the knowledge I have acquired.

I would like to thank my advisor Dr. James J. Brown for his help and valuable suggestions and comments. I am grateful to my upgrade examiner Professor Nyovani J. Madise for her useful comments and suggestions during the upgrade exam.

I would like to say thank you to my current employers in the EPSRC Care Life Cycle (CLC) Research Programme, Professor Jane Falkingham and Professor Maria Evandrou, who supported me during the final stage of my PhD when I was preparing the final draft for the submission.

I am grateful to Professor Li Bohua from the China Population and Development Research Centre in Beijing, China for access to the China National Population and Reproductive Health Survey 1997 and the China National Family Planning and Reproductive Health Survey 2001 and for answering my questions regarding these and other Chinese data sources which I used in my thesis for the analysis.

I am also grateful to all my friends and colleagues from the Division of Social Statistics and from other divisions in the University of Southampton, from the S3RI and from the CLC Research Programme and to all my friends from outside academia for their help, valuable advice, moral and practical support and encouragement during my PhD.

I would like to thank my family back in Belarus for their support. My special thanks to my husband, Georgi Pavlovski, for his love, kindness, moral support, patience and understanding during the last three years, and especially during the final stages of my PhD.

I am also grateful to Kay Channon for proofreading this thesis.

Thank you all!

List of Acronyms

AIC	Akaike Information Criterion
AIDS	Acquired Immune Deficiency Syndrome
ARRM	AIDS Risk Reduction Model
ARV	Antiretroviral
BIC	Bayesian Information Criterion
CAIC	Consistent Akaike Information Criterion
CDC	Centre for Disease Control
CHAMP	China AIDS Media Partnership
CIA	Central Intelligence Agency
CPDRC	China Population Development and Research Centre
CP4	Country Programme Four
CP5	Country Programme Five
CSW	Commercial Sex Worker
DF	Degree of Freedom
DHS	Demographic and Health Survey
EFA	Exploratory Factor Analysis
FPD	Former Plasma Donors
FP	Family Planning
HIV	Human Immunodeficiency Virus
ICA	Intelligence Community Assessment
ICD	International Statistical Classification of Diseases
ICPD	International Conference on Population and Development
IDU	Intravenous Drug User
IEC	Information, Education and Communication
IIPS	International Institute for Population Sciences
IMB	Information-Motivation-Behavioural Skills Model
IRT	Item Response Theory
KAB	Knowledge, Attitudes and Behaviour
KABP	Knowledge, Attitudes, Beliefs and Practices
LCA	Latent Class Analysis

LTA	Latent Trait Analysis
MAIC	Modified Akaike Information Criterion
MDG	Millennium Development Goals
MSM	Men who have Sex with Men
MTCT	Mother-to-Child Transmission
NFHS	National Family Health Survey
NGO	Non-Governmental Organisation
OR	Odds Ratio
PSU	Primary Sampling Unit
RH	Reproductive Health
RSE	Robust Standard Error
SARS	Severe Acute Respiratory Syndrome
SE	Standard Error
SPSS	Statistical Package for the Social Sciences
STD	Sexually Transmitted Disease
STI	Sexually Transmitted Infection
S3RI	Southampton Statistical Sciences Research Institute
TV	Television
UK	United Kingdom
UN	United Nations
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNFPA	United Nations Population Fund
UNGASS	United Nations General Assembly Special Session
US	United States
UV	Underlying Variable
WHO	World Health Organisation

Chapter 1: Introduction

The main purpose of this chapter is to present the introduction to the thesis which contains necessary background material, the rationale for the study, and the common for all three papers literature review. The main aims and research questions of the thesis, structure of the thesis as well as conceptual framework are also presented in this chapter.

1.1 HIV Epidemics in the World

It has been almost 30 years since the first cases of HIV/AIDS were reported in the USA. Since that time HIV/AIDS has been recorded in all parts of the world. According to the 2009 HIV estimates, 33.3 million [31.4-35.3 million][1] people are living with HIV in the world (UNAIDS 2010).

One of the eight Millennium Development Goals (MDGs) formulated at the Millennium Summit in 2000 in the United Nations Millennium Declaration[2] is to combat HIV/AIDS, malaria and other diseases by 2015 (Goal 6). Many efforts and measures to control HIV epidemic have been taken across the world. Some positive results have been achieved: according to UNAIDS (2010) by the year 2009 in many countries of the world (with the exception of some countries in Eastern Europe and Central Asia) HIV epidemics "have either stabilized or are showing signs of decline" (p. 8). However, HIV is still spreading worldwide and humanity is still very far from the end of HIV/AIDS and not everyone in the world who needs the life-prolonging antiretroviral treatment has access to it (Lancet 2010). In order to avoid the HIV epidemic moving in the other direction in places where some successes have been already observed and in order to achieve successes in countries where HIV epidemics are still not declining, comprehensive efforts should be

---

[1]This range and ranges below are confidence intervals which "reflect the degree of uncertainty associated with estimates and define boundaries within which the actual numbers lie" (UNAIDS 2008b, p.4).
[2] http://www.un.org/millennium/declaration/ares552e.pdf [Accessed 28 July 2009]

employed and sustained. A vaccine against HIV is still not available and, therefore, the main force against the spread of HIV/AIDS worldwide is HIV prevention.

According to the International Statistical Classification of Diseases (ICD-10) [3] HIV/AIDS is classified as an infectious disease. However, in contrast to many other infectious diseases, the main two vectors of HIV transmission are not environmental but behavioural: through risky sexual behaviours and through sharing of injection equipment among drug users. The mainly behavioural nature of HIV transmission makes it difficult to develop and implement necessary and effective prevention programmes. One of the main priorities for HIV prevention to date was to develop and implement strategies to modify risk behaviours. However, years of experience suggest that "no simplistic or even simple solutions exist for HIV prevention" (Coates *et al.* 2008, p.669).

HIV prevention strategies have several dimensions. Merson *et al.* (2008), Padian *et al.* (2008), Coates *et al.* (2008) and Gupta *et al.* (2008) argue that HIV prevention efforts can only be effective if a combination of behavioural, structural and biomedical dimensions is considered when HIV prevention interventions and programmes are designed and implemented. The structural dimension of HIV prevention deals with changes in the economic, social, political, legal, cultural, community and other contexts which contributes to the reduction of HIV vulnerability and risk (Gupta *et al.* 2008; Coates *et al.* 2008). Examples of actions within this dimension of HIV prevention include implementation of policies, creation of supportive environment, enforcement of human rights and gender equality, and minimisation of poverty. The biomedical dimension interventions aim to block HIV infection and to reduce infectiousness of people living with HIV through availability of tests and antiretroviral (ARV) therapy, through the use of male condoms, through male circumcision, through treatment of sexually transmitted infections and other methods (Padian *et al.* 2008; Coates *et al.* 2008). The behavioural dimension strategies "attempt to motivate behavioural change within individuals and social units by use of a range of educational, motivational, peer-group, skills-building approaches, and community normative approaches" (Coates *et al.* 2008, p. 670). The main components of the behavioural dimension are "knowledge, stigma reduction, access to services, delay of onset of first intercourse, decrease in number of partners, increases in condom sales or use, and decreases in sharing of

---

[3] http://apps.who.int/classifications/apps/icd/icd10online/ [Accessed 05 March 2011]

contaminated injection equipment" (Coates *et al*. 2008, p. 669). All these three dimensions should operate in synergy in order to achieve positive results in the area of HIV prevention (UNAIDS 2010).

To increase HIV awareness and knowledge is one of the main components of the behavioural dimension jigsaw, and this component is especially useful in countries where HIV epidemic is still at an early stage. Understanding how HIV awareness and knowledge are spatially distributed across different population groups is crucial to formulate appropriate interventions.

Regions and countries of the world differ by prevalence and incidence rates as HIV epidemics are at different stages in different parts of the world. It can be argued that countries usually follow the same trajectory with the HIV epidemic starting among high-risk groups (men who have sex with men (MSMs), intravenous drug users (IDUs), commercial sex workers (CSWs)), and then moving to the general population predominantly through heterosexual transmission. Sub-Saharan Africa is the part of the world which is hardest hit by the HIV epidemic with the generalised HIV epidemics[4] and the highest HIV prevalence rates in the world. The main route of HIV transmission in sub-Saharan Africa is heterosexual transmission.

The assessment by the National Intelligence Council in the US highlighted the rising HIV/AIDS problem in five countries that have large populations at risk for HIV infection: Nigeria, Ethiopia, Russia, India, and China (ICA 2002). These five countries in Africa and Eurasia have been labelled the "next-wave countries" (ICA 2002). Eberstadt (2002) argued that "Eurasia[5] will likely be home to the largest number of HIV victims in the decades ahead" (p.22). Eberstadt (2002) also stated that "it is quite possible that the centre of the global HIV/AIDS crisis, in terms of absolute numbers, will shift from Africa to Eurasia over the coming generation" (p.22). It can be argued with a high degree of confidence that by 2007 HIV/AIDS has become established in Eurasia. Russia together with some countries in Eastern European and Central Asia (including countries of the Former Soviet Union) are still experiencing the most rapidly expanding and increasing HIV/AIDS epidemics in the world (UNAIDS 2010). According to recent UNAIDS estimates, the HIV epidemic in India is decreasing

[4] An HIV epidemic is considered to be generalised when the prevalence rate is above 5% in urban areas among women attending antenatal clinics (World Bank 1997).
[5] Here "territory encompassing the continent of Asia, plus Russia" (Eberstadt 2002, p.22).

(UNAIDS 2010). Unfortunately, China together with some other countries with complex HIV epidemics was not included in the UNAIDS 2010 epidemic estimation round. Therefore, it is not clear whether the HIV epidemic in China stabilised or not (UNAIDS 2010). In the whole region of Eurasia the epidemics are still at expansion stages, i.e. are still contained within high-risk groups, but it is just a matter of a short period of time before the epidemics can become generalised and HIV can affect not only people from high-risk groups but also general populations. It could be argued that the whole region is on the brink of an HIV catastrophe which could lead to infection rates comparable with those in the hardest-hit parts of sub-Saharan Africa, if effective measures are not implemented and sustained. The main difference between the HIV catastrophe in Africa and the potential one in Eurasia is the absolute numbers of people infected with HIV due to large populations in countries in Eurasia. The HIV catastrophe, if advanced, will bring devastations at individual, household and country levels in Eurasia, and will affect the population structure and economic development which might negatively affect the prospects of the nations and of the whole region.

1.2 HIV Epidemic in China

1.2.1 Current situation

The population of China was estimated to be 1,338,100,000[6] in mid-2010 by the Population Reference Bureau. According to the most recent UNAIDS estimates, 740,000 [540,000-1,000,000] people (adults and children) lived with HIV/AIDS in China in 2009 (UNAIDS 2010). New estimations were produced in 2009 by using better estimation methods and more representative data collection with improved data quality (Wang *et al*. 2010). The estimated prevalence rate in China in 2009 was 0.05% [0.04-0.07%], and the estimated HIV prevalence among adults of reproductive age (15-49 years) was 0.1%[7] (UNAIDS 2010). The prevalence rate is not high but because of the large population of China, it translates into a large number of people. The number of people living with HIV in China is comparable to, for example, 41% of the population of Botswana (1,800,000 people)[8] or more than half of the population of

---

[6] http://www.prb.org/Datafinder/Geography/Summary.aspx?region=164&region_type=2 [Accessed 25 November 2010]
[7] Range for this estimate was not available in the UNAIDS 2010 report.
[8] http://www.prb.org/Datafinder/Geography/Summary.aspx?region=64&region_type=2 [Accessed 23 November 2010]

Swaziland (1,200,000 people)[9]. These two countries have one of the highest HIV prevalence among adults of reproductive age in the world (24.8% [23.8-25.8%] and 25.9% [24.9-27.0%] respectively) but estimated numbers of people living with HIV in these two countries are lower than in China (320,000 [300,000-350,000] in Botswana and 180,000 [170,000-200,000] in Swaziland) (UNAIDS 2010)[10]. This suggests that even in a non-generalised HIV epidemic context in China if the epidemic is not controlled appropriately, it can become a huge problem due to the fact that it affects a large number of people.

It was estimated that 230,000 [160,000-300,000] women (15+) lived with HIV in China in 2009 (UNAIDS 2010) which means that the estimated male to female ratio of HIV infections is currently 2.2:1 and this ratio becomes smaller with time (for example, in 2007 the ratio was 2.5:1 (UNAIDS 2007)). It was also estimated that in China by the end of 2009 26,000 [24,000-49,000] people have already died from AIDS-related conditions (UNAIDS 2010).

The most recent estimates of number of people living with HIV are lower than the ones which were produced in 2003 and suggested that 840,000 people in China were living with HIV at that time (State Council 2004). As mentioned earlier, all these new estimations were produced by using better estimation methods and more representative data collection. However, despite the fact that a smaller number of people is thought to live with HIV, the number of cases is growing and the virus is spreading from the traditionally high-risk groups to the general population.

By 2006 not more than 22% of people infected with HIV had been identified (Gill and Okie 2007; Wu *et al*. 2007b). Zhang (2004) provides a number of reasons to explain why it is difficult to obtain reliable data on HIV/AIDS in China. Some of the reasons are "the country's underdeveloped disease reporting system and epidemic surveillance networks; lack of knowledge and low awareness about HIV among the general public and medical professionals, especially in the countryside; shortage of diagnostic facilities in hospitals and health centres; intense stigma and discrimination against those living with HIV/AIDS and their families; and cover-ups by local authorities, who fear that

[9] http://www.prb.org/Datafinder/Geography/Summary.aspx?region=68&region_type=2 [Accessed 23 November 2010]

[10] These figures are presented not to suggest that the epidemic is more severe in China than in African countries but to illustrate the extent of the epidemic in the Chinese context due to the large population in the country.

their localities will lose external investment if the full extent of the problem were disclosed, or that they would be demoted for failing to meet the envisaged GDP growth targets because of the loss of business and financial inflows due to AIDS" (Zhang 2004, p. 1158). All these issues should be addressed in order to be able to identify people with HIV and to be able to provide medication and social support to people infected with and affected by HIV/AIDS in China and to slow down the HIV epidemic.

1.2.2 History of HIV/AIDS in China

Zhang (2004) and Wu *et al*. (2004) define the three major stages of the spread of HIV epidemic in China. These stages are similar to stages in other countries, where generalised HIV epidemics had been recorded, experienced in the past. HIV/AIDS was reported in China slightly later than in other regions of the world. The first stage is thought to have started in 1985 when the first case of AIDS was diagnosed in Beijing in a US citizen (Zhang and Beck 1999; Zhang 2004; Wu *et al*. 2007b). This stage is described as "sporadic cases amongst relatively better-off individuals and groups" mainly in coastal cities (Zhang 2004, p.1156). Wu *et al*. (2004) define this stage as "the introduction phase".

The second stage started in 1989 when the virus started spreading "amongst injecting drug users in western and south-western provinces of Yunnan, Guangxi and Xinjiang situated along the main drug trafficking routes" (Zhang 2004, p.1156). According to Wu *et al*. (2004) this stage can be described as "the spreading phase".

The third stage is demarcated by the rapid spread of the virus in the whole country from 1995. During this stage the virus began to spread beyond the high-risk groups. This stage is defined by Wu *et al*. (2004) as "the expansion phase". This stage is also characterised in China by a major change in the governmental attitude towards HIV/AIDS and its priority in the country's political agenda which will be discussed below.

In the 1980s and the 1990s two main HIV epidemics in China were identified: an epidemic which was fuelled by blood donation mainly in Central China, and the second one affected intravenous drug users in Southern China. The first outbreak of HIV epidemic was recorded among intravenous drug users in Yunnan province in China in the late 80s (Wu *et al*. 2004). Later in the 90s the second HIV epidemic which affected commercial plasma donors in rural areas of Central China was documented. During recent years the third major HIV epidemic has been observed in China – an epidemic which progressed through the sexual transmission: commercial sex workers, their client and migrant workers are major contributors to this particular epidemic as well as sexual partners of intravenous drug users and those people who were infected with HIV during the blood trade scandal in China. The virus now is moving to the mainstream heterosexual population and many new cases acquired the virus via heterosexual contacts. Homosexual transmission is also present in China. However, due to the cultural taboo which surrounds homosexuality there are problems with estimating the number of HIV cases that can be attributed to the homosexual transmission. The main vulnerable groups in China will be discussed in detail below.

As mentioned earlier the gender distribution of HIV-infected people in China follows the same pattern as other countries and regions in the world: at the start of epidemic the majority of infected people were males. However, the ratio is changing over time and the number of infected females is growing persistently. According to Zhang (2004) at the beginning of the epidemic the male-to-female ratio was 5:1, in 2001 it was estimated to be 4:1, and as mentioned earlier according to the latest UNAIDS estimates, it was 2.2:1 in 2009 (UNAIDS 2010). The increasing number of infections among women means that the cases of mother-to-child transmission (MTCT) increase and as a result of that there is also a rise in the number of children infected by HIV.

Geographically HIV spreads all over the country and some parts of China are more affected than others. The most affected provinces in China are: Henan, Yunnan, Guangxi, Xinjiang, Guangdong (Gill and Okie 2007). In some areas mainly drug users are affected (Yunnan, Guangxi, Xinjiang) whereas in other areas former blood and plasma donors are affected (Henan, Hebei, Anhui, Shanxi) (Wu *et al.* 2007b).

China has a specific feature of the HIV epidemic: until recently it predominately affected rural poor people (IDUs in Southern part of China and farmers in Central China) and only later it started spreading to urban areas (Wu *et al*. 2004; Wu *et al*. 2007b). In African countries the HIV epidemic is more concentrated in urban areas than in rural areas. According to Wu *et al*. (2004), in 2000 80% of HIV infected individuals lived in rural areas of China. In rural areas residents are generally poorer than in urban areas and have worse access to health care because of the rural health crisis (Gill and Okie

2007; Dummer and Cook 2007). Rural areas suffer from lack of trained medical personal and from the lack of good quality medical infrastructure (Gill and Okie 2007). Lee (2004) argues that "[a]t present, the allocation of health resources focuses on city and treatment and ignores rural areas and prevention" which exacerbates the problem even further (p.335). This suggests that even if the antiretroviral treatment is available in China, people in rural areas on many occasions are excluded from services because of the reasons mentioned above. Moreover, the level of stigma and discrimination is higher in rural areas due to the fact that communities are more traditional and less open-minded.

The HIV epidemic predominantly affects young (Gill and Okie 2007) and sexually active people in China which means that once sexual transmission becomes the main route of HIV infection, HIV will spread quickly to the general population as happened in sub-Saharan Africa, and then it will be much harder to control the epidemic.

Merli and Hertog (2010) argue that China is characterised by demographic features unique in the world, such as oversupply of males in the market for sexual partners and rapid fertility decline since the 1970s, which affects the Chinese population age structure. This oversupply of males in the society and associated potential squeeze in the marriage and sexual partnership markets will lead to an increase in occurrence of unprotected sex with commercial sex workers and will become important contributors to the further spread of HIV through heterosexual transmission (Merli and Hertog 2010).

Thompson (2005) argues that "other sexually transmitted infections (STIs) are increasing significantly in the general population" (p.7). Untreated STIs can facilitate the transmission of HIV through genital ulceration (Shapiro 2002; Bongaarts 1996; Buve *et al.* 2001). The increased level of STIs in the country suggests the risky sexual behaviour and the increased risk of HIV transmission.

Globally, four main routes of HIV transmission are recognised: sharing contaminated equipment for intravenous drug use, unprotected sexual intercourse (heterosexual or homosexual), transmission from an infected mother to her baby or so called mother-to-child transmission (MTCT), and transfusion of contaminated blood or blood products in medical settings.[11]

Currently the main routes of HIV transmission in China are intravenous drug use and homosexual and heterosexual intercourses, with MTCT becoming another route of transmission as the number of women infected with HIV increases. The main transmission modes have changed over time: in 2003 50-60% of the total reported cases were IDUs and only 7-8% cases were attributed to sexual transmission (Zhang 2004), but "[a]mong new infections in China, 48.6% are caused by drug use, 49.8% by sexual transmission" (Gill and Okie 2007, p.1802). In the 90s, blood donation was one of the main routes of HIV transmission in Central China. At the beginning of the 90s, the Chinese government provided an opportunity to poor farmers to earn some extra money by donating blood (Zhang 2004; Wu *et al*. 2007b). Mobile blood donation stations were organised in many villages in Central China. In many blood stations the safe procedure of blood collection was violated: instruments were reused, blood of different people was mixed together and after plasma was extracted, the rest of the mixture was injected back to donors in order to prevent anaemia and to enable them to give blood again within a short time (Zhang 2004; Wu *et al*. 2007b). These violations caused an HIV epidemic in Central China (Zhang 2004; Wu *et al*. 2007b). According to Zhang (2004), a survey in selected areas of Henan province revealed that as many as 43% of the adult population were former plasma donors (FPDs). She also argued that "the prevalence rate amongst FPDs reaching 10-20% on average, with the highest being over 60% in certain communities" (Zhang 2004, p. 1160). For a long time the government was trying to cover up the blood scandal which had a bad impact on the people who were infected or affected by HIV/AIDS. The Chinese government has initiated efforts to control blood donation system: voluntary blood donation is now encouraged and commercial blood donation is banned in China. According to Adams *et al*. (2009), China's public health system managed to successfully control blood donation infrastructure in urban China. However, there are still reported cases of HIV infections through blood products in the country.

Drug use is illegal in China. It is a criminal offence to use or to sell drugs. Drug users belong to a hard-to-reach group which results in the lack of knowledge about HIV/AIDS among the group and creates difficulties in accessing this group with HIV

[11] http://www.cdc.gov/hiv/resources/factsheets/PDF/transmission.pdf [Accessed 28 July 2009]

prevention messages. It is a common practice in China that needles and syringes are shared between people. Lack of knowledge about HIV/AIDS and its routes of transmissions fuels the increase in number of people who are infected with HIV through this mode of HIV transmission. It was estimated that in 2009 HIV prevalence among intravenous drug users in Beijing was 9.3% (UNAIDS 2010).

Harm reduction programmes include needle exchange, provision of methadone to drug users, and outreach educational programmes for IDUs. However, while international experience demonstrates that harm reduction is an effective form of HIV prevention programmes for IDUs, access to clean needles and other components of harm reduction are not yet a common practice in all parts of China. The argument against harm reduction programmes is that these programmes contribute to and even encourage drug use. Recent change in attitude towards needle-exchange programmes was reported in China and the Chinese government is now supporting needle-exchange programmes and setting up methadone-maintenance therapy sites for IDUs (Gill and Okie 2007). However, those policies are not always implemented at local levels and, therefore, not all people have access to the programmes and as a result of this the number of people infected with HIV increases.

Methadone (drug replacement therapy) has proved to be very successful in reducing HIV transmission among IDUs in countries where it is available. Pilot programmes were organised in China and proved to be successful and as a result of "the successes of the pilot, the programme began scaling-up in 2004 and plans are in place to open an additional 1500 methadone maintenance treatment clinics for about 300,000 heroin users by 2008" (Wu *et al*. 2007b, p.682). All these positive changes can be interpreted as good progress in response to HIV epidemic in the country.

Men who have sex with men (MSM) form another risk group and this group is hidden underground in China as until 2001 homosexuality was considered as a psychiatric disorder (Zhang 2004; Wu *et al*. 2007b). Because of this, homosexual people are still highly stigmatised and experience social pressure, and this stigma fuels the HIV epidemic as it does in many other parts of the world. As this group is a hard-to-reach group, prevention messages if they exist can not easily reach people from the group. "There are strong taboos against homosexual behaviour in China, where men are under enormous pressure to marry and produce male heirs" (Gill and Okie 2007, p.1804). In this situation homosexual men can become a bridge population and if infected with HIV, can spread the virus to their girlfriends and wives (Thompson 2005; He and Detels 2005). "Chinese health officials estimate that nationwide about 1% of men who have sex with men are HIV-positive" (Gill and Okie 2007, p.1804). However, in Beijing alone in 2009 HIV prevalence among MSMs was estimated to be 5% (UNAIDS 2010). It is important not to ignore this group of people and their needs in order to be able to control the epidemic in the country.

HIV transmission through heterosexual sex has so far mainly taken place in the context of vulnerable groups such as IDUs, commercial sex workers (CSWs), migrant population and commercial blood donors in China. Commercial sex work is becoming a more important feature of HIV epidemic in China. For some CSWs, selling sex is the only way to earn money and to support themselves and their families. The overlap between drug use and commercial sex work exists in different countries' contexts and China is not an exception. For some IDUs, commercial sex is the only way to access drugs, either through exchange of sex for drugs, or because it is the only way to earn enough money to buy the drugs. According to the 2007 AIDS epidemic update document produced by UNAIDS, "[i]ncreasing numbers of women are injecting drugs and in some places as many as half of those also sell sex. Many male injecting drug users also buy sex, often without using a condom" (UNAIDS 2007, p.21).

Since 1987 prostitution is prosecuted under the law in China. "Commercial sex workers operate out of places of entertainment (e.g., karaoke bars), hotels, hair-dressing salons, or on the street" (Wu *et al*. 2007b, p.683). In 2003 homosexual prostitution also became a crime in China. Clients can also be prosecuted under the same law. It was estimated that in Beijing HIV prevalence among CSWs was 0.6% which is higher than the HIV prevalence rate in the general population (UNAIDS 2010).

Sex workers have been found to not use condoms consistently in China (Van den Hoek *et al*. 2001; Lau *et al*. 2002; Yang *et al*. 2005; Tucker *et al*. 2005; Wang *et al*. 2005; Zhu *et al*. 2005; Gu and Renwick 2008), and coupled with a large number of sexual partners and in some cases with drug use, this increases the vulnerability of this group to HIV infection (Zhu *et al*. 2005; Chu and Levy 2005).

In recent years because of the increased freedom of movement in China new phenomena appeared in the Chinese labour market, the so-called "floating" population or migrants. Young, unmarried and sexually active people are attracted to urban areas as these areas provide better job opportunities (Zhang and Beck 1999; Anderson *et al.* 2003; Yang 2004; Yang 2006; Li *et al*. 2007). Young males are at higher risk of being involved in risky sexual behaviour as they have more disposable income and they can be involved in casual sexual relations, some times with commercial sex workers (UNAIDS 2002). They can acquire HIV or other STIs through unprotected sexual intercourse and then when they go back to their home towns and villages they are likely to pass an infection on to their partners (He and Detels 2005). This facilitates the spread of HIV from high-prevalence to low-prevalence areas (Thompson 2005) and from high risk groups to the general population.

Commercial sex work is closely linked with migration, and the overlap between the two does exist. One of the reasons that women become commercial sex workers is economic: "young and ill-educated women from rural areas took up sex work because they could not find another job" (MAP 2005, p.16). This situation with commercial sex is not unique to China.

It can be concluded that at the moment the most vulnerable groups in China are migrants, commercial sex workers, their clients and drug users.

1.2.3 The Chinese context

In order to better understand conditions in which HIV epidemic is spreading in China, it is important to understand the cultural and political environment and cultural peculiarities of the Chinese context.

Sex, sexual relations and other sex-related matters such as contraception are taboo topics in the Chinese culture (Gao *et al.* 2001 cited in Li *et al.* 2004c; Zhang *et al*. 2004). Sexual matters are not usually discussed in families, especially by people from older generations (Zhang *et al*. 2004; Thompson 2005). This cultural attitude towards sex can explain the absence of sexual education in schools until recently, the reluctance of teachers or parents to discuss sexually related issues with their children, and the absence of educational programmes on TV. This long-standing tradition is difficult to eliminate. The situation is changing slowly. However, not all people, especially not people of the older generation, are comfortable with discussing issues related to sex. This attitude towards sex creates an obstacle in the task of controlling of HIV epidemic as knowledge about safe sex is one of the crucial components of HIV prevention.

It is also important to understand attitude towards HIV/AIDS in the Chinese context in order to understand better the HIV epidemic in China. Kutcher (2003) argues that "AIDS is considered a shameful disease – for the people who have it, for their families, and even for China" (p.281). It obtained this status because it usually is associated with drug use, commercial sex work and homosexuality, and these behaviours are highly stigmatised in the Chinese context. Therefore, some people find it shameful to discuss HIV/AIDS openly. This attitude towards HIV/AIDS fuels stigma and discrimination against HIV positive people and their families. Stigma and discrimination of HIV positive people are big problems in China. They negatively affect people who live with HIV/AIDS and their relatives. Gill and Okie (2007) argue that "[p]eople from "HIV villages", such as in Henan, where many plasma donors became infected, cannot find jobs, and agricultural produce from these locales cannot find markets" (p.1802). For those, who contracted HIV through intravenous drug use or commercial sex work, the stigma and discrimination are even higher. As mentioned earlier, homosexual people are highly stigmatised in the Chinese society. According to Kutcher (2003) "many people in China, including government officials, claim that the country has virtually no gay people" (p.282). All the above mentioned issues create obstacles in the fight against HIV/AIDS in China.

Low level of knowledge about HIV/AIDS also fuels stigma and discrimination against people living with HIV. HIV knowledge is poor even within some medical specialists such as doctors and nurses (Cai *et al*. 2007). The lack of knowledge among the general population also affects HIV positive children who have problems being accepted to nurseries and schools. This situation is not unique to China; other countries in the world experience similar problems. Lack of awareness and knowledge about HIV/AIDS not only affects people living with HIV, but also fuels the HIV epidemic in the country. Lack of knowledge about main routes of HIV transmission and the risks involved, contributes to people engaging in risky sexual behaviour and unsafe practices.

It is a well-documented fact that education has a positive impact on health (Gregson *et al*. 2001; Jejeebhoy and Sathar 2001; Snelling *et al*. 2007). In order to understand the context in which HIV epidemic is expanding in China, it is important to know the level of literacy of people in the country and the educational structure as it might have positive impacts on the prevention campaigns and programmes. According to the Chinese Census conducted in 2000, 95.1% of males and 86.5% of females are literate (aged 15 and over and can read and write), and the total literacy level was estimated to be 90.9%[12]. This level of literacy is high in comparison to many developing countries in the world and might have a positive effect on the effectiveness of campaigns which are designed to increase the level of HIV awareness and knowledge in the country. A high level of literacy helps to communicate messages effectively, and it provides people with opportunities to obtain and comprehend educational information. The educational system in China provides free primary and secondary education (six years of primary education and six years of secondary education which is split between the junior secondary (three years) and senior secondary (three years)). According to Compulsory Education Law of the People's Republic of China (Article 2) which was adopted at the Fourth Session of the Sixth National People's Congress on the 12th of April 1986 and promulgated by Order No. 38 of the President of the People's Republic of China and which became effective on the 1st of July 1986[13], nine years of education is compulsory for all Chinese children (primary and junior secondary education). The quality of education in China is also high when compared with many other parts of the developing world. The high level of access to education across the whole country and the existence of a large number of educational establishments can provide the mechanism of communicating information to people starting from a young age.

1.2.4 The Chinese government and HIV/AIDS

Political commitment plays a crucial role in the fight against HIV in every country of the world[14]. In China, after years of denial and neglect, the problem of AIDS finally received an official recognition and the spread of HIV has gained serious attention from Beijing (Gill *et al.* 2002). The Government's goals and objectives for dealing with the problem of AIDS were formulated in November 1998 in the "China Long/Medium Term Plan for AIDS Control 1998-2010"[15]. In 2002 the Chinese Centre for Disease Prevention and Control (CDC) was established. This centre coordinates HIV related

[12] https://www.cia.gov/library/publications/the-world-factbook/print/ch.html [Accessed 1 December 2008]
[13] http://www.edu.cn/20050114/3126820.shtml [Accessed 28 July 2009]
[14] http://icad-cisd.com/pdf/publications/e_political.pdf [Accessed 18 August 2009]
[15] http://hivaidsclearinghouse.unesco.org/search/resources/NationalPlan1998-2010_HIV_China.pdf [Accessed 1 December 2008]

programmes at the national level (Zhang 2004; Shen and Yu 2005; Gu and Renwick 2008). In 2001 a crucial report "HIV/AIDS: China's Titanic Peril" was produced by the UN Theme Group on HIV/AIDS in China (UNAIDS 2002). This document provided an update of the HIV/AIDS situation in China and concluded that China was on the brink of a serious HIV epidemic and that the government should take the situation seriously, otherwise it could be too late (UNAIDS 2002). During that time the Chinese government denied the existence of the problem in the country. At the beginning of the epidemic "the disease was constructed as a foreign, particularly Western one, resulting from a 'decadent bourgeois lifestyle' marked by promiscuity and sexual obsession" (Zhang 2004, p.1161). Other countries experiences suggested that it is difficult to achieve any results in the fight against HIV epidemic without the strong political commitment of the government.

The significant change in the governmental attitude towards the problem of HIV/AIDS happened in December 2003. A symbolic hand shake between the Premier Wen Jiabao and an AIDS patient took place in Beijing Ditan Hospital on the World AIDS Day on the 1st of December 2003 (Wu *et al*. 2007b). Wen Jiabao, the Chinese Premier, has brought HIV/AIDS to the governmental agenda and the problem of HIV/AIDS has been given the highest priority. This change was driven by the situation with severe acute respiratory syndrome (SARS) (Wu *et al*. 2010). When the SARS epidemic went out of control, the existence of the problem was denied and, as a result, necessary measures were not taken in time. The Chinese government learned the lesson from the SARS epidemic, health officials "were tainted by the SARS experience", and they now "want to do the right things with AIDS" (Park 2003). The Chinese government pledged to control the HIV/AIDS situation in the country by urgent implementation of new mechanisms and effective measures which will help to combat the epidemic. According to Thompson (2003), "the Chinese government has expressed a willingness to work with the international community to create policies and programmes that will prevent HIV/AIDS from becoming a disaster" (p.2). The Chinese government pledged to keep the total number of people living with HIV by 2010 under 1.5 million (Gu and Renwick 2008). In December 2004 "A Joint Assessment of HIV/AIDS Prevention, Treatment and Care in China" was launched by joint efforts by the Chinese government (State Council AIDS Working Committee) and the UN Theme Group on HIV/AIDS in China (State Council 2004). Many positive changes have happened since December 2003. One of the main recent changes in the attitude towards HIV/AIDS is the growth of the budget allocated to tackle HIV/AIDS in the country (Zhang 2004; Gill and Okie 2007). Also a multi-sectoral approach towards tackling the epidemic was adopted and civil society and NGOs became involved in the HIV response (Zhang 2004; Wu 2005; He and Rehnstorm 2005; Wu *et al*. 2010; Li *et al*. 2010).

A number of successful policies and programmes which addressed HIV prevention, treatment and care were introduced. In 2003 so-called "Four Frees and One Care" policy was announced. This policy guaranteed free life-saving antiretroviral treatment for all rural residents and poor urban residents who were not covered by the health insurance. It also guaranteed free voluntary HIV counselling and testing for anybody who needs it, free antiretroviral therapy for HIV positive pregnant women to prevent MTCT, as well as free HIV testing for their new born babies, free schooling for HIV/AIDS orphans, and economic support for the households with people living with HIV/AIDS (Wu *et al*. 2007b). Another important HIV prevention and treatment programme was launched in 2003. It was called "The China Comprehensive AIDS Response" (China CARES). The main purpose of the programme was to expand access to HIV treatment and care in areas most affected by HIV/AIDS (127 high-prevalence counties) (He and Detels 2005; Wu *et al*. 2007b). In 2006 the first legislation directly related to HIV/AIDS was passed: the AIDS Prevention and Control Regulations (Wu *et al*. 2007b). All these changes were important and timely. However, it was impossible to obtain the instant results in all spheres of the HIV response.

China has gone through serious institutional changes in the recent past. They resulted in political decentralisation which brings a high degree of autonomy to local governments. According to Zhang (2004), "political decentralization has tended to dilute power of the central government to implement and enforce AIDS-related policies formulated at the top level" (p.1166). This decentralisation and autonomy of local governments "resulted in a mixed response and inconsistent enforcement of HIV/AIDS policy" (Wu *et al.* 2007b, p.687). Sometimes local policies adopt different approach to tackling HIV epidemic. For instance, some local governments emphasize "elimination of drug use and prostitution" and use of "compulsory 'education camps' and rehabilitation centres for CSWs and IDUs" instead of promotion of "the central government new 'harm-reduction' projects comprising of free methadone, needle exchange and condom distribution" (Zhang 2004, p.1164). In order to control the epidemic, local governments should be consistent with the policies and programmes which are implemented at the national level.

Despite the problems associated with the decentralisation in China mentioned above, a number of successful initiatives was launched at national and local levels in China. There are many governmental and non-governmental, domestic and international organisations and groups seeking to tackle the potential crisis in China (Renwick 2002; He and Rehnstrom 2005; Wu 2005; Shen and Yu 2005; Gu and Renwick 2008). It can be argued that significant progress has been made in China since 2003. The Chinese Deputy Minister of Health, Dr Wang Londe, and Professor Zhang Beichuan from Qingdao University received special awards from the UNAIDS which suggests that the UNAIDS "acknowledged the good progress that has been made in China's response to AIDS"[16]. However, there are still many issues which should be addressed in China in order to be able to fully control the HIV epidemic. For example, in order to be able to maintain success in educational campaigns, it should not be a one-off measure. Measures should be implemented on a regular basis because new generations are growing up and these new generations should be armed with an adequate level of HIV awareness and knowledge to be able to protect themselves.

1.2.5 Interventions by UNFPA: An overview

The priority to advance HIV awareness and knowledge among all groups of people was already stressed in the 1994 Cairo International Conference on Population and Development (ICPD 1994). During the 1994 International Conference on Population and Development (ICPD) in Cairo a programme of action was outlined, and chapter VIII of the programme called Health, Morbidity and Mortality had a section on HIV/AIDS (Section D). In this section it was stated that governments should take various actions to control the HIV epidemic: sex education and necessary information should be made available to all segments of society, especially to young people, HIV awareness and knowledge should be increased among all groups of people, and behavioural changes should be encouraged (ICPD 1994).

According to Fang and Kaufman (2008), in response to this programme of action and specifically this section, family planning services in China have expanded their agenda

[16] http://www.unaids.org/en/MediaCentre/PressMaterials/FeatureStory/20070717_China_UNAIDS_awards_leadership_excellence.asp [Accessed 29 November 2007]

and have started providing education on HIV/AIDS for rural women and young people as well as the floating population or migrants.

UNFPA conducted country programmes in China in order to assess the state of reproductive health and family planning in the country as well as to help design necessary interventions in order to be able to achieve goals formulated at the ICPD. Two surveys were conducted within the Fifth Country Programmes (CP5) in China: a baseline survey was conducted in September 2003 and an endline survey was conducted in November 2005 with the aim of evaluating the impact of CP5 intervention programme.

According to the document[17] which formulated the main goals of the CP5 in China (UNFPA 2003), the new strategies for the CP5 were to further improve family planning and reproductive health services and the quality of reproductive health care for men, women and adolescents, with a special emphasis on the underserved groups including migrants, young people, unmarried people and people living with HIV/AIDS. HIV/AIDS prevention was one of the core themes for the CP5. Information regarding HIV prevention in the form of information on modes of HIV transmission was recommended to be provided during client-provider interactions (UNFPA 2003). It was also recommended to update Information, Education and Communication (IEC) materials for more effective communication of the messages about HIV prevention. One of the CP5 project targets was to improve the level of HIV transmission knowledge among general population with the aim that "[a]t least 50% of women and 40% of men of reproductive age (15-49) in the general population are aware of 3 methods of prevention of HIV" (UNFPA 2003, p. 29).

A key recommendation from the CP5 baseline report was that intervention measures should place an additional emphasis on issues related to HIV/AIDS (Li *et al.* 2004b). The report concluded that there was an urgent need to increase the level of HIV awareness and knowledge among different strata of Chinese population but particularly among young men and women. About 30% of unmarried and currently married women from the Western region of the country (which is the least developed part of China) had never heard of HIV/AIDS at the time of the baseline survey and, therefore, they constituted another important group which required improvement of the level of HIV

[17] This document is an agreement between the government of the People's Republic of China and United Nations Population Fund.

awareness and knowledge (Li *et al*. 2004b). Another finding suggested that among never-married men and women from the Central and Western regions the knowledge of at least three possible ways of HIV transmission was low (Li *et al*. 2004b). According to the recommendations from the Fourth Country Programme (CP4), more efforts should be made to improve reproductive and sexual health knowledge, including HIV knowledge, of unmarried young people (CPDRC and S3RI 2004). According to the recommendations which were based on the findings from the 2003 baseline survey, intervention efforts needed to be strengthened in the Western and Central regions of China (Li *et al*. 2004b). The level of HIV knowledge should be increased in all lagging behind groups (Li *et al*. 2004b). These groups of people were then specifically targeted in the areas covered by the CP5 interventions with the help of IEC materials.

A major focus of the CP5 interventions was on disseminating standard education materials on HIV/AIDS prevention including wall posters, leaflets, pamphlets and video materials explicitly focusing on prevention of HIV transmission and reducing stigma and discrimination and providing information about care and treatment of people living with HIV/AIDS. In addition, health providers including doctors, nurses and community health workers were provided with training sessions on HIV prevention and treatment compliance strategies including partner referral and follow-up and counselling as well as on emphasising the relevance of reducing stigma and discrimination against those living with HIV.

1.2.6 Challenges ahead

Progress in the fight against HIV/AIDS in China has been observed in recent years. However, in order to be able to fully control the epidemic in China a number of existing challenges should be addressed such as human and financial resources, training, medical infrastructure, rural health crisis, political decentralisation, stigma and discrimination of HIV positive people and their families, as well as lack of HIV awareness and knowledge among the general population. Addressing these issues can help control the spread of HIV in China and can prevent the country from the potential negative consequences which HIV/AIDS catastrophe might bring. HIV/AIDS can affect negatively the economic and political prospects of the country and can bring tragedies at individual level, if the problem of HIV is not continued to be taken seriously.

1.3 HIV Awareness and Knowledge

1.3.1 Importance of studying HIV awareness and knowledge

In the situation where the number of people living with HIV is growing and in the absence of cure or vaccine against HIV, prevention plays crucial role in the fight against the disease in every cultural context. As mentioned earlier, HIV prevention consists of different dimensions. One of the important tasks of HIV prevention is to reduce risky behaviours as two of the main routes of HIV transmission, intravenous drug use and unprotected sex, homosexual and heterosexual, are related to risky behaviours. According to Snelling *et al*. (2007), "[a] fundamental public health strategy to reduce the risk of HIV/AIDS is to increase levels of awareness and knowledge about the disease" (p. 422). A number of educational campaigns were introduced in different parts of the world and focused on the improvement of levels of HIV awareness and knowledge. According to UNAIDS (2010), "[b]ehaviour change and increased comprehensive correct knowledge reduce HIV incidence and prevalence in most countries with high HIV prevalence" (p. 68). However, even in high prevalence countries such as Mozambique (where HIV prevalence among adults was estimated to be 11.5% [10.1-12.1%] (UNAIDS 2010)) many people's knowledge about HIV is still very low according to a recent survey (2009) conducted by the Ministry of Health and the National Statistic Institute[18]. Therefore, it is important to assess needs in HIV knowledge in different contexts and to address these needs effectively.

Level of HIV knowledge is assessed in different countries of the world within the context of the Millennium Development Goals targets. Young people's (15-24 years of age males and females) knowledge of HIV indicator (13) is one of the UNGASS indicators for Millennium Development Goal target 6.A[19] (to halt and begin to reverse the spread of HIV/AIDS by 2015) (UNAIDS 2010). This indicator[20] shows the percentage of young women and men aged 15-24 years who both correctly identified ways of preventing the sexual transmission of HIV and who reject major misconceptions about HIV transmission (UNAIDS 2008a; UNAIDS 2010). For this indicator data were collected in 2004 in 38 countries, in 2006 only in 16 countries, in

[18] http://allafrica.com/stories/201004080798.html [Accessed 15 April 2010]
[19] http://www.un.org/millenniumgoals/aids.shtml [Accessed 26 November 2010]
[20] "Numerator: number of respondents aged 15-24 years who gave the correct answer to all five of the following questions: 1. Can the risk of HIV transmission be reduced by having sex with only one uninfected partner who has no other partners? 2. Can a person reduce the risk of getting HIV by using a condom every time they have sex? 3. Can a healthy-looking person have HIV? 4. Can a person get HIV from mosquito bites? 5. Can a person get HIV by sharing food with someone who is infected? Denominator: number of all respondents aged 15-24" (UNAIDS 2010, p. 222).

2008 in 110 countries and in 2010 in 119 countries[21] (UNAIDS 2010). In 2001 the following targets for comprehensive HIV knowledge were set: 90% of young people aged 15-24 having adequate HIV knowledge by 2005 and 95% by 2010 (Coates *et al*. 2008). Coates *et al*. (2008) reports that the 2005 target for HIV knowledge has not been achieved and they also argued that "it is highly doubtful that the 2010 targets can be reached" (p. 681). According to the UNAIDS 2008 and 2010 Report on the Global AIDS Epidemic, there are still large groups of people in the world such as young people aged 15-24 with an inadequate level of knowledge about HIV and its routes of transmission (UNAIDS 2008a; UNAIDS 2010). Since 2008 the level of comprehensive and correct knowledge about HIV among young men and women has increased slightly but in 2010 the percentages of young people with comprehensive HIV knowledge was only around one third of the UNGASS target of 95% and despite gradual improvement of knowledge about HIV many countries will "still fall short of the global targets for comprehensive knowledge set in 2001" (UNAIDS 2010, p.68). As targets for comprehensive knowledge were not achieved in 2010, there is an urgent need to improve HIV awareness and knowledge worldwide. An identification of groups with inadequate HIV knowledge, which would help designing effective targeted interventions, is necessary to make a progress in the area of HIV awareness and knowledge.

As mentioned earlier, HIV awareness and knowledge are integral and important components of a successful multi-dimensional HIV prevention strategy. Therefore, it is important to assess levels of not only young people but also general population's HIV awareness and knowledge in different cultural and epidemiological contexts.

Nutbeam and Blakey (1990) applied general health promotion framework which was formulated in the Ottawa Charter (WHO 1986) to the area of HIV prevention. This framework has five main themes: promoting health through public policy, creating a supportive environment, developing personal skills, strengthening community action and reorienting health services (WHO 1986; Nutbeam and Blakey 1990). HIV awareness and knowledge belong to the third theme – development of personal skills. Nutbeam and Blakey (1990) argues that "[r]educing discrimination, taking up opportunities for testing, and adopting safer sexual practices all depend on levels of

[21] Limited data on comprehensive knowledge of HIV is available from Western and Central Europe, the Middle East and North Africa but complete data in available from Sub-Saharan Africa because of the Demographic and Health Surveys (UNAIDS 2010).

knowledge and understanding" (p. 237). However, they also acknowledge that "knowledge is generally unrelated to behaviour change, without modification of attitudes and beliefs" (Nutbeam and Blakey 1990, p. 237). However, despite the fact that knowledge does not have a direct effect on behaviour change, HIV knowledge is still an important component of the third theme.

A number of different theoretical frameworks have been used to study HIV risk behaviours, such as the health belief model (Becker 1974; Janz and Becker 1984), the theory of reasoned action (Ajzen and Fishbein 1980), the theory of planned behaviour (Ajzen and Fishbein 1980; Ajzen 1991), and social cognitive theory (Bandura 1986; Bandura 1990). The information-motivation-behavioural skills model (IMB) (Fisher and Fisher 1992) integrates key components from other models in a single model of HIV risk behaviour. Another model that was developed to study specifically AIDS risk reduction is an AIDS Risk Reduction Model (ARRM) (Catania *et al.* 1990) which is also a three-stage model and which consists of recognition and labelling, commitment, and enactment stages. In the IMB and the ARRM models knowledge of HIV transmission and prevention as well as consequences of HIV were "identified as determinants of HIV-related risk-taking and protective behaviours" (Carey *et al.* 1997, p.73). In all the theoretical frameworks mentioned above HIV awareness and knowledge have a role to play. In the IMB model the first component of the model is information, and it is one of the three major components and contains HIV awareness and knowledge. According to Fisher and Fisher (1992), Ross and Kelaher (1993), Ingham (1995), Ambati *et al.* (1997), and UNAIDS (2010), information, i.e. awareness and knowledge, is a prerequisite of risk reduction behaviour as in order to change behaviours it is necessary to know that the behaviour is risky. However, it is very important to mention that this information is an essential and necessary component for the risk-reduction behaviour but not a sufficient one (Ingham 1995; Helweg-Larsen and Collins 1997; UNAIDS 2010). Ambati *et al.* (1997) argues that "[w]hile knowledge is a prerequisite, true success in this epidemic is, of course, to be measured in behavioural change" (p. 327). According to Ross and Kelaher (1993) and Helweg-Larsen and Collins (1997), it was demonstrated in a number of studies that the relationship between knowledge and behavioural change is weak and that "the potency of knowledge to change behaviour is limited" (Ross and Kelaher 1993, p.5). There are a number of mediating variables between knowledge and behaviour, and attitudes and beliefs are some of them (Nutbeam and Blakey 1990; Ross and Kelaher 1993). However, as HIV awareness and knowledge are important prerequisites and they have a role to play, it is important to study HIV awareness and knowledge and their dynamics, transitions and changes in different contexts across the world.

Data on HIV awareness and knowledge are collected frequently and consistently as a part of larger surveys which also cover other areas such as reproductive health and family planning in different cultural and epidemiological contexts. Many researches use HIV awareness and knowledge variables either as response or as explanatory variables in their models. Studies which use HIV awareness or HIV knowledge variable as a response are interested in predictors of HIV awareness and knowledge (Manchester 2002; Chen *et al*. 2003a; Dong *et al*. 2003; Li *et al*. 2004c; Zhang *et al*. 2006; Tan *et al*. 2006; Wu *et al*. 2007a; Stephenson 2009). Other studies control for HIV awareness and knowledge when modelling, for example, AIDS-related stress (Pleck 1998), ever using a condom to protect against STDs including HIV/AIDS and ever having sex with only one partner to protect against AIDS (Snelling *et al*. 2007), recent fertility patterns in Kenya (Magadi and Agwanda 2010), HIV-related stigma and HIV testing (Do and Guend 2009), multiple sexual partnership (Buszin *et al*. 2009) and others.

HIV knowledge has various components and cannot be collected as a single variable or measured directly and, therefore, it is important to analyse and compare different approaches to measuring HIV knowledge and to study advantages and disadvantages of different approaches.

1.3.2 Relationship between HIV awareness and knowledge

The majority of surveys which collect data on HIV awareness and knowledge assume that HIV knowledge follows HIV awareness and, as a result of this, only people who report being aware of HIV are asked further questions on HIV knowledge. Figure 1.1 shows composition of any sample of people by their HIV awareness and knowledge. All people in the sample can first be partitioned between those who are aware of HIV (group 2 including groups 3 and 4) and those who are not aware of HIV (group 1 excluding groups 2, 3 and 4). Then people who are aware of HIV (group 2 including groups 3 and 4) can be further split between those who are aware of HIV but do not have adequate knowledge (group 2 excluding groups 3 and 4) and those who are aware of HIV and have adequate knowledge of correct routes of HIV transmission or ways to prevent HIV (group 3) and incorrect routes of HIV transmission or misconceptions about HIV (group 4). The intersection between the two groups (group 3 and group 4) represents a group of people who are aware of HIV and who have adequate knowledge of both correct and incorrect routes of HIV transmission or ways to prevent HIV and misconceptions about HIV.

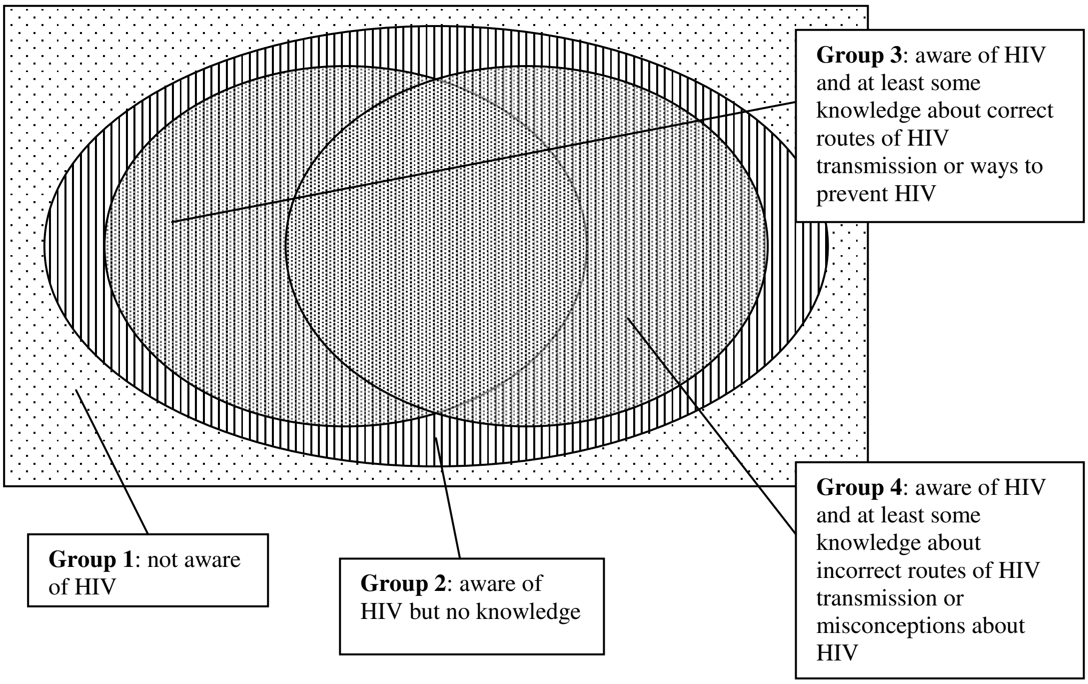


Figure 1.1: Venn diagram: A composition of a sample of women by HIV awareness and knowledge.

1.3.3 Literature review: HIV awareness and knowledge in China

A number of studies investigated HIV awareness and knowledge in China. The results of the studies were published in the Western and Chinese journals. The vast number of articles on HIV knowledge in different groups appeared in Chinese journals after 2003, when HIV/AIDS received serious attention from the Chinese government. However, the majority of studies examined the HIV/AIDS knowledge of specific groups of the population such as university students (Li *et al*. 2004c; Tan *et al*. 2006; Wu 2006; Tan *et al*. 2007; Zuo *et al*. 2007; Tan 2008), health care workers (Wu *et al*. 1999 cited in Li *et al*. 2004c), criminal suspects (Ping *et al*. 2006 cited in Tan *et al*. 2007), STD clinic patients (Wang *et al*. 2001), migrants or so-called floating population (Dong *et al*. 2003; Guo and Xu 2006; Zhou *et al*. 2007), specifically road construction workers (Zhu and Tan 2006), miners (Yang *et al*. 2007), and service industry workers (Zhu and Tan 2006). Other studies investigated the HIV knowledge of pregnant women (Yuan *et al*. 2003 cited in Wu *et al*. 2007a), female sex workers (Lau *et al*. 2002; Huang *et al*. 2004; Hesketh *et al*. 2005; Qi *et al*. 2007), market vendors (Cao *et al*. 2006), and entertainment industry workers (Ding *et al*. 2006; Zhang *et al*. 2006; Pei and Wang 2007). Some of these studies did not use appropriate sample designs and, therefore, the generalisation of the results even within these groups is questionable. In 2008 a survey called "AIDS-Related Knowledge, Attitudes, Behaviour, and Practices: A Survey of 6 Chinese Cities" was conducted in China. This survey collected data on HIV knowledge (among other variables) for different groups of people (males and females, migrant workers, youth, blue-collar workers, white-collar workers) in the Chinese context (CHAMP 2008). The main aim of the survey was to study HIV/AIDS knowledge, attitude and behaviour among different groups of people in the six cities in order to formulate strategies for AIDS media campaigns (CHAMP 2008).

A small number of studies examined the level of HIV knowledge in China in the general population at one point in time (Manchester 2002; Holtzman *et al*. 2003; Zhu and Tan 2006). To date there has not been any research that systematically examined the changes in HIV awareness and knowledge in the general population in China over time.

The studies mentioned above were conducted in different geographic locations in China, both urban and rural. The majority of studies were quantitative but some of them were qualitative (Huang *et al*. 2004; Hesketh *et al*. 2005).

The studies, which will be discussed below, examined the factors that are associated with HIV awareness and knowledge through univariate and bivariate analysis and through further statistical modelling. The main factors which were found to be associated with HIV/AIDS knowledge in the Chinese context were gender, age, education, income/expenditure, type of university degree, place of residence, type of job or occupation, media exposure, and perceived risk in the neighbourhood.

A number of studies investigated differences in HIV knowledge by gender (Manchester 2002; Dong *et al*. 2003; Li *et al*. 2004c; Tan *et al*. 2006; Yang and Xia 2006; Zhang *et al*. 2006; Tan *et al*. 2007; Tan 2008; CHAMP 2008). The majority of studies did not find difference in the levels of HIV awareness and knowledge by gender. However, three studies found a significant difference in the level of HIV awareness and knowledge between males and females with males being more knowledgeable about HIV than females (Dong *et al*. 2003; Li *et al*. 2004c; Zhang *et al*. 2006). The first and the third studies reported this finding on the basis of only a descriptive analysis, whereas in the second study the results still hold when other characteristics of the respondents were controlled for. The findings of Manchester (2002) and Tan *et al*. (2007) suggest that gender differences do not exist in relation to HIV knowledge in general but do exist with regard to the specific knowledge about the role of condom use in the prevention of HIV/AIDS.

Age is reported to be a significant predictor of HIV knowledge (Lau *et al*. 2002; Chen *et al*. 2003a; Dong *et al*. 2003; Guo and Xu 2006; Wu *et al*. 2007a). People from the age group between 25 and 30 obtained the highest scores on HIV knowledge when compared with younger and older people (Dong *et al*. 2003; Wu *et al*. 2007a). The age of respondents was also significantly associated with the condom use (Wang *et al*. 2001) and with the awareness about the role of condom in the HIV/AIDS prevention (Manchester 2002).

A number of studies examined the educational or schooling level and HIV knowledge score achieved (Lau *et al*. 2002; Chen *et al*. 2003a; Dong *et al*. 2003; Hesketh *et al*. 2005; Guo and Xu 2006; Zhang *et al*. 2006; Wu *et al*. 2007a). In the study that examined the female sex workers in Yunnan, an association between the HIV knowledge score and educational level was not established (Hesketh *et al*. 2005). This might be due to the fact that the majority of respondents had similar and quite low level of education. In the studies that found the significant association between the level of HIV knowledge and education or schooling, positive association between the two was found: the level of HIV knowledge would increase together with the educational level (Manchester 2002; Chen *et al*. 2003a; Wu *et al*. 2007a; CHAMP 2008).

Tan *et al*. (2006), Tan *et al*. (2007) and Tan (2008) argue that HIV knowledge varies by the type of university degree students are studying: students who study medicine have better HIV knowledge than non-medical students.

Two studies found that the level of knowledge depends on the job or occupation of the respondents but they did not provide further details of their findings (Chen *et al*. 2003a; Dong *et al*. 2003). CHAMP's (2008) descriptive findings suggest that the level of HIV knowledge was higher among white collar workers than blue collar workers.

Some studies investigated the association between the HIV knowledge and income or expenditure (Manchester 2002; Tan *et al*. 2006; Tan *et al*. 2007; Tan 2008). Manchester (2002) argues that higher average monthly income is associated with the higher level of HIV awareness and knowledge. According to Tan *et al*. (2006), Tan *et al*. (2007), and Tan (2008), higher monthly expenditure is also associated with higher level of HIV knowledge and safer behaviours.

A number of studies examined differences in HIV knowledge by residence. A difference was found between urban and rural areas in China with people from urban areas having a better HIV knowledge than people from rural areas (Chen *et al*. 2003a; Chen *et al*. 2003b; Chen *et al*. 2003c).

Comparisons were made between different cities and the findings suggest that the level of knowledge is not homogenous across urban areas either: in some cities, such as Beijing, Shanghai, Guangzhou, Xiamen, people are more knowledgeable about HIV/AIDS than in others, such as Wuhan, Changdu, Shenyang, Jinan (Manchester 2002). CHAMP's (2008) findings suggest that the level of knowledge about HIV was highest in Shanghai and Kunming, followed by Wuhan and Beijing and the lowest level of knowledge about HIV was observed in Zhengzhou and Shenzhen. Wu *et al*. (2007a) found difference in the levels of knowledge in different, but comparable by type of residence, counties of several provinces: the level of knowledge was higher in Henan province and lower in Guangdong, Guangxi and Gansu.

Zhu and Tan (2006) compared HIV knowledge of the general population and migrant population in Central China. They found that the general population had better HIV knowledge than migrants (road construction workers and service industry workers). CHAMP's (2008) findings also suggested that the level of knowledge among migrants was lower than the level of knowledge of non-migrants included in the analysis.

Chen *et al*. (2003a) studied an association between media exposure (TV, radio, newspapers and magazines) and the level of HIV knowledge and found a positive association between the two. They also studied the association between the HIV knowledge and perceived risk of HIV/AIDS in the neighbourhood. The findings suggest that the level of knowledge is higher in areas where the perceived risk of acquiring HIV infection is higher among the population (Chen *et al*. 2003a).

A few studies investigated changes in HIV knowledge in the context of the assessment of interventions (Xiaoming *et al*. 2000; Wang and Keats 2005; Cao *et al*. 2006; Pei and Wang 2007). These four studies demonstrated the effect of educational interventions on HIV knowledge. In all four studies the level of HIV knowledge was higher after the interventions were completed.

Some of the studies also examined the sources of information about HIV/AIDS (Lau *et al*. 2002; Chen *et al*. 2003a; Chen *et al*. 2003c; Dong *et al*. 2003; Li *et al*. 2004c; Cheng *et al*. 2005; Wu 2006; Zhu and Tan 2006; Tan *et al*. 2007; Zhou *et al*. 2007; Tan 2008). The main and most important sources of information were found to be TV, newspapers and radio (Chen *et al*. 2003a; Chen *et al*. 2003c; Cheng *et al*. 2005; Tan *et al*. 2007; Zhou *et al*. 2007; Tan 2008).

Generally the level of HIV knowledge was reported to be low or moderate among STD patients, floating population, among some female sex workers and among rural women, and higher among university students (Wang *et al*. 2001; Dong *et al*. 2003; Li *et al*. 2004c; Cheng *et al*. 2005; Wu 2006; Zhu and Tan 2006; Qi *et al*. 2007; Tan *et al*. 2007; Yang *et al*. 2007; Zhou *et al*. 2007). According to CHAMP (2008), "the average level of knowledge of AIDS and HIV transmission is relatively low and serious misperceptions, such as transmission occurring from mosquito bites are fairly common" (p.23). It was found that people usually had better knowledge of the correct routes of HIV transmission but not of the incorrect routes of HIV transmission (Lau *et al*. 2002; Cheng *et al*. 2005; Guo and Xu 2006; Zhu and Tan 2006; Wu *et al*. 2007a; Yang *et al*. 2007; Zuo *et al*. 2007). These results are expected as the HIV epidemic in China is at its early stage and at the beginning of HIV epidemics interventions and campaigns are focused on the messages about correct routes of HIV transmission and once the epidemic progresses further and more people become affected by HIV, stigma becomes an important issue. At this point interventions and campaigns are more focused on the main misconceptions about HIV or in other words, on the knowledge about incorrect routes of HIV transmission.

No studies examined changes in HIV awareness and knowledge in the general population in China over time or compared levels of knowledge in China with those in other countries and, therefore, the current research will address these gaps.

1.4 Conceptual Framework

The development of the conceptual framework for the analysis is directed by the literature review. This conceptual and analytical framework identifies the main response variables for the analyses presented in this thesis: HIV awareness and knowledge. HIV knowledge cannot be measured directly and can only be measured through different ways of combining separate components of HIV knowledge such as knowledge of separate routes of HIV transmission, knowledge of ways of prevention of HIV infection and knowledge about misconceptions about HIV transmission. HIV knowledge is a composite construct and it can be measured in different ways. In this research HIV knowledge is measured in six different ways (they will be discussed in details in section 1.6). Figure 1.2 presents groups of variables which will be considered in the analysis. It also shows possible associations between different groups of explanatory variables and the response variables based on the existing literature. The graph shows response variables used for the analysis (HIV awareness and knowledge) and in case of HIV knowledge it also shows the components of the HIV knowledge variable, which construct the HIV knowledge response variable, and their relations with the response variable. The components of the HIV knowledge used for the analyses are discussed in detail in Chapter 2 and presented in Figures 2.1 and 2.2. The explanatory variables included in the analysis presented in the thesis belong to the three main groups. The first group consists of demographic and socio-economic characteristics of participants (individual level variables) such as age, education, ethnicity, region, marital status, income and others. The second group of variables includes the macro-level variables such as year of the surveys or country, HIV prevalence rate, type of HIV epidemic, female literacy, and HIV awareness rate. The third group represents factors which are important for improvement of HIV awareness and knowledge such as political commitment and effective interventions.

Figure 1.2 shows that demographic and socio-economic characteristics of women as well as macro-level characteristics of a country are predictors of HIV awareness and knowledge. All the variables included in the analysis will be discussed in detail in Chapter 2.

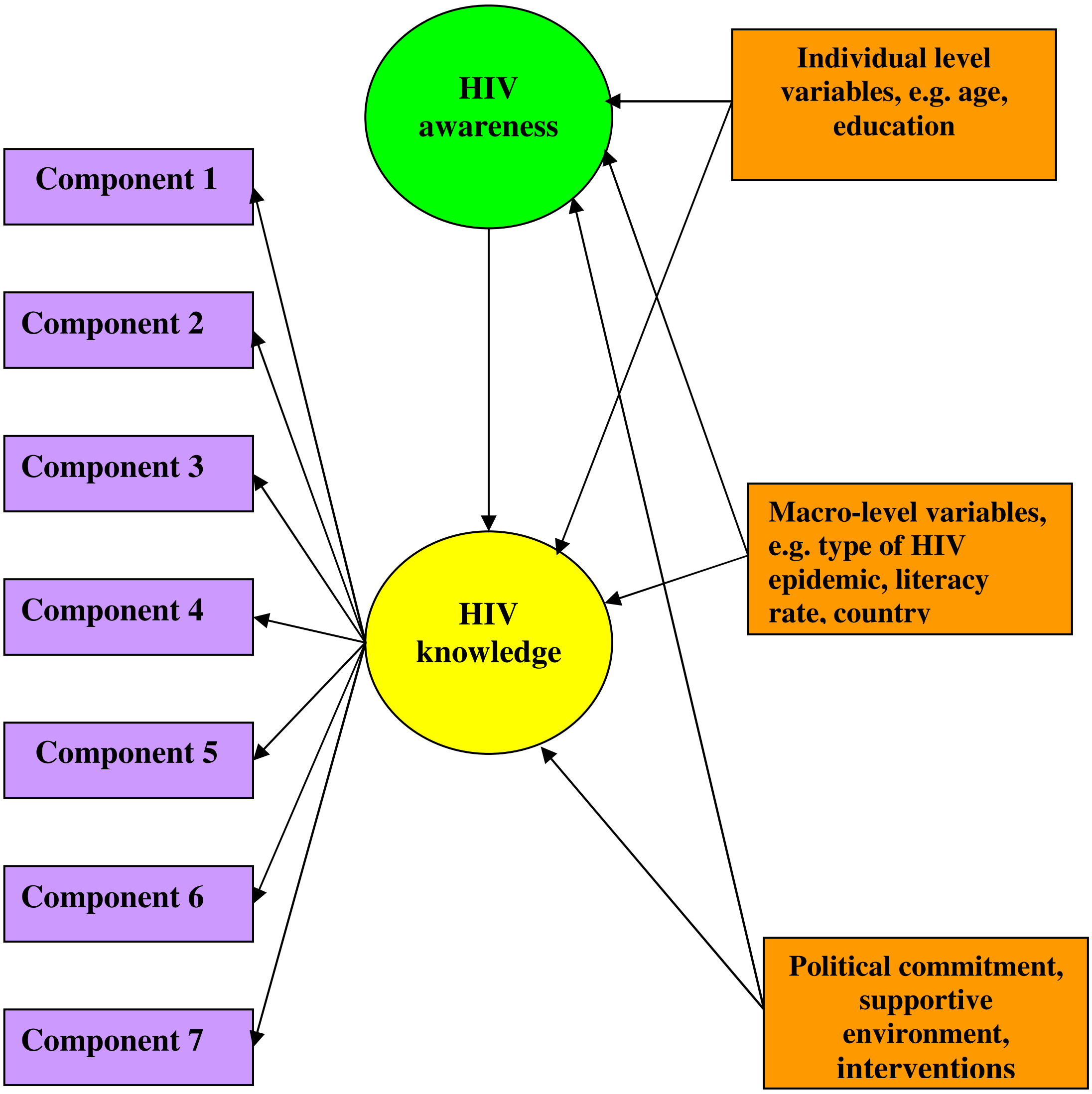


Figure 1.2: Conceptual framework.

1.5 Study Relevance and Rationale

As mentioned earlier, in 2002 China was labelled "the next wave country" by the National Intelligence Council in the US as China together with four other countries was expected to have a serious HIV problem with a large population being at risk of HIV infection by 2010 (ICA 2002). It was projected that by the year 2010 ten to twenty million people could become infected with HIV in China if no effective measures are taken (Eberstadt 2002; ICA 2002; Zhang 2004; Chou 2004; Gu and Renwick 2008). The HIV prevalence would reach 1.3-2.5% (Eberstadt 2002). Eberstadt (2002) and ICA (2002) produced different HIV prevalence scenarios. According to ICA (2002) with the low estimates of future, 10 million people will be affected by HIV by 2010, with the high estimates – 15 million people. Eberstadt (2002) provided three scenarios: "mild", "intermediate" and "severe" which would correspond to low, medium and high levels of HIV infections with HIV prevalence of 1.5%, 3.5% and 5.0% respectively. However, in contrast according to UNGASS and UNAIDS (2010), it was estimated that in 2009 740,000 [540,000-1,000,000] people were living with HIV in China, and the prevalence among adults of reproductive age (15-49) was estimated to be 0.05% [0.04-0.07%]. This large difference between projections and the real situation could suggest that the projections were based on unrealistic assumptions and that effective measures were taken in the country or even that China is a success story which managed to control the HIV epidemic within a short period of time. If China is indeed a success story, it is important then to find out how China became successful with its HIV prevention strategies and how it managed to control the spread of HIV in the country as these findings could help formulate effective interventions and policies especially for countries where HIV epidemics are still not generalised. It could be argued that China is an interesting case for studying changes in HIV awareness and knowledge in a general population due to its large and diverse population. It is important to mention that HIV prevention has many dimensions but this research focuses specifically on HIV awareness and knowledge.

As mentioned earlier, HIV awareness and knowledge levels are increasing over time in different parts of the world. The change in HIV awareness and knowledge can be attributed to a number of different reasons such as increased political commitment, effective interventions, educational campaigns, but also to the level of literacy and access to education in specific contexts as well as to other factors. It is important to assess the change in HIV awareness as well as to develop different measures of HIV knowledge and then assess the evolution of HIV knowledge in the Chinese context and to compare it to other cultural and epidemiological contexts. The literature review presented above suggests that no studies systematically examined these issues and, therefore, the current research will address these gaps. This analysis will help to identify the groups which are HIV awareness and knowledge deprived and will help to compare the levels of HIV knowledge in countries with similar and different characteristics. The results of these analyses might have important policy implications and might help design more effective educational campaigns not only in China but also in other countries in the world.

China's early stage of the HIV epidemic, its large and diverse population as well as its unique demographic features make China an interesting case for studying changes in HIV awareness and knowledge in a general population. This thesis systematically examines the evolution of HIV awareness and knowledge among women in China between 1997 and 2005.

As mentioned earlier, many studies in the area of HIV/AIDS use HIV knowledge variables in their analyses. HIV knowledge has different components and cannot be measured directly. Different measures of HIV knowledge can be obtained and used, and it is important to analyse and compare different approaches to measuring HIV knowledge and to study the relative advantages and disadvantages of different approaches. No studies examined and/or compared different approaches to measuring HIV knowledge and, therefore, this thesis will also contribute to addressing this research gap in the literature.

1.6 Structure of Thesis and Main Research Questions

The thesis has two main aims: methodological and substantive. The methodological aim is to develop and compare six different measures of HIV knowledge as well as to compare two different approaches to measuring HIV knowledge – a simple score approach and a latent variable approach. The substantive aim is to assess HIV awareness and knowledge in China across time and to compare HIV knowledge in China to other countries in order to be able to answer the question, has China succeeded in improving women's HIV awareness and knowledge?

Therefore, the main research questions of the thesis are:

1. Has China succeeded in improving women's HIV awareness and knowledge?

2. How can we better measure HIV knowledge in diverse cultural and epidemiological contexts?

The first research question is the main substantive research question of the thesis, whereas the second research question is the main methodological research question of the thesis.

In order to achieve these two main aims and to answer the two main research questions this thesis is developed in three distinct but related papers. A variety of data and methods are employed to fully explore the topic. The first paper of the thesis is HIV Awareness in China: A Decomposition Analysis (1997-2005). This paper is presented in Chapter 3 of the thesis. The following data sources[22] are used for the analysis: the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001 and the UNFPA Reproductive Health and Family Planning Surveys 2003 and 2005. The response variable in this analysis is HIV awareness[23]. As mentioned in section 1.3.2, every sample of women used for the analysis is split between women who reported being aware of HIV (group 2 including group 3 and group 4) and those who reported being unaware of HIV (group 1 excluding groups 2, 3 and 4) (Figure 1.1). A regression decomposition analysis technique is used in this paper in order to be able to disentangle the two main components where a change in HIV awareness comes from: a change in a population structure and a change in the relationships or effect sizes (coefficients) due to external factors such as environment, interventions and programmes.

The second paper focuses on a so-called simple score approach to measuring HIV knowledge. This paper is presented in Chapter 4 of the thesis. This paper analyses the evolution of HIV knowledge in China across time and compares HIV knowledge in China with HIV knowledge in other countries in the world (India, Kenya, Malawi and the Ukraine) using four simple score measures of HIV knowledge. The following data sources are used for the analysis: the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001, the UNFPA Reproductive Health and Family Planning Survey 2005, India DHS 2006, Kenya DHS 2003, Malawi DHS 2004 and the Ukraine DHS 2007. Four response variables are used for analysis: Score One for HIV knowledge of correct routes of HIV transmission in the Chinese context or ways to prevent HIV in the five country context, Score Two for HIV knowledge of incorrect routes of HIV transmission in China or knowledge about misconceptions about HIV in the five countries, Score Three for combined HIV knowledge (combination of Score One and Score Two) and Score Four is a variable which groups different options of Score Three. Advantages and

[22] Details about data sources will be provided in section 2.1.
[23] All response variables used in three papers will be discussed in section 2.3.

limitations of different simple score measures of HIV knowledge are discussed and different measures are compared. As mentioned earlier, most surveys which collect data on HIV awareness and knowledge assume that HIV knowledge follows HIV awareness and as a result of this, only those people who reported being aware of HIV are asked further questions on HIV knowledge. Group 2 including groups 3 and 4 will be used for the analysis in this paper (Figure 1.1). Partial proportional odds models are used for the analysis in this paper.

The third paper focuses on latent variable approach to measuring HIV knowledge. This paper is presented in Chapter 5 of the thesis. This paper also analyses the evolution of HIV knowledge in China across time and compares HIV knowledge in China with HIV knowledge in other countries (India, Kenya, Malawi and the Ukraine) in the world but the difference between Papers Two and Three is that this paper uses two latent variable measures of HIV knowledge as response variables for the analysis. This paper uses the same data sources as Paper Two. Similar to Paper Two, group 2 including groups 3 and 4 will be used for the analysis in this paper (Figure 1.1). Two latent variables are developed and used for the analysis: a continuous latent measure of HIV knowledge and a categorical latent measure of HIV knowledge. Latent trait analysis (LTA) is used to develop a continuous latent measure of HIV knowledge. Latent class analysis (LCA) is used to develop a categorical latent measure of HIV knowledge. The different latent variable measures of HIV knowledge are compared and their advantages and limitations are discussed. The latent variable measures of HIV knowledge are also compared to the so-called simple score measures used in Paper Two. Multinomial logistic regression models are used for the analysis in this paper.

This thesis is presented in six main parts. The first chapter of the thesis presents the introduction to the thesis which contains necessary background material, the rationale for the study, and the common for all three papers literature review. The main aims and research questions of the thesis, the structure of the thesis as well as the conceptual framework are also presented in Chapter 1. The second chapter presents data used in the three papers. Chapters 3 to 5 present the three individual papers. Each paper is presented in five or six main sections: introduction to the specific topic together with specific to each paper literature review, methodology used, results obtained and discussion of limitations of the specific analysis together with the conclusions drawn. The sixth chapter presents results of the study as a whole, the main contributions are discussed, the main policy recommendations are proposed and suggestions for possible further investigations are offered. All tables and figures presented in the thesis are produced by the author unless stated otherwise.

Chapter 2: Data

The main purpose of this chapter is to introduce the data which are used for the analysis in three papers of the thesis in order to investigate HIV awareness and knowledge in China and to compare the levels of HIV knowledge in China over time and across countries. The first section describes surveys which are used for the analysis. The second section discusses samples used for the analysis. The third section presents response variables which are used for the analysis. Further to this, the explanatory variables and the rationale behind the choice of these variables are discussed. At the end of the chapter, data quality and limitations are considered.

2.1 Datasets Used

There are no Demographic and Health Surveys (DHS) in China. Therefore, in order to study HIV awareness and knowledge in China different data sources which are available and which ask a standard HIV awareness question (have you heard of HIV/AIDS?) followed by a set of questions about correct and incorrect routes of HIV transmission or ways to prevent HIV infection and misconceptions about HIV are used for the analysis. Four cross-sectional Chinese datasets for women of reproductive age (15-49) are analysed in this thesis: the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001, as well as the UNFPA Reproductive Health and Family Planning Surveys 2003 (CP5 baseline survey) and 2005 (CP5 endline survey). The first two surveys are nationally representative of the whole of China. The main purpose of these two surveys was to find out details about fertility, contraception and reproductive health, including HIV knowledge, of women of reproductive age, as well as to find out about Family Planning (FP) and Reproductive Health (RH) services that were provided in China at these times. Another China National Family Planning and Reproductive Health Survey was conducted in 2006 but unfortunately data have not been made available for public use and, therefore, this dataset is not included in the analysis.

The 1997 survey used a two-stage sample: first 1041 communities were selected, then 15213 women were selected within those communities (Jiang 2000). The sample of women in this survey is self-weighted (Jiang 2000). According to Jiang (2000), the quality of the data collected met all objectives the research team had.

The 2001 survey was conducted in the same 1041 communities which were selected for the 1997 survey (Pan *et al*. 2003). All women of reproductive age in these communities were interviewed for this survey. The sample of women in this survey is not self-weighted but unfortunately weights are not available in the dataset. According to Pan *et al*. (2003), the quality of data was also satisfactory.

The last two surveys (CP5 baseline and endline surveys) are not nationally representative. They were conducted in 30 purposely selected counties from the 30 provinces in China, and they are representative of those 30 counties in China. In every province the best performing county was put forward for the participation in the programme. Reproductive health and family planning knowledge, including HIV awareness and knowledge of individuals and knowledge about FP services, were the primary interests of the survey. Intervention programmes conducted by UNFPA which were discussed in detail in section 1.2.5 were introduced in these counties between 2003 and 2005. These interventions were based on the recommendations from the CP5 baseline survey report (Li *et al*. 2004b).

The 2003 survey used stratified multi-stage sample of women aged 15 to 49 years (Li *et al*. 2004a). Thirty counties were stratified into the three regions: Eastern, Central, and Western and then regions were stratified by urban and rural areas. The sample size for each region was equal. Within each of the three regions, 35 townships were first selected (Li *et al*. 2004a). Systematic random sampling with probabilities proportional to population of women aged 15 to 49 within each township was used for the selection of townships within each stratum in the baseline survey (Li *et al*. 2004a). Then within each township four communities were selected using systematic random sampling with probabilities proportional to the population of women aged 15 to 49 years within each community. At the final stage of the sample selection a systematic random sample of 20 women was selected from a list ordered by age of all women aged 15 to 49 years within each selected community (Li *et al*. 2004a).

In the 2005 survey the main change in sampling occurred: the sampling procedure was moved from the sample of individuals to the sample of households (Li *et al*. 2008). The 2005 survey used stratified multi-stage sample of households (Li *et al*. 2008). Thirty counties were first stratified into the three regions: Eastern, Central, and Western and then regions were stratified by urban and rural areas. The sample size for each region was equal. Within each of the three regions, 50 townships were first selected (Li *et al*. 2008). Systematic random sampling with probabilities proportional to the number of households within each township was used for the selection of townships in the endline survey (Li *et al*. 2008). Then within each township four communities were selected using systematic random sampling with probabilities proportional to the number of households within each community. In the final stage, a systematic random sample of 15 households was selected for a list ordered by local geography, and all women of reproductive age were considered eligible for the survey in these households (Li *et al*. 2008). Women were asked similar question at two points of time (2003 and 2005) which allows trends to be studied over time in the context of the 30 selected counties. However, the surveys are cross-sectional in nature and different samples of women were selected for the two surveys.

The sample designs of the baseline survey (2003) and endline survey (2005) are not a simple random sample and, therefore, weights were necessary in order to produce 'unbiased' estimates of means and proportions (Li *et al*. 2004a; Li *et al.* 2008). The designs of the surveys within regions and the urban and rural strata of each region are self-weighted. Therefore, there is no variation in weights across women in these groups (Li *et al*. 2004a; Li *et al*. 2008). The raw weights for each woman are calculated as the inverse of the selection probability in the 2003 survey (Li *et al*. 2004a). In the 2005 survey first raw weights for each woman are calculated as the inverse of the selection probability, and then this weight was multiplied with the raw household weight to calculate the overall weight for a woman (Li *et al*. 2008).

These four Chinese datasets are used for the analysis in Paper One of the thesis. Due to the different representativeness levels of the two sets of surveys (1997-2001 and 2003-2005) and the similarity of the representation levels within these two sets, these two periods of time are considered separately.

One of the aims of Papers Two and Three is to compare the levels of HIV knowledge at various points in time across population sub-groups between 1997 and 2005 in China and then to compare HIV knowledge in China with the levels of HIV knowledge in four other countries. The following countries are selected to enable comparisons of the levels of HIV knowledge in different cultural and epidemiological contexts: India, Kenya, Malawi and the Ukraine. Kenya is a country with high HIV prevalence around 6.3% among adults aged 15-49 (UNAIDS 2010) with a long history of educational interventions and campaigns. Malawi is another country with a high HIV prevalence around 11% among adults aged 15-49 (UNAIDS 2010) and also with a long history of educational interventions and programmes. These two countries have generalised or mature HIV epidemics and, therefore, they are interesting cases for comparative purposes. India is a country with HIV prevalence (around 0.3% among adults (15-49) (UNAIDS 2010)) comparable to China and with a large population at HIV risk due to the size of the population in the country. The Ukraine is a country with a slightly higher HIV prevalence and the highest HIV prevalence in Eastern Europe and Central Asia (around 1.1% among adults (15-49) (UNAIDS 2010)), with a smaller number of people living with HIV due to the smaller size in the country but an educational system comparable to Chinese which is a heritage from the Former Soviet Union.

Three out of four Chinese datasets discussed earlier are used for the analysis in Papers Two and Three: the China National Population and Reproductive Health Survey 1997 and the China National Family Planning and Reproductive Health Survey 2001, as well as the UNFPA Reproductive Health and Family Planning Survey 2005. UNFPA Reproductive Health and Family Planning Survey 2003 is excluded from the analysis because of the problems with the data collection of HIV knowledge questions. In the survey respondents were given the list of potential routes of HIV transmission and were asked to list those which they think might transmit HIV[24]. When the data were coded, “yes” and “no” were the only two possible options for answers to questions on the knowledge of routes of HIV transmission. If people did not mention a route, their answers were recorded as “no” because of the absence of “do not know” category and, therefore, the level of knowledge about HIV misconceptions is potentially overstated in this survey (as for the misconceptions questions “no” is the correct answer).

[24] Private communication with Professor Li Bohua from the China Population and Development Research Centre, Beijing, China (August 2008).

Data for the other countries used for Papers Two and Three come from DHSs which were conducted between 2003 and 2007. "Demographic and Health Surveys (DHS) are nationally-representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition"[25]. They consistently collect data on HIV knowledge, attitudes and behaviour which are linked to a wide range of demographic, economic, social, and health characteristics of individuals. In India a DHS is not conducted but a survey with questionnaires and sampling methodologies similar to DHS was carried out, the National Family Health Survey (NFHS-3). This survey is a part of a DHS programme and, therefore, will be referred hereafter as a DHS. For this study four different surveys were selected (India 2006, Kenya 2003, Malawi 2004 and Ukraine 2007) in order to be able to compare HIV knowledge in China with HIV knowledge in other parts of the world. Countries with generalised HIV epidemics (Kenya and Malawi) as well as with non-generalised epidemics and similar to China characteristics, e.g. population size (India), educational system (the Ukraine) were selected for the analysis. The choice of years of surveys was determined by data availability as well as by the attempt to use the most recent available data which are chronologically closer to the most recent Chinese survey used for the analysis (the UNFPA Reproductive Health and Family Planning Survey 2005). An attempt was made to include India 1999 data in order to assess changes in HIV knowledge in India across time with the aim of comparing them with changes observed in China. However, due to the data limitations[26] this dataset was excluded from the analysis.

The India DHS, Kenya DHS, Malawi DHS and Ukraine DHS are all samples of households as are all other DHS surveys. The Kenyan, Malawian and Ukrainian surveys have two-stage sample designs. At the first stage clusters were sampled, from urban and rural areas. During the second stage of selection, households were systematically sampled. All women aged 15-49 years in these households could be interviewed in the survey.

In India urban and rural samples within each state were drawn separately (IIPS 2007). The same sample design was used in all states in India. However, sample selection procedures differ in urban and rural areas. In rural areas a two-stage sample design was

[25] http://www.measuredhs.com/aboutsurveys/dhs/start.cfm [Accessed 14 May 2009]
[26] It was not possible to create comparable across countries scores for measuring HIV knowledge as a question about mother-to-child transmission was not asked in 1999 (IIPS 2000).

used. At the first stage primary sampling units (PSUs) or villages with probability proportional to population size were selected, at the second stage households were randomly selected within each PSU. In urban areas a three-stage procedure was adopted. At the first stage wards were selected with probability proportional to population size, at the second stage one census enumeration block was randomly selected from each ward and at the third stage households were randomly selected from each census enumeration block (IIPS 2007).

2.2 Samples Used

In four datasets used for the analysis in Paper One women were asked a question if they have ever heard of HIV/AIDS. This will be referred to as the HIV awareness question. Table 2.1 contains the distributions of answers to this question and sample sizes in the four datasets included in the analysis in Paper One. The three datasets out of four have two possible categories for answers: “yes” and “no”, and women who answered “yes” to the question were asked further HIV knowledge questions. However, in the China 1997 dataset there were three possible answers: “no”, “yes but do not understand it”, “yes and understand it”. All women from the second and third categories were asked further HIV knowledge question and, therefore, these two categories were combined and created “yes” category. Table 2.1 shows that HIV awareness increases over time in China.

Table 2.1: Number of respondents reporting HIV awareness in different countries and at different times (in China) and sample sizes of datasets.

Country and year of survey	Number who have heard of HIV (%)	Number who have not heard of HIV (%)	Total number of women
China 1997	9653 (63.5)*	5560 (36.5)	15213
China 2001	28781 (72.7)	10805 (27.3)	39586
China 2003	6836 (81.4)	1564 (18.6)	8400
China 2005	6855 (93.2)	501 (6.8)	7356

*Note: The figures in parentheses show the percentages. *- this category is a combination of two categories – heard of it but do not understand it (7968 women) and heard of it and understand it (1685).*

The four Chinese datasets which contained all women are used for the analysis in Paper One. Separate datasets are cleaned. The four datasets do not contain any missing values.

In Papers Two and Three seven datasets are used for the analyses. The datasets used for the analysis contain only respondents who were asked HIV knowledge related questions. Table 2.2 presents stages of preparation of datasets for the analyses in Papers Two and

Three and shows percentages of women who are excluded from the final datasets due to missing values. In China 1997, China 2005, India 2006 and Ukraine 2007 all women who answered “yes” to the HIV awareness question were asked further HIV knowledge questions. In the 2001 survey in China only women who answered “yes” to the HIV awareness question and to the following question “is AIDS a communicable disease?” were asked further HIV knowledge questions. In Kenya 2003 and Malawi 2004 only women who answered “yes” to the HIV awareness question and to the following question “is there anything a person can do to avoid getting AIDS or the virus that causes AIDS?” were asked all further HIV prevention questions. Only women who answered all seven selected HIV knowledge component questions which are discussed in section 2.3 are retained in the datasets used for analysis in Papers Two and Three.

In Paper Two it would be possible to include all women in analysis. If a woman answered “no” to the HIV awareness question her HIV knowledge could be assumed to be inadequate and she would score 0 on HIV knowledge. This approach was used in the literature (see section 4.1.1). However, this approach is not appropriate for a latent variable approach as latent variables can only be derived for women who answered the separate HIV knowledge questions. Therefore, in order to enable comparison of final results from Papers Two and Three, only women who are asked HIV knowledge related questions were included in analysis in both papers.

The separate datasets are then further cleaned and observations with missing values are removed. Missing values were assumed to be missing at random[27]. Table 2.2 shows that proportions of removed observations with missing values are zero in the Chinese context and are low in all other contexts and, therefore, observations with missing values are not examined further.

[27] When values are missing at random, it is assumed that the probability of a value to be missing cannot depend on any unobserved value of either the predictors or any outcome values but can depend on any observed data (Little and Rubin 2002; Singer and Willett 2003).

Table 2.2: Percentage of women who were excluded from the analysis due to missing values among those who are aware of HIV in both contexts.

	All women who are aware of HIV (A)	All women who were asked HIV knowledge questions (B)	Women who answered all components of HIV knowledge questions (C)	Final datasets used for the analysis in Papers Two and Three (D)	Percentages of women with missing values removed (%) (E)
Chinese context					
China 1997	9653	9653	9653	9653	0.0
China 2001	28781	26138	26138	26138	0.0
China 2005	6837	6837	6837	6837	0.0
Five country context					
China 2005	6837	6837	6837	6837	0.0
Kenya 2003	8052	6912	6877	6877	0.5
Malawi 2004	11548	11024	11002	11002	0.2
India 2006	88395	88395	88232	88223	0.2
Ukraine 2007	6729	6729	6638	6638	1.3

*Notes: The only difference in frequencies between columns C and D is observed in India 2006 dataset. Nine women who answered all components of HIV knowledge questions have missing values for variable education. Percentages in column E are derived using the following formula: (D/B)*100.*

Then the separate datasets are pooled together into two pooled datasets to enable comparisons in Papers Two and Three. The first pooled dataset is produced to study HIV knowledge in the Chinese context over time and includes China 1997, China 2001 and China 2005. The second dataset is created to compare levels of HIV knowledge in China with levels of knowledge in other countries (the five country context) as well as to compare HIV knowledge in countries with generalised and with non-generalised HIV epidemics and it includes China 2005, Kenya 2003, Malawi 2004, India 2006 and Ukraine 2007. Table 2.3 presents sample sizes of all separate datasets together with pooled datasets which are used for the analysis in Papers Two and Three.

Table 2.3: Sample sizes of separate datasets and of pooled datasets used for the analysis in Papers Two and Three.

Country and year of survey	Total number of women	Percentage of women in pooled dataset 1	Percentage of women in pooled dataset 2
China 1997	9653	22.6	
China 2001	26138	61.3	
China 2005	6837	16.1	5.7
India 2006	88223		73.8
Kenya 2003	6877		5.8
Malawi 2004	11002		9.2
Ukraine 2007	6638		5.6
Pooled dataset 1	42628		
Pooled dataset 2	119577		

2.3 Response Variables

HIV awareness is the response variable in Paper One. HIV awareness is a binary variable with the value of 1 for women who reported being aware of HIV and 0 otherwise.

HIV knowledge variables are response variables in Papers Two and Three. As mentioned earlier, HIV knowledge cannot be measured directly but can be measured through different components of HIV knowledge (answers to different questions about routes of HIV transmission or ways of HIV prevention) and HIV misconceptions. An attempt was made to obtain a comparable measure of HIV knowledge for the Chinese and the five country contexts but unfortunately variables on knowledge of different routes of HIV transmission were collected in two different ways in the China and in DHS surveys. In China each respondent was asked about every possible route of HIV transmission and was expected to answer every question about different routes of HIV transmission. However, in the DHS surveys, knowledge of routes of HIV transmission (with the exception of the questions about MTCT and transmission of HIV through sharing food or utensils with an HIV positive person) was collected within the knowledge of ways to avoid HIV section of the questionnaire. Respondents were first asked if they know ways to avoid HIV and then if the answer to this introductory question was “yes”, they were expected to spontaneously mention ways to avoid HIV they know. The respondents were not further probed on the ways they did not mention. Therefore, it makes it difficult to compare scores which were created on the basis of these variables across countries.

As the result of this, two sets of questions are used to enable comparison of HIV knowledge: one set for investigating HIV knowledge in the Chinese context over time and the second set for comparing HIV knowledge in the five country context.

The first set of questions (Figure 2.1) for studying the evolution of HIV knowledge in China over time consists of the following questions:

1. Can HIV be transmitted through blood transfusion?
2. Can HIV be transmitted through contaminated needles?
3. Can HIV be transmitted through sexual intercourse?
4. Can HIV be transmitted from the infected mother to her child?
5. Can HIV be transmitted through a handshake?

6. Can HIV be transmitted through sharing food and utensils with an HIV infected person?
7. Can HIV be transmitted through kissing?

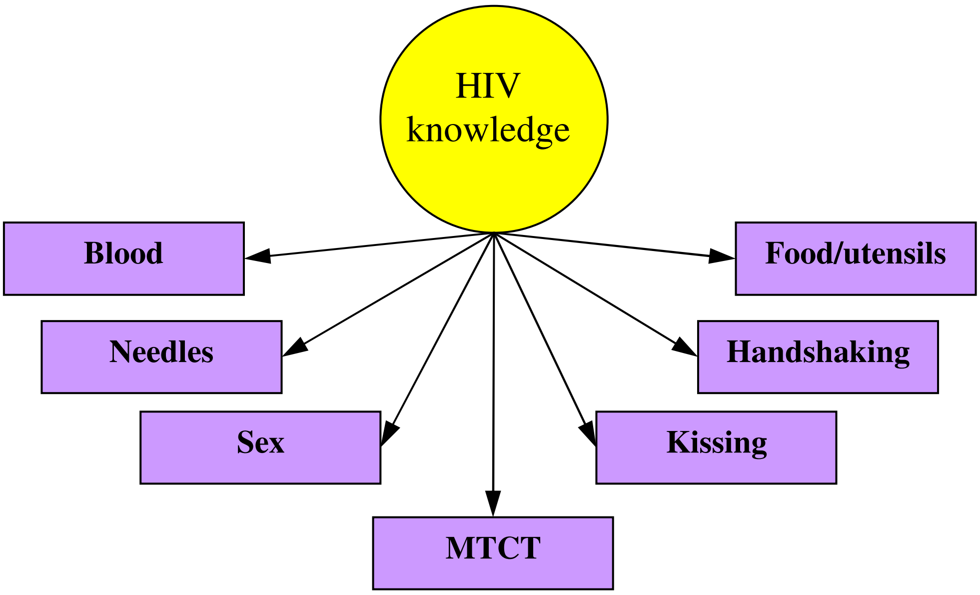


Figure 2.1: The components of the HIV knowledge variable in the Chinese context.

The first four questions are included to measure the knowledge of correct routes of HIV transmission. These four routes are considered to be the main routes of HIV transmission according to the Centre for Disease Control (CDC)[28]. The last three questions are included to measure the knowledge of incorrect routes of HIV transmission in China. These three questions give good coverage from simple interactions between people (handshaking) to more close daily contact (sharing utensils or food) to personal contact (kissing). Other misconceptions were mentioned in different surveys (HIV transmission through haircut, through using public bathrooms, through mosquito bites). However, in order to ensure comparability across time in the Chinese context, they are not included in the current analysis.

Table A.1 in Appendix A presents the availability of comparable variables in the different Chinese surveys. Blood transfusion, mother-to-child transmission, handshaking, sharing food or utensils and kissing questions were asked in the same way in all three Chinese surveys which will be used for the analysis in Papers Two and Three. However, not all the surveys asked questions about needles sharing and sexual

[28] http://www.cdc.gov/hiv/resources/factsheets/PDF/transmission.pdf [Accessed 29 July 2009]

transmission of HIV/AIDS in the same way in the Chinese context (Table A.1 in Appendix A). A question about transmission of HIV through injections in 1997 survey is used as a proxy for needle sharing as the mechanism of transmission of HIV is the same in these two cases: septic instruments which transmit HIV through blood products. The "having multiple sexual partners can transmit HIV" variable is used as a proxy for sexual route of transmission in 1997 and 2001. The way this question about the sexual route of transmission of HIV was asked in the Chinese context has limitations. It is a well known fact that HIV can be transmitted even when people have one sexual partner when this partner is an HIV positive person. According to UNAIDS (2009), "it is estimated that more than 90% of the 1.7 million women living with HIV in Asia became infected from their husbands or partners while in long-term relationships" (p.1). However, despite this limitation, this question is used for the calculation of different HIV knowledge measures which are used in Papers Two and Three.

The following questions (Figure 2.2) are used in the construction of HIV knowledge variables for cross-country comparisons of HIV knowledge in the five country context:

1. In your opinion, can people reduce their chances of getting HIV/AIDS by having just one uninfected sex partner who has no other sex partners?
2. In your opinion, can people reduce their chances of getting HIV/AIDS by using a condom every time they have sex?
3. In your opinion, can people reduce their chance of getting HIV/AIDS by abstaining from sexual intercourse?
4. Can HIV/AIDS be transmitted from a mother to her baby?
5. Is it possible for a healthy-looking person to have HIV/AIDS?
6. In your opinion, can people get HIV/AIDS from mosquito bites?
7. In your opinion, can people get HIV/AIDS by sharing food with a person who has AIDS?[29]

[29] Questionnaires are accessed through reports available on http://www.measuredhs.com/accesssurveys/search/start.cfm [Accessed 15 July 2009]

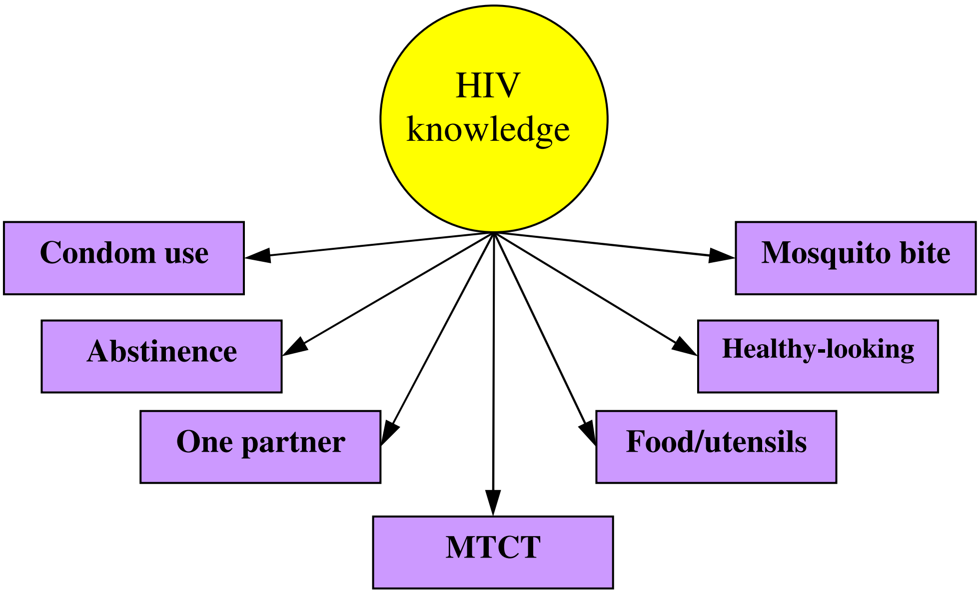


Figure 2.2: The components of the HIV knowledge variable in the five country context.

The wordings of questions differ slightly from survey to survey. The exact wordings of questions can be found in Table A.2 in Appendix A. The first four HIV knowledge questions are included in the calculation of Score One in the five country context. For simplicity these questions are further referred to as one sexual partner, condom use, abstinence and MTCT questions. These four questions are concerned with the HIV prevention knowledge about two main routes of HIV transmission: sexual transmission and mother-to-child transmission. As heterosexual transmission is a main contributor to the HIV epidemics in sub-Saharan Africa, as well as in China (UNGASS 2008), it is important to measure knowledge about prevention of heterosexual transmission through the knowledge about abstinence, condom use and sexual relations with one partner. The number of women infected with HIV increases once the HIV epidemic progresses. The increasing number of infections among women means that the cases of mother-to-child transmission (MTCT) increase, and as a result of that, the number of children infected by HIV, increases too. However, mother to child transmission of HIV can be prevented and according to UNAIDS (2010), "[v]irtual elimination of mother-to-child transmission of HIV is possible" (p.9). Knowledge about prevention of HIV from mother to a child is one of the important pre-requisites for successful prevention of MTCT in every epidemiological context. Therefore, it is important to include knowledge about this route of HIV prevention into the measurement about HIV prevention knowledge.

The last three questions are included in the calculation of the Score Two in the five country context and for simplicity are further referred to as healthy-looking person, mosquito bites and sharing food questions. The same set of three questions about misconceptions about HIV is used in the UNGASS indicator 13 which was discussed in detail in section 1.3.1.

A question about ways to prevent HIV transmission through intravenous drug use (IDU) is not included in the measure of knowledge about ways to prevent HIV knowledge because of unavailability of comparable questions in all five contexts used for the analysis in Papers Two and Three. The blood transfusion question is not included in the measure of HIV knowledge in the five country context either because of unavailability of comparable questions but also due to the fact that, in general, changes at institutional level rather than at individual level are more important for effective prevention of HIV through this route of HIV transmission. An exception is China because of the blood donation scandal discussed in section 1.2.2 where not only institutional but also individual changes were of crucial importance.

Table A.2 in Appendix A presents the availability of comparable questions in different countries. The majority of questions were asked in a similar way in all five surveys with slight variations in wordings as mentioned earlier (Appendix A, Table A.2). The questions about healthy-looking persons and MTCT represent exceptions. The healthy-looking person question was asked slightly differently in China when compared to other countries included in the analysis in Papers Two and Three in the five country context. In China the following question was asked: "Can you tell if a healthy-looking person has the HIV virus?", whereas in other counties the question was: "Is it possible for a healthy-looking person to have HIV/AIDS?".

MTCT questions were asked in the same way in all countries apart from the Ukraine. One question about MTCT was available in four contexts (Appendix A, Table A.2). However, in the Ukraine three related questions were asked (Appendix A, Table A.2). The MTCT variable was derived in the Ukrainian context on the basis of three available variables: if a respondent correctly answered at least one question, it was assumed that this respondent had an adequate knowledge about MTCT. All other questions were asked in the same way in the different surveys.

In all datasets used for the analysis in Papers Two and Three for each component of HIV knowledge there were three possible answer options available to respondents: "yes", "no" and "do not know". All answers were recoded into "correct" and "incorrect" categories. "Do not know" answer options were coded as "incorrect" as people who answered "do not know" were assumed to have no knowledge about a specific question. This approach was used in other studies (see section 4.1.1). Details about derivations of specific HIV knowledge measures used for the analyses in the thesis are presented in Papers Two and Three.

2.4 Explanatory Variables

The literature discussed in section 1.3.3 suggests that a number of characteristics of individuals are associated with HIV awareness and knowledge in the Chinese context, and the most frequently mentioned ones are education, age, type of place of residence, income, occupation, exposure to media and gender (Manchester 2002; Chen *et al*. 2003a; Dong *et al*. 2003; Li *et al*. 2004c; Zhang *et al*. 2006; Tan *et al*. 2006; Wu *et al*. 2007a; Tan *et al*. 2007; Tan 2008). In some cultural contexts (especially in Asia, Africa and Middle East) unmarried women are excluded from family planning and reproductive health services due to the cultural attitude to sexual relations which are only considered within the contexts of wedlock. Traditionally, age of marriage in these contexts is young for women (Coale 1992; Singh and Samara 1996). However, the marital age increases in different cultural contexts and, therefore, young unmarried women are exposed to unsafe sexual behaviour. These women usually lack information about safe sex practices; they might also not have access to contraception, including condoms. According to Zhang *et al*. (2004), in China there were no specific reproductive health programmes for unmarried women that would provide education, information and services. Since HIV knowledge in China was promoted through family planning services, it was considered to be important to include marital status as an explanatory variable when modelling HIV awareness and knowledge. As this thesis first examines changes in HIV awareness and then HIV knowledge at different points in time as well as in different countries, comparable control variables have been chosen across all different contexts. To ensure comparability and consistency, only a limited number of variables are used in this thesis.

In Paper One age, education, residence, ethnicity and marital status variables are available in all datasets and these five variables were standardised where necessary to have the same categories for comparative purposes. Some extra variables are available for the analysis of changes in HIV awareness during the second period of time (2003-2005) and they are also included into the analysis: region, occupation, and media exposure variables, as they were found significant in explaining variability of HIV awareness in the literature (see section 1.3.3). Demographic and other characteristics of respondents in the four Chinese datasets used for the analysis in Paper One are presented in Table B.1 of Appendix B.

In Papers Two and Three the following comparable explanatory variables are available and are used in both study contexts: age, education, residence and marital status. It is hypothesised that HIV awareness and level of HIV knowledge might differ by region and by ethnicity. Regional and ethnic differences in HIV awareness and knowledge can be explained by the fact that regions might have different levels of development and different level of access to education (e.g., in China the Eastern region is the most developed and the Western is the least developed, in Kenya the North Eastern Province is the least developed part of the country when compared with other parts of Kenya) and people of the same ethnicity might live in specific regions of a country. As ethnicity and region are not comparable across countries, these variables can only be used in the Chinese context for comparison of HIV knowledge in China across time. Unfortunately, a region variable is not available in China 1997 but ethnicity variable are used in the Chinese context for comparison of HIV knowledge in China across time. All variables were standardised where necessary to have the same categories for comparative purposes.

A number of macro-level variables is considered in the five country context when comparison between HIV knowledge in China and in other countries is conducted. The following macro-level variables are hypothesised to be predictive of the level of HIV knowledge: HIV prevalence, HIV awareness rate, female literacy rate, and type or severity of epidemic. HIV prevalence data are created on the basis of the estimates available in annual UNAIDS Reports on the Global AIDS Epidemic[30] and from the relevant DHS reports for Kenya, Malawi and India[31]. The HIV awareness rate is calculated from the separate original datasets, subsamples of which are then further used in the second pooled dataset. Female literacy rate indicators for appropriate years are

[30] http://www.unaids.org/en/resources/unaidspublications/2010/ [Accessed 05 March 2011]
[31] http://www.measuredhs.com/topics/hivprev/data.cfm#2 [Accessed 20 November 2010]

obtained from the CIA World Factbooks education statistics[32]. The type of epidemic was hypothesised to be important for predicting HIV knowledge (Snelling *et al*. 2007) and, therefore, type of epidemic variable which has two categories "generalised epidemic" and "non-generalised epidemic" is created on the basis of HIV prevalence variable and conventional definition of generalised epidemic mentioned in section 1.1. The last macro-level variable which is considered in the new pooled dataset is a country with year of survey variable.

The demographic and other characteristics of respondents together with their HIV related knowledge in the Chinese context and in the five country context which are used for the analysis in Papers Two and Three are presented in Table F.1 of Appendix F.

2.5 Data Quality and Limitations

The results of any study are dependent on the quality of data which are used for the analysis. Therefore, it is important to discuss limitations of data used for the analysis. The China 2001 national data are not self-weighted data but the dataset do not contain weights. However, the absence of weights in the dataset should not cause a problem for the study as the main interest of the study is to establish relationships. Unfortunately, 1997 survey does not have region, province and PSU identifiers, and 2001 data do not have province and PSU identifiers. Therefore, it is impossible to control for the effects of clustering by analysing potential random effects which might be significant. In contrast the China 2003 and China 2005 data have all variables to enable multilevel modelling but these data are not nationally representative and, therefore, the results can be interpreted only for the 30 selected countries in China and inferences cannot be made to the whole female population of reproductive age in China.

Comparability issues exist because of differences in questions used in different surveys in the Chinese context and the five country context discussed in sections 2.3 and 2.4. Comparability issues might also exist due to differences in cultural contexts as, for example, rural areas in China might not be comparable to rural areas in Kenya, or standards, for example, of primary education, might vary in different parts of the world. However, all studies which use a number of countries for comparisons will face these limitations (Van de Vijver and Leung 1997; Harkness 1999; Harkness *et al*. 2003).

[32] http://www.nationmaster.com/graph/edu_lit_fem-education-literacy-female [Accessed 30 November 2010]

The usage of pooled datasets presents another limitation as pooled datasets might mask peculiarities of some specific contexts. However, the pooled datasets enable comparisons over time and across countries and this is the main interest in Papers Two and Three. Peculiarities of specific contexts might be investigated further by examining separate datasets if this is of interest for the research.

Box 2.1: Key points in Chapter 2

- Data for this thesis are taken from the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001, the UNFPA Reproductive Health and Family Planning Surveys 2003 and 2005, India DHS 2006, Kenya DHS 2003, Malawi DHS 2004 and Ukraine DHS 2007.
- Samples of all women who answered the HIV awareness question in four Chinese datasets are used for the analysis in Paper One.
- HIV awareness is a response variable in Paper One.
- Samples of women who answered all HIV knowledge questions selected for the analysis in seven datasets are used for the analysis in Papers Two and Three.
- Six different HIV knowledge measures are used as response variables in Papers Two and Three.
- Two pooled datasets are created: one to compare HIV knowledge in the Chinese context across time, the second one to compare HIV knowledge across countries in the five country context.

Chapter 3: HIV Awareness in China: A Decomposition Analysis

Abstract

This paper examines the evolution of HIV awareness among women in China between 1997 and 2005. The aim of this paper is to compare the levels of HIV awareness at various points in time between 1997 and 2005. For the analysis, the following data sources are used: the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001 and the UNFPA Reproductive Health and Family Planning Surveys 2003 and 2005. Changes in levels of HIV awareness in different data sources can be attributed to the time at which the data were collected, different sampling techniques, different representativeness levels, as well as changes in the governmental agenda and the effect of programmes, interventions and campaigns. A regression decomposition analysis technique is used in this paper in order to disentangle the two main components driving a change in HIV awareness: change in population structure and change in effect sizes due to external factors such as political environment, interventions and programmes. The results show that HIV awareness increased over time in China. Education, ethnicity and media exposure to TV and newspapers are the main factors that are associated with HIV awareness in China in 2005. With time, lower awareness groups are catching up and gaps between groups with initially different awareness levels are narrowing. The increases in HIV awareness observed between 1997 and 2001 are similar between groups of women with different demographic characteristics whereas between 2003 and 2005 increases are more pronounced among different groups that were targeted by the various interventions including UNFPA interventions. The change in HIV awareness between two sets of surveys can also be partly attributed to different representativeness levels of surveys. The results suggest that the smaller part of the observed change in HIV awareness is attributed to the change in population structure but the main driver of

the observed change in HIV awareness over time in China is change in the environment such as in political commitment, interventions and campaigns.

3.1 Introduction

HIV awareness consistently increases all over the world (Snelling *et al*. 2007). China's early stage of the HIV epidemic, its large and diverse population as well as its unique demographic features make China an interesting case for studying changes in HIV awareness in a general population. The change in HIV awareness can be attributed to a number of different factors such as change in political commitment, interventions, individual change and change in population structure or population characteristics. The change can also be due to the differences in sampling technique used in a particular survey and in the representativeness levels of different surveys. Individual change can be affected by political commitment and interventions but could also be independent of these two and could just be a function of changed characteristics of an individual, for example, a person moved from a rural area to an urban area or obtained a higher educational level. It is important to establish what drives the change in HIV awareness in different contexts as well as what brings larger change to HIV awareness. These findings might be helpful for designing as well as for assessing the effectiveness of interventions for countries where HIV epidemic is still at an early stage and where there are still gaps in HIV awareness. Decomposition analysis can help to find out what part of the change is determined by a change in relationship between HIV awareness and characteristics of people and what part is driven by a change in a population structure. It is also important to understand what specific characteristics of the people drive the total change in relationships and effect sizes between HIV awareness and individual characteristics.

As mentioned earlier, no studies have examined changes in HIV awareness in the general population in China over time. Therefore, the analysis presented in this paper will fill this gap. As HIV awareness is an important component in the HIV risk reduction frameworks discussed in section 1.3.1, it is important to know how the levels of awareness are changing over time in general population in China. The results of this investigation can feed directly into the evaluation of existing HIV prevention programmes in China and might suggest implications for both general policy as well as targeted interventions. These findings might also be helpful in designing interventions for countries where the HIV epidemic is still at an early stage.

3.1.1 Aims, research questions and structure of Paper One

This paper systematically examines the evolution of HIV awareness among women in China between 1997 and 2005. The main aim of the paper is to decompose the total change in the proportions of HIV awareness at different points of time (1997 and 2001; 2003 and 2005) into the change due to difference in population structure and the change due to change in HIV awareness among women whose characteristics are similar or in other words the change in HIV awareness due to the external or environmental factors which can affect HIV awareness such as interventions and political commitment.

The main research questions are:

1. Has HIV awareness changed over time in China?
2. What are the main drivers and how large is their influence on the observed change in HIV awareness over time in China?
3. What are the main characteristics of women which are contributing to the change in HIV awareness over time in China?

In order to answer the main research questions, this paper will

1. study HIV awareness in China at different points in time;
2. investigate the determinants of HIV awareness at different points in time;
3. examine the change in HIV awareness between 1997 and 2001, and then between 2003 and 2005;
4. study contributions to the total change in HIV awareness that come from the change in population structure, and from the change in relationship or effect sizes between HIV awareness and characteristics of women between 1997 and 2001, and between 2003 and 2005;
5. compare these contributions;
6. and finally, will examine the main explanatory variables contributing to the change in HIV awareness between 1997 and 2001 and between 2003 and 2005.

To address the research questions and the main aims and objectives discussed above, the four Chinese datasets discussed in section 2.1 are used for the analysis. Due to the different representativeness levels of the two sets of surveys (1997-2001 and 2003-2005) and the similarity of the representation levels within these two sets, these two periods of time are considered separately in this analysis.

This paper consists of five main sections. The second section presents the methodology used for the analysis. Section three discusses and summarises all the results. In section four, the limitations of this study are discussed. Finally in section five the main conclusions are presented.

3.2 Methodology

3.2.1 Exploratory data analysis

Exploratory data analysis is conducted in order to investigate initially the relationship between HIV awareness and the socio-economic and demographic explanatory variables available in the four datasets used for the analysis. Bivariate associations between the outcome and explanatory variables are explored graphically (using line graphs) and statistically (using Pearson's chi-squared tests).

3.2.2 Logistic regression

Logistic regression analysis is then used to model HIV awareness at different years in China and to initially estimate the effects of different factors on the probability of being aware of HIV in China at four points in time. When the outcome variable is measured on a binary scale logistic regression can be used (Agresti 1996). Therefore, it is appropriate to use a logistic regression to model HIV awareness as it is a dichotomous variable with the value of 1 for women who reported being aware of HIV and 0 otherwise. The general equation of a logistic regression model is

$$\log(\text{odds}) = \text{logit}(\text{probability}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k, \qquad (3.1)$$

where α is a constant term (intercept), β_1, β_2, …, β_k are coefficients of the regression equation, X_1, X_2,…, X_k are explanatory variables, including dummies for categorical variables, and *k* is a number of explanatory variables/categories in the model.

Four logistic regression models are fitted to the four different datasets separately. The statistical package SPSS version 15 (SPSS 2006) will be used for the logistic regression modelling.

3.2.3 Model selection and interactions

As the comparability between surveys is more important than model selection process, all variables that are comparable in two sets of surveys are included in the analysis. No model selection is performed. Only a limited number of variables is used in the analysis for comparability reasons: age, residence, education, ethnicity, and marital status. For the 2003-2005 period of time, some extra variables are available and are also used for the analysis: occupation, region and exposure to media (TV, radio and newspapers). As comparability between surveys is of the main importance, only the main effects discussed above are included in the models and interactions are not tested for significance or added to the final models.

3.2.4 Decomposition analysis

Decomposition analysis is a widely used technique in different disciplines (Oaxaca 1973; Gomulka and Stern 1990; Nielsen 1998; Charasse-Pouélé and Fournier 2002; Canudas Romo 2003). Decomposition analysis helps to separate the essential components which can explain the phenomena under study in more detail. This technique is widely used in the field of demography. Standardization techniques both direct and indirect are a form of decomposition analysis. Canudas Romo (2003) summarises different ways in which decomposition analysis can be used when studying demographic phenomena. One of the decomposition methods he mentions within the field of demography is so-called regression decomposition. The main aim of this regression decomposition analysis is to disentangle the two main components of where a change comes from: a change in a population structure and a change in relationships or effect sizes (coefficients), i.e. do more people have a university degree in this year than in a previous year versus do more people with primary education know about HIV this year than in a previous year. The total change contains both elements and decomposition analysis allows the separation of these two different components.

Initially regression decomposition analysis was used in the field of labour economics, and specifically in studies on discrimination. Ronald Oaxaca used the so-called standardization analysis or decomposition methodology in 1973 to study male-female wage differentials and discrimination against female workers in urban areas of the United States (Oaxaca 1973). He decomposed the wage differential into the effects of differences in individual characteristics and the effects of discrimination.

Blinder (1973) used decomposition analysis to study wage differential among white and black men and among white men and women. Gomulka and Stern (1990) examined change in the proportion of married women in employment in the UK between 1970 and 1983. In their paper they referred to decomposition analysis as "growth accounting" and analysed change in proportion of wives working over time. In their decomposition analysis they separated the change in the structure of the population at different years and changes in coefficients which they attributed to changes in behaviour and the environment. They also used detailed decomposition analysis in order to be able to answer the question: "…which coefficients seem to be central among those that are changing and which aspects of the sample or population are responsible for the change arising from the sample" (Gomulka and Stern 1990, p.188). They pointed out the limitation of the detailed decomposition analysis approach which is the inability to calculate standard errors for the predicted participation rates calculated from 'composite' models (Gomulka and Stern 1990). Therefore, they suggested that conclusion based on the detailed decomposition analysis should be tentative but they also stated that their conclusions based on the detailed decomposition are strongly suggestive (Gomulka and Stern 1990).

Nielsen (1998) used decomposition analysis to study gender discrimination and tendency to be employed in the formal sector in urban Zambia. She also used a two-step approach in her decomposition analysis. At the first step Nielsen decomposed differences in probabilities in employment in the formal sector into differences in characteristics between males and females and discrimination. The second step Nielsen used was detailed decomposition analysis and showed how discrimination effects could be further decomposed into effects which were associated with particular coefficients.

Regression decomposition analysis then moved to other discipline such as, for example, health research. Charasse-Pouélé and Fournier (2002) used decomposition analysis to study health disparities among South-Africans. Jacobson *et al*. (2007) employed a multilevel decomposition analysis to investigate racial disparities in alcohol treatment completion. Stewart Williams (2009) used non-linear decomposition analysis to examine discrimination in referral to a cardiac rehabilitation programme. González Álvarez and Clavero Barranquero (2009) applied decomposition analysis to study inequalities in health care utilization in Spain.

Decomposition analysis has been applied to different models. Initially the regression decomposition analysis was used for linear models (Oaxaca 1973; Blinder 1973); this decomposition method is referred to as Oaxaca-Blinder decomposition. Later Gomulka and Stern (1990) extended the approach and used it for probit models, and Nielsen (1998) further extended the approach for logit models. Charasse-Pouélé and Fournier (2002) and Stewart Williams (2009) used the extension of the so called Oaxaca-Blinder decomposition analysis for non-linear models. Jackobson *et al*. (2007) applied decomposition analysis in a multilevel framework.

In this study, decomposition analysis is performed for logistic regression. This extension Oaxaca-Blinder decomposition is possible as log-odds ratios are linear in the parameters and, therefore, Oaxaca-Blinder decomposition analysis can be easily transferred from linear models to logistic models (Nielsen 1998). Nielsen (1998) argues that log-odds ratios can be used for interpretation but as it is easier to interpret difference in probabilities it is preferably to use them.

The use of decomposition analysis for the logistic model has advantages when compared to linear models as further decomposition analysis is possible. Oaxaca and Ransom (1999) in their note highlighted that detailed decomposition analysis is misleading for linear regressions due to the fact that the arbitrary choice of reference category in categorical variables will lead to differences in estimated intercepts and coefficients. However, this problem does not affect decomposition analysis using logistic regression as calculated predicted probabilities for different observations are not affected by the parameterisation of the model as attention is not specifically on coefficients which might change because of changes in baseline categories.

The decomposition method which was used by Nielsen (1998) can be used to account for the growth in the proportion of HIV awareness for cross-sectional data at different years. The main interest of the paper is to decompose the total change in the proportions of HIV awareness at different points of time (1997 and 2001; 2003 and 2005) into the change due to difference in population structure and the change due to the change in HIV awareness due to the external or environmental factors which can affect HIV awareness such as interventions and political commitment. Due to the short period of time that exists between the two sets of consecutive surveys, the main interest of the study is not the change in HIV awareness due to the change in population structure, as this change is expected to be small, but the change due to external factors and, therefore, detailed decomposition analysis is employed in order to determine which coefficients contribute most to the total change in relationship or effect sizes between HIV awareness and characteristics of women. If no interventions are implemented or interventions are not effective, similar change in HIV awareness is expected to be observed across different groups of women. However, if interventions are effective, we expect to see larger changes within the groups which were targeted by specific interventions.

The first step in the current decomposition analysis (Nielsen 1998) will be to decompose the difference in probabilities into the one part which is caused by differences in population structure between 1997 and 2001 (*S*) and the second part which is caused by change in relationships and might be partially attributed to change in the environment in a form of, for example, programmes and interventions (*E*). Define the following probabilities:

$$\overline{P}(\hat{\beta}_{1997}x_{1997}) = \sum_{i=1}^{N_{1997}} F[x_{1997i}\hat{\beta}_{1997}]/N_{1997} \tag{3.2}$$

$$\overline{P}(\hat{\beta}_{2001}x_{2001}) = \sum_{i=1}^{N_{2001}} F[x_{2001i}\hat{\beta}_{2001}]/N_{2001} \tag{3.3}$$

$$\overline{P}(\hat{\beta}_{1997}x_{2001}) = \sum_{i=1}^{N_{2001}} F[x_{2001i}\hat{\beta}_{1997}]/N_{2001} \tag{3.4}$$

$$\overline{P}(\hat{\beta}_{2001}x_{1997}) = \sum_{i=1}^{N_{1997}} F[x_{1997i}\hat{\beta}_{2001}]/N_{1997}\ , \tag{3.5}$$

where $F(x\beta) = e^{x\beta} / (1 + e^{x\beta})$ is the logistic cumulative distribution function, *x* is a row vector of explanatory variables and *N* is the sample size, subscripts 1997 and 2001 indicate the year of survey, and subscript *i* is an index of individuals. $\hat{\beta}$ is the estimated parameter vector or a vector of coefficients for the year 1997 and for the year 2001 depending on subscript.

$\overline{P}(\hat{\beta}_{1997}x_{1997})$ and $\overline{P}(\hat{\beta}_{2001}x_{2001})$ are the average probabilities of being aware of HIV for 1997 and 2001 respectively, and $\overline{P}(\hat{\beta}_{1997}x_{2001})$ is the average probability for the year 2001 if the respondents were treated like the year 1997, $\overline{P}(\hat{\beta}_{2001}x_{1997})$ is the average probability for the year 1997 if the respondents were treated like the year 2001.

The difference in the average probabilities for HIV awareness between the year 2001 and the year 1997 is then

$$\overline{P}(\hat{\beta}_{2001}x_{2001}) - \overline{P}(\hat{\beta}_{1997}x_{1997}). \tag{3.6}$$

By subtracting and adding the term $\overline{P}(\hat{\beta}_{1997}x_{2001})$, the difference in HIV awareness between the year 1997 and 2001 can then be decomposed into the two components:

$$\begin{aligned}
&\overline{P}(\hat{\beta}_{2001}x_{2001}) - \overline{P}(\hat{\beta}_{1997}x_{1997}) = \\
&[\overline{P}(\hat{\beta}_{2001}x_{2001}) - \overline{P}(\hat{\beta}_{1997}x_{2001})] + [\overline{P}(\hat{\beta}_{1997}x_{2001}) - \overline{P}(\hat{\beta}_{1997}x_{1997})] = \\
&E + S.
\end{aligned} \tag{3.7}$$

The first term in the square brackets (*E*) is the contribution to the change which comes from the change in coefficients between the two different years or in other words the term *E* is a difference between the average probability of being aware of HIV in the year 2001 and the average probability of being aware of HIV assuming that in the year 2001 relationship between HIV awareness and people's characteristics were the same as in 1997, i.e. we use the same sample but different logistic regression equation coefficients.

The second term in the second square brackets (*S*) is the contribution to the total change which comes from the change in the population structure between the two years or in other words the term *S* is a difference between the average probabilities of being aware of HIV assuming that in the year 1997 and in the year 2001 the relationships between the response variable and the explanatory variables were the same as in the year 1997.

Alternatively, by subtracting and adding the term $\overline{P}(\hat{\beta}_{2001}x_{1997})$ we get,

$$\overline{P}(\hat{\beta}_{2001}x_{2001}) - \overline{P}(\hat{\beta}_{1997}x_{1997}) =$$
$$[\overline{P}(\hat{\beta}_{2001}x_{2001}) - \overline{P}(\hat{\beta}_{2001}x_{1997})] + [\overline{P}(\hat{\beta}_{2001}x_{1997}) - \overline{P}(\hat{\beta}_{1997}x_{1997})] = \quad (3.8)$$
$$S' + E',$$

These both approaches to decomposition analysis lead to the same results. Equation (3.7) is used for the analysis in this paper.

The same equations are also applied to the decomposition analysis for 2003 and 2005 surveys.

Detailed decomposition analysis

The second step in the process is the detailed decomposition of the *E* component which will help to determine which specific explanatory variables contribute to the total change in HIV awareness among women with the same characteristics or to the change which partly can be attributed to influential changes in environment and interventions and campaigns.

All explanatory variables used for the analysis are categorical. The equation for decomposition analysis for the model with only one categorical variable which has two categories as explanatory variables is presented below. This equation can be easily extended to more categorical explanatory variables with two or more categories:

$$E =$$
$$\sum_{i=1}^{N_{2001}} \{F[\hat{\beta}_{0,2001} + \hat{\beta}_{1,2001}x_{i,2001}] - F[\hat{\beta}_{0,1997} + \hat{\beta}_{1,1997}x_{i,2001}]\} / N_{2001} =$$
$$\frac{1}{N_{2001}} \sum_{i=1}^{n_1} \{F[\hat{\beta}_{0,2001}] - F[\hat{\beta}_{0,1997}]\} + \frac{1}{N_{2001}} \times \quad (3.9)$$
$$\sum_{i=1}^{n_2} \{F[\hat{\beta}_{0,2001} + \hat{\beta}_{1,2001}x_{i,2001}] - F[\hat{\beta}_{0,1997} + \hat{\beta}_{1,1997}x_{i,2001}]\} =$$
$$E_1 + E_2,$$

where n_1 is a size of sub-sample of people who belong to the category 1 and n_2 from the category 2, E_1 is the change among people from the category 1 and E_2 from the category 2. $x_i = 0$ for the reference category, category 1, of an explanatory variable, and $x_i = 1$ for the category 2 of an explanatory variable.

In equation (3.9), $F(\beta_0 + \beta_1 x_i) = e^{\beta_0+\beta_1 x_i} / (1 + e^{\beta_0+\beta_1 x_i})$. (3.10)

Equation (3.9) helps to identify the contribution of each group of people to the total change in relationship between the response variable and the explanatory variables.

STATA 9.0 (STATA 2005) is used for calculation of four different sets of predicted probabilities which are used for the decomposition analysis.

The same equations are also applied to the detailed decomposition analysis for 2003 and 2005 surveys.

3.2.5 Bootstrap method for estimation of confidence intervals

In order to establish whether changes in probabilities of being aware of HIV within groups of women are significant between two points in time, confidence intervals for the predicted probabilities estimations should be constructed. The bootstrap method for estimation of the confidence intervals is employed to construct confidence intervals. The bootstrap is a resampling method which helps to produce different summary statistics or to perform hypothesis tests (Efron and Tibshirani 1993). The bootstrap method allows estimation of confidence intervals from distributions where parametric assumptions do not hold. Therefore, this method is suitable for estimation of confidence intervals for the fitted probabilities of being aware of HIV as their distributions are not normal (see section 3.3.4 for details).

1000 bootstrap samples are used which is sufficient according to general guidance for bootstrap methods (Efron and Tibshirani 1993). In order to construct confidence intervals for the mean to test if there is a significant difference between the means at different years or in other words if a significant change in the level of awareness in general population as well as in some specific groups of women has happened over a certain period of time (between 1997 and 2001 and between 2003 and 2005), the whole dataset is resampled 1000 times. (Code for 1997 and 2001 as well as for 2003 and 2005 for obtaining of means of predicted probabilities for 1000 replications can be found in the Appendix D.) During every replication a logistic regression model is fitted to the appropriate part of the data and predicted probabilities are estimated for the observations which are selected in the sample during the current replication. Then the total mean and appropriate group means are calculated and recorded in the new dataset which by the end of the 1000 replications will have distributions of means of predicted probabilities which will have 1000 items in each distribution. As 2003 and 2005 data has clustered sample design, bootstrap samples, therefore, are selected within the six clusters (three regions, and two types of residence) in order to better represent the original dataset.

Confidence intervals for two separate predicted probabilities are constructed using bootstrap method. These confidence intervals are produced by identifying 2.5^{th} and 97.5^{th} percentiles from the distributions of 1000 means of predicted probabilities. If two confidence intervals overlap, the difference between the two points in time is considered not to be significant.

3.3 Results

3.3.1 Exploratory data analysis

As mentioned earlier, all four surveys have an introductory (or filter) HIV awareness question asking if a person has heard of HIV and, therefore enabling comparison of HIV awareness among women in China across time.

The literature discussed in section 1.3.3 suggests that the following factors are associated with HIV awareness in China: gender, age, education, income, place of residence, type of job or occupation, and media exposure (Manchester 2002; Dong *et al*. 2003; Li *et al*. 2004c; Wu *et al*. 2007a; Tan et al. 2007). Demographic characteristics together with other characteristics of the respondents can be found in Appendix B.

Variations in HIV awareness by the selected characteristics (if available) are also presented in Appendix B, Table B.1. As discussed in section 2.4, since in China HIV awareness was initially promoted through family planning services which tend to focus mostly on married women, it was considered to be important to include marital status as an explanatory variable when modelling HIV awareness and, therefore, HIV awareness by marital status is also presented in Table B.1.

The results in Table B.1 suggest that, as expected, HIV awareness increases with time among different groups of women.

Figure 3.1 and Table 2.1 in Chapter 2 show the proportion of people who have heard of HIV at different times in China. The level of awareness increases over time as expected. Due to the major change in the government attitude towards HIV in 2003, the level of HIV awareness is expected to increase quicker after that year and Figure 3.1 and Table 2.1 confirm this expectation. However, part of this increase can be also explained by the representativeness level of 2003 and 2005 surveys discussed in section 2.1. A steeper increase in HIV awareness between 2003 and 2005 might partly be attributed to the effectiveness of various interventions, including the UNFPA interventions, which were introduced in the country after 2003. However, as there were other interventions in place in China and the design of the surveys do not allow controlling for them (no controls are available), it is unfortunately impossible to assess the specific effect of the UNFPA intervention in this study.

Figure 3.1 also shows proportions of people who have not heard of STIs or HIV, who have heard of STIs, who have heard of HIV and who have heard of both. Figure 3.1 shows that the levels of awareness about both HIV and STIs gradually increase with time (red line) but there are still people who have heard of HIV but not about STIs (green line) at every point in time. This might suggest that the knowledge of the fact that HIV is a sexually transmitted infection is lacking among some people in the population and HIV is considered being mainly an infection which affects intravenous drug users and people who use blood products or who are blood donors. As mentioned earlier in the section 1.2.2, at the beginning of HIV epidemic in China the main two routes of HIV transmission were through blood products and through intravenous drug use and not through sexual intercourse. Given the fact that by 2007 sexual transmission of HIV was responsible for more than 50% of HIV infections in China (40.6% were infected through heterosexual transmission, and 11.0% through homosexual transmission) (UNGASS 2008) and in 2009 sexual transmission caused 75% of new HIV infections (Wu *et al*. 2010), it is important to increase awareness among the population that HIV is a sexually transmitted infection.

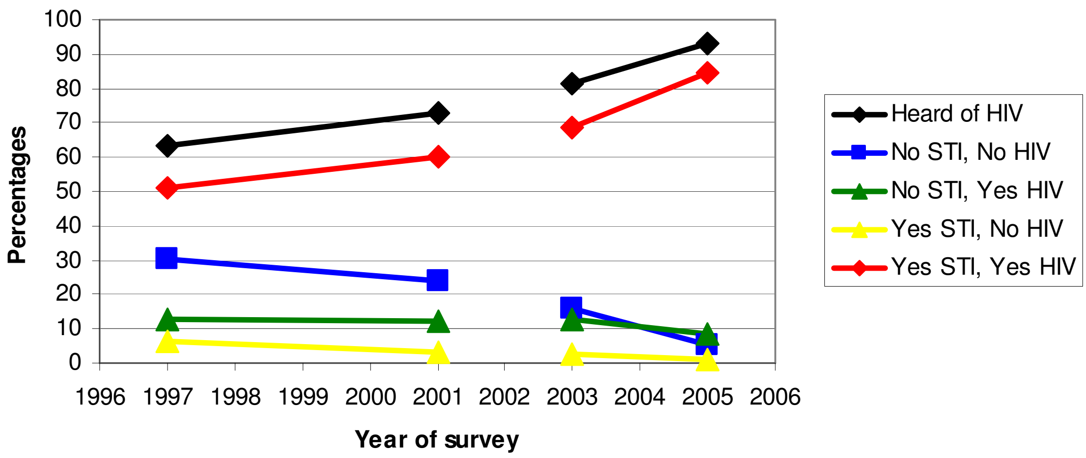


Figure 3.1: Awareness about HIV and STIs in China over time.

Different groups of women have different levels of HIV awareness and these levels also change over time. Figure 3.2 shows that women who belong to Han ethnicity have consistently higher HIV awareness than women who belong to ethnic minorities. This might be explained by the fact that women from Han ethnicity have better access to education and information on a wide range of issues, including HIV. The level of awareness increases quickly among ethnic minorities, especially between 2003 and 2005. Figure 3.2 also shows that the difference between the Han and the ethnic minorities reduced substantially between 2003 and 2005. This change might be explained by the range of interventions in place including the UNFPA interventions as these interventions targeted specific groups of women in China including minority groups.

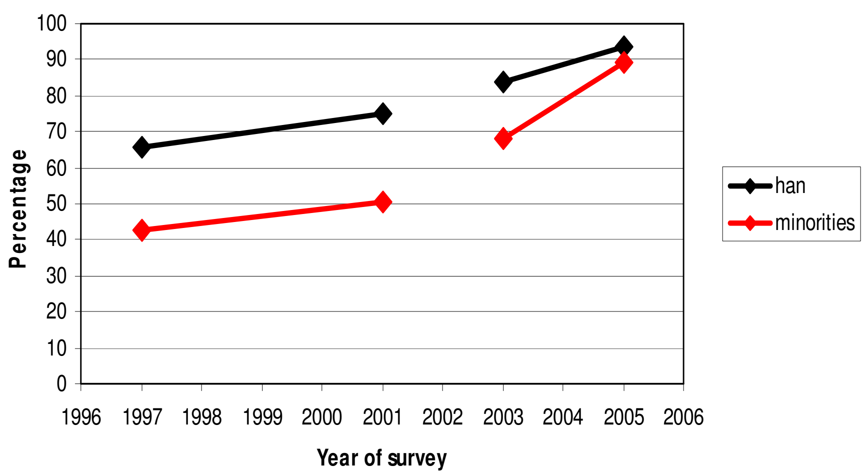


Figure 3.2: HIV awareness by ethnicity in China.

Figure 3.3 shows that urban women had consistently higher level of HIV awareness in comparison to rural women at all times. The gap is reducing with time as the level of awareness among rural women increases quickly between 2003 and 2005. This trend can be explained by the fact that women who live in urban areas have better access to education and information on a wide range of issues, including HIV. The rapid increase of HIV awareness among rural women between 2003 and 2005 might also be explained by the effective interventions including the UNFPA interventions which targeted rural women in their education campaigns. It is of a crucial importance to increase level of awareness among women in rural areas due to the specific feature of HIV epidemic in China (HIV until recently affected more people in rural areas than in urban areas) discussed in section 1.2.2.

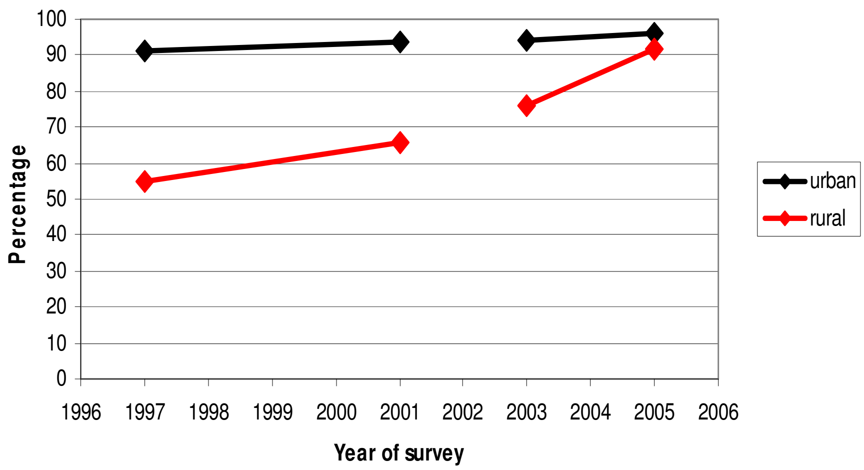


Figure 3.3: HIV awareness by residence in China.

Figure 3.4 shows differences in the level of HIV awareness based on the marital status of women. Figure 3.4 suggests that those divorced and never married women have higher HIV awareness than other groups, whereas widowed women have the lowest HIV awareness when compared to other groups of women. The level of HIV awareness increases consistently over time for all groups of women. The level of HIV awareness among divorced women is consistently high. As it was already high in 1997, it did not increase much over time.

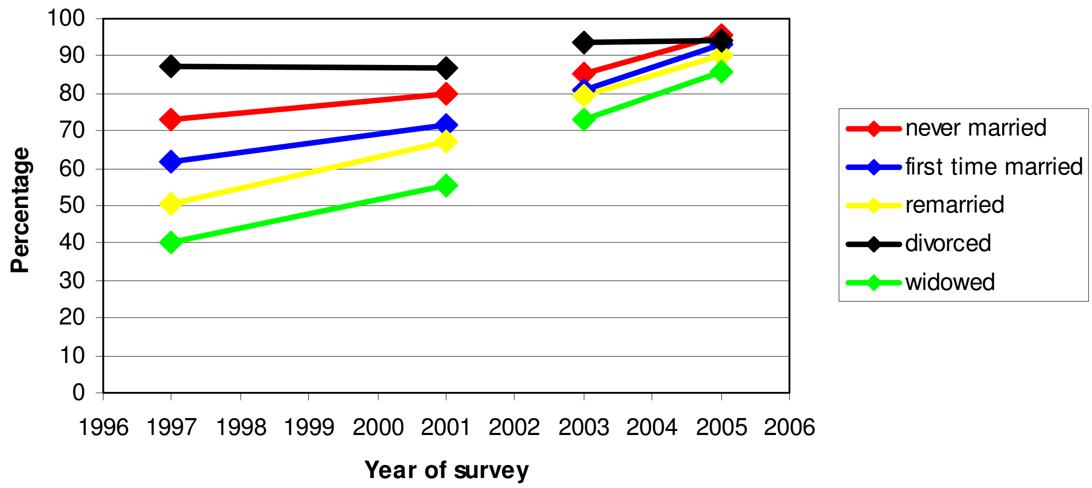


Figure 3.4: HIV awareness by marital status in China.

Figure 3.5[33] shows differences in HIV awareness level for women in different regions of China: Eastern, Central and Western. The level of awareness increases in all three regions over time. The Eastern region has the highest level of HIV awareness compared with the other two regions whereas the Western region has the lowest HIV awareness at all times when compared with other regions. These results are expected. The reason for this difference might be the level of development of the regions and the access to education and information for women in these regions. The Eastern region is the most developed part of China whereas the Western is the least developed one. However, by 2005 there is almost no difference in HIV awareness by regions.

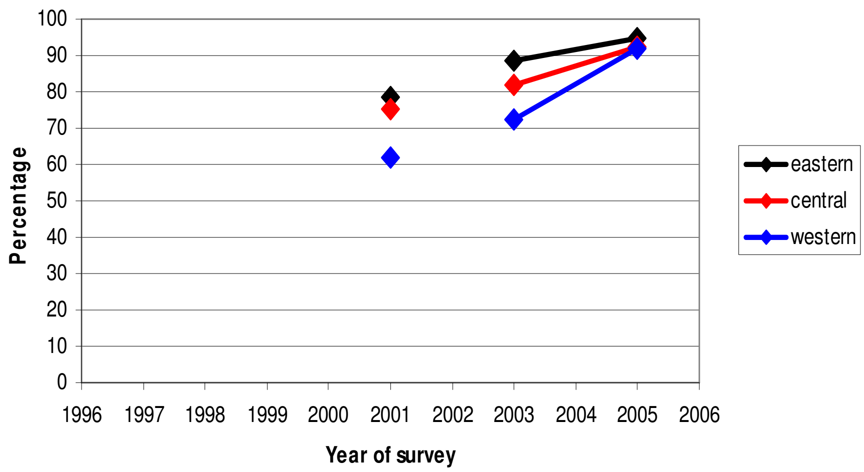


Figure 3.5: HIV awareness by region of residence in China.

[33] No region variable is available in 1997 survey.

Figure 3.6 shows differences in the level of HIV awareness by women's age. The proportion of women from the 20 to 29 years old group who have heard of HIV is the highest when compared with other age groups at all four points in time. Women from the group of 40 to 49 years have the lowest HIV awareness when compared with other groups at all times as well. This can be explained by the fact that those aged between 20 and 29 might be more exposed to sex education related information than people at other ages. Older people are possibly not considered being at risk so the interventions are not focused on this group of people and those people have less chances of being involved in the education and training course unless their work is related to HIV/AIDS.

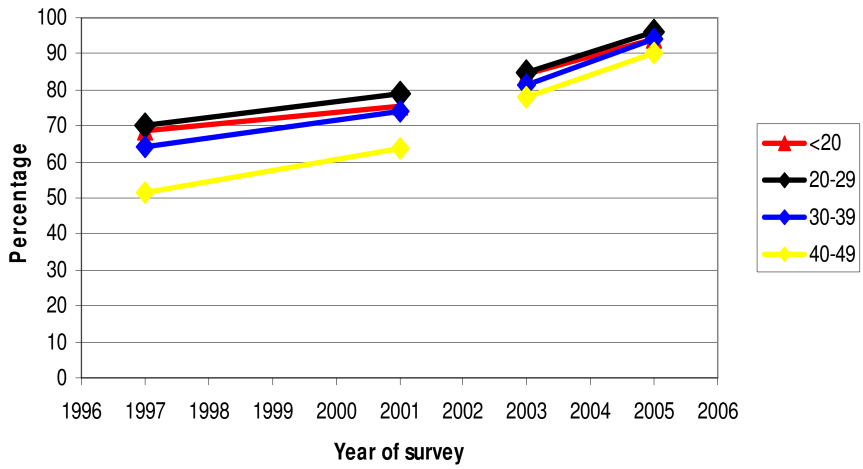


Figure 3.6: HIV awareness by age in China.

Figure 3.7 shows the differences in the HIV awareness by education level. The higher the level of education, the higher the chance that a respondent had heard of HIV/AIDS. The level of awareness increases in all groups with different degrees of increase but it is consistently very high for those respondents with secondary education and above. Given the political attitude towards HIV/AIDS in China before the 2003, it is interesting that the level of awareness was very high even in 1997 for the top group which suggest that some information about HIV/AIDS was already available in the mid-90s in China.

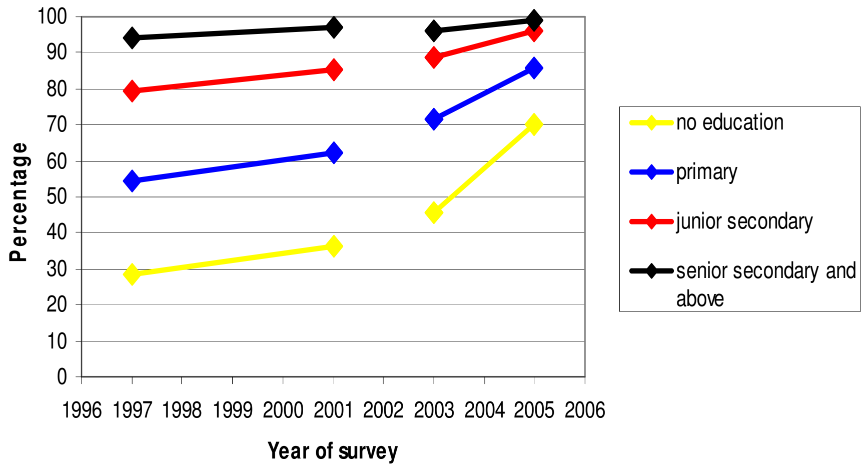


Figure 3.7: HIV awareness by education in China.

Figure 3.8 shows differences in HIV awareness by occupation in 2003 and 2005. HIV awareness increases among all types of occupations in China between 2003 and 2005. The most pronounced increase was observed among women who are involved in agricultural work. For women with intellectual type of work the level of HIV awareness was already very high in 2003 and, therefore, did not increase much between the two points in time.

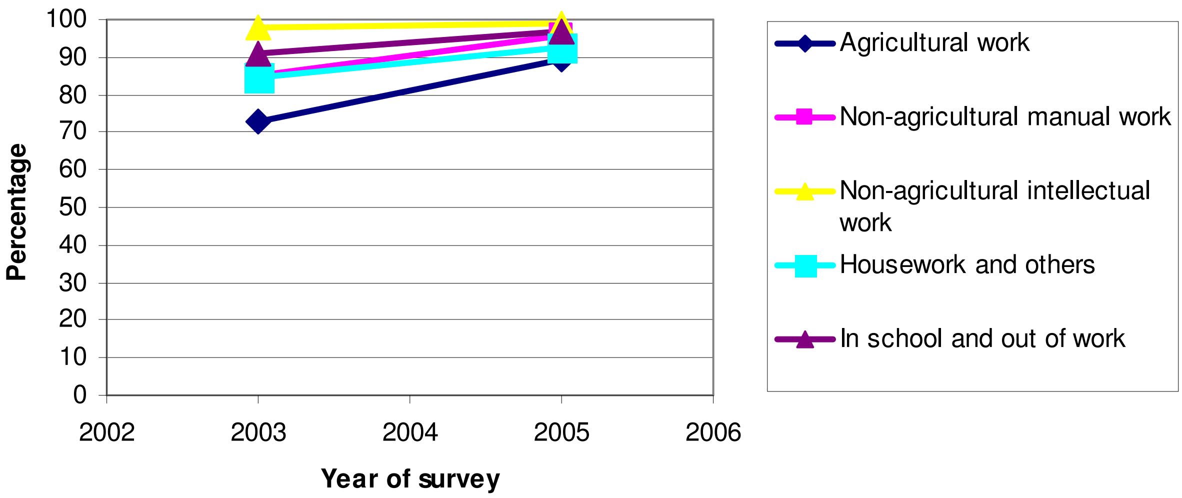


Figure 3.8: HIV awareness by occupation in China.

Figures 3.9-3.11 show HIV awareness by different levels of exposure to TV, radio and newspapers. For all 9 groups of women HIV awareness increases between 2003 and 2005. Women who have been exposed to any of the media regularly have the highest

HIV awareness at both points in time when compared to other levels of exposure. Women who are never exposed to any of three media have the lowest HIV awareness at both points in time when compared to other levels of exposure. Figure 3.10 shows that a substantial increase in HIV awareness is observed among women who are never exposed to TV which suggests that these people obtained information about HIV through sources of information other than TV.

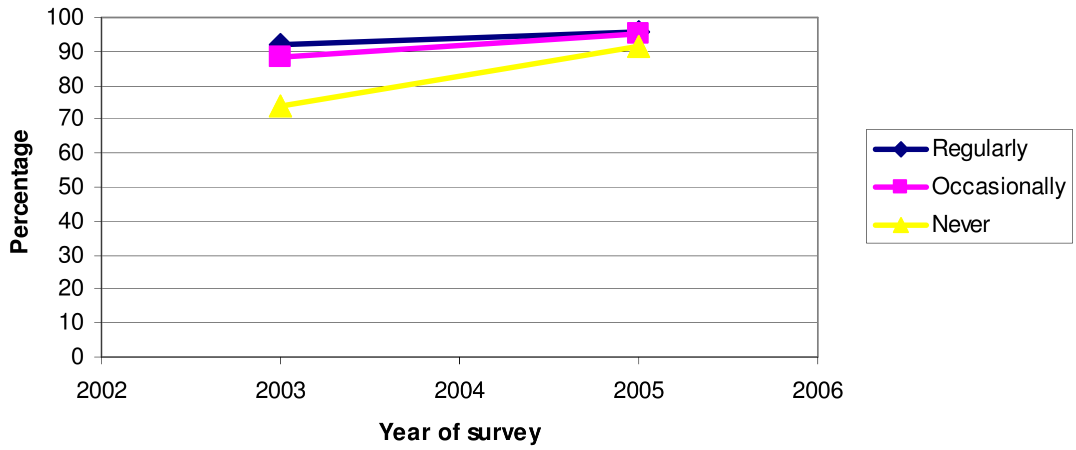


Figure 3.9: HIV awareness by exposure to radio in China.

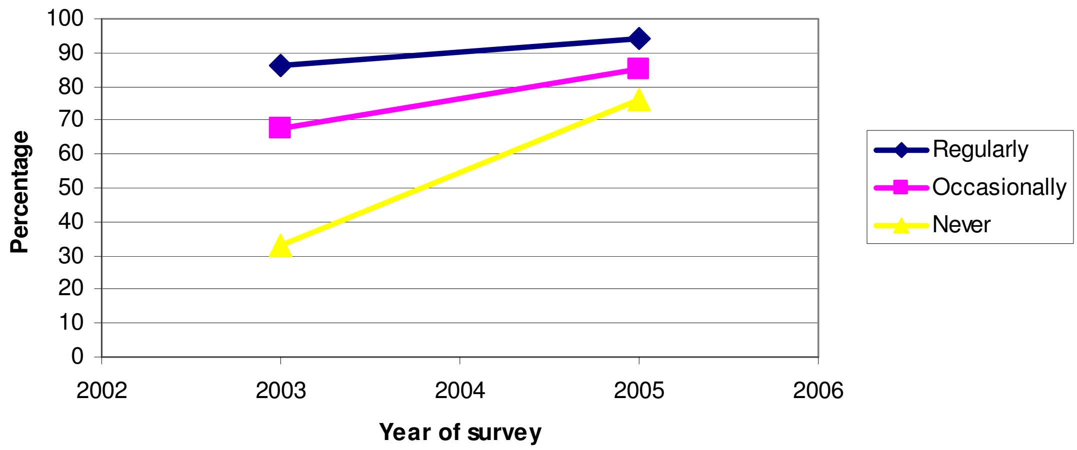


Figure 3.10: HIV awareness by exposure to TV in China.

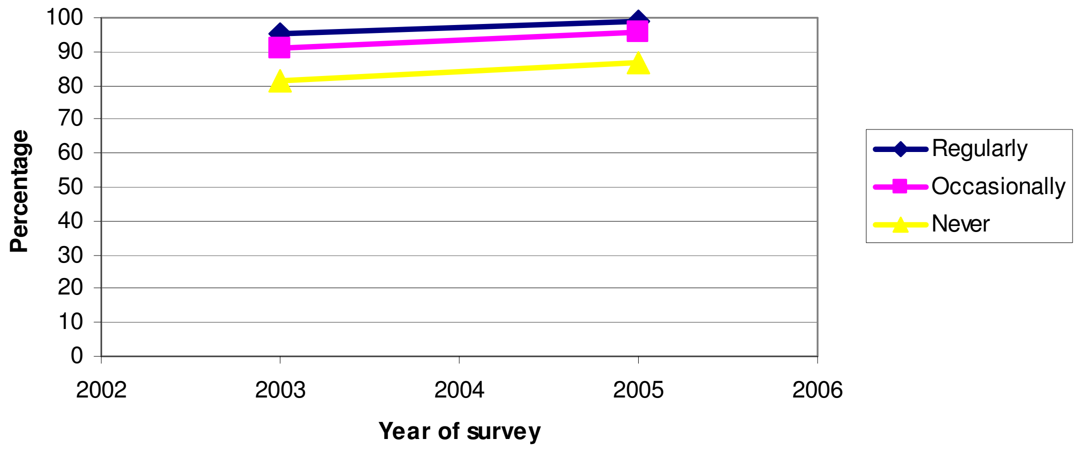


Figure 3.11: HIV awareness by exposure to newspapers in China.

Bivariate associations between the outcome and explanatory variables were explored using Pearson's chi-squared tests. The results of chi-squared tests suggest that in all four surveys used for the analysis all explanatory variables which are included in the modelling are significantly associated with HIV awareness ($p<0.05$).

3.3.2 Logistic regression

After descriptive analysis, it is important to understand if the relationships between the response variable and explanatory variables hold in conditional models when other explanatory variables are taken into account.

The main purpose of the logistic regressions is to enable the decomposition analysis, the results of which are discussed in the next section. However, the results from the logistic regressions also provides useful insights into the evolution of HIV awareness in China.

As mentioned in section 2.4, only a small number of variables are used for the analysis due to the limited number of comparable variables in the two sets of surveys available. For Models 1 and 2 (for 1997 and 2001 surveys respectively) the following variables are used as explanatory variables: residence, ethnicity, age, education and marital status of women; for Models 3 and 4 (for 2003 and 2005 surveys respectively) the same five variables are used for the analysis together with region, occupation and exposure to media variables (TV, radio and newspapers) as these variables are considered as being important in the literature discussed in section 1.3.3 and are available in the both surveys.

Tables 3.1 and 3.2 present results of the logistic regression models for four different surveys. HIV awareness is the outcome variables. In Models 1 and 2 (Table 3.1) the same variables were found to be significantly associated with HIV awareness: respondent's residence, age, ethnicity and education. Marital status was not significant in both models. The fact that the same variables are found to be significant suggests that between 1997 and 2001 there were no interventions which were targeting specific groups of women or if interventions were in place, they were not effective. If targeted interventions were in place and were effective, we would expect HIV awareness to become more homogenous among groups of women and less variables to be associated with HIV awareness in Model 2 when compared to Model 1. Education as expected is found to be an important predictor of HIV awareness among women in both models. The higher the education, the higher the odds of being aware of HIV. Women who belong to Han ethnicity are found to be more likely to be aware of HIV in both models. The same is true for women from urban areas. Slightly different relations between age and HIV awareness are observed at different points in time: in 1997 women of 20-29 have significantly higher probability of being aware of HIV than younger women, whereas in 2001 all women above 20 have higher probability of being aware of HIV than younger women.

Table 3.1: Results of logistic regression (1997 and 2001).

	Model 1 (1997)			**Model 2 (2001)**		
Variable	**β**	**SE**	**OR**	**β**	**SE**	**OR**
Constant	-1.804	0.092		-1.928	0.062	
Residence						
Rural (ref)	0.000		1.000	0.000		1.000
Urban	1.250	0.071	3.490***	0.980	0.047	2.664***
Age group						
<20 (ref)	0.000		1.000	0.000		1.000
20-29	0.465	0.101	1.592***	0.649	0.077	1.914***
30-39	0.170	0.111	1.185	0.534	0.084	1.706***
40-50	0.076	0.114	1.079	0.373	0.086	1.452***
Ethnicity						
Non-Han (ref)	0.000		1.000	0.000		1.000
Han	0.727	0.065	2.069***	0.921	0.040	2.512***
Education						
No education(ref)	0.000		1.000	0.000		1.000
Primary	0.983	0.051	2.672***	0.998	0.034	2.713***
Junior secondary	2.013	0.058	7.486***	2.168	0.039	8.741***
Senior secondary and above	3.002	0.103	20.126***	3.578	0.080	35.802***

Marital status						
Never married(ref)	0.000		1.000	0.000		1.000
Married or remarried	0.015	0.093	1.015	0.071	0.073	1.073
Divorced or widowed	-0.333	0.202	0.717	-0.106	0.126	0.899

***p<0.001; **p<0.01; *p<0.05; ref – reference category.

In Model 3 and Model 4 (Table 3.2) different variables are found to be significantly associated with HIV awareness. In Model 3 it is found that HIV awareness was significantly associated with women's residence, ethnicity, education, marital status, region, occupation, exposure to radio, exposure to TV and exposure to newspapers whereas in Model 4 it is found that HIV awareness is significantly associated only with ethnicity, education and exposure to TV and newspapers. In 2005 significant difference at 5% level in HIV awareness is also observed between women who are 20-29 and younger women, with younger women having lower probability of being aware of HIV than 20-29 year old women. Education as expected is found to be an important predictor of HIV awareness among women in both models. The higher the education, the higher the odds are of being aware of HIV. In both models women who belong to Han ethnicity have higher odds of being aware of HIV. Region, residence, marital status, occupation and exposure to radio effects are significant in 2003 but they are not significant in 2005. The trend in reduction of the number of significant variables from Model 3 to Model 4 suggests that HIV awareness becomes more homogeneous across the 30 counties with the main differential factors being the level of education, ethnicity and exposure to TV and newspapers. These results might suggest that various interventions including the UNFPA interventions which were implemented specifically in these 30 countries and in the whole country which targeted specific groups of women were effective.

Table 3.2: Results of logistic regression (2003 and 2005).

	Model 3 (2003)			Model 4 (2005)		
Variable	**β**	**SE**	**OR**	**β**	**SE**	**OR**
Constant	-0.145	0.233		1.775	0.442	
Residence						
Rural (ref)	0.000		1.000	0.000		1.000
Urban	0.590	0.106	1.804***	-0.218	0.135	0.804
Age group						
<20 (ref)	0.000		1.000	0.000		1.000
20-29	0.033	0.210	1.033	1.061	0.461	2.890*
30-39	0.031	0.227	1.031	0.889	0.479	2.433
40-50 (ref)	-0.033	0.227	0.967	0.526	0.479	1.692
Ethnicity						

Non-Han (ref)	0.000		1.000	0.000		1.000
Han	0.843	0.090	2.323***	0.593	0.134	1.809***
Education						
No education (ref)	0.000		1.000	0.000		1.000
Primary	0.720	0.096	2.054***	0.644	0.149	1.904***
Junior secondary	1.327	0.109	3.770***	1.691	0.173	5.425***
Senior secondary and above	1.835	0.175	6.265***	2.399	0.289	11.012***
Marital status						
Never married (ref)	0.000		1.000	0.000		1.000
Married or remarried	0.768	0.183	2.155***	-0.139	0.439	0.870
Divorced or widowed	0.827	0.329	2.286*	-0.421	0.524	0.657
Region						
Western (ref)	0.000		1.000	0.000		1.000
Eastern	0.391	0.087	1.478***	-0.019	0.128	0.981
Central	0.220	0.083	1.246**	-0.098	0.124	0.907
Occupation						
Agricultural work (ref)	0.000		1.000	0.000		1.000
Non-agricultural manual work	-0.041	0.098	0.960	0.162	0.171	1.176
Non-agricultural intellectual work	1.054	0.278	2.869***	0.424	0.365	1.528
Housework and others	0.348	0.088	1.416***	0.085	0.125	1.089
In school and out of work	0.201	0.206	1.223	0.250	0.315	1.284
Radio						
Regularly (ref)	0.000		1.000	0.000		1.000
Occasionally	-0.100	0.132	0.905	-0.028	0.227	0.972
Never	-0.789	0.123	0.454***			
TV				-0.122	0.206	0.885
Regularly (ref)	0.000		1.000	0.000		1.000
Occasionally	-0.545	0.073	0.580***	-0.711	0.112	0.491***
Never	-1.412	0.168	0.244***	-0.894	0.299	0.409**
Newspaper						
Regularly (ref)	0.000		1.000	0.000		1.000
Occasionally	0.004	0.132	1.004	-1.082	0.290	0.339***
Never	-0.872	0.131	0.418***	-1.542	0.294	0.214***

***p<0.001; **p<0.01; *p<0.05; ref – reference category.

Results of logistic regressions are consistent with the existing literature discussed in section 1.3.3.

3.3.3 Decomposition analysis

Descriptive data analysis suggests that HIV awareness increases over time. A similar pattern was observed in other parts of the world and it is not surprising knowing the amount of effort which was put in educational campaigns and interventions across the world. As mentioned earlier, the change in HIV awareness can be attributed to a number of different reasons such as change in political commitment, interventions, individual change and change in population structure or population characteristics. In order to be able to assess the effectiveness of the efforts directed to the increasing of the level of HIV awareness it is important to find out what the change in HIV awareness can be attributed to. Decomposition analysis can be helpful in identifying components of the total change in HIV awareness over time.

Changes in the level of HIV awareness at different points of time might be partly explained by differences in population structure at different times. Descriptive data analysis suggests that, for example, the percentage of women with no education decreased from 21.4% of women being illiterate in the 1997 dataset, to 16.6% in the 2001 dataset, and also from 9.2% in the 2003 data and to only 4.3% in the 2005 data (Appendix B, Table B.1). This change can be attributed to the policies related to eradication of illiteracy in China[34]. Large difference in the proportions between the two sets of surveys might be attributed to the differences in sample designs between surveys discussed in section 2.1. Another example is the change in female rural-urban composition: in 1997 data 76.6% of women reported being from rural areas, whereas in 2001 only 70.9% of women were from rural areas; 71.4% in 2003 in comparison with only 68.5% of women in 2005 reported being from rural areas of China (Appendix B, Table B.1). This change can be attributed to the process of urbanisation in China. Therefore, differences in the level of HIV awareness can be attributed partially to the variations in population structure. However, due to the short time period between two sets of consecutive surveys, the difference in population structure is not expected to be large between 1997 and 2001 or between 2003 and 2005 and, therefore, changes in population structure may not be the main driver of the change in HIV awareness between two consecutive surveys.

[34] http://www.admissions.cn/education/17668.shtml [Accessed 28 July 2009]

Another part of the change can be explained by the change which happened to people with the same characteristics between two consecutive surveys, e.g. more people with no education are aware of HIV in 2001 than in 1997 or in other words by the change in the relationship or size of the effect between HIV awareness and educational levels due to different factors such as interventions, educational campaigns, change in political commitment and others.

The differences in the predicted probabilities of being aware of HIV between 1997 and 2001, and between 2003 and 2005, which are obtained using the five- and ten-covariate logistic regression models respectively, are decomposed. Tables 3.3 and 3.4 show the main results of the decomposition analysis. Average predicted probabilities during each year of interest can be found along the highlighted diagonal of both tables. When going across the columns, there are the average predicted probabilities of HIV awareness obtained by applying different logistic regression equation coefficients to the given sample. The difference between these two predicted probabilities is due to the change in effect sizes. When moving down the rows, there are average predicted probabilities of HIV awareness obtained by applying the same logistic regression equation coefficients to different samples of women. The difference between these two predicted probabilities is due to the change in population structure. Average predicted probabilities in Table 3.3 are obtained by first estimating logistic regression coefficients for the 1997 sample and then applying 1997 logistic regression equation to the 2001 data. The processes discussed above are repeated to complete the table: first 2001 logistic regression coefficients are obtained and then they are applied to 1997 data. The same process is conducted to obtain average predicted probabilities for 2003 and 2005 data for the Table 3.4.

Table 3.3: Decomposition analysis (1997 and 2001).

	1997 βs	**2001 βs**
1997 xs	**0.635**	0.700
2001 xs	0.660	**0.727**

Table 3.4: Decomposition analysis (2003 and 2005).

	2003 βs	**2005 βs**
2003 xs	**0.812**	0.912
2005 xs	0.837	**0.929**

Tables 3.3 and 3.4 present main results of decomposition analysis and show that the average predicted probabilities of HIV awareness increases with time as expected. The total absolute change in HIV awareness between 1997 and 2001 is 0.092 or 9.2 percentage points. The total change in HIV awareness between 2003 and 2005 is 11.7 percentage points. The change which can be attributed to the change in population structure between 1997 and 2001 is 2.5 percentage points, and between 2003 and 2005 is also 2.5 percentage points. As expected, this change is not large due to the short period of time between the two consecutive surveys and not enough time for the population structure to have substantially changed. The rest of change can be attributed to the change in relationships between HIV awareness and explanatory variables over time. Between 1997 and 2001 this change is 6.7 percentage points and between 2003 and 2005 it is 9.2 percentage points. The change due to change in relationships between HIV awareness and explanatory variables is larger than the change due to change in population structure and, therefore, this change due to change in relationships is the main driver of the overall change in HIV awareness in both sets of surveys.

Detailed decomposition analysis

In order to identify which coefficients are driving the change in effect sizes between HIV awareness and characteristics of women, detailed decomposition analysis is conducted. The results of the detailed decomposition analysis are presented in Appendix C, Tables C.1-C.2 and Figures 3.12-3.13. Figure 3.12 show change in relationships or effect sizes between 1997 and 2001 by the five main demographic characteristics of Chinese women. Figure 3.12 presents predicted probabilities of women being aware of HIV for specific groups of women which are obtained by applying the two different logistic regression equation coefficients (1997 and 2001) to the 2001 sample. No groups experienced substantial changes in awareness between 1997 and 2001 and changes are more or less evenly spread across different groups with slightly bigger changes observed in groups which were lagging behind (rural women, women with no education and primary education and women between 30 to 49 years). These relatively similar and not substantial changes can be attributed to the fact that the government was denying the existence of the problem at that time and, therefore, not much was happening in the field of HIV prevention. Greater change among rural women (Figure 3.12a) might be attributed to the information about dangerous blood donation in rural parts of China which existed at that time. The average probability of being aware of HIV is greater for women who are married or remarried and widowed and divorced in 2001 than in 1997 when compared with women who were never married. This might be explained by the fact that information regarding HIV was available for women mainly through family planning services (Fang and Kaufman 2008) and never married women were not using them.

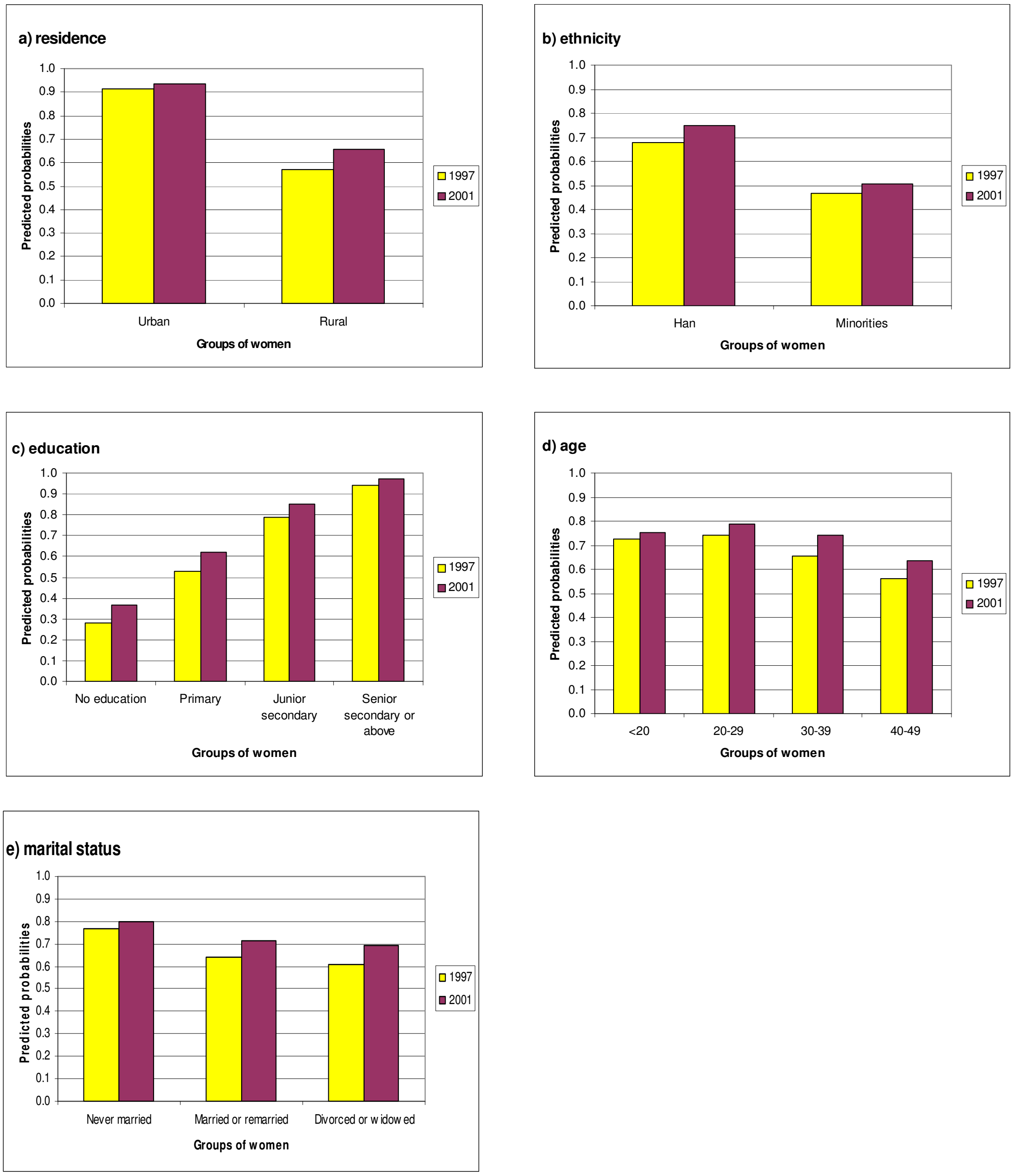
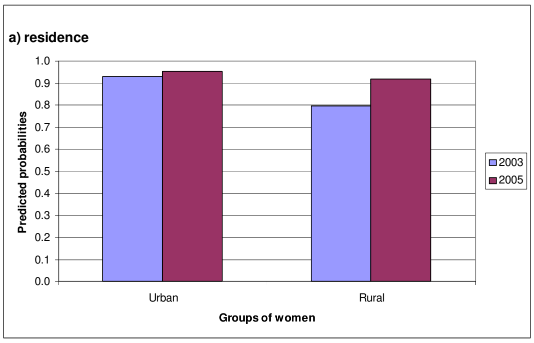
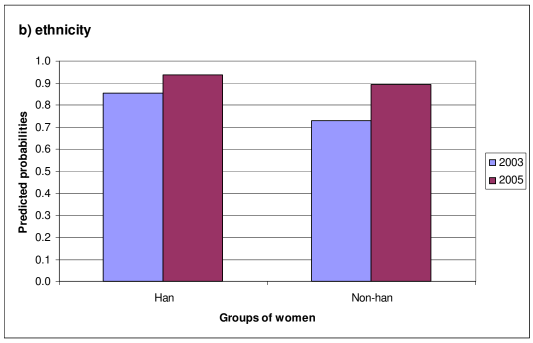


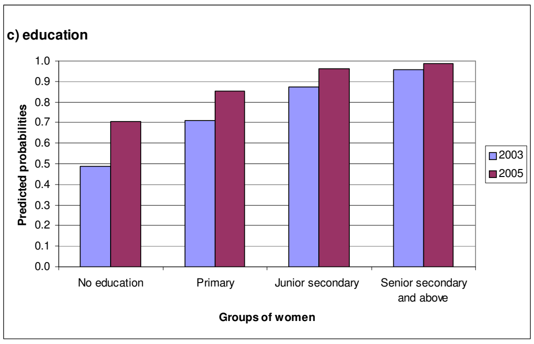
Figure 3.12: Predicted probabilities of being aware of HIV by different demographic characteristics of women obtained by applying different logistic regression coefficients to the 2001 sample.

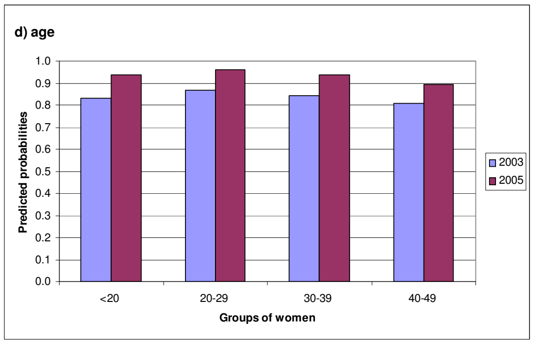
Figure 3.13 presents predicted probabilities obtained by applying the two different logistic regression coefficients (2003 and 2005) to the 2005 sample. This figure shows change in effect sizes between 2003 and 2005 by the ten main demographic characteristics of Chinese women in 30 purposefully selected counties in China. These figures show that substantial change in HIV awareness between 2003 and 2005 happened to rural women (Figure 3.13a), to women who belong to minorities (Figure 3.13b), to those with no education (large change of 22 percentage points) and with primary education (Figure 3.13c) when compared to other categories of the variables. Roughly similar change for all age groups might suggest that the interventions were not targeting specific age groups of women. With time, lower awareness groups are catching up and the gaps between the groups are closing. The fact that the government changed its attitude towards HIV and many interventions and campaigns started in the country in the year 2003 might have contributed to the change in effect sizes. The progress in HIV awareness is more pronounced among specific groups of women mentioned above and might suggest the effectiveness of interventions which were in place in all China as well as specific interventions designed for these 30 counties selected for the survey, including the UNFPA interventions discussed in section 1.2.5. (As mentioned earlier, the main target by UNFPA interventions for HIV awareness were women who were lagging behind in terms of knowledge, specifically, unmarried young women, rural women, and women who live in the Western region of China.) Figure 3.13f shows the change in size effects between 2003 and 2005 by regions. We can see the biggest change in effect sizes in the Western region when compared with the other two regions which also might be attributed to governmental programmes and campaigns as well as to the UNFPA interventions, as special attention was directed to women from Western part of the country, as was mentioned earlier in section 1.2.5.

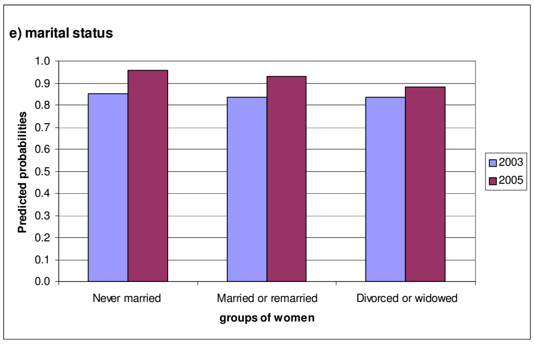
A substantial change in HIV awareness between these two years also happened to women who reported to have no exposure to radio, TV and newspaper (Figure 3.13h-j). This finding suggests the effectiveness of interventions other than ones which were introduced at the local and country levels to reach women through methods of mass media.

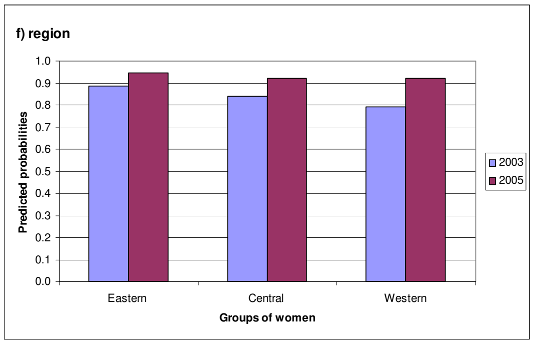

a) residence
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Urban
Rural
Groups of women
2003
2005

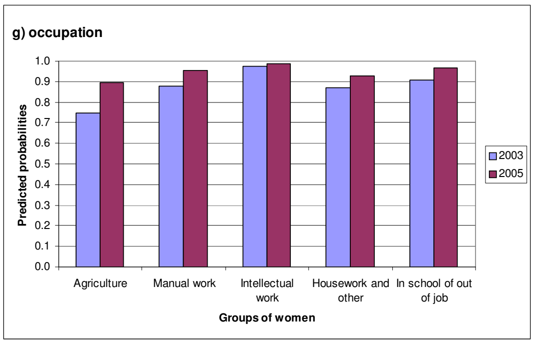

b) ethnicity
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Han
Non-han
Groups of women
2003
2005

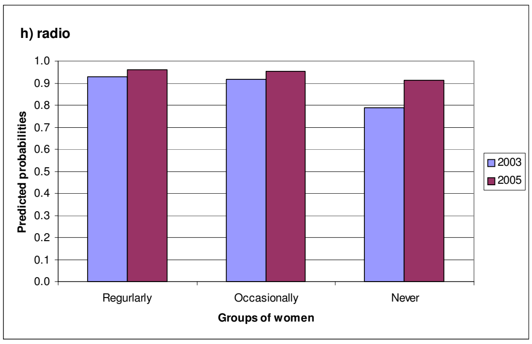

c) education
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
No education
Primary
Junior secondary
Senior secondary and above
Groups of women
2003
2005

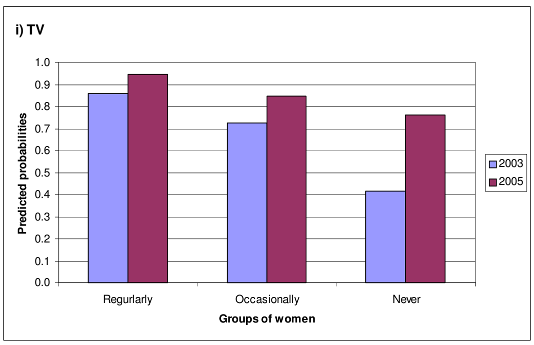

d) age
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
<20
20-29
30-39
40-49
Groups of women
2003
2005


e) marital status
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Never married
Married or remarried
Divorced or widowed
groups of women
2003
2005


f) region
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Eastern
Central
Western
Groups of women
2003
2005


g) occupation
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Agriculture
Manual work
Intellectual work
Housework and other
In school of out of job
Groups of women
2003
2005


h) radio
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Regurlarly
Occasionally
Never
Groups of women
2003
2005


i) TV
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Regurlarly
Occasionally
Never
Groups of women
2003
2005

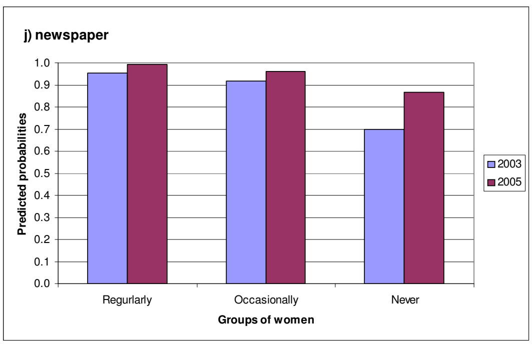

j) newspaper
Predicted probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Regurlarly
Occasionally
Never
Groups of women
2003
2005

Figure 3.13: Predicted probabilities of being aware of HIV by different demographic characteristics of women obtained by applying different logistic regression coefficients to the 2005 sample.

3.3.4 Confidence intervals

It is important to produce confidence intervals and to assess if differences in fitted probabilities of being aware of HIV for 1997 and 2001 models and for 2003 and 2005 models are significant or not. The four predicted probabilities which are of the main interest of the decomposition analysis are not normally distributed (see Table 3.5). Therefore, the test of difference between two means for normal distributions is not appropriate. In order to overcome the problem of non-normality of distributions, bootstrap technique, discussed in section 3.3.4, is used. The bootstrap method was used to estimate confidence intervals for the probabilities of being aware of HIV in order to determine if observed changes in probabilities of being aware of HIV at different points in time are statistically significant or if they are observed simply by chance.

Table 3.5: Means and medians of the predicted probabilities of being aware of HIV for different models used in decomposition analysis.

Probability	Mean	Median
$\overline{P}(\hat{\beta}_{1997} x_{2001i})$	0.660	0.723
$\overline{P}(\hat{\beta}_{2001} x_{2001i})$	0.727	0.804
$\overline{P}(\hat{\beta}_{2003} x_{2005i})$	0.843	0.885
$\overline{P}(\hat{\beta}_{2005} x_{2005i})$	0.929	0.966

The confidence intervals constructed using bootstrap suggest that almost all differences in probabilities of being aware of HIV for the whole samples as well as for sub-samples for different demographic characteristics of women between 1997 and 2001 as well as between 2003 and 2005 are significant with the exception of widowed or divorced women for both sets of surveys, women who were in the age group below 20 years old for the years 1997 and 2001 and women who reported having an intellectual job for the years 2003 and 2005. Constructed confidence intervals are presented in Appendix E, Table E.1.

3.3.5 Comparison of change in HIV awareness between the two studied periods (1997-2001 and 2003-2005)

It is expected that the change will be larger between 2003 and 2005 than between 1997 and 2001 due to change in political agenda, and interventions and campaigns which were introduced in the supportive political environment after 2003.

Figure 3.14 shows the differences between the changes in effect sizes at two points in time for specific groups of women for two sets of surveys. This figure helps to study crude differences between two sets of surveys. When conducting this comparison it is important to bear in mind that different variables were used for the two decomposition analyses (region, occupation and exposure to media variables are missing from the analysis of 1997 and 2001 surveys) and the two sets of surveys represented different populations. However, the main purpose of this graphical investigation is to show the general pattern in the level of change of the size of the effects between the two sets of consecutive surveys. As expected, Figure 3.14 shows substantial differences for specific sub-groups of women between the two sets of surveys. For all groups of women except women with senior secondary education and above and divorced and widowed women, the change between 2003 and 2005 was higher than between 1997 and 2001. This figure shows substantial difference in changes for women from non-Han backgrounds and from rural areas, for women with no or primary education, for women of 15-29 years and for women who are never married which happened between two sets of surveys. The findings suggest that more changes happened to these specific groups between 2003 and 2005 than between 1997 and 2001 which might in turn suggest effectiveness of interventions which were introduced in the country after 2003. The results might also suggest that either never married women started being included in reproductive health services provision or they were targeted specifically through some interventions. As mentioned earlier, young people were specifically targeted by the UNFPA interventions as they were least aware of HIV. Figure 3.14 shows that for women below 20 the change is substantially larger between 2003 and 2005 than between previous years which might suggest that some of the existing interventions, including the UNFPA interventions, might have reached younger women and succeeded in increasing the HIV awareness among this group.

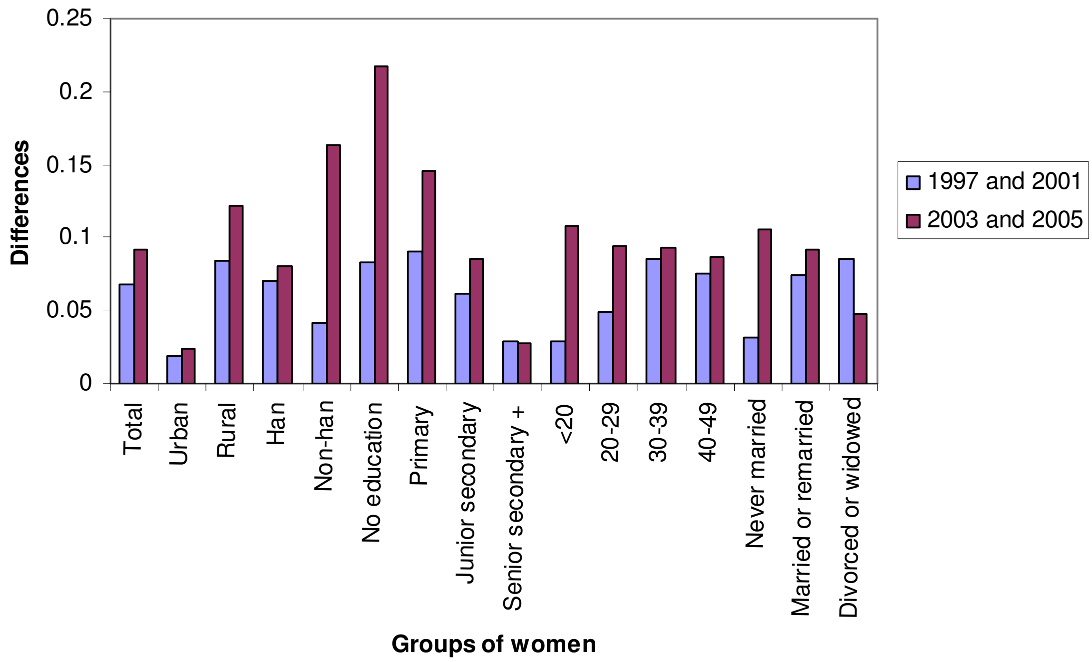


Figure 3.14: Differences between changes in effect sizes at two points in time for specific groups of women for two sets of surveys.

It can be speculated that Figure 3.14 shows the effect of interventions in 30 counties selected for the CP5 together with other efforts directed at the increase of HIV awareness in the whole China after 2003. If interventions and campaigns were not effective, we would have expected more or less the same changes being observed across all groups. If efforts and measures triggered by the political commitment to fight HIV were not effective, we would have expected the level of change to have been less substantial between 2003 and 2005 and to have been similar to the level between 1997 and 2001.

3.4 Limitations of This Study

The main limitation of the results obtained through the decomposition analysis is the small number of variables which were available for the analysis, especially for the period between 1997 and 2001. It is possible that there are other factors which can explain variation in HIV awareness but were not included in the logistic regression models due to the lack of data. However, despite this limitation the results yielded a better understanding of the evolution of HIV awareness in China. Also, the decomposition analysis is a good technique which allows separating different components that contribute to the total change in HIV awareness between different points in time. The analysis suggests the importance of political commitment, and the importance of educational campaigns and interventions for the increase in the level of HIV awareness.

Another limitation is that the 2003 and 2005 surveys only represented 30 counties of China which suggest that results obtained for 2003 and 2005 surveys cannot be generalised to the whole China. However, the results obtained are indicative of the possibility that if the right interventions are in place, the level of HIV awareness can be substantially improved within a short period of time.

The 2001 survey does not have self-weighting designs but the datasets do not contain weights. However, the absence of weights in the datasets should not introduce a problem for the study as the main interest of the study is to establish relationships.

Absence of access to the national survey which was conducted in 2006 represents another limitation of the study as the comparison between 1997, 2001 and 2006 would provide more up-to-dated results of the HIV awareness in China, and it would be possible to produce results generalisable to the whole country.

Finally, another limitation of the study is that it is not possible to isolate the specific effects of the UNFPA interventions due to the limitations of data (absence of control observations).

3.5 Conclusions

This paper analysed HIV awareness among Chinese women between 1997 and 2005. The analysis confirms that China has responded well to generating population level awareness of HIV in a short period of time. The results of the study suggest that the situation with HIV awareness in China is similar to other parts of the world (Ingham 1995) with the level of HIV awareness increasing over time. Education remains one of the main factors which are associated with HIV awareness, the other main factors HIV awareness is associated with in 2005 are ethnicity, exposure to TV and newspapers. These results are consistent with the literature discussed in section 1.3.3.

A smaller part of the observed change in HIV awareness over time is due to change in population structure but a larger part is due to change in environment such as in political commitment, successful interventions, and health promotion campaigns. However, part of the change in HIV awareness between the two sets of surveys analysed in this paper is also due to the differences in representativeness levels of surveys and sampling techniques used for their collection.

With time, lower awareness groups are catching up and gaps between groups with initially different awareness levels are closing. The UNFPA interventions discussed in section 1.2.5 together with other interventions and programmes introduced in China are effective in closing the gaps between groups between 2003 and 2005 as suggested by the larger increase in HIV awareness in specific groups of women between 2003 and 2005 than between 1997 and 2001. The increases observed between 1997 and 2001 are more evenly spread between groups of women with different demographic characteristics whereas between 2003 and 2005 increases are more pronounced among specific groups of women. These differences also suggest the existence and the effectiveness of interventions which targeted specific groups of women between 2003 and 2005.

The fact that the government changed its attitude towards HIV in 2003 and many interventions and campaigns started in the country in the year 2003 might have contributed to the change in HIV awareness among women in China. The analysis suggests the importance of political commitment, and the importance of educational campaigns and interventions for the increase in the level of HIV awareness.

The results of the analysis also indicate the importance of increasing awareness among the population in China that HIV is a sexually transmitted infection.

Decomposition analysis provided a useful tool for assessing the effect of changes in environment using cross-sectional data. It also allowed the assessment of macro-level change in HIV awareness in China over time by isolating the change in HIV awareness which was not due to change in individual demographic characteristics of women or, in other words, not due to changes in population structure which are observed over time.

It could be speculated that increased HIV awareness among different groups of women in China might have contributed to the adoption of protective behaviours such as avoidance of illegal blood donation, increased rate of condom use and others which in turn might have influenced the decrease in speed of spread of HIV infection across the country. However, in order to establish these links, further work needs to be conducted.

Box 3.1: Key findings in Paper One

- HIV awareness has increased in China over time in different groups of women.
- Education, ethnicity and exposure to TV and newspapers are associated with HIV awareness in China.
- Gaps between groups with initially different HIV awareness levels are narrowing over time and lower awareness groups are catching up with higher HIV awareness groups.
- A smaller part of change in HIV awareness between 1997 and 2001 and between 2003 and 2005 is due to the change in population structure.
- A larger part of change in HIV awareness between 1997 and 2001 and between 2003 and 2005 is due to the change in effect sizes which can be explained by the changes in environment due to changes in political commitment and effectiveness of interventions.
- Decomposition analysis is a useful technique which helps to disentangle different components of change in HIV awareness and allows the assessment of macro-level change in HIV awareness in China over time.
- Various interventions introduced in the whole China and specifically in the 30 counties, including the UNFPA interventions, proved effective in improving HIV awareness of specific groups of women in China between 2003 and 2005.
- Results suggest the need to improve awareness that HIV is a sexually transmitted infection.

Chapter 4: Measuring HIV Knowledge: A Simple Score Approach

Abstract

This paper focuses on the evolution of HIV knowledge among women in China. The aim of this paper is to assess whether China has succeeded in improving women's HIV knowledge between 1997 and 2005, and if China is a relative success story in improving women's HIV knowledge when compared to other countries in the world with generalised as well as with non-generalised HIV epidemics. The following data sources are used for the analysis: the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001, the UNFPA Reproductive Health and Family Planning Survey 2005, and recent DHS surveys (India 2006, Kenya 2003, Malawi 2004 and Ukraine 2007). In order to assess whether China has succeeded in improving women's knowledge of HIV, HIV knowledge is first compared in China across times. To establish relative success of China in improving HIV knowledge, HIV knowledge in China is then compared with HIV knowledge in India, Kenya, Malawi and the Ukraine. A simple score approach is used to measure HIV knowledge in this paper. A partial proportional odds modelling technique is employed for the analysis of patterns of HIV knowledge in the Chinese context and in the five country context. The main findings indicate that China has succeeded in improving women's HIV knowledge. Knowledge about ways to prevent HIV transmission in China is comparable to the levels in other countries with both generalised and non-generalised epidemics. The levels of knowledge about misconceptions about HIV transmission in China are comparable to those in the countries with non-generalised epidemics. The HIV knowledge in China has become more homogeneous over time as it improved among different groups of women in the country. However, the gap in knowledge between different groups still exists and, therefore, more efforts should be directed towards improvement of knowledge about HIV among women in China.

4.1 Introduction

4.1.1 Simple score measures of HIV knowledge in literature

Data on HIV knowledge are collected frequently and consistently as a part of larger surveys which sometimes also cover other areas such as reproductive health and family planning (DHS surveys). HIV knowledge data are also collected in surveys which are either interested in just HIV knowledge or in HIV knowledge in combination with HIV attitudes and behaviour. The latter can be used for conducting analyses within, for example, Knowledge, Attitudes, Behaviour (KAB) framework which was first used in the context of the evaluation of national family planning programmes (Mauldin 1967). One of the examples of KAB surveys in China is the survey called "AIDS-related knowledge, attitudes, behaviour, and practice: A 2008 survey of 6 Chinese cities" (CHAMP 2008). HIV knowledge data have been and are collected in countries and in regions where the HIV epidemic is generalised as well as in countries and regions where HIV prevalence is still relatively low and where HIV does not affect general populations. Some surveys target general populations (for example, DHS surveys), whereas in some surveys specific groups of people such as high-risk groups (intravenous drug users, commercial sex workers) or professions who deal with people living with HIV (doctors, nurses) or groups which are believed to be at higher risk of HIV infection (STI clinics patients, young people) are targeted for the purpose of studying their HIV knowledge. HIV knowledge has various components and cannot be collected as a single variable and, therefore, it is important to review how HIV knowledge data are used and how such data have been analysed so far in previous studies. There are two main approaches to measuring HIV knowledge which exist in literature: a so-called simple score approach and a latent variable approach. This paper focuses on simple score approach to measuring HIV knowledge. This section will review the simple score approach to measuring and analysing HIV knowledge in the existing literature. The second approach to measuring HIV knowledge will be reviewed, applied and discussed in Chapter 5 of the thesis.

In their book on sexuality-related measures, Davis *et al*. (1998) devoted a section to different HIV/AIDS measures including measures of HIV knowledge and some of the measures from the book are discussed below. Carey *et al*. (1997) presented a review of different measures of HIV knowledge which were used in the literature. The measures reviewed in the two publications mentioned above, together with other studies which used HIV knowledge simple score measures, are discussed below.

In some studies different routes of HIV knowledge are studied separately (Ingham 1995; Lau *et al*. 2002; Chen *et al*. 2003a; Benotsch *et al*. 2004; Wong *et al*. 2008), their main focus is on descriptive statistics about knowledge of different routes of HIV transmission in a given population (Lau *et al*. 2002). However, in the majority of studies a scoring approach to measuring HIV knowledge is adopted. While constructing a score, the majority of studies include only questions about HIV knowledge. However, some studies include broader questions from the area of reproductive health into the score (Cai *et al*. 2008), but there are not many studies which use a broader measure. In some studies HIV knowledge is measured as a combination of answers to questions about correct and incorrect routes of HIV transmission together (Ambati *et al*. 1997; Carey *et al*. 1997; Lagarde *et al*. 1997; Zimet 1998; Ford *et al*. 2000; Braithwaite and Thomas 2001; Barden-O'Fallon *et al*. 2004; Li *et al*. 2004c; Hesketh *et al*. 2005; Wade *et al*. 2006; Tan *et al*. 2007). However, other studies either create two or even more separate measures, for example, one for HIV prevention knowledge (or knowledge of correct routes of HIV transmission) and another one for knowledge about HIV misconceptions (Fitzgerald *et al*. 2003). Some studies are concerned only with one type of HIV knowledge, for example, studies which focus on stigma and discrimination and would only collect data on knowledge of misconceptions about HIV/AIDS (Chen *et al*. 2007; Genberg *et al*. 2008; Genberg *et al*. 2009).

The majority of studies use a simple score approach (Barden-O'Fallon *et al*. 2004; Lou *et al*. 2006; Stephenson 2009; Feldman *et al.* 2011). They add up correct answers to HIV knowledge questions in order to calculate a score of HIV knowledge for each respondent. Studies which use a simple score approach to measuring HIV knowledge use either a true/false scale (Ambati *et al*. 1997; Zimet 1998; Lau *et al*. 2002; Loue *et al*. 2003; Pinkerton *et al*. 2003; Barden-O'Fallon *et al*. 2004; Benotsch *et al*. 2004; Li *et al*. 2004c; Loue *et al*. 2004; Hu *et al*. 2006; Wu *et al*. 2007a; Stephenson 2009) or a Likert-type scale when questions about HIV knowledge are presented in a form of attitudinal questions (Fisher *et al*. 1996; Fisher *et al*. 2002). One point is usually given for a correct answer when a true/false scale is used; incorrect answers as well as answers "do not know" are usually given no points (Lagarde *et al*. 1997; Carey *et al*. 1998; Zimet 1998; Ford *et al*. 2000; Lau *et al*. 2002; Chen *et al*. 2003a; Pinkerton *et al*. 2003; Barden-O'Fallon *et al*. 2004; Hu *et al*. 2006; Wu *et al*. 2007a; Cai *et al*. 2008). In the Wang and Keats (2005) study, they asked 28 questions on HIV knowledge and the maximum score was 56, but they do not provide any explanation on how the score was created. When a Likert-type scale is used, there is a possibility of having different answers considered as correct (for example, either “agree” or “strongly agree” can be considered as being a correct answer) (Shrum *et al*. 1998; Snell *et al*. 1998; Li *et al*. 2004c). For a Likert-type scale, correct answers are either given one point or every possible answer has certain number of point assigned to it (Koch and Singer 1998; Luquis and Koch 1998; Shrum *et al*. 1998; Braithwaite and Thomas 2001). In some studies both scales are used together depending on a specific question (Brown and Bocarnea 1998; Koch and Singer 1998; Koopman and Reid 1998; Tan *et al*. 2007). In some of the studies reviewed those respondents who answered “no” to a filter question - if they had ever heard of HIV/AIDS - scored 0 on HIV knowledge as their HIV knowledge is assumed to be inadequate (Lagarde *et al*. 1997), in other studies only women who were asked HIV knowledge questions were included in analysis.

In the Snelling *et al*. (2007) study, the researchers used a different approach to producing an HIV knowledge score. The score on knowledge of specific routes of HIV transmission was created on the basis of six indicator variables, then these scores for individual women were used to derive score distributions for each country included in the analysis, and then these country-level scores were used for analysis.

Once a score variable is calculated, it is either used as a continuous variable (Lagarde *et al*. 1997; Benotsch *et al*. 2004; Stephenson 2009) or it is converted into a categorical variable with two or more categories such as “low score” and “high score” (Loue *et al*. 2003; Li *et al*. 2004c) for the analysis. In the Chen *et al*. (2007) study they used an HIV knowledge variable as a binary categorical variable which was coded 1 if a respondent ticked five or more listed misconceptions about HIV/AIDS and 0 otherwise.

In all studies reviewed above a higher score was regarded as an indicator of better HIV knowledge, so the main assumptions in all these studies is the higher the score, the better the HIV knowledge.

Many studies, especially those published in Chinese sources, performed only descriptive analysis of separate routes of HIV transmission or of the obtained score (Wang *et al*. 2001; Chen *et al*. 2003a; Chen *et al*. 2003c; Zhang *et al*. 2006; Zhou *et al*. 2007; Wong *et al*. 2008), but other studies use HIV knowledge score for further statistical modelling

(Pleck 1998; Snelling *et al*. 2007). In some studies HIV knowledge is used as a response variable in order to determine predictors of a good HIV knowledge in different contexts (Manchester 2002; Chen *et al*. 2003a; Dong *et al*. 2003; Li *et al*. 2004c; Zhang *et al*. 2006; Tan *et al*. 2006; Wu *et al*. 2007a; Tan *et al*. 2007; Tan 2008; Feldman *et al*. 2011). Linear and logistic regression models are fitted to data depending on the HIV knowledge variable being continuous or categorical. In other studies the HIV knowledge variables were used as predictors of different outcomes. For example, HIV knowledge scores were used as a predictor variable for AIDS-related stress in the Pleck (1998) study. HIV knowledge score was also used as a predictor variable for ever using a condom to protect against STDs including HIV/AIDS variable and ever having sex with only one partner to protect against AIDS variable in the Snelling *et al.* (2007) study, for HIV-related stigma and HIV testing (Do and Guend 2009), for multiple sexual partnership (Buszin *et al*. 2009) and others.

Some studies used HIV knowledge measures in order to assess effectiveness of interventions and collected HIV knowledge data at two points in time (baseline survey and follow-up survey) (Misovich *et al*. 1998; Tian *et al*. 2007).

The study by Feldman *et al.* (2011) examined trends in HIV knowledge over time. In this study trends in HIV knowledge were assessed among Israeli army releases between 1993 and 2005 using simple score approach to measuring HIV knowledge.

As mentioned earlier, no studies examined the changes in HIV knowledge in the general population in China over time or compared the level of knowledge in China to those in other countries and, therefore, the current research will fill these gaps as it aims to study changes in HIV knowledge in the general population of China over time and to compare HIV knowledge in China with HIV knowledge in India, Kenya, Malawi and the Ukraine.

As HIV knowledge is an important prerequisite and component in the HIV risk reduction framework discussed in section 1.3.1 and important component of effective HIV prevention strategies, it is important to know how the levels of HIV knowledge were changing over time in the general population in China. In order to validate the levels and changes in HIV knowledge in China, a comparison is made with other countries in the world such as India, Kenya, Malawi and the Ukraine. These comparisons might provide useful insights into understanding of the pace and the evolution of HIV knowledge in China.

4.1.2 Aims, research questions and structure of Paper Two

This paper aims to:

1. assess change in HIV knowledge in China across time (1997-2005);
2. compare HIV knowledge in China with other countries with generalised (Kenya and Malawi) and with non-generalised HIV epidemics (India and the Ukraine);
3. compare different simple score measures of HIV knowledge.

The main research questions are:

1. Has China succeeded in improving women's knowledge of HIV/AIDS over time?
2. Is China a relative success story in improving women's knowledge of HIV/AIDS when compared with other countries with generalised and non-generalised HIV epidemics?
3. What are advantages and limitations of different simple score measures of HIV knowledge?

In order to assess the level of improvement of HIV knowledge in China, general patterns of HIV knowledge among two specific groups of women (women with potentially the best and potentially the worst knowledge) will be assessed and compared. To compare the levels of HIV knowledge across time in China and across countries and to place China in an international context, two main assumptions are introduced. It is assumed that levels of HIV knowledge are higher in generalised epidemics than in countries where the epidemic is in its initial stages (Snelling *et al.* 2007) due to time, extensive educational campaigns and targeted interventions. Therefore, it is expected to observe the highest HIV knowledge scores in Kenya and Malawi in comparison with other countries included in the current analysis. It is also assumed that the higher levels of HIV knowledge about both ways to prevent HIV and misconceptions are observed in countries with generalised HIV epidemics but higher levels of HIV knowledge about ways to prevent HIV and lower levels of knowledge about misconceptions are observed in countries with non-generalised HIV epidemics. This can be explained by the order of the focus of interventions appearing in different countries: at the beginning of HIV epidemic the focus is on the HIV prevention and knowledge of correct routes of HIV transmission, once the epidemic progresses and more people are affected by the virus, interventions regarding the reduction of stigma and discrimination against people living with HIV are introduced and, therefore, the level of knowledge about misconceptions or incorrect routes of HIV transmission increases. Therefore, it is expected that in Kenya and Malawi the levels of both types of knowledge about HIV are high, whereas in China, India and the Ukraine, the levels of knowledge about ways to prevent HIV are higher than the levels of knowledge about misconceptions about HIV.

It is hoped that the results of the analysis could be helpful for interventions and programmes in order to be able to identify gaps in knowledge, and to identify low knowledge groups and to be able to make better use of limited resources for prevention programmes. According to Carey *et al*. (1997), reliable assessment of HIV knowledge will "allow for more precise evaluation of programmes designed to reduce risk-conferring behaviours" (p.73). Results of the analysis might help to inform and strengthen policies and programmes in China and in other countries of the world.

It is also hoped that the comparison of different score options can help researchers make decisions about more appropriate measures of HIV knowledge as well as other health related concepts, for example, knowledge about cot death or knowledge about resistance to antibiotics and other types of health-related knowledge.

This paper consists of six main sections. In section two different HIV knowledge simple score measures used for the analysis are discussed. In section three the methodology used for the analysis is presented. In section four all the results of the analysis are presented and summarised. In section five, the limitations of this study are discussed. Finally, in section six the main conclusions are presented.

4.2 HIV Knowledge Scores

The approach used in this paper to measure HIV knowledge will be referred to as a simple score approach. It has been widely used in different studies (see section 4.1.1). A set of four scores is created for both research contexts: the Chinese context for the analysis of HIV knowledge in China over time and the five country context for the comparison of HIV knowledge in different cultural and epidemiological contexts. In order to assess the level of HIV knowledge in China and other countries included in analysis, as well as to compare different simple score measures, the following measures for HIV knowledge are created: Score One for HIV knowledge of correct routes of HIV transmission in the Chinese context and for knowledge about ways to prevent HIV in the five country context; Score Two for incorrect routes of HIV transmission in China and for knowledge about HIV misconceptions in the five countries; Score Three for combined HIV knowledge in both contexts and finally, Score Four which groups different options of a score for combined HIV knowledge in both contexts. Variables which were used to derive different scores in both contexts are discussed in detail in section 2.3.

Respondents are given one point for every correct answer to knowledge of correct routes of HIV transmission or HIV prevention questions. Respondents were given 0 for an incorrect answer or for a "do not know" answer (those who answered "do not know" were assumed to have inadequate knowledge about the specific component(s) of HIV knowledge). The highest possible score for Score One is four and this score indicates the perfect knowledge of the correct routes of HIV transmission in the Chinese context or perfect knowledge about ways to prevent HIV in the five country context. The highest possible score for Score Two is three and this score indicates the perfect knowledge of incorrect routes of HIV transmission in China or perfect knowledge about misconceptions about HIV in the five countries. The highest possible score for Score Three is seven and this score indicates perfect overall HIV knowledge in both study contexts. Score Four was created in order to have a more manageable categorical variable for combined HIV knowledge as Score Three has eight response categories. Score Four has four response categories. Category one combined respondents who either had no knowledge about HIV or answered correctly to 1 component of the combined HIV knowledge. This category was labelled as "poorest knowledge". Category two combined respondents who answered correctly to 2 or 3 components of the combined HIV knowledge. This category was labelled "bad knowledge". Category three combined respondents who answered correctly to 4 or 5 questions. This category was labelled "ok knowledge". And the fourth category combined respondents who answered correctly to 6 or to all questions. This category was labelled "best knowledge".

Score One and Score Two help to conduct comparison of separate types of knowledge in China over time as well as between countries. Score Three or Score Four help to conduct comparison of combined HIV knowledge in China across times as well as in different countries.

4.3 Methodology

4.3.1 Scores

HIV knowledge cannot be observed or measured directly but it can be measured with the help of the HIV knowledge components or indicators (Figures 1.2, 2.1 and 2.2) which can be measured directly such as knowledge of specific routes of HIV transmission or knowledge about ways to prevent HIV. In order to measure HIV knowledge, a scale or score, which is based on the information derived from several questions or indicators, is created. Creation of a score simplifies the analysis as it is easier to analyse just one variable instead of many variables (De Vaus 2002) and allows studying HIV knowledge as a concept rather than separate components of it. An HIV knowledge score consists of answers to a number of questions and for each question every individual receives a score depending on their answer, the scores for each question then is added up together to provide an overall score for every individual (De Vaus 2002). The highest possible score indicates that this specific individual has a perfect HIV knowledge whereas the lowest score (0) indicates the absence of an adequate HIV knowledge. The overall score is an ordinal variable (a variable where there is an obvious natural ordering of categories (Agresti 2002)) and, therefore, tests of associations for ordinal data should be used as an exploratory technique, and ordinal logistic regression can be used for modelling.

A number of studies treated HIV knowledge score as a continuous variable (Lagarde *et al*. 1997; Benotsch *et al*. 2004; Stephenson 2009). This approach has a number of limitations as it assumes a normal distribution for the residuals as well as equal distances between different score options (for example, distance between score 1 and score 2 should be equal to the distance between score 4 and score 5). These assumptions are very strong and, therefore, treating score variables as ordinal provides a better framework for the analysis of HIV knowledge scores.

4.3.2 Exploratory data analysis

Bivariate associations: Test of associations for ordinal data

Associations between score variables and explanatory variables in different temporal and epidemiological contexts can be visually explored using bar charts and stacked bar charts. Stacked bar charts can be used for the analysis as the sum of percentages of different score options adds up to 100% in every dataset that is used for the analysis.

According to Agresti (1996), the Pearson's chi-squared test of association is inappropriate when one of the variables is ordinal. In the situation when one of the variables is ordinal and the second one is categorical but without ordered categories, the most useful test is the Kruskal-Wallis test which is a generalisation of the Wilcoxon rank-sum test (Mehta and Patel 1998). This test does not assume normality and, therefore, can be used for ordinal variables. The test is similar to the F-test in a way that it does not tell us how groups are different from each other, it only tells us if they are different or not.

Under the Kruskal-Wallis test the null hypothesis (H_0) is that there is no association between the two variables, or in other words that all populations' distribution functions are equal; the alternative hypothesis (H_A) is that at least one of the populations differs from at least one of the other populations (Mehta and Patel 1998). A high value of the Kruskal-Wallis statistic and an associated low p-value (e.g., $p<0.05$) provide the evidence for rejection of the H_0, and suggest that there is an association between the two variables.

When both variables are ordinal, there are two other tests which can be used to test for bivariate associations: the Jonckheere-Terpstra test and the Linear-by-linear association test (Mehta and Patel 1998). H_0 and H_A are the same as above and conclusions from the test results are drawn similarly.

4.3.3 Logistic regression

When a response variable has more than two categories the choice of the logistic regression depends on the type of categories in the response variable, if they are ordered or unordered. If categories in the response variable have natural ordering, ordinal logistic regression models such as the proportional odds model or the partial proportional odds model can be used for modelling the response variable. Multinomial logistic regression is a more general form of logistic regression and can be used for modelling of both unordered and ordered (when the assumptions of ordinal logistic regression do not hold) response variables. The proportional odds approach uses a smaller number of parameters and, therefore, has more compact representation and this represents the main advantage of the proportional odds model over the partial proportional odds model or multinomial logistic regression approach. However, the proportional odds models are more restrictive when compared to other models such as multinomial logistic regression models, and may not fit the data well (Fox 2002). Soon (2010) conducted the assessment of all three models (proportional odds model, partial proportional odds model and multinomial logistic regression model) on the basis of parsimony in parameters and flexibility in assumptions and concluded that "the partial proportional odds model compromises between the parameter parsimony and assumption flexibility features" (p. 95). Proportional odds model scores high on parsimony in parameters but scores low on flexibility of assumptions, whereas multinomial logistic regression in contrast scores high on flexibility of assumptions but scores low on the parsimony in parameters (Soon 2010). All three models will be discussed in detail in the sections below.

4.3.3.1 Proportional odds model

Ordinal logistic regression is a special case of logistic regression when outcome variable is an ordinal variable. The cumulative logit model is a special case of ordinal logistic regression (McCullagh 1998). The proportional odds model is a special case of a cumulative logit model. The main assumption of this model is the proportional odds assumption, which implies that for different values of *j*, the equations are identical except for the intercept, so the difference between the logits for different values of *j* is constant, resulting in a constant odds ratio (Fox 2002).

The proportional odds model can be expressed as:

$$\text{logit}\{P(Y \leq j)\} = \log\{\frac{P(Y \leq j)}{P(Y > j)}\} = \alpha_j - (\beta_1 x_1 + \cdots + \beta_p x_p), \qquad (4.1)$$

for j=1, …. , R – 1, where R is the number of ordered categories ordinal variable Y has, j is a category of ordinal variable Y, $\alpha_1 < \alpha_2 < \cdots < \alpha_{R-1}$ are intercepts or cutpoint parameters which are not usually of the direct interest (Agresti and Finlay 1997; Agresti 2002).

This model assumes that a variable effect (β) on the odds of response below category j is the same for all j. A unit increase in x multiplies the odds of response being less than or equal to j by exp (β). Each cumulative logit has its own intercept and the model has the same effect β for each cumulative logit (Agresti 2002).

The negative sign of the linear predictor (equation (4.1)) is a convention ensuring that high values of the linear predictor lead to increased probability for the higher categories[35]. However, this might vary in different software packages as some of them use a positive sign which can make the interpretation less intuitive.

According to Fox (2002), proportional odds model may not fit the data well as the main assumption of the model, proportional odds assumption, is quite restrictive. According to Agresti (1996), if the ordinal logistic regression does not fit the data well, the following changes to the model might improve the model fit and might provide a better model:

1. inclusion of interaction or quadratic (if appropriate) terms;
2. using alternative link functions other than logit link;
3. using a multinomial logistic regression.

If the proportional odds model does not fit the data well, before considering using a multinomial logistic regression, a partial proportional odds model should be considered (Peterson and Harrell 1990; Ananth and Kleinbaum 1997; Williams 2006). This model is more parsimonious than multinomial logistic regression and, therefore, has a number of advantages when ordinal variables are modelled. The partial proportional odds model is discussed in detail in section 4.3.3.2.

In order to perform the model diagnostics for proportional odds models, SPSS or STATA software packages can be used. SPSS provides a tool which is called the test of parallel lines. This test helps to assess if the parameters are the same for all categories. This test compares a proportional odds model with one set of coefficients with a multinomial logistic regression model with separate sets of coefficients for all categories. The null hypothesis (H_0) for the test is that the parameters are the same for each category; the alternative hypothesis (H_A) states that the parameters are not the same for all categories and, therefore, the model does not fit the data well. A low p-value ($p<0.05$) suggests that the parameters are not the same for all categories, that the model does not fit the data well and that modifications to the model mentioned above should be applied in order to improve the model fit.

[35] http://www.norusis.com/pdf/ASPC_v13.pdf [Accessed 30 July 2009]

The Brant test is available in STATA to assess if the proportional odds assumption holds or not. This test "provides both a global test of whether any variable violates the parallel-lines assumption, as well as tests of the assumptions for each variable separately" (Williams 2006, p. 61). In order to find out for which slopes the proportional odds assumption do not hold, the Brant command can be used in STATA (Brant 1990; Ellis *et al*. 2007; Fotso *et al*. 2009). The test compares "the separate (correlated) fits to the binary logistic models[36] underlying the overall model" (Brant 1990, p. 1171). In other words it tests the parallel line assumption for each variable separately by comparing the slope coefficients for separate binary logistic regressions (Long and Freese 2006). The Brant command returns a chi-squared statistic for the test with associated p-values. The H_0 for this test is that the parallel lines assumption holds for the specific explanatory variable, whereas the H_A is that the parallel lines assumption does not hold for this variable. When the test statistic is significant and the associated p-value for the specific variable is low ($p<0.05$), it suggests that the parallel lines assumption has been violated for this specific variable and that one set of coefficients or just one model is not enough to describe relationships between the response and an explanatory variable and, therefore, different coefficients for different levels of response variable are required.

Proportional odds models are first fitted to the two pooled datasets in order to investigate determinants of HIV knowledge at different points of time in China as well as in different cultural and epidemiological contexts.

4.3.3.2 Partial proportional odds model

When the test for parallel lines in SPSS or Brant test in STATA suggest that the H_0 (the parameters are the same for each category) should be rejected and that the parameters are not the same for all categories and that the proportional odds model does not fit the data well, there is a need to consider less restrictive model which is called partial proportional odds model (Peterson and Harrell 1990; Ananth and Kleinbaum 1997; Williams 2006). As mentioned earlier, this model is less restrictive than proportional odds model but more parsimonious than multinomial logistic regression model (multinomial logistic regression model is discussed in detail in section 4.3.3.3). The

[36] According to Brant (1990), the proportional odds model can be represented by a series of binary logistic regressions.

main advantage of the partial proportional odds models over multinomial logistic regression models is that they are still treating the response variable as an ordinal variable, whereas the multinomial model completely discards the ordered nature of the response variable and, therefore, leads to the loss of efficiency of the model (Peterson and Harrel 1990; Fotso *et al.* 2009). The partial proportional odds model takes into account ordinal nature of the response variable but also allows for the violation of the parallel lines assumption for some of the slopes. Or in other words, it relaxes the parallel lines assumption for a subset of slopes (Soon 2010). According to Williams (2006), multinomial logistic regression models estimate more parameters than necessary and this can "cause some effects to become statistically insignificant" (p. 66). Therefore, multinomial logistic regression should be used with caution in the context of ordinal response variable and only when necessary. By using partial proportional odds models which estimate non-proportional parameters only for covariates for which the parallel line assumptions did not hold, the problem of estimation of unnecessary parameters can be overcome.

An example of the partial proportional odds model, in which slopes for x_1 and x_2 variables are the same for all values of j but for x_3 they are free to vary, can be expressed as:

$$\log \frac{P(Y=j)}{P(Y=R)} = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \beta_{3j} x_3 \, , \tag{4.2}$$

j=1,2,…,R-1, where R is number of categories of the ordinal response variable and j is a category of the ordinal response variable.

The probability of an individual i being in a category j can be calculated as

$$P(Y_i > j) = \frac{\exp(\alpha_j + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{3i}\beta_{3j})}{1+\exp(\alpha_j + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{3i}\beta_{3j})}, \tag{4.3}$$

j=1,2,…, R-1, where R is number of categories of the ordinal response variable and j is a category of the ordinal response variable (Williams 2006).

Partial proportional odds models can be fitted to data using the STATA gologit2 command (Williams 2006). This command is a user-written programme and can be used for modelling ordinal response variables. The gologit2 command can fit all three models: proportional odds model, partial proportional odds model and multinomial logistic regression model (Williams 2006).

Not many researchers use partial proportional odds when model ordinal response variables. Examples of studies which used this method are Ellis *et al*. (2007), Fotso *et al*. (2009), Soon (2010). Fosto *et al*. (2009) used the partial proportional odds model to investigate factors associated with use of appropriate maternal health services in slums in Kenya. Ellis *et al*. (2007) used the model to investigate male baldness. Soon (2010) studied the determinants of students' return intentions with the help of partial proportional odds models. According to Ananth and Kleinbaum (1997) and Soon (2010), the partial proportional odds models have been underutilized by researchers. Given the advantages of the model when compared to both proportional odds model and multinomial logistic regression, this method should be used wider in the context of ordinal response variables.

The partial proportional odds method is employed to modelling of HIV knowledge in both pooled datasets when the parallel line test or Brant test for the proportional odds model rejects H_0. Only if the parallel line test fails for every slope or in other words when there is a need to relax the parallel line assumption for all slopes, the multinomial logistic regression model will be considered and employed.

4.3.3.3 Multinomial logistic regression

If the response variable has R categories, then the multinomial logistic regression can be expressed as a set of $R(R\text{-}1)$ possible logistic model equations. However, only R-1 is necessary and the rest are redundant. For the reference category the most meaningful category should be chosen. The multinomial logistic regression can be expressed as:

$$\log \frac{P(Y = j)}{P(Y = R)} = \alpha_j + \beta_{1j} x_1 + \cdots + \beta_{pj} x_p, \quad (4.4)$$

where j=1,…, R-1, R is the reference category placed last and p is the number of variables in the model.

Multinomial logistic regression can be specified in terms of probabilities which are better for interpretational purposes. If the third category is the reference category then logit models for multinomial logistic regression are:

$$\log(p_1 / p_3) = \alpha_1 + \beta_1 x_1 \tag{4.5}$$

$$\log(p_2 / p_3) = \alpha_2 + \beta_2 x_2 \tag{4.6}$$

On probability scale they can be expressed as:

$$p_1 = \frac{e^{\alpha_1 + \beta_1 x}}{1 + e^{\alpha_1 + \beta_1 x} + e^{\alpha_2 + \beta_2 x}} \tag{4.7}$$

$$p_2 = \frac{e^{\alpha_2 + \beta_2 x}}{1 + e^{\alpha_1 + \beta_1 x} + e^{\alpha_2 + \beta_2 x}} \tag{4.8}$$

$$p_3 = \frac{1}{1 + e^{\alpha_1 + \beta_1 x} + e^{\alpha_2 + \beta_2 x}} \text{ (the probability for the reference category)} \tag{4.9}$$

As mentioned earlier, multinomial logistic regression can be used to model ordinal response variables if the proportional odds model and partial proportional odds model do not fit the data well (Agresti 1996; Williams 2006).

Multinomial logistic regression will only be fitted to the two pooled datasets if needed in order to be able to investigate differences in patterns of HIV knowledge at different years of surveys for different groups of women in the Chinese context and to investigate differences in patters of HIV knowledge in different countries.

4.3.4 Model selection

Forward selection procedure is used to select the final proportional odds models. Forward selection is a process of model selection which starts from a simple model. Explanatory variables are added one by one (the most significant ones), followed then by adding interactions between the explanatory variables. The selection process stops when either satisfactory goodness of fit is reached or the improvement in fit is not significant any more.

Forward stepwise model selection process starts as a forward selection but then alternates between backward elimination and forward selection and the process continues until no terms left which can meet either entry or removal criteria (SPSS 2007b). Forward stepwise automatic model selection process is employed within multinomial logistic regression framework to identify all main effect and all interactions which are entered into the final models. Once final models are identified in SPSS, they are then transferred to STATA which will allow obtaining more parsimonious results for the final models within partial proportional odds models' framework by placing constrains on some of the slopes for which proportional odds assumption hold.

Likelihood ratio tests (using the change in the L^2 goodness-of-fit statistic) often termed the deviance for proportional odds models and multinomial logistic regression models is used to test significance of model terms and to inform the decision about the inclusion of a term or an interaction in the final model. According to Fox (2002), proportional odds models are not properly nested and, therefore, likelihood ratio tests cannot be performed properly, but they still can be useful while comparing relative fits of the models.

4.3.5 Interactions

Potential interaction effects should always be investigated in any study during the model selection process as they could be important in describing the associations between variables and, if not included when needed, the results of the analysis might not be reliable. Another reason for testing for interactions is that interactions might improve the fit of the proportional odds model and, therefore, are used during the modelling process if needed (Agresti 1996). Potential interactions are investigated in this paper.

The SPSS software package version 16 (SPSS 2007a) and STATA software package version 9.2 (STATA 2005) are used for the analysis presented in this paper.

4.3.6 Random effects

The effect of clustering should be accounted for when observations are not independent because the data have structure either due to the data collection process or due to the natural structure which might exist within the population. If this hierarchical structure is not accounted for, standard errors for the model parameters can be underestimated. As a result of this, "estimates of effects may appear to be significant when in fact they are not" (Madise *et al.* 1999, p.333). The effect of clustering can be controlled for by the two main methods: random effect modelling can be employed or corrections to standard errors can be applied (robust standard errors can be calculated in, for example, STATA). Robust standard errors estimation procedure is available in STATA for both ordinal and multinomial logistic regression. However, unfortunately, this procedure is not available within partial proportional odds modelling framework.

4.4 Results

Demographic and other characteristics of women together with their HIV knowledge in seven separate datasets used for the analysis in this paper are presented in Appendix F.

4.4.1 Exploratory data analysis

In this section the results of the descriptive analysis and the results of the inferential bivariate tests will be presented and discussed.

4.4.1.1 Levels of knowledge of separate routes of HIV transmission or separate ways of HIV prevention by years of survey and countries

In this section knowledge about individual components of HIV knowledge scores is studied separately.

The Chinese context

Figures 4.1 and 4.2 show trends in the two types of HIV knowledge among Chinese women: knowledge about correct and about incorrect routes of HIV transmission. These figures suggest that the levels of knowledge about separate routes of HIV transmission increase with time. Figure 4.1 shows that knowledge of correct routes of transmission, especially for some routes such as sexual transmission, was already high in 1997 among women who were aware of HIV. It also shows that the knowledge about the sexual route of HIV transmission is the highest in comparison to the knowledge about the other correct routes of HIV transmission in China at all times. The knowledge about blood transfusion is also high at all times which is not surprising due to the specific feature of HIV epidemic in China related to blood donation discussed in section 1.2.2. Surprisingly, the level of knowledge about needle sharing is slightly lower than the knowledge about sex or blood transfusion. This route of HIV transmission was expected to be better known among women because of the IDUs epidemic in the Southern China (discussed in section 1.2.2). The level of knowledge about mother to child transmission (MTCT) is still slightly lower than knowledge about sex and blood transfusion routes of HIV transmission at every point in time but it can be explained by the fact that at the beginning of epidemic more males are affected by HIV than females and, therefore, not many children are born to women who are HIV positive.

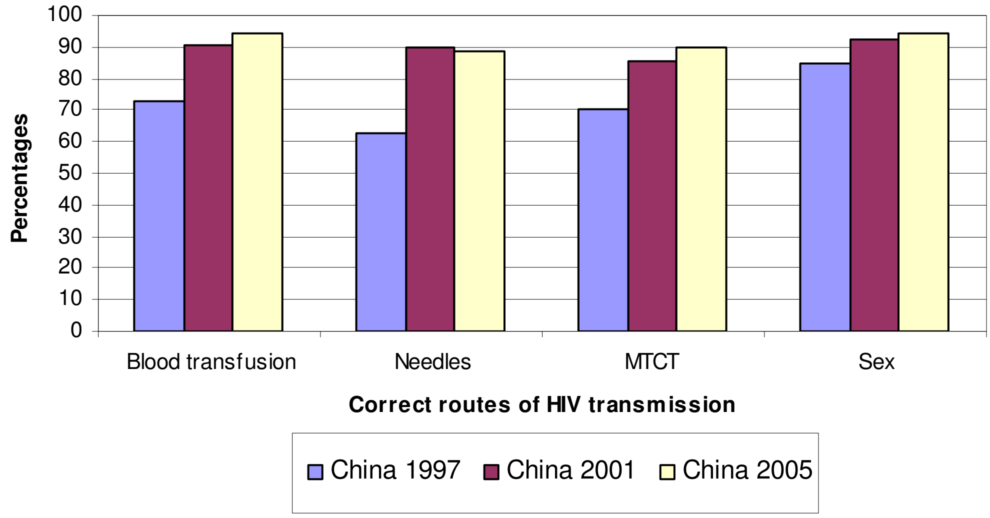


Figure 4.1: Knowledge of correct routes of HIV transmission over time in China[37].

Figure 4.2 suggests that the levels of knowledge about the incorrect routes of HIV transmission in 1997 were quite low (below 50%). Even in 2005 the level of knowledge about incorrect routes of HIV transmission is lower than the level of knowledge about correct routes of HIV transmission (see Figure 4.1). However, this finding can be explained by the early stage of HIV epidemic in China.

[37] Differences in representation levels for three Chinese surveys used for the analysis should be taken into account when results are interpreted hereafter.

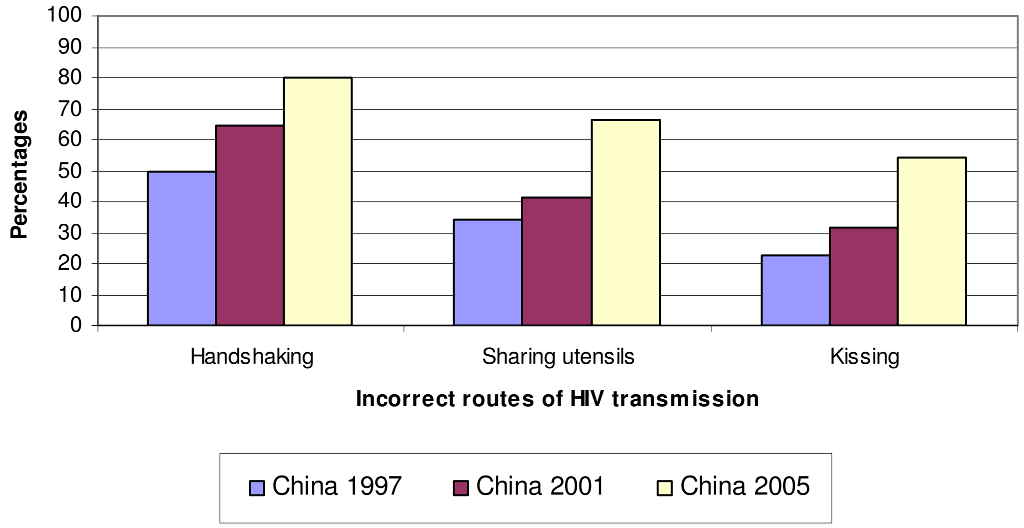


Figure 4.2: Knowledge of incorrect routes of HIV transmission over time in China.

The five country context

Figures 4.3 and 4.4 show proportions of women who gave correct answers to the separate questions about ways of HIV prevention and about misconceptions respectively in the five country context. Both figures suggest that the levels of knowledge about separate ways of HIV prevention and about misconceptions vary across countries.

The levels of knowledge about MTCT, condom use and one partner in China are relatively high. However, the level of knowledge about abstinence is relatively low. Figure 4.3 shows that levels of knowledge about condom use are quite high in the Ukraine and China but substantially lower in India and, surprisingly, relatively low in the two countries where HIV epidemics are generalised and the main route of HIV transmission is through heterosexual intercourse (Kenya and Malawi). Levels of knowledge about one partner, abstinence and MTCT are also relatively low in Malawi. These findings suggest the need in sustainable interventions to increase knowledge about ways to prevent HIV transmission even in countries where HIV epidemics are generalised, as knowledge about HIV is an important pre-requisite in strategies directed at the reduction of risky behaviours.

The level of knowledge about the misconception of healthy-looking people is high in China but the level is not as high as for the knowledge about sharing utensils. Additionally, the level is very low for the knowledge about mosquito bites in China when compared with other countries in the five country context. As expected, the level of knowledge about misconceptions in Kenya and Malawi is high.

These results suggest that the level of knowledge in China about separate ways to prevent HIV is comparable with, and on some occasions higher than in countries where the HIV epidemic is generalised. The level of knowledge about misconceptions is comparable with other countries where the HIV epidemic is non-generalised.

The results of this part of descriptive analysis suggest that China has succeeded in improving HIV knowledge about separate routes.

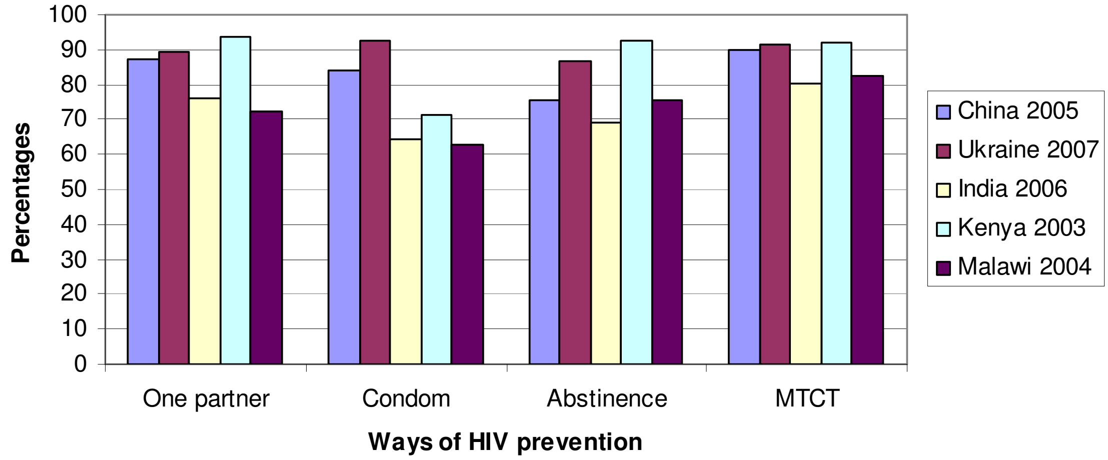


Figure 4.3: Knowledge about ways to prevent HIV transmission in the five countries.

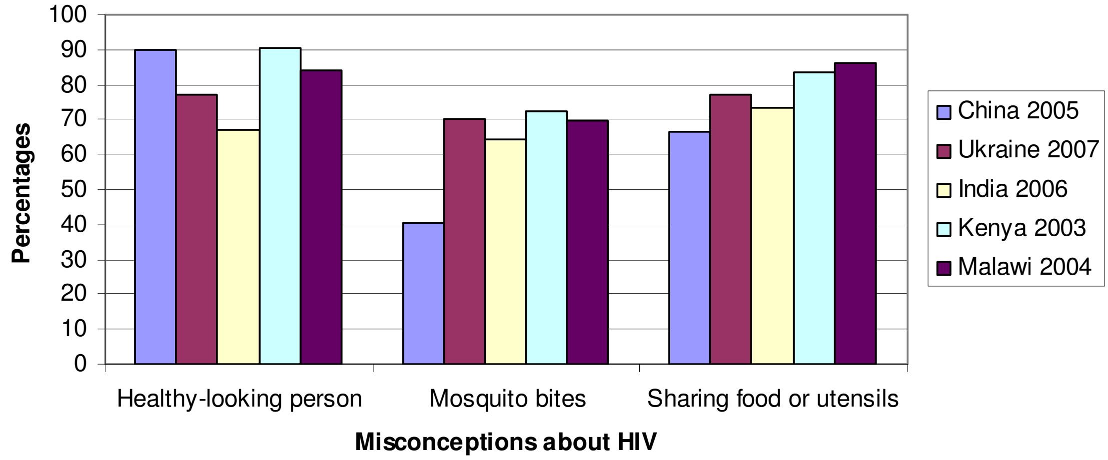


Figure 4.4: Knowledge about misconceptions about HIV transmission in the five countries.

4.4.1.2 Distributions of HIV knowledge scores

The Chinese context

Figure 4.5-4.8 shows the distributions of all four scores derived for measuring HIV knowledge in China over time. Figure 4.5 shows the distributions of scores for knowledge of correct routes of HIV transmission in China over time.

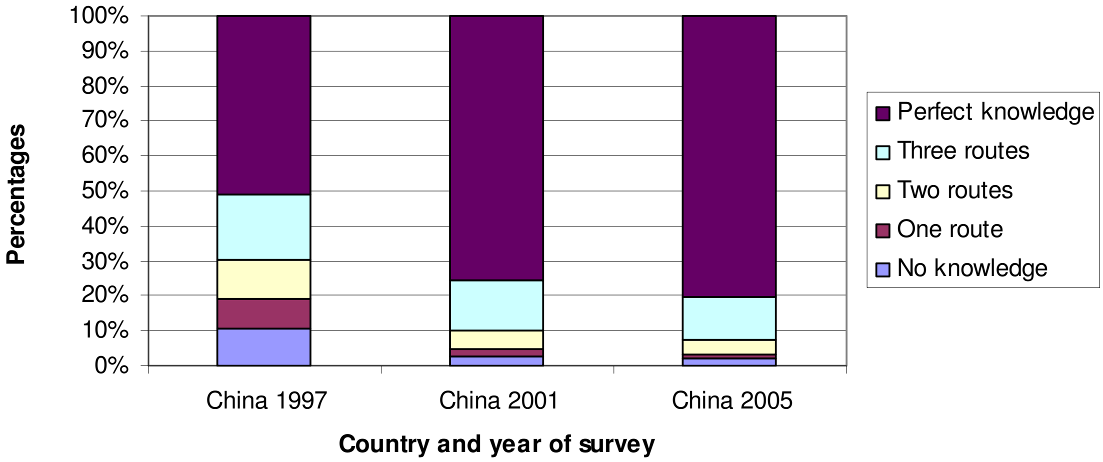


Figure 4.5: Distributions of scores for knowledge of correct routes of HIV transmission in China.

Figure 4.6 shows distributions of scores for knowledge of incorrect routes of HIV transmission in China over time.

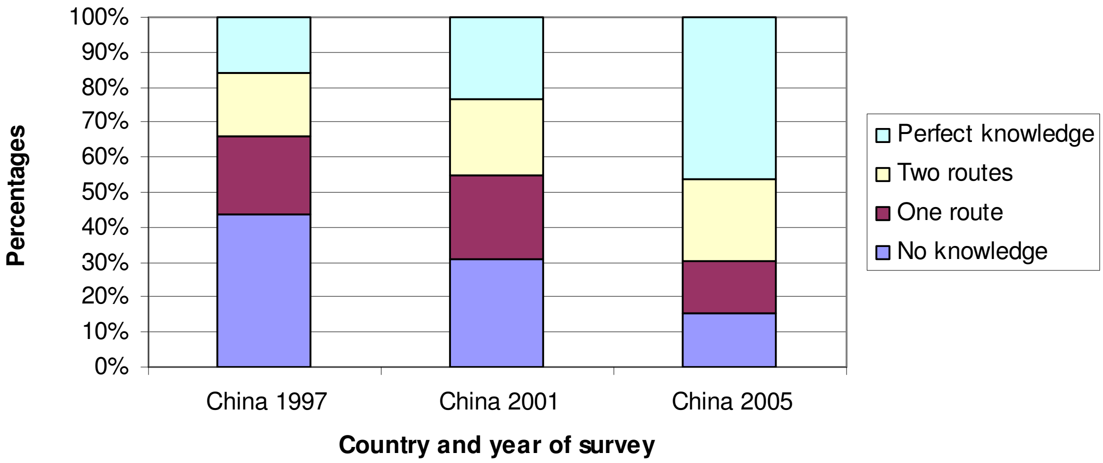


Figure 4.6: Distributions of scores for knowledge of incorrect routes of HIV transmission in China.

Figure 4.7 and 4.8 show distributions of scores for overall knowledge about HIV transmission in China over time. Figure 4.8 presents a simplified representation of the score that is presented in the Figure 4.7.

Figures 4.5-4.8 suggest that the situation in China improves with time and that a large proportion of women in 30 selected counties in China in 2005 have perfect knowledge of correct and incorrect routes of HIV transmission as well as of overall HIV knowledge. Larger groups have perfect knowledge about the correct routes of HIV transmission than about the incorrect routes of HIV transmission. This finding is expected as China is a country with a non-generalised HIV epidemic and more efforts were taken to improve the level of knowledge of correct routes than of incorrect routes of HIV transmission. It is also important to mention that even in 1997 a large proportion of women already had perfect knowledge about the correct routes of HIV transmission.

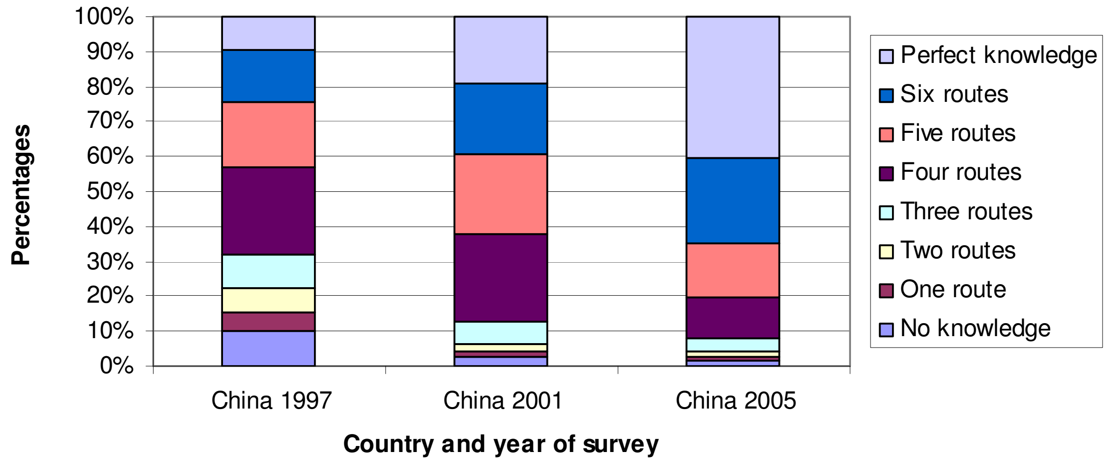


Figure 4.7: Distributions of scores for overall knowledge of HIV transmission in China.

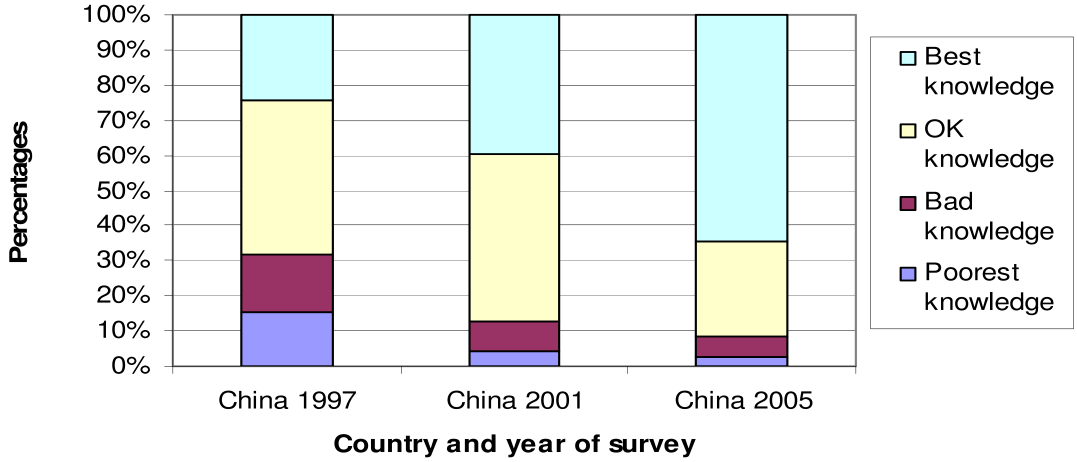


Figure 4.8: Distributions of scores for grouped overall knowledge of HIV transmission in China.

The five country context

Figures 4.9-4.12 show the distributions of all four scores derived for measuring HIV knowledge in the five countries. Figure 4.9 shows the distributions of scores for knowledge of ways of prevention of HIV transmission in the five countries.

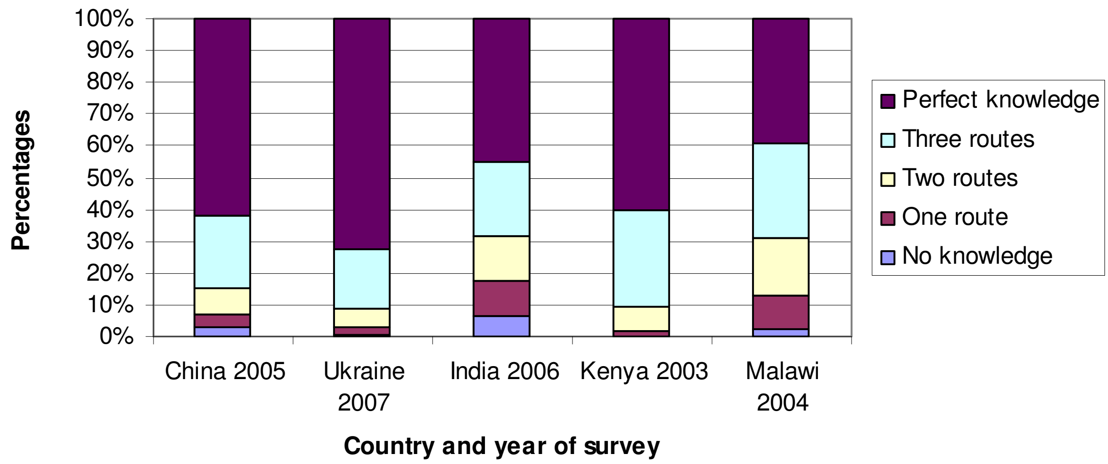


Figure 4.9: Distributions of scores for knowledge of ways to prevent HIV transmission in the five countries.

Figure 4.10 shows the distributions of scores for knowledge of misconceptions about HIV in the five countries.

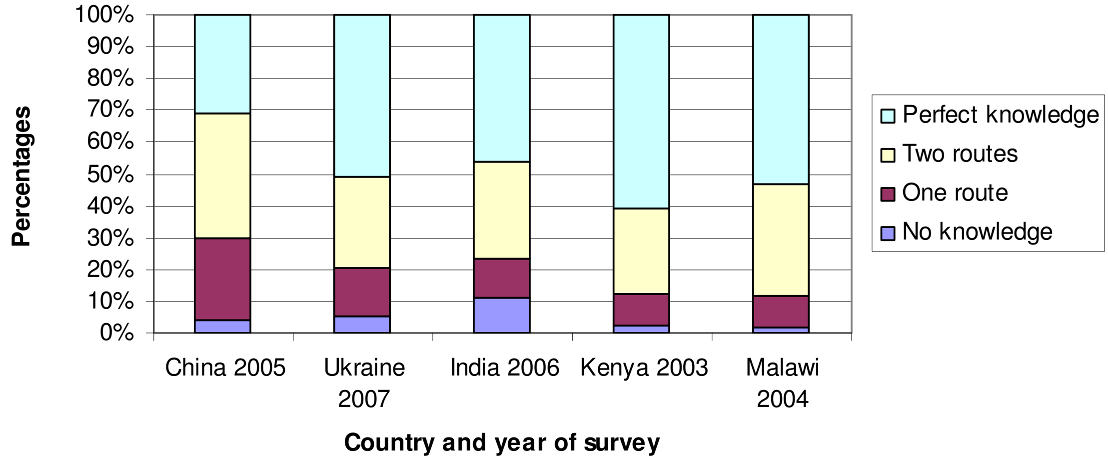


Figure 4.10: Distributions of scores for knowledge of misconceptions about HIV transmission in the five countries.

Figures 4.11 and 4.12 show the distributions of scores for overall knowledge about HIV. Figure 4.12 presents a simplified representation of the score that is presented on the Figure 4.11.

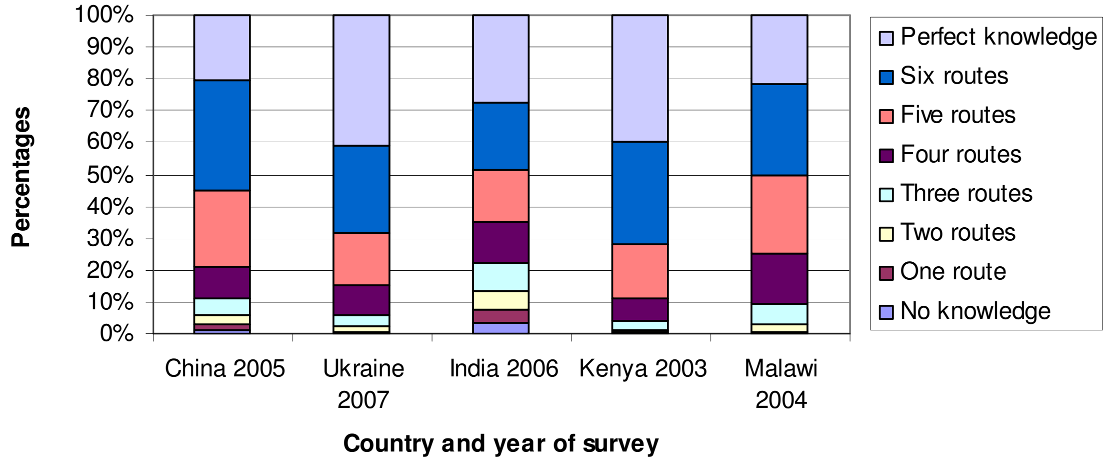


Figure 4.11: Distributions of scores for overall knowledge about HIV in the five countries.

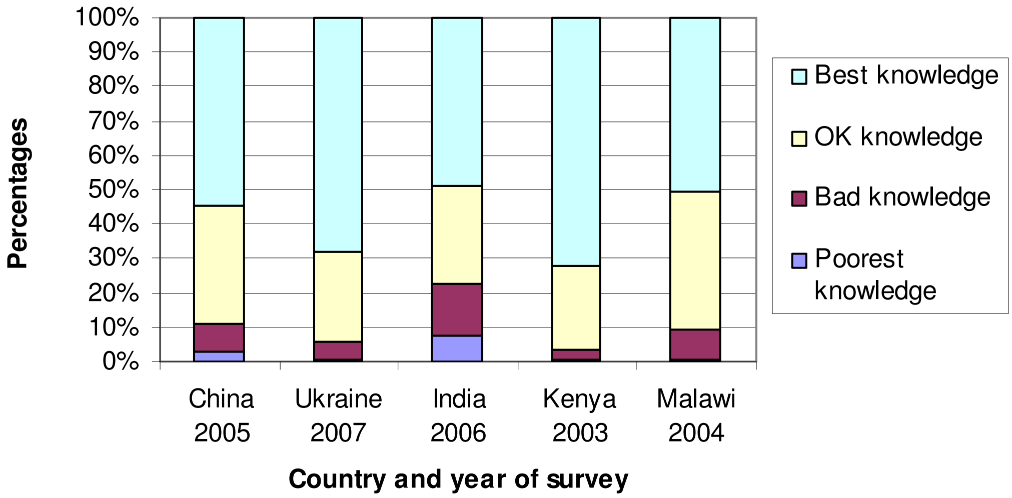


Figure 4.12: Distributions of scores for grouped overall knowledge about HIV in the five countries.

Figures 4.9-4.12 suggest that in China a large proportion of women have perfect knowledge about ways to prevent HIV but the smallest proportions of women have perfect knowledge of misconceptions about HIV transmission or of overall HIV knowledge when compared with other countries. In Kenya, Malawi and the Ukraine the groups of women with perfect knowledge about misconceptions are the largest. All four figures suggest that in China there are still women with no knowledge. In Kenya, Malawi and the Ukraine this group of women is almost non-existent (below 0.2%) in the score of overall HIV knowledge (see Figure 4.11).

This part of descriptive analysis suggests that in China, women's knowledge increased between 1997 and 2005. This might be attributed to the macro-level characteristics of China such as the effectiveness of China's education system and level of literacy (see section 1.2.3). Literacy enables people to obtain and comprehend educational information, and access to education can provide a mechanism to transfer HIV related knowledge. Another important reason is political commitment. This increase in HIV knowledge might also be attributed to the effective interventions and campaigns. However, the higher levels of HIV knowledge in 2005 might also be attributed to difference in representativeness level as this survey is not nationally representative and, therefore, the results should not be generalised to the whole of China. The increase in the level of HIV knowledge between 2001 and 2005 suggests that it is possible to achieve this increase quickly and potentially the increase might be transferred to the whole country.

It is important to explain the differences in observed patterns of scores in different countries. The following macro-level variables might be responsible for these differences: stage of HIV epidemic, effectiveness of educational campaigns, level of education and literacy in a specific country, level of political commitment, and others. Macro-level variables will be included in the analysis of HIV knowledge across countries and this will be discussed in section 4.4.2.1.

4.4.1.3 Distributions of Score Three

As mentioned earlier, in some studies HIV knowledge score variables are treated as continuous variables. This approach can potentially be applied to Score Three as it has 8 score options (other scores have fewer options and, therefore, can be easily modelled as categorical variables). However, in order to be able to model Score Three as a continuous variable, the distribution of the score in the Chinese and in the five country contexts should be normal. Figures 4.13 and 4.14 show distributions of Score Three in both study contexts. These two figures suggest that the distributions of Score Three are not normal in either the Chinese or the five country context and, therefore, it is not appropriate to model Score Three as a continuous variable as it is unlikely that the residuals will follow a normal distribution.

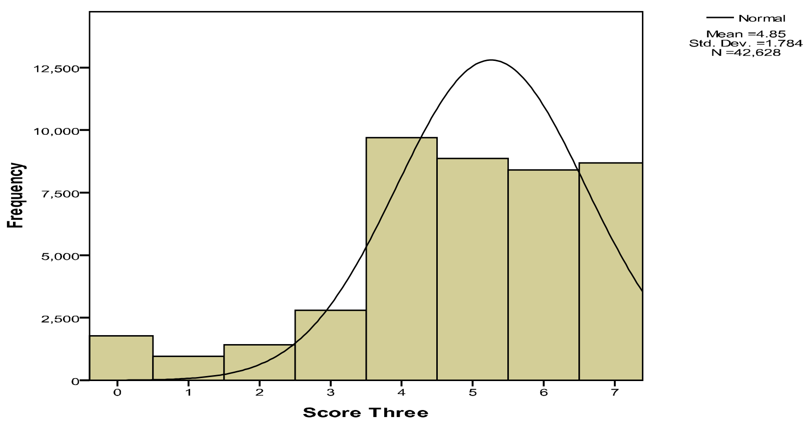


Figure 4.13: Distribution of Score Three in China.

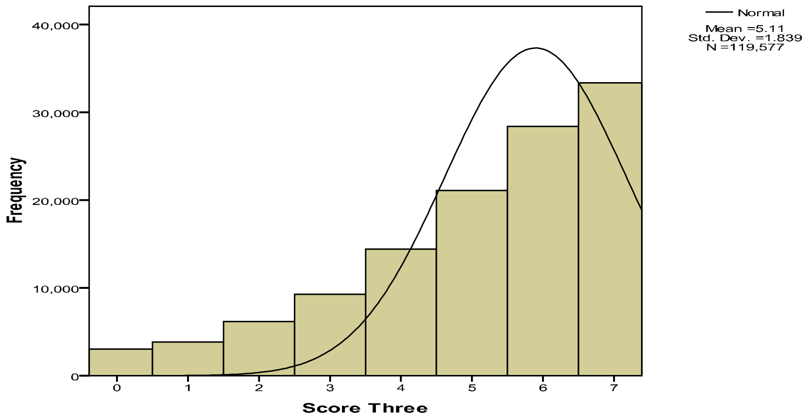


Figure 4.14: Distribution of Score Three in the five countries.

4.4.1.4 Women with perfect HIV knowledge by demographic characteristics

The Chinese context

Figures 4.15 to 4.19 show proportions of women with perfect knowledge of correct routes (those who answered all four questions correctly), incorrect routes of HIV transmission (those who answered all three questions correctly), and overall HIV knowledge (those who answered all seven questions correctly and those who belong to the "best knowledge" category (knowledge of 6-7 routes) in Score Four) by main demographic characteristics of women such as age, education, residence, marital status and ethnicity in the Chinese context respectively. Results for Score One are presented in the upper left panels, for Score Two – in the upper right panels, for Score Three – in the lower left panels and for Score Four - in the lower right panels of the Figures 4.15-4.19.

Figures 4.15-4.19 show the increase in proportions of women with perfect knowledge over time in China. Figure 4.15 shows that the proportions of women with perfect knowledge of different measures of HIV increase over time in all age groups in China. Age group 20-29 almost always has the highest proportion of women with the perfect knowledge when compared with other age groups (with the exception of the proportion of women with the perfect knowledge of incorrect routes in China 1997, and of all routes grouped in China 1997), whereas the 40-49 age group has the smallest proportion of women with the perfect knowledge (with the exception of knowledge of correct routes in China 2001 and 2005 (when the smallest proportions were observed among the 15-19 age group)).

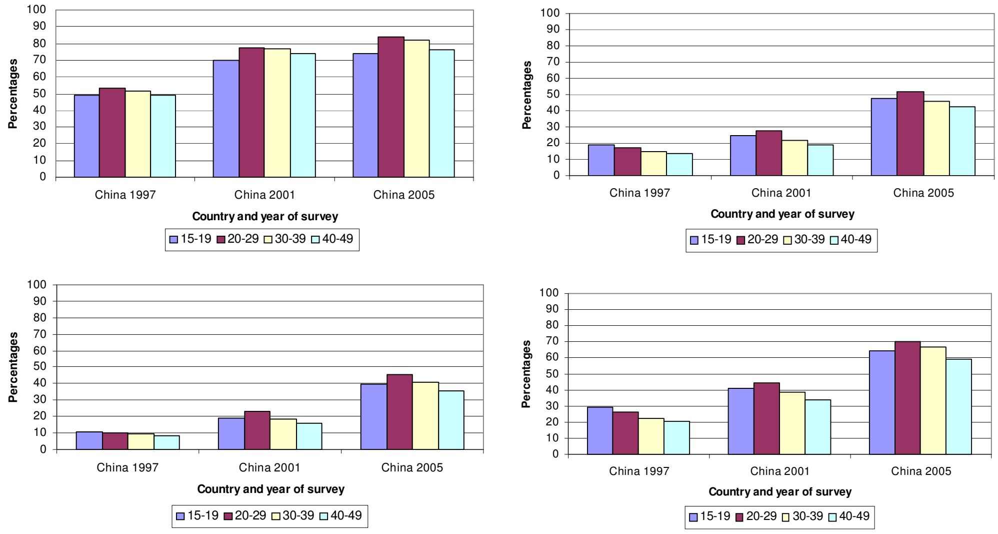


Figure 4.15: Proportions of women with perfect knowledge of correct routes (upper left panel), incorrect routes (upper right panel), all routes (lower left panel) and all routes grouped (lower right panel) by age in China.

Figure 4.16 suggests that the level of perfect knowledge of the correct routes, incorrect routes and all routes of HIV transmission either increases over time for different education groups of women or stays roughly the same. Women with higher than secondary education have the highest level of knowledge of correct routes of HIV transmission and the increase is not very high across time due to the high level of knowledge among this group already in 1997. Figure 4.16 also suggests that the higher the level of education, the higher the proportion of women with perfect knowledge for all four scores. This finding is consistent with the literature as knowledge is positively associated with education (Ingham 1995; Manchester 2002; Chen *et al*. 2003a; Wu *et al*. 2007a; Snelling *et al*. 2007).

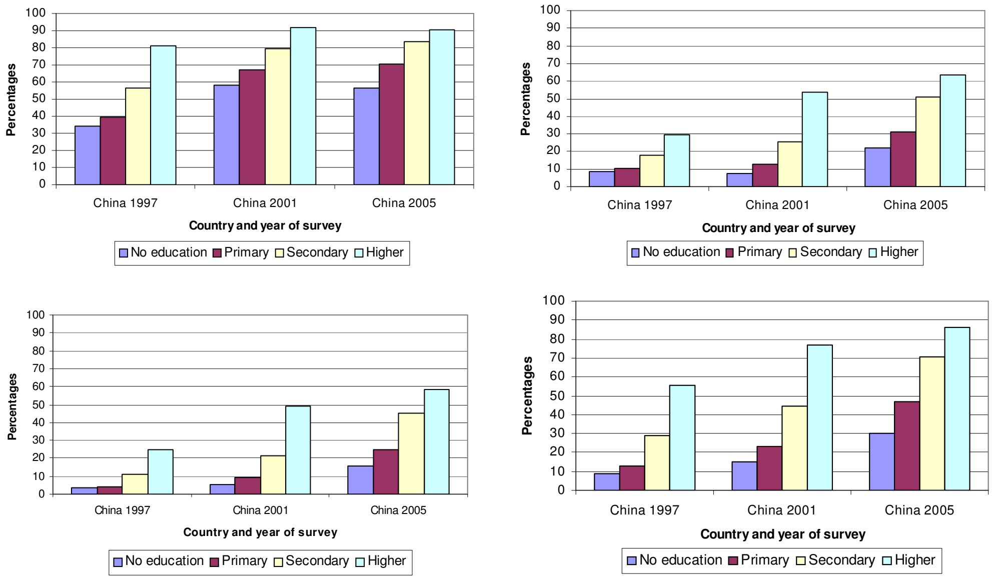


Figure 4.16: Proportions of women with perfect knowledge of correct routes (upper left panel), incorrect routes (upper right panel), all routes (lower left panel) and all routes grouped (lower right panel) by education in China.

Figure 4.17 shows the proportions of women with perfect knowledge for the four scores of HIV knowledge by residence in China. The proportion of women with perfect knowledge is consistently higher among women from urban areas than among women from rural areas. The level of perfect knowledge was substantially higher in urban areas than in rural areas in 1997 and in 2001. The level of perfect knowledge increases in both groups and by the year 2005 the gap between the two groups is smaller than in the previous years which might be attributed to the success of different interventions including the UNFPA interventions as women from rural areas were specifically targeted in these interventions.

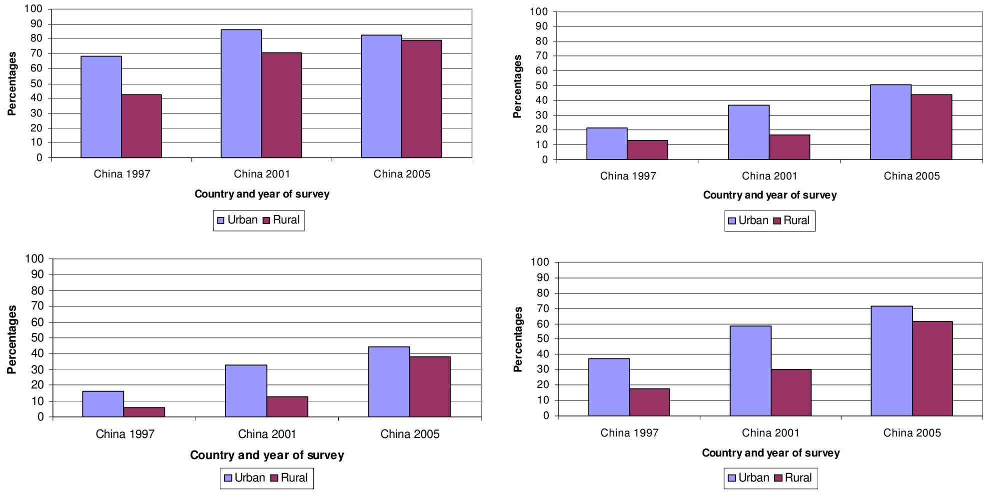


Figure 4.17: Proportions of women with perfect knowledge of correct routes (upper left panel), incorrect routes (upper right panel), all routes (lower left panel) and all routes grouped (lower right panel) by residence in China.

Figure 4.18 shows that patterns of women with perfect knowledge differ by measures and by year of surveys for different marital statuses.

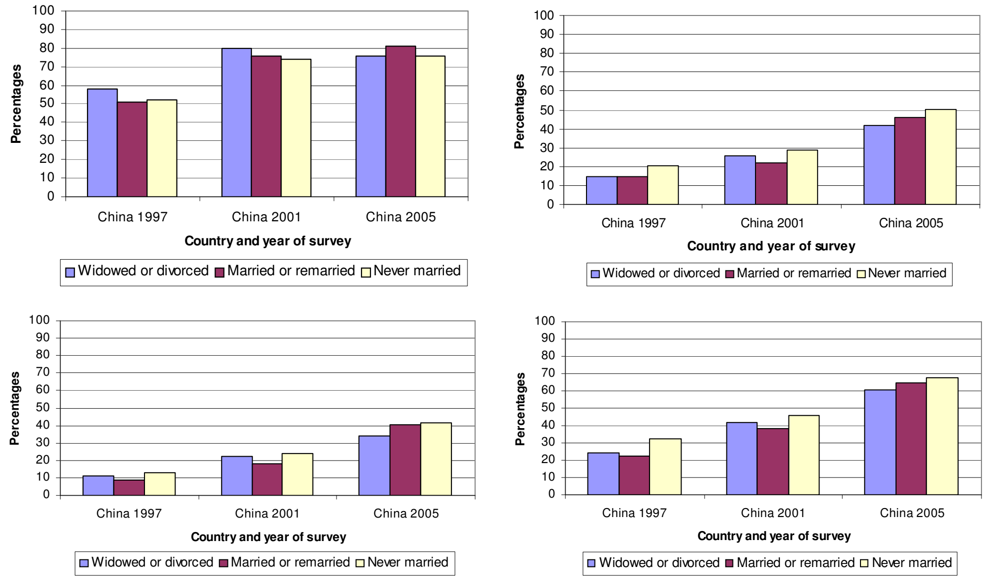


Figure 4.18: Proportions of women with perfect knowledge of correct routes (upper left panel), incorrect routes (upper right panel), all routes (lower left panel) and all routes grouped (lower right panel) by marital status in China.

Figure 4.19 suggests that women who belong to Han ethnicity have consistently higher or equal (in 2001 for incorrect routes and all routes grouped and in 2005 for all routes) level of HIV knowledge. The level of knowledge increases in the both groups of women and the difference in the proportion of women with perfect knowledge between the two groups decreases over time which also might be attributed to the successes of different interventions including interventions by the UNFPA who targeted women from ethnic minority groups.

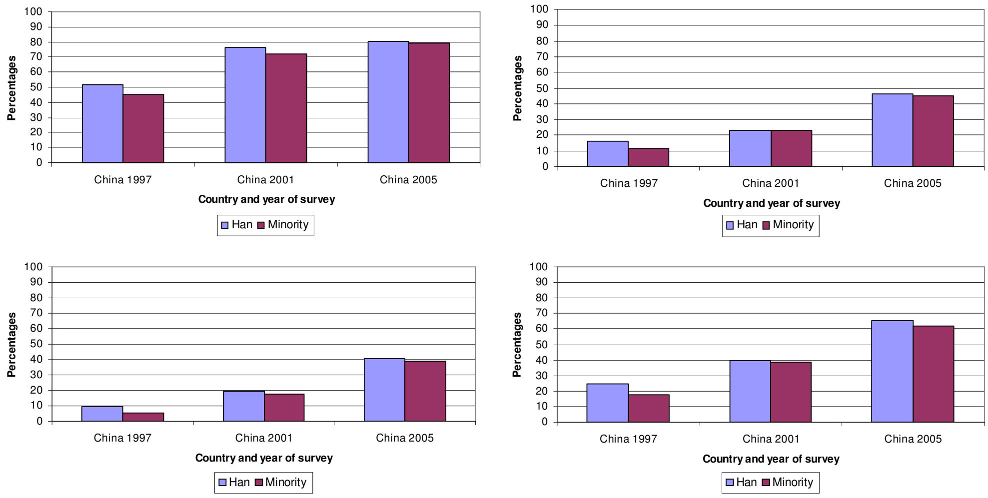


Figure 4.19: Proportions of women with perfect knowledge of correct routes (upper left panel), incorrect routes (upper right panel), all routes (lower left panel) and all routes grouped (lower right panel) by ethnicity in China.

Figures 4.15-4.19 show that the proportions of women with perfect knowledge by different demographic characteristics are substantially higher for correct routes of HIV transmission when compared to the knowledge of incorrect routes and as a result for all routes of HIV transmission. This observation supports the second assumption used in this paper that in a country with a non-generalised HIV epidemic people first learn about correct routes of HIV transmission, and once the epidemics progress further, the level of knowledge about incorrect routes starts improving.

The five country context

Figures 4.20-4.23 show proportions of women with perfect knowledge of ways to prevent HIV (those who answered all four questions correctly), of misconceptions about HIV transmission (those who answered all three questions correctly), and all components (those who answered all seven questions correctly and those who belong to the "best knowledge" category (correct knowledge about 6-7 questions) in Score Four) by main demographic characteristics of women such as age, education, residence and marital status for the five country context. Results for Score One are presented in these figures in the upper left panels, for Score Two – in the upper right panels, for Score Three – in the lower left panels and for Score Four - in the lower right panels.

Figure 4.20 shows that in all countries the proportions of women with perfect knowledge are almost always the highest among women of 20-29 years for different measures of HIV knowledge when compared with women from other age groups. The only exceptions are observed in China 2005 and Ukraine 2007 for knowledge about ways to prevent HIV. In these two countries the highest proportions of women with perfect knowledge are observed in the 30-39 age group. However, the differences between proportions in this group and between the 20-29 age group proportions are not very large. The proportions of women with perfect knowledge are the smallest in the 15-19 and 40-49 age groups when compared with the other two groups.

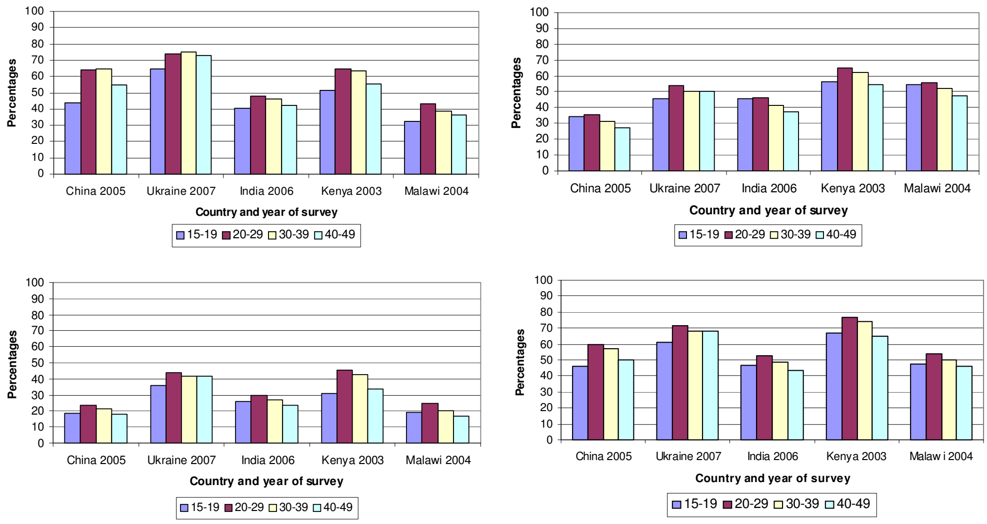


Figure 4.20: Proportions of women with perfect knowledge of ways to prevent HIV (upper left panel), misconceptions about HIV transmission (upper right panel), all components (lower left panel) and all components grouped (lower right panel) by age in the five countries.

Figure 4.21 shows that the higher the level of education, the higher the proportion of women with perfect knowledge for all four scores. This finding is consistent with the literature as knowledge is also positively associated with education (Ingham 1995; Manchester 2002; Chen *et al*. 2003a; Wu *et al*. 2007a; Snelling *et al*. 2007). In Malawi the proportions of women with perfect knowledge about ways to prevent HIV transmission do not differ as much as in other countries between the two extreme groups (women with no education and women with higher than secondary education). However, for the knowledge of misconceptions about HIV transmission the gap between these two groups is much larger in Malawi.

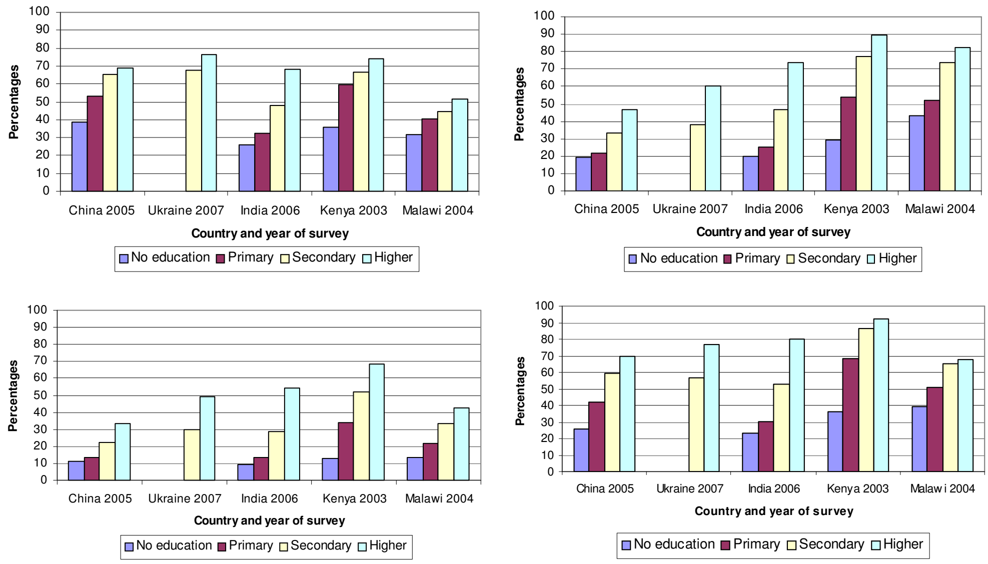


Figure 4.21: Proportions of women with perfect knowledge of ways to prevent HIV (upper left panel), misconceptions about HIV transmission (upper right panel), all components (lower left panel) and all components grouped (lower right panel) by education in the five countries.
Note: In the Ukraine a small number of women (see Appendix F) belongs to either no education or primary level of education groups and, therefore, proportions for these two groups in the Ukraine are not produced here.

Figure 4.22 shows proportions of women with perfect knowledge for different HIV knowledge measures by residence in the five countries. This figures shows that these proportions are consistently higher for women from urban areas than for women from rural areas. The only exception to this is the proportion of women with perfect knowledge about ways to prevent HIV (upper left panel) in Malawi with proportions for urban and rural residences being almost equal. In China for all measures, differences in the proportions are not very large between rural and urban areas, whereas in the Ukraine, India and Kenya these differences are larger. In Malawi these differences are larger for the knowledge of misconceptions about HIV and all components of HIV knowledge selected for this study.

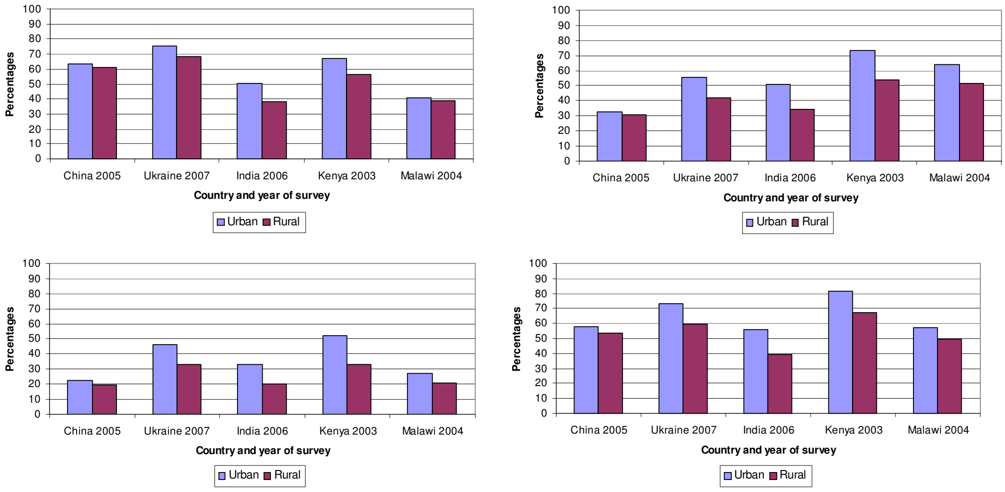


Figure 4.22: Proportions of women with perfect knowledge of ways to prevent HIV (upper left panel), misconceptions about HIV transmission (upper right panel), all components (lower left panel) and all components grouped (lower right panel) by residence in the five countries.

Figure 4.23 shows proportions of women with perfect knowledge for different measures of HIV knowledge by marital status in the five countries. The patterns differ by countries and by measures.

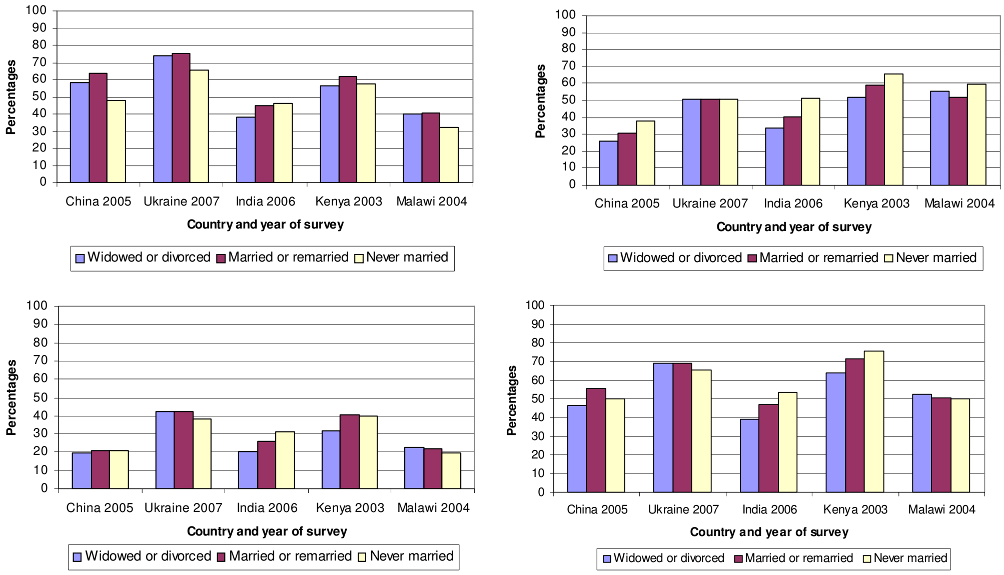


Figure 4.23: Proportions of women with perfect knowledge of ways to prevent HIV (upper left panel), misconceptions about HIV transmission (upper right panel), all components (lower left panel) and all components grouped (lower right panel) by marital status in the five countries.

Figures 4.20-4.22 show that in Kenya the proportions of women with perfect knowledge by different demographic characteristics are nearly similar for knowledge about ways of HIV prevention and about misconceptions about HIV transmission. This result is expected as Kenya is a country with generalised HIV epidemic with a long history of educational campaigns. In Malawi, these proportions are substantially higher for misconceptions than for knowledge about ways to prevent HIV. This result is surprising as Malawi is also a country with generalised HIV epidemic and, therefore, the level of HIV knowledge is expected to be higher for all components and especially for ways to prevent HIV transmission than in other countries. This result suggests that efforts should be directed at improving the level of knowledge about HIV not only in the countries where HIV epidemics are still at early stages but also efforts should be sustained in countries where HIV epidemics are mature. There is a need to continue interventions even in countries with generalised HIV epidemics as knowledge about ways to prevent HIV is an important pre-requisite for risk reduction behaviours. In China and the Ukraine these proportions are higher for ways to prevent HIV and lower for misconceptions and those results are expected given the non-generalised HIV epidemics in the country. In India these proportions are similar for Score One and Score Two and are quite low and, therefore, also not expected. The observed proportions are the lowest for knowledge about ways to prevent HIV transmission in India and Malawi when compared with other countries. In Kenya and Malawi the proportions for knowledge about misconceptions about HIV transmission are the highest when compared with other countries.

The graphical exploratory analysis conducted in this section suggests that the scores vary by the demographic characteristics of women and further more formalized bivariate analysis of the association between all measures of HIV knowledge and demographic characteristics is necessary and is discussed in the following section.

4.4.1.5 Tests of association (Kruskal-Wallis test)

Prior to statistical modelling, bivariate tests of association between all four measures of HIV knowledge and all available demographic characteristics of women are performed. The Kruskal-Wallis test is a bivariate test and, as was mentioned in the methodology section (see section 4.3.2), it is a test of association between two categorical variables one of which is an ordinal variable. It is important to mention that this test has limitations and might be affected by large sample sizes in the same way as a chi-squared test, i.e. when the size of sample is large, the association might be significant even if there is no substantial association between the two variables.

The results of bivariate test of association, the Kruskal-Wallis tests, suggest that in the Chinese context the following variables are significantly associated ($p<0.05$) with all four scores: country, age, place of residence, education and marital status (Table 4.1). Ethnicity variable was significantly associated only with Score One but not with other scores. As we are interested in studying differences in different measures of HIV knowledge at different points in time, the test was conducted between all four scores and ethnicity variable for 1997, 2001 and 2005 surveys separately. The results of these tests can be found in Table 4.2. The results suggest that in 1997 ethnicity was significantly associated with all four scores, whereas in 2001 with only Score One and in 2005 with Score Three and Score Four. Given that the ethnicity variable is significantly associated with some scores at some points in time, this variable is included in the modelling together with other variables mentioned above.

Table 4.1: Kruskal-Wallis tests of association in the Chinese context.

Variable	Score One Chi-square statistic (df), p-value	Score Two Chi-square statistic (df), p-value	Score Three Chi-square statistic (df), p-value	Score Four Chi-square statistic (df), p-value
Country	2869.045 (2), 0.000	2603.295 (2), 0.000	3793.558 (2), 0.000	3414.521 (2), 0.000
Age	125.435 (3), 0.000	283.882 (3), 0.000	184.036 (3), 0.000	136.441 (3), 0.000
Education	2148.258 (3), 0.000	4119.107 (3), 0.000	5157.494 (3), 0.000	4533.963 (3), 0.000
Residence	1321.886 (1), 0.000	2201.028 (1), 0.000	2953.447 (1), 0.000	2541.665 (1), 0.000
Marital status	19.072 (2), 0.000	239.395 (2), 0.000	110.434 (2), 0.000	69.273 (2), 0.000
Ethnicity	6.713 (1), 0.010	2.238 (1), 0.135	0.035 (1), 0.852	0.001 (1), 0.979

Table 4.2: Kruskal-Wallis tests of association between four measures of HIV knowledge and ethnicity at different points in time in China.

	Score One Chi-square statistic (df), p-value	Score Two Chi-square statistic (df), p-value	Score Three Chi-square statistic (df), p-value	Score Four Chi-square statistic (df), p-value
China 1997	11.699 (1), 0.001	27.105 (1), 0.000	27.957 (1), 0.000	26.471 (1), 0.000
China 2001	12.093 (1), 0.001	0.030 (1), 0.862	2.101 (1), 0.147	2.048 (1), 0.152
China 2005	2.410 (1), 0.121	2.548 (1), 0.110	4.117 (1), 0.042	4.749 (1), 0.029

The results of bivariate test of association, the Kruskal-Wallis tests, suggest that in the five country context the following variables were significantly associated ($p<0.05$) with all four scores: country, age, place of residence, education and marital status (Table 4.3). Therefore, all these variables are included in the modelling.

Table 4.3: Kruskal-Wallis tests of association in the five country context.

Variable	Score One Chi-square statistic (df), p-value	Score Two Chi-square statistic (df), p-value	Score Three Chi-square statistic (df), p-value	Score Four Chi-square statistic (df), p-value
Country	4095.782 (4), 0.000	2385,039 (4), 0.000	2920.130 (4), 0.000	3136.631 (4), 0.000
Age	671.894 (3), 0.000	805.913 (3), 0.000	579.330 (3), 0.000	510.092 (3), 0.000
Education	9353.569 (3), 0.000	11376.337 (3), 0.000	14916.570 (3), 0.000	13150.335 (3), 0.000
Residence	1299.541 (1), 0.000	2269.829 (1), 0.000	2675.684 (1), 0.000	2096.051 (1), 0.000
Marital status	29.560 (2), 0.000	1257.800 (2), 0.000	286.614 (2), 0.000	240.092 (2), 0.000

Descriptive analysis suggests that education is a significant predictor of four measures of HIV knowledge at all times in China and in different countries. The higher the level of education of women, the higher the scores for all four measures used in the study, and this finding is in agreement with the existing literature. Residence is also a significant predictor of all four measures; women in urban areas have higher levels of knowledge than women in rural areas at different points in time in China and in different cultural and epidemiological contexts. Age is also significantly associated with all four measures of HIV knowledge used in this analysis and women of 20-29 years have higher scores than women of 40-49 years. Marital status is significantly associated with all four measures of HIV knowledge but observed patterns are different at different points in time, in different cultural and epidemiological contexts and for different scores. Ethnicity is also a significant predictor of scores in the Chinese context and the levels of knowledge are higher for Han women when compared to women who belong to minority groups. These findings are consistent with existing literature discussed in section 1.3.3.

4.4.2 Results from regression analysis

Scores One to Four are all ordinal variables and, therefore, it is appropriate to model them using proportional odds models. If the proportional odds assumption does not hold, then the scores will be modelled using partial proportional odds models.

All models in this analysis are first fitted to the Chinese pooled dataset to compare HIV knowledge scores in China over time and then to the five country pooled dataset to compare HIV knowledge scores across countries with generalised and non-generalised HIV epidemics.

The response variables and the explanatory variables used for the analysis in this paper are discussed in sections 2.3, 2.4 and 4.2.

The main aim of fitting proportional odds models to the pooled datasets is to compare the probability of having higher scores of HIV knowledge at different times in China and to compare the probability of having higher scores of HIV knowledge across countries. The main aim of fitting partial proportional odds models to the Chinese pooled dataset is to compare the probability of being in different parts of distributions of HIV knowledge scores at different times in China and to understand if HIV knowledge level increases in China over time among women. The main aim of fitting the models to the five country pooled dataset is to compare the probability of being in different parts of distributions of HIV knowledge scores across countries to understand how China performs relative to other countries with generalised and non-generalised HIV epidemics.

The main two aims of fitting models to four separate scores are

1. to investigate HIV knowledge in China over time and across countries to compare different score options and
2. to understand advantages and limitations of using different options for measuring HIV knowledge.

4.4.2.1 Proportional odds models

Forward selection is used to identify the final proportional odds models. Proportional odds models are fitted to the Chinese pooled dataset first and five individual characteristic variables (age group, education, residence, marital status, and ethnicity) and one macro-level variable (year of survey), are used during the model selection process. Proportional odds models are then fitted to the five country pooled dataset and four individual characteristic variables are used during the model selection process (age group, education, residence, and marital status) and five macro-level variables are also considered (country, type of HIV epidemic, HIV prevalence, HIV awareness rate, and female literacy rate).

Once the country variable is entered into models in the five country context, it is not possible to include type of HIV epidemic, HIV prevalence, HIV awareness rate or female literacy rate variables due to collinearity between the country variable and other macro-level variables. As we are interested in differences across five countries, we cannot remove the country variable. This variable controls for the country specific context including type of HIV epidemic, literacy rate among females and other country characteristics. Unfortunately, it is not possible to disentangle specific relationships between the components behind country variable but it is possible to speculate about different potentially important aspects within different country contexts. If the main interest is to disentangle those differences, removal of the country variable would be required. However, this is not the main aim of the current analysis. Therefore, only one macro-level variable (country) is used during the analysis in the five country context.

Four proportional odds models are selected in the Chinese data (one for each score) and four proportional odds models are selected in the five country data (one for each score). The results of the parallel line test suggest that none of the eight models fitted data well (p-values are lower than 0.05) as the main model assumption of proportional odds does not hold for any of the models. Therefore, partial proportional odds models will be fitted to the two pooled datasets in order to investigate differences in patterns of the probabilities of having different score options for four scores in China at different points in time and in the five countries.

4.4.2.2 Partial proportional odds models

Automatic forward stepwise model selection conducted in SPSS within multinomial logistic regression framework is used to identify the final models. Models are fitted to the Chinese pooled dataset first and five individual characteristic variables (age group, education, residence, marital status, and ethnicity) and one macro-level variables (year of survey) are used during the model selection process. Multinomial logistic regression models are then fitted to the five country pooled dataset and four individual characteristic variables (age group, education, residence, and marital status) and one macro-level variables (country) are used during the model selection process. Two-way interactions between the variables are also tested for inclusion in the final models for both study contexts. Once the final models are identified within multinomial framework in SPSS, they are transferred to STATA and more parsimonious models are obtained within partial proportional odds framework. Four models are selected in the Chinese context (one for each score) and four models are selected in the five country context (one for each score). The results for the eight partial proportional odds model are presented in Appendix G, Tables G.1-G.8.

The Chinese context

Table 4.4 summarises the results of partial proportional odds models in the Chinese context. Detailed results of the modelling can be found in Appendix G, Tables G.1-G.4.

Table 4.4: Summaries of partial proportional odds models in the Chinese context.

Model	Number of significant main effects	Number of significant two-way interactions	Number of constraints for parallel lines imposed	Total number of parameters in the final model
Score One	6	3	13	23
Score Two	6	8	30	49
Score Three	6	4	10	25
Score Four	5	4	20	30

Predicted probabilities are calculated for the four scores at different points in time for women with potentially the best HIV knowledge and potentially the worst HIV knowledge in China in order to enable the assessment of changes in general patterns of HIV knowledge among different groups of women over time. Women with potentially the best HIV knowledge in China are assumed to have the following characteristics: they are from 20-29 age group, with higher than secondary education, living in urban areas, married or remarried and belong to Han ethnicity. Women with potentially the worst knowledge in China are assumed to be from the 40-49 age group, with no education, living in the rural area, widowed or divorced and belong to ethnic minority groups. Figures 4.24 and 4.25 respectively present predicted probabilities for different options of all four scores for women with potentially the best and potentially the worst HIV knowledge in China.

Figure 4.24 shows that for women with potentially the best knowledge, the probability of having perfect knowledge for all four scores increases over time in China with one exception for Score One in 2005, it does not increase when compared to 2001 but this can be explained by the fact that in 2001 this probability was already very high. The probability of having perfect knowledge for all four scores is the highest in China at almost every point in time (the exceptions are the knowledge of incorrect routes and the overall knowledge in 1997 when the probability of knowing two and six routes was higher than the probability of having perfect knowledge). However, the probabilities of having perfect knowledge of the incorrect routes of HIV transmission are still much lower than the probabilities of having perfect knowledge of the correct knowledge among women with potentially the best knowledge about HIV at every point in time. Figure 4.24 also shows that the probability of having no knowledge among women with potentially the best knowledge of correct routes and of all routes of HIV transmission is low even in 1997 and nearly non-existent at other points in time, whereas this probability was slightly higher at every point in time for the knowledge of incorrect routes of HIV transmission (upper right panel). The probability of having no knowledge as well as other score options apart from the perfect knowledge decreases over time.

Figure 4.25 shows that for women with potentially the worst knowledge, the probability of having no knowledge for all four scores decreases with time and the probability of having perfect knowledge also increases with time in China. The probability of having no knowledge for all four scores is the highest in China in 1997 when compared with the other score options. This probability is also the highest probability in 2001 and 2005 for the knowledge of incorrect routes of HIV transmission for women with potentially the worst knowledge (upper right panel). Figure 4.25 shows that the probability of having knowledge of five and six routes for the Score Three (lower left panel) increases with time in China. Figure 4.25 also shows that for women with potentially the worst knowledge in 2001 and 2005 the probability of having perfect knowledge of correct routes of HIV transmission is higher than having no knowledge among this group of women (upper left panel) and the differences are quite substantial. These findings suggest that the level of knowledge of correct routes of HIV transmission increases over time among this group of women. However, for the knowledge of incorrect routes of HIV transmission, the progress is not as visible. In 2005, the probability of having the perfect knowledge is also higher than the probability of having no knowledge for Scores Three and Four (lower left and right panels).

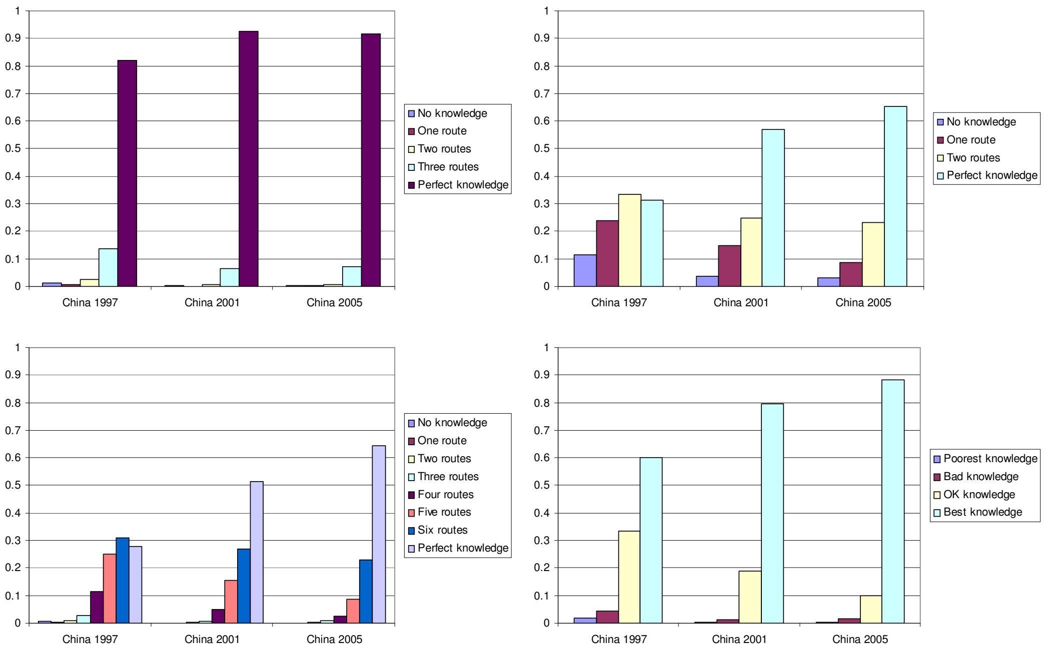


Figure 4.24: Predicted probabilities of having different score options for knowledge about correct routes of HIV transmission (upper left panel), incorrect routes (upper right panel), overall knowledge (lower left panel) and grouped overall knowledge (lower right panel) for women with potentially the best knowledge in China.

Figures 4.24 and 4.25 show that the level of knowledge about correct routes of HIV transmission is higher than the level of knowledge about incorrect routes of HIV transmission, among both groups of women at every point in time. This result is expected as the HIV epidemic in China is still not generalised and at an earlier stage than in countries where HIV epidemics are mature.

Both Figures 4.24 and 4.25 suggest that the HIV knowledge increased in China with time in all groups of women. A larger increase is observed in knowledge of correct routes of HIV transmission than in knowledge of incorrect routes, and this is expected as the epidemic in China is still at an early stage and, therefore, HIV prevention knowledge is still more important than anti-discrimination knowledge. However, there is also progress in improvement of anti-discrimination knowledge in the Chinese context among different groups of women. The progress of the second group is not as visible as for the first group of women used for the assessment of changes in patterns of HIV knowledge.

These two figures (4.24 and 4.25) suggest that in the Chinese context we observe a big contrast in HIV knowledge between women with potentially the best and potentially the worst knowledge. Different measures of HIV knowledge suggest that HIV knowledge is not homogenous in these two groups of women yet. The gap in knowledge between women from these two groups still exists. The level of knowledge increases among these two groups but the rate of the increase is different. However, it is also important to mention that the level of all types of HIV knowledge has improved over time in China.

It can be concluded that China has succeeded in improving women's knowledge about HIV as we observe an increase in all for scores in different groups of women over time. Due to the representativeness level of China 2005 dataset (see Chapter 2), the results for 2005 modelling cannot be generalised to the whole China. However, it can be argued that if the right interventions are in place in the whole countries, similar results (improved levels and increased homogeneity of HIV knowledge among different groups of women) could be observed in the other parts of the country as well.

If achievements in other dimensions of HIV prevention are equally successful, China might become a success story in controlling HIV epidemic and might provide useful lessons for other countries in the world. However, in order to investigate the relative success of China in the achieved levels of HIV knowledge, it is essential to conduct cross-country comparison. The results from the cross-country comparison are presented below.

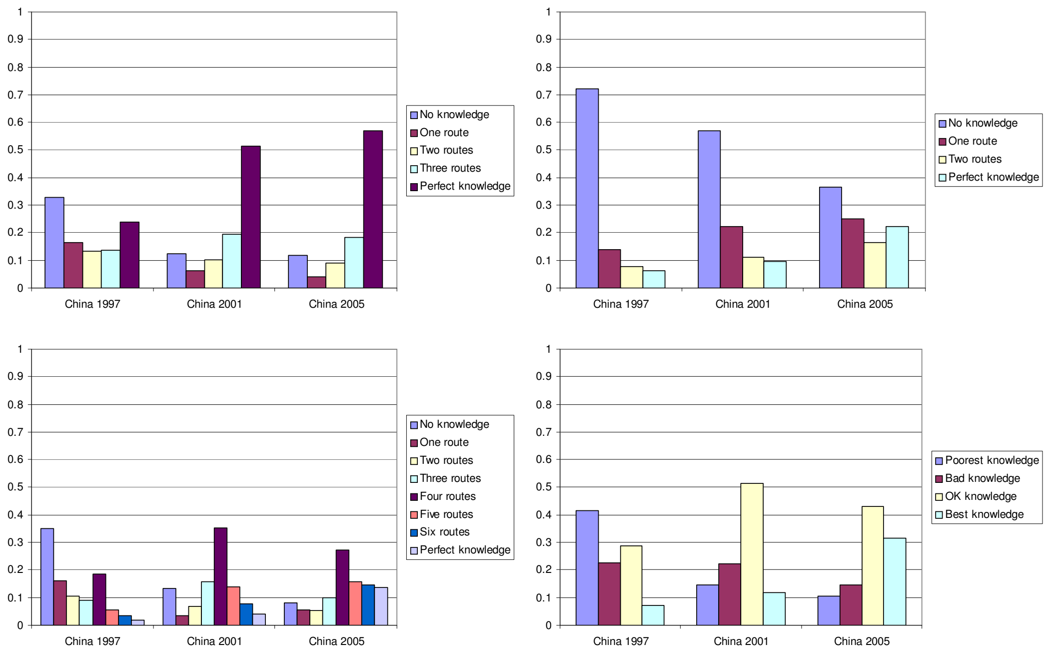


Figure 4.25: Predicted probabilities of having different score options for knowledge about correct routes of HIV transmission (upper left panel), incorrect routes (upper right panel), overall knowledge (lower left panel) and grouped overall knowledge (lower right panel) for women with potentially the worst knowledge in China.

The five country context

The results of the analysis suggest that in all four models all five main effects were found to be significantly associated with scores. Table 4.5 summarises the results of partial proportional odds models in the Chinese context. Detailed results of the modelling can be found in Appendix G, Tables G.5-G.8.

Table 4.5: Summaries of partial proportional odds models in the five country context.

Model	Number of significant two-way interactions	Number of constraints for parallel lines imposed	Total number of parameters in the final model
Score One	6	28	49
Score Two	4	16	34
Score Three	3	13	29
Score Four	2	22	55

Predicted probabilities are also calculated for the four scores in the five countries for women who are assumed to have potentially the best HIV knowledge and potentially the worst HIV knowledge to compare general patterns in HIV knowledge across countries for these two groups. Women with potentially the best HIV knowledge in the five country context are assumed to have the following characteristics: they are from 20-29 age group, with higher than secondary education, living in urban areas, never married. Women with potentially the worst knowledge in the five country context are assumed to be from the 40-49 age group, with no education, living in the rural area, married or remarried. Figure 4.26 and 4.27 respectively present predicted probabilities for different options of all four scores for women with potentially the best and potentially the worst HIV knowledge in the five countries.

Figure 4.26 shows that for women with potentially the best knowledge, the probability of having perfect knowledge is the highest in all five countries. The only exception is observed in China 2005 for Score Three (lower left panel): the probability of having knowledge of six routes is slightly higher than the probability of having perfect knowledge. The probability of having perfect knowledge for all four measures of HIV knowledge is the highest in Kenya when compared with other countries. The probability of having perfect knowledge for Score One is the lowest in Malawi (and this result is surprising given the stage of the HIV epidemic in the country), for Scores Two and Three is in China and for Score Four is in India. In both Kenya and Malawi the probabilities of having the prefect knowledge are the highest for Scores Two to Four when compared with other countries. In all five countries the second highest probability is the probability of having knowledge of three out of four routes of HIV transmission for Score One, of two out of three for Score Two, in six out of seven for Score Three and to be in the third category out of four for Score Four. The probability of having no knowledge for all four measures in the five countries is low and in some contexts are even close to zero (Score One in Kenya, the Ukraine and Malawi, Score Two in Kenya and Malawi, Score Three in Kenya, Malawi, the Ukraine and China, Score Four in Kenya, Malawi, China and the Ukraine). These results suggest that the level of HIV knowledge is high among women with potentially the best knowledge in all five countries.

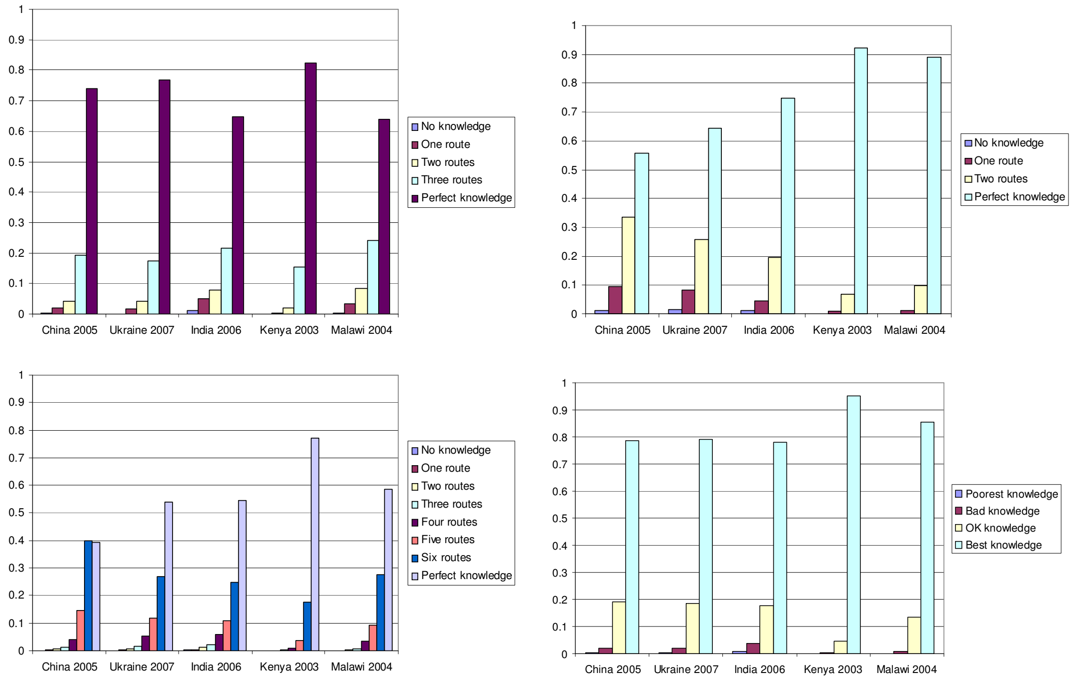


Figure 4.26: Predicted probabilities of having different score options for knowledge about ways to prevent HIV transmission (upper left panel), about misconceptions (upper right panel), all components of HIV knowledge selected for the study (lower left panel) and grouped knowledge of all components (lower right panel) for women with potentially the best knowledge in the five countries.

Figure 4.27 shows that for women with potentially the worst knowledge the probability of having no knowledge for all components of HIV knowledge selected for the study (lower left panel) is nearly non-existent in Kenya and Malawi which suggests that even women with potentially the worst HIV knowledge have some knowledge about HIV in these two countries. In the Ukraine and China these probabilities are low with India having a slightly higher probability of having no knowledge about HIV. The probability of having perfect knowledge among women with potentially the worst knowledge is the highest in Kenya and Malawi for Score Two, Score Three and Score Four when compared with other countries and this is expected as these countries experience generalised epidemics and have had educational campaigns in place for a long time. For Score One the probability of having perfect knowledge is the highest in China, the Ukraine and India when compared with other score options. In Kenya and Malawi it is the second highest probability after the knowledge about three components. The probability of having perfect knowledge of Score One is the highest in Kenya, followed by the Ukraine and China and then followed by Malawi with the lowest probability observed in India. In India the probabilities of having different score options for Score One are roughly similar. For Score Two the probability of having perfect knowledge is the highest in Kenya and Malawi when compared with other countries. This is expected due to the different stages of HIV epidemics in five countries. For Score Three in Kenya, Malawi and China, the probabilities of having knowledge of five or six routes are the highest when compared with other score options, whereas in India and the Ukraine, the probabilities of having knowledge of four or five routes are the highest when compared with other score options within countries. No score option stands out in India for Score Three as differences between probabilities of different options are not large. Figure 4.27 suggests that the level of knowledge for all four scores is higher in Kenya and Malawi among women with potentially the worst knowledge than in other countries. Figure 4.27 also suggests that despite the fact that some women among those with potentially the worst knowledge in all five countries have some knowledge about HIV, the level of overall knowledge is still not adequate for many women in all cultural contexts and these women should be targeted for educational campaigns and interventions.

These two figures (4.26 and 4.27) suggest that in the five countries a big contrast in HIV knowledge between women with potentially the best and potentially the worst knowledge is observed. Different measures of HIV knowledge suggest that HIV knowledge is not homogenous yet in these two groups of women in all countries used for the analysis.

The level of knowledge about ways of prevention of HIV transmission is similar in different country contexts within different groups of women. However, the level of knowledge about misconceptions of HIV transmission is higher in Kenya and Malawi within different groups of women when compared with other countries, and this is expected as the epidemics in these two countries are already mature whereas in China, the Ukraine and India they are still in the early stages and, therefore, HIV prevention knowledge is still more important than anti-discrimination knowledge.

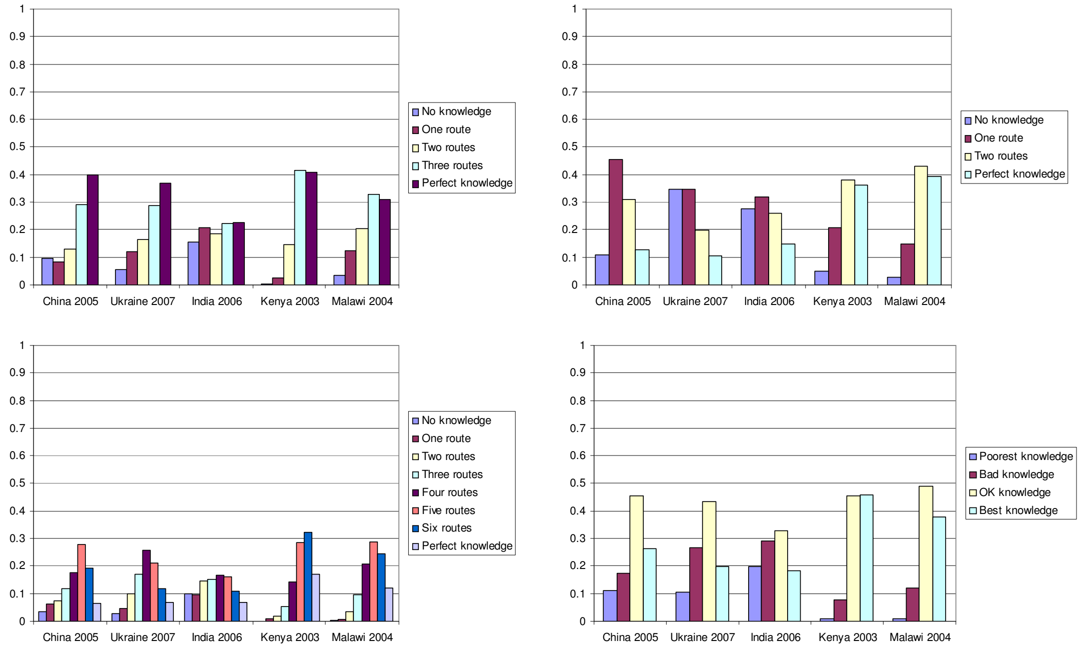


Figure 4.27: Predicted probabilities of having different score options for knowledge about ways to prevent HIV transmission (upper left panel), about misconceptions (upper right panel), all components of HIV knowledge selected for the study (lower left panel) and grouped knowledge of all components (lower right panel) for women with potentially the worst knowledge in the five countries.

Figures 4.26 and 4.27 suggest that China, the Ukraine and India have similar patterns in the four measures of HIV knowledge. However, it is important to notice that the performance in China is slightly better among women with potentially the worst knowledge than in India (Figure 4.27). It is also important to mention that the proportion of women with perfect knowledge about ways to prevent HIV is slightly higher in China in both groups of women when compared with Malawi where the HIV epidemics is generalised, but not better than in Kenya. Figures 4.26 and 4.27 suggest that for both groups of women the level of knowledge about ways to prevent HIV in China is comparable to those in Kenya. However, the levels of knowledge about misconceptions are not yet comparable to those in Kenya and Malawi.

Therefore, it can be concluded that China has achieved a relative success in improving knowledge about ways to prevent HIV. However, given the respresentativeness level of the Chinese dataset used for the analysis, it can be argued that some aspects of HIV knowledge in China in 2005 in the general population of women of the entire country would be worse as only the best performing counties were put forward for participation in the survey (see section 2.1).

China went a long way in improving women's knowledge about HIV/AIDS; knowledge of both correct and incorrect routes of HIV transmission increased over time between 1997 and 2005. The results of the analysis suggest that the level of different aspects of HIV knowledge increases over time in China. However, the level of knowledge still differs by groups of women and, therefore, HIV knowledge is not homogenous in the country. The level of HIV knowledge about ways of HIV prevention is comparable to all other countries used in the analysis, even to the countries with generalised HIV epidemics (on some occasions, it is even higher than in Malawi where epidemic is generalised). The level of knowledge about misconceptions and about all components of HIV knowledge selected for the study in China is comparable to those in other countries with non-generalised HIV epidemics (India and the Ukraine). Therefore, it can be concluded that China has succeeded in improving HIV knowledge among different groups of women.

4.4.3 Simple score measures of HIV knowledge: Their advantages and limitations

Different approaches exist to measuring HIV knowledge in the literature. This paper focused on a simple score approach to measuring HIV knowledge. Four different simple score measures of HIV knowledge were used for the analysis in this paper. In this section these four different simple score measures are compared and their advantages and limitations are discussed. As HIV knowledge cannot be measured directly, researchers always face the choice between different measures they can use in their analysis. Assessment of different measures of HIV knowledge available can help researchers to decide which measure is most suitable in their study context. Different aspects of measure suitability can be taken into account when decision about usage of a certain measure is taken. The summary of different simple score measures, which is provided below, can be used not only within the context of measures of HIV knowledge but also within contexts of any other health-related knowledge measures which also cannot be measured directly and have to be derived on the basis of different components.

Examples of other areas of health-related knowledge in which this summary can be useful are knowledge about antibiotics resistance among general public, knowledge of routine prenatal screening tests for foetal abnormality among obstetricians and midwives, knowledge about cardiovascular health-related diet and exercise behaviours, knowledge about eating habits among health professionals, knowledge about risk factors among diabetic patients, knowledge about health hazards among schoolchildren, knowledge about breastfeeding among health professionals, knowledge about symptoms of heart attack among women above 50, cervical cancer prevention knowledge and others.

Each HIV knowledge measure can be assessed from the point of view of its so-called qualitative and quantitative meanings. Quantitative or ordinal meaning of any measure is the number of components of the measure known by a respondent. Or in other words, a quantitative meaning of a measure can help to identify the place of a respondent on a continuous spectrum of HIV knowledge. Qualitative or nominal meaning of any score measure is a specificity of knowledge of each respondent in each category, i.e. which specific elements of HIV knowledge are known by a respondent in a specific group. Or in other words, a qualitative meaning of a measure can help to identify specific groups of respondents which exist in a population on the basis of their HIV knowledge. Ideally a perfect measure of HIV knowledge should have a precise qualitative and a precise quantitative meaning. However, this might not be achievable and, therefore, more suitable measure should be selected. In some specific research contexts only one of the two requirements or even a part of one requirement might be sufficient.

Score One is a simple score measure for knowledge of correct routes of HIV transmission in the Chinese context and a measure for knowledge about ways to prevent HIV in the five country context. Score Two is a simple measure for knowledge of incorrect routes of HIV transmission in China and a measure for knowledge about misconceptions about HIV transmission in the five countries. Score Three is a simple score measure for combined knowledge about HIV (combination of Score One and Score Two components). Score Four is a variable which simplified and combined score options of Score Three into four groups.

The first three measures have very clear and precise quantitative or ordinal meaning. These measures are constructed in order to quantify the number of correct answers to the HIV knowledge component questions. Therefore, each respondent has a value for each of the first three scores and this value means the number of components of each measure which is known by a respondent. This information can help to place each respondent on a continuum between no knowledge about HIV and perfect knowledge.

The assessment of the qualitative meaning of the first three measures used for the analysis suggests that only extreme categories have a precise qualitative meaning (the first and the last categories). The first score option always suggests that a respondent does not have any knowledge and the last score option suggests that a respondent has knowledge of all components of HIV knowledge. All other score options in the first three scores suggest the number of known component but they do not indicate which specific components are known by a respondent. For different individuals having the same number of correct answers can represent different patterns of HIV knowledge, e.g. respondents who have knowledge about one route, either MTCT or handshaking will end up in the same group in Score Three (knowledge of one route of HIV transmission).

For Score Four, the quantitative or nominal meaning is not very precise. Each respondent can have one out of two options of known components and the specific number of options is not known from this variable. However, rough position on the continuum between no knowledge and perfect knowledge can be identified. The qualitative meaning of this score is also not very precise for all categories (even for the extreme ones (as the extreme groups in this score contain two quantitative possibilities of responses each). However, despite the fact that this variable does not have very precise qualitative and quantitative meaning, this variable might be a better option for some research projects in comparison to Score Three as this variable has a lower number of categories and, therefore, can be modelled easier.

All these four measures can be used as response variables in proportional odds models, partial proportional odds models, and multinomial logistic regressions. The first three HIV knowledge measures can also be used as a response variable in a linear regression if HIV knowledge measure has a normal distribution. All four measures can also be used as an explanatory variable in different models.

Score One measure is useful when the main interest is specifically placed on correct routes of HIV transmission. Score Two measure is useful when the main interest is specifically placed on incorrect routes of HIV transmission. When the main interest is a combined knowledge about HIV, it is better to use either Score Three or Score Four. However, it is important to remember that these measures will not provide any information about specificity of knowledge unless a respondent has no knowledge or has a perfect knowledge for Score Three. For example, for respondents who scored 5 on Score Three measure, the measure will not provide details whether this respondent has a good level of knowledge about correct routes or incorrect routes or a mixture of both. The decision about choosing Score Three or Score Four depends on the required level of quantitative specificity about combined knowledge of HIV for interpretational purposes.

When Score One and Score Two are modelled separately as response variables, they will mask the information about combined knowledge of HIV for respondents. If the interest is in separate types of knowledge as well as on general knowledge, three measures (two measures for different types of knowledge and a measure of overall knowledge) should be used together.

In this study in order to compare levels of overall knowledge about HIV, Score Three or Four both were useful. In order to compare levels of two separate types of HIV knowledge, Scores One and Two were useful. Different scores can serve different purposes and sometimes they are not interchangeable and need to be used together.

The main limitation of Scores Three and Four is their inability to discriminate between different patterns of knowledge for the majority of groups of respondents apart from quantitative patterns. Therefore, they are best used when respondents should be located on a continuum of HIV knowledge.

It is a well-known fact in survey research that there will always be people answering "yes" to all questions. This effect is called acquiescence or agreeing-response bias, or "[i]n other words, it is the tendency of some respondents to agree with an item irrespective of the content of that item" (Billiet and Davidov 2008, p.543; Hamilton 1968; Cloud and Vaughan 1970; Bentler *et al*. 1971; Watson 1992). Response bias is a threat to every measurement regardless of cultural or epidemiological context (Moors 2004). However, the degree to which this effect is observed in different cultural contexts might have a link with a culture in a particular country.

In the Chinese context the correct answers to Score One are "yes", whereas for Score Two are "no". In the five country context the correct answers to Score One and one question from Score Two (healthy-looking person question) are "yes", whereas for the remaining questions for Score Two the correct answers are "no". Unfortunately, Score

Three in both contexts cannot help identifying individuals who had a tendency to display acquiescence as Score Three does not provide precise qualitative meaning for each individual as for different individuals having the same number of correct answers can represent different patters. This is another limitation of a simple score approach.

The main limitation of all four measures is the inability of the simple score measures to take weights into account for each contributing component which might be important in some research contexts.

All four measures used in the analysis are easily obtained in any statistical software package. This is the main advantage of a simple score approach to measuring HIV knowledge. On many occasions the information which is provided by these measures is sufficient for the researchers. These measures provide general picture of the levels of HIV knowledge and can be used successfully for the assessment of the evolution of HIV knowledge as well as for studying positions of different groups on the continuum between no knowledge and perfect knowledge. If more detailed information about patterns of HIV knowledge is required, other approaches to measuring HIV knowledge can be used, and this will be discussed in Chapter 5 of the thesis. Appendix H, Table H.1 summarises advantages and limitations of each of four scores used for the analysis in this paper.

4.5 Limitations of This Study

The current analysis has a number of limitations. The main limitation of the study is comparability across countries, especially for score variables, as slightly different questions were used for the calculation of scores in different countries and contexts (see section 2.3 for details). In some situations when a question of interest was not available in a survey it was replaced by another question which was assumed to be comparable. An example of this was the usage of a question about injections instead of sharing needles in China in 1997. However, the major problem was the way questions were asked in the Chinese surveys and in the DHS surveys which made it impossible to conduct cross-county comparison of levels of HIV knowledge using the same set of questions as for the comparison of HIV knowledge in China across times. As a result of this, two sets of scores were created to conduct the analysis of HIV knowledge.

Comparability issues exist also due to differences in questions used in different surveys across countries caused by translation of questions or cultural specificities.

Comparability issues might also exist owing to differences in cultural contexts and differences in, for example, standards of primary education in different contexts. Another problem which is related to comparability issues is that surveys used for the analysis were not conducted at the same point in time in different cultural contexts. However, all studies which use a number of countries for comparisons will face the limitations mentioned above (Van de Vijver and Leung 1997; Harkness 1999; Harkness *et al*. 2003).

The level of representativeness of the survey in China 2005 is another limitation of the study, as well as the lack of access to the data which were collected in 2006 for the National Family Planning and Reproductive Health Survey. The results of the modelling for China 2005 can not be generalised to the whole country, but they are indicative of the possibility that if the right interventions are in place, the level of HIV knowledge can be substantially improved in the whole country.

Another limitation of this study is the inability to control for potential random effects due to the limitations in the Chinese data (no cluster identifiers are available in 1997 and 2001) and due to the limitations of the gologit2 procedure in the five country context which was used to fit the partial proportional odds models in STATA. It is possible that standard errors are underestimated in the models presented in this paper and potentially more variables and interaction effects than necessary are included in the final models. However, this limitation will not present a problem for this study as the main interest in the analysis is to identify levels of HIV knowledge among specific groups of women (women with potentially the best and potentially the worst HIV knowledge) rather than identify separate effects of specific variables on the response variables.

Although simple scores of HIV knowledge provide us with general patterns of the distributions of HIV knowledge and despite the wide use of the simple score approach in the literature, it has a number of limitations. It treats all elements of the constructed score as equally-weighted, it considers the score as a single dimension construct (Ferguson *et al*. 1995), it does not take into account the fact that respondents might have guessed correct answers (Ferguson *et al*. 1995; Carey *et al*. 1998) or might have answered “yes” to every question, and it assumes that all answers represent the best and true level of knowledge of respondents. However, in every dataset a subset of the sample will be guessing and some people will be answering "yes" to all questions (for both correct and incorrect routes of HIV transmission). For example, in the 1997 survey 12.8% of respondents answered "yes" to all questions. In some research projects it might be important to identify these groups of people. However, the simple score approach does not allow this.

The simple score approach also assumes that perfect knowledge or in other words correct answers to all questions included into the score is the only best option for representation of an adequate HIV knowledge. However, it is possible that in some contexts the knowledge of one or two correct routes of HIV transmission (e.g., knowledge of sexual transmission and needle sharing) might be sufficient for effective HIV prevention.

Despite all these limitations, results of the analysis presented in this paper provide useful insights into understanding the evolution of HIV knowledge in China. Despite the limitations of a simple score approach to measuring HIV knowledge, it provided general patterns of HIV knowledge in different cultural and epidemiological contexts. Simple score measures can also be easily obtained in different statistical software packages and, therefore, this approach to measuring HIV knowledge should not be discarded because of the limitations mentioned above as it can provide useful information for assessment of progress in the area of HIV knowledge and for the design of general interventions and educational campaigns in different cultural and epidemiological contexts.

4.6 Conclusions

This paper analysed the evolution of HIV knowledge in China using a simple score approach to measuring HIV knowledge.

The descriptive analysis suggests that HIV knowledge differs by education, residence, age, ethnicity, marital status and year of survey in the Chinese context and by education, residence, age, marital status and country in the five countries used for the analysis.

The results of the regression analysis show that the level of HIV knowledge has increased in all groups of women in China between 1997 and 2005. Higher levels are observed in the knowledge of correct routes of HIV transmission than in the knowledge of incorrect routes of HIV transmission among different groups of women, and this is expected as the HIV epidemic in China is still at an early stage and not generalised and, therefore, HIV prevention knowledge is still more important than anti-discrimination knowledge. However, there is also progress in improvement of anti-discrimination knowledge in China among different groups of women, but this progress is not as pronounced as the progress with the knowledge about correct routes.

The current analysis suggests that the homogeneity of HIV knowledge across groups has increased in China over time which might be attributed to effective interventions and campaigns. However, despite this improvement in the levels of knowledge among women with potentially the best knowledge and women with the potentially the worst knowledge, the gap between these groups' levels of knowledge still exist and, therefore, more efforts are needed to ensure a high level of knowledge about HIV among different groups of women.

The results of comparison of HIV knowledge in the five countries show that the level of knowledge about ways to prevent HIV is comparable in China to those in countries with generalised HIV epidemics and sometimes the levels are even higher. However, the level of knowledge about misconceptions is still only comparable to the countries where the HIV epidemics are not generalised. These results suggest that more efforts need to be invested in improving knowledge about HIV misconceptions in countries where HIV epidemics are still not generalised, including China.

The results of the analysis conducted in the five country context also suggest that even women with potentially the worst knowledge in all five countries have some knowledge about HIV. However, the level of overall knowledge is still not adequate for many women in this group in all cultural contexts and these women should be targeted for educational campaigns and interventions.

The results of the analysis of the separate types of knowledge found that the level of knowledge about ways to prevent HIV transmission in Malawi was surprisingly low when compared with other countries. However, it is not clear from this research if the level was always low or if it decreased due to reduction in intervention efforts in the specific area of HIV knowledge. If the latter is true, this result suggests that efforts should be sustained to ensure a high level of knowledge about HIV prevention in different cultural and epidemiological contexts including mature epidemics and among different groups of women.

It can be concluded that China has succeeded in improving women's knowledge about HIV as we observe an increase in all four scores in different groups of women over time. However, there is still room for further improvements. Due to the representativeness level of the China 2005 data, the results for 2005 modelling cannot be generalised to the whole China. However, it can be argued that if the right interventions are in place in the whole countries, similar results (improved level and increased homogeneity of HIV knowledge between different groups of women) could be observed in the other parts of the country.

If achievements in other dimensions of HIV prevention are equally successful, China might become a success story in controlling HIV epidemic and might provide useful lessons for other countries in the world (for example, for countries in Easter Europe and Central Asia where HIV epidemics are still growing). In order to fully understand overall successes and failures of HIV prevention in China, other aspects of HIV prevention should be studied in order to be able to answer the question how the HIV epidemic in China has become controlled. However, as mentioned earlier, those questions are outside of the remit of this study.

This paper also studied the simple score approach to measuring HIV knowledge. Simple score measures proved to be useful measures when HIV knowledge evolution is studied. Simple score measures can help place groups of respondents on a continuum of HIV knowledge between no knowledge and perfect knowledge. However, simple score measures have number of limitations and the main limitation is that they do not provide detailed information about specific patterns of HIV knowledge which can be observed in a population. Therefore, it might be difficult to use the results of this analysis to design specific interventions in which specific types of knowledge were to be addressed if separate types of knowledge are not studied separately. However, the results would still be useful for designing general information interventions in which all aspects of information is provided at the same time. Simple score measures can be easily derived in any conventional statistical software package and can provide useful insights into different aspects of phenomena which surround HIV/AIDS in the world including HIV knowledge.

Box 4.1: Key findings in Paper Two

- China has succeeded in improving women's knowledge of HIV over time.
- HIV knowledge of both correct and incorrect routes of HIV transmission is increasing in China over time in all groups of women.
- HIV knowledge differs by education, residence, age, ethnicity, marital status and year of survey in China and education, residence, age, marital status and by country in the five countries.
- The level of knowledge about correct routes is higher than the level of knowledge about incorrect routes of HIV transmission in China.
- The level of knowledge increased in different groups of women with time but the gap in knowledge between different groups of women still exists in China. Therefore, more efforts are needed to improve levels of knowledge of those groups with low levels of knowledge.
- The level of knowledge about ways to prevent HIV transmission in China is comparable to the levels in countries with both generalised and non-generalised HIV epidemics.
- The level of knowledge about misconceptions about HIV transmission in China is comparable to the levels in countries with non-generalised HIV epidemics but not yet with generalised HIV epidemic.
- The level of knowledge about ways to prevent HIV transmission in Malawi (a country where the HIV epidemic is generalised) is relatively low. Therefore, this result suggests the need for sustainable HIV knowledge interventions even in countries where HIV epidemics are mature.
- A simple score approach produced an easily derivable HIV knowledge measure which can be successfully used when the evolution of HIV knowledge is assessed. It can also be useful in many different analyses in the area of HIV/AIDS.

Chapter 5: Measuring HIV Knowledge: A Latent Variable Approach

Abstract

This paper focuses on the evolution of HIV knowledge among women in China. The main aim of this paper is to assess whether China has succeeded in improving women's HIV knowledge between 1997 and 2005, and if China is a relative success story in improving women's HIV knowledge when compared with other countries in the world with generalised as well as with non-generalised HIV epidemics. This paper also aims to evaluate advantages and limitations of the different latent variable measures of HIV knowledge and then compares these latent variable measures to simple score measure discussed in Chapter 4 of the thesis. The following data sources are used for the analysis: the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001, the UNFPA Reproductive Health and Family Planning Survey 2005, and recent DHS surveys (India 2006, Kenya 2003, Malawi 2004 and Ukraine 2007). In order to assess whether China has succeeded in improving women's knowledge of HIV, HIV knowledge is first compared across times in China. To establish the relative success of China, HIV knowledge in China is then compared with HIV knowledge in India, Kenya, Malawi and the Ukraine. A latent variable approach is used for measuring HIV knowledge in this paper. Multinomial logistic regression modelling is employed for the analysis of patterns of HIV knowledge in China and in the five countries. The main findings indicate that China has succeeded in improving women's HIV knowledge. Knowledge about ways to prevent HIV in China is comparable to knowledge in other countries even with generalised epidemics. Knowledge about misconceptions is comparable to knowledge in other countries where epidemics are non-generalised. Knowledge in China has improved and has become more homogeneous over time but differences in levels between different groups of women still exist and, therefore, more efforts should be directed towards improvement of knowledge about HIV among women in China.

The main methodological findings show that both simple score and latent variable approaches to measuring HIV knowledge are useful in studying the evolution of HIV knowledge and provide unique insights into the topic along with complementary information about the evolution of HIV knowledge in China.

5.1 Introduction

HIV knowledge cannot be observed or measured directly but it can be measured with the help of the various components which can be observed and measured directly such as knowledge of specific routes of HIV transmission (see Figures 1.2, 2.1 and 2.2). Therefore, HIV knowledge should be treated as a latent variable. There are two main types of latent variables – continuous and categorical. Continuous latent variables measure a concept on a continuous scale. Sometimes a latent concept can be uni-dimensional and it would be represented as one continuous latent variable but in some cases a latent concept can be multi-dimensional and this dimensionality is represented by more than one continuous latent variable (Bartholomew *et al*. 2008; Everitt and Dunn 1991). Categorical latent variables split samples into classes in accordance with participants' response patterns and identify more prevalent classes which exist in populations.

5.1.1 Latent variable measures of HIV knowledge in literature

This section reviews two approaches to measuring and analysing HIV knowledge employed in the existing literature: the latent trait approach and the latent class approach.

Latent trait analysis (LTA) and latent class analysis (LCA) are widely used in different health-related disciplines and areas: HIV-related research (HIV knowledge and attitudes, HIV-related stigma and discrimination, HIV-related fear, HIV counselling and testing) (Carey *et al*. 1997; Davis *et al*. 1999; Burkholder *et al*. 1999; Nyamathi *et al*. 2000; Dias 2001; Ezedinachi *et al*. 2002; Beadnell *et al*. 2003; Kattumuri 2003; Stein and Li 2008; Genberg *et al*. 2008; Collins 2009; Genberg *et al*. 2009), sexually transmitted infections and risky behaviours (Jaworski and Carey 2007; Stuart and Hinde 2009), smoking and tobacco dependence (Chen *et al*. 2004; Storr *et al*. 2004; Storr *et al*. 2005), drug use (Monga *et al*. 2007), depression (Sullivan *et al*. 1998), eating disorders (Bulik *et al*. 2000), and others.

In some studies these analyses are the main interest of the studies and help to determine dimensionality of unobserved concepts (latent trait analysis) and to refine scales or to identify number of homogeneous classes in heterogeneous populations (latent class analysis) (Ferguson *et al*. 1995; Ambati *et al*. 1997; Carey *et al*. 1997; Carey *et al*. 1998; Koopman and Reid 1998; Sullivan *et al*. 1998; Davis *et al*. 1999; Bulik *et al*. 2000; Nyamathi *et al*. 2000; Dias 2001; Ezedinachi *et al*. 2002; Kattumuri 2003; Chen *et al*. 2004; Storr *et al*. 2004; Storr *et al*. 2005; Jaworski and Carey 2007; Monga *et al*. 2007; Stein and Li 2008; Genberg *et al*. 2008; Stuart and Hinde 2009; Genberg *et al*. 2009). In other studies variables created with the help of either latent trait or latent class analysis are used for further analysis within regression modelling or structural equation modelling frameworks (Burkholder *et al*. 1999; Beadnell *et al*. 2003; Collins 2009).

Latent trait analysis is frequently referred to as factor analysis or exploratory factor analysis in the literature. By latent trait analysis we mean here the analysis when unobserved latent variable is continuous and observed variables are categorical.

One of the main aims of the studies which used latent trait analysis and examined HIV knowledge in different contexts was to determine if HIV knowledge has a single dimension or if it is a multi-dimensional construct (Ferguson *et al*. 1995; Ambati *et al*. 1997; Carey *et al*. 1997; Carey *et al*. 1998; Koopman and Reid 1998; Davis *et al*. 1999; Genberg *et al*. 2008; Genberg *et al*. 2009). Different studies used different sets of questions while constructing HIV knowledge factors and, therefore, slightly different solutions were obtained. Carey *et al*. (1997) and Carey *et al*. (1998) obtained a uni-dimensional instrument for measuring of HIV knowledge. Ambati *et al*. (1997) obtained a three-dimensional solution (true transmission items, false transmission items group 1 (sneezing, mosquitoes, dirty water) and false transmission items group 2 (kissing, sharing utensils, toilet and clothes, handshaking)). Koopman and Reid (1998) also obtained a three-dimensional solution with slightly different factors (medical/scientific knowledge, myths of HIV transmission, knowledge of high-risk and prevention behaviours). Davis *et al*. (1999) obtained a four-dimensions solution with the following factors identified: transmission myths, attitudes, personal vulnerability and facts. In each of the multi-dimensional solutions reviewed here there are two common factors which can be found: true transmission routes and myths about HIV transmission.

Only two of the above mentioned studies were conducted using Chinese data (Davis *et al*. 1999; Stein and Li 2008). However, only one study was interested specifically in HIV-related knowledge and attitudes (Davis *et al*. 1999).

The main aim of the studies which used latent class analysis was to determine the number of homogenous segments in a heterogeneous population and to split population into these homogenous groups or classes. The majority of studies that used latent class approach are from areas other than HIV/AIDS in general or HIV knowledge in particular (Sullivan *et al*. 1998; Bulik *et al*. 2000; Chen *et al*. 2004; Storr *et al*. 2004; Storr *et al*. 2005; Monga *et al*. 2007). However, three studies which investigated HIV-related issues and used latent class analysis were found (Dias 2001; Beadnell *et al*. 2003; Kattumuri 2003). The first study investigated the question of attrition from an HIV/STD group counselling (Beadnell *et al*. 2003). Two other studies examined HIV-related knowledge and the main aim of these two studies was to allocate respondents to classes of knowledge based on their responses to the HIV transmission routes questions (Dias 2001; Kattumuri 2003). Kattumuri (2003) investigated patterns in HIV knowledge among HIV positive patients in India (Tamil Nadu), whereas Dias (2001) studied patterns in HIV knowledge among the general population in Brazil using the DHS 1996 survey. The first study extracted three mutually exclusive and exhaustive classes: low, medium and high class of HIV knowledge (Kattumuri 2003), and the second study obtained a four-class solution and classes were defined on the basis of distance from the biomedical model[38]: large lack of knowledge or the most distant group from the biomedical model, some dissonance with the biomedical model (above average knowledge but some problems will need to be handled), largest distance to the biomedical model (lack of information about specific routes of transmission), and closer to the biomedical model group (Dias 2001).

As mentioned earlier, no studies examined HIV knowledge in the general population in China using latent class analysis approach or studied HIV knowledge in general population in China over time or compared levels of HIV knowledge in China to the levels in other countries in order to place China into the world context. Therefore, the current research will fill these gaps as it aims to study HIV knowledge in the general

[38] "By biomedical model, we mean the medical truth according to the western medical science" (Dias 2001, p.10).

population in China, and to compare the levels of knowledge with other countries with generalised and non-generalised HIV epidemics by applying latent variable approach to measuring HIV knowledge.

5.1.2 Aims, research questions and structure of Paper Three

The main aims of this paper are to derive latent measure(s) of HIV knowledge and then to model these measures in order to identify predictors and to compare patterns of HIV knowledge in China over time and in the cross-country context among different groups of women. This paper also aims to identify if HIV knowledge is a uni-dimensional or multi-dimensional construct and what are the predominant HIV knowledge classes which exist in two study contexts. This paper then focuses on the comparison of the latent variable approach to measuring HIV knowledge with the simple score variable approach which was discussed in Chapter 4 of the thesis.

Therefore, the main research questions are:

1. Has China succeeded in improving women's knowledge of HIV/AIDS over time?
2. Is China a relative success story in improving women's knowledge of HIV/AIDS when compared with other countries with generalised and non-generalised HIV epidemics?
3. Are the results about the evolution of HIV knowledge in China similar when two different approaches (simple score and latent variable approaches) are applied to measuring HIV knowledge?
4. What are advantages and limitations of different latent variable measures of HIV knowledge?
5. How do the simple score and latent variable approaches to measuring HIV knowledge compare?

In order to answer the research questions, this paper will

1. derive HIV knowledge latent trait continuous measure in the Chinese and the five country contexts and to examine if HIV knowledge is a uni-dimensional or multi-dimensional construct;
2. derive HIV knowledge latent class categorical measure and to identify how many classes better represent HIV knowledge level observed in the Chinese context and in the five country context;

3. compare the patterns of HIV knowledge at various points in time across population sub-groups between 1997 and 2005 in China;
4. compare HIV knowledge in China with the levels of HIV knowledge in four other countries: Kenya and Malawi (countries where HIV epidemics are generalised), and India and the Ukraine (countries with non-generalised HIV epidemics);
5. and finally, will compare the results from the latent variable approach to measuring HIV knowledge with results from the simple score variable approach discussed in Paper Two and to discuss advantages and limitations of both approaches.

The main motivation for using the latent variable approach to measuring HIV knowledge is to obtain a complete set of possible measures of HIV knowledge in order to be able to compare the available range of measures (simple score measures and latent variable measures) and to compare results about HIV knowledge evolution obtained when two different approaches to measuring HIV knowledge are applied. This approach to measuring HIV knowledge might help to overcome some limitations of the simple score approach discussed in section 4.4.3 and to understand the evolution of HIV knowledge in China and across different populations better.

In order to compare HIV knowledge across time and in a cross country context, the same assumptions as in Paper Two (see section 4.1.2 for details) are used. It is assumed that in the countries where HIV epidemic is mature the level of knowledge about HIV is higher than in countries where the epidemic is still at early stage. It is also assumed that the higher levels of both types of HIV knowledge are observed in countries with generalised HIV epidemics but a higher level of HIV knowledge about ways to prevent HIV and a lower level of knowledge about misconceptions are observed in countries with non-generalised HIV epidemics.

To investigate progress in China with improving women's knowledge about HIV, latent trait (continuous) and latent class (categorical) variables are first derived and then used for further analysis in a regression framework. As discussed in section 2.3, two sets of questions are used to enable comparison of HIV knowledge and to ensure consistency and comparability: one set for investigating HIV knowledge in China over time and the second set of questions for comparing HIV knowledge in different cultural and epidemiological contexts. Unfortunately, it was not possible to use the same set of questions for two parts of the analysis due to differences in data collection processes in different surveys.

This paper consists of six main sections. In section two the methodology used for the analysis is presented. In section three all results of the analysis are presented and summarised. Section four discusses comparison of the simple score approach and the latent variable approach to measuring HIV knowledge. In section five, the limitations of latent variable approach to measuring HIV knowledge are discussed. Finally, in section six, the main conclusions are presented.

5.2 Methodology

5.2.1 Latent variables

There are numerous concepts in social sciences which cannot be measured directly: intelligence, political attitudes, HIV knowledge, socio-economic status. However, these variables, which are called latent variables in the literature can be derived with the help of observed variables which can be measured directly in the surveys. These observed variables are called either indicator variables or manifest variables (Bartholomew *et al.* 2008).

As mentioned earlier, HIV knowledge is a latent variable as it cannot be measured directly but only through the indicator variables such as, for example, knowledge of separate routes of HIV transmission (see Figures 1.2, 2.1 and 2.2). Latent variables help to reduce the number of dimensions and to summarise data which come from a number of indicator variables into one (or more) summary measure(s) without significant loss of information (Bartholomew *et al.* 2008). Latent variables can be either continuous or categorical. Manifest variables can also be either continuous or categorical. In the situation when all manifest variables are categorical, latent trait analysis is used in order to obtain a continuous latent variable or latent class analysis is used in order to obtain a categorical latent variable. These latent variables once derived can later be used for further analysis.

This paper focuses on both analyses: latent trait analysis and latent class analysis. Conducting both analyses helps to understand how many dimensions can better represent HIV knowledge and how many and what predominant groups of HIV knowledge exist in the population selected for analysis.

In order to conduct latent trait and latent class analysis and to extract continuous and categorical latent variables the M-Plus statistical software version 5.1 (Muthén and Muthén 2009) is used. This statistical software package was specifically designed for use of continuous and categorical latent variables in cross-sectional as well as longitudinal contexts for single- and multi-level data (Muthén and Muthén 2009).

5.2.1.1 Latent trait analysis

When manifest variables are categorical and latent variable is continuous, the relationship between manifest variables and latent variable are established with the help of latent trait models. Latent trait analysis is a model-based technique which allows inferences to be made about a population and helps to explain data in a smaller number of dimensions (through extraction of a smaller number of factors) without significant loss of information (Everitt and Dunn 1991; Hoyle and Duvall 2004; Bartholomew *et al*. 2008). One of the main tasks of the latent trait analysis is the determination of the number of necessary underlying factors which have to be extracted (Hoyle and Duvall 2004). After determining the number of necessary underlying factors, latent trait models assign score(s) to each respondent for each latent variable (these variables will be referred to as factors and associated scores as factor scores) on the basis of their responses (Bartholomew *et al*. 2008). These scores can then be further used either as a response or as an explanatory variable. The main aims of the current analysis are first to determine dimensionality of HIV knowledge in the contexts analysed and then to extract scores for each individual in the datasets in order to be able to use them for identification of predictors of different components of HIV knowledge and to study changes in patterns of HIV knowledge among two specific groups of women (women with potentially the best and potentially the worst HIV knowledge).

According to Bartholomew *et al*. (2008), there are two main approaches for constructing and fitting a latent trait model: underlying variable (UV) approach and item response theory (IRT) approach. The results which are obtained through these two approaches are the same as these two types of model are equivalent for binary data (Bartholomew *et al*. 2008).

The main assumption of the UV approach is that observed manifest variables "…are assumed to be realisations of fictitious continuous underlying variables" (Bartholomew *et al*. 2008, p.224). These variables are incompletely observed variables as the main interest of the model is to check if they exceed some threshold or not (Bartholomew *et al*. 2008).

The underlying variable method can be mathematically expressed as follows:

$$x_i^* = \alpha_{i1}^* f_1 + \alpha_{i2}^* f_2 + \ldots + \alpha_{iq}^* f_q + e_i \ , \qquad (5.1)$$

where α_{ij}^* are the factor loadings, the f_j are the latent variables, the e_i are the residuals with zero mean and variance δ^2, *i*=1,…,*p*; *j*=1,…,*q* (where *p* is the number of manifest variables and *q* is the number of extracted factors), for each binary variable x_i it is assumed that there is an incompletely observed continuous variable x_i* which is normally distributed with mean μ_i and variance δ_i^2. The parameter τ_i is called threshold parameter and when x_i^* takes values below this parameter, the binary item x_i takes the value of 1, otherwise, it takes the value of 0 (Bartholomew *et al*. 2008).

Item response theory models use the logit model and can be mathematically expressed as follows:

$$\text{logit}\pi_i(f) = \log_e \frac{\pi_i(f)}{1-\pi_i(f)} = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} f_j \ , \qquad (5.2)$$

where the α_{ij} is factor loading known as the discrimination parameter, *j*=1,…, q and $f = (f_1, \ldots, f_q)$ (Bartholomew *et al*. 2008). "The larger the value of α_{ij}, the greater is the effect of factor *j* on the probability of a positive response to item *i* (Bartholomew *et al*. 2008).

Factor loadings are standardised and can be interpreted as correlation coefficients or weights as in factor analysis and, therefore, relative importance of components can be assessed (Bartholomew *et al*. 2008).

It is important to determine number of underlying factors in the latent trait model which can be further used for the analysis as response or explanatory variables. In order to determine number of factors, model selection procedure and global goodness-of-fit test can be used (Bartholomew *et al*. 2008). Two tests can be used to test goodness-of-fit of the model: the log-likelihood ratio test statistic

$$L^2 = 2\sum_{r=1}^{2^p} O(r)\log_e \frac{O(r)}{E(r)}\ , \qquad (5.3)$$

or the Pearson chi-squared goodness-of-fit statistic

$$\chi^2 = \sum_{r=1}^{2^p} \frac{(O(r)-E(r))^2}{E(r)}\ , \qquad (5.4)$$

where *r* is a response pattern, *O(r)* are observed frequencies of response pattern, *E(r)* are expected frequencies of response pattern, and *p* is a number of manifest variables (Bartholomew *et al*. 2008).

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) can also be used in determining of the numbers of underling factors. These two approaches will be discussed in more detail in the latent class analysis (LCA) subsection as they are primary methods for determination of the number of groups in LCA.

When more than one factor is extracted, the solution is not unique. Rotation of factors can be used. The likelihood statistic does not change if a rotation of the factors is applied. Rotation can help to express the original solution in a more interpretable way (Bartholomew *et al*. 2008). There are two main types of rotation: orthogonal and oblique. The main difference between the two is that the first one keeps the underlying factors uncorrelated whereas the second one allows the factors to be correlated (Bartholomew *et al*. 2008; Field 2009). According to Bartholomew *et al*. (2008), “the Varimax procedure attempts to find an orthogonal rotation that is close to simple structure by finding factors with few large loadings and as many near-zero loadings as possible” (p. 191).

Once a latent trait model is fitted and underlying factors are determined, factor scores should be extracted. M-Plus obtains factor scores for each individual (Muthén 1998-2004; Muthén and Muthén 2009).

Once extracted, the scores can be used for further analysis, for example, for identification of determinants of different dimensions of HIV knowledge in different cultural and epidemiological contexts.

5.2.1.2 Latent class analysis

Latent class modelling was first introduced by Lazarsfeld and Hendry (1968) and later the methodology was extended by Goodman (1974a; 1974b). In latent class analysis observed variables are categorical and the single latent variable is also categorical.

Similar to latent trait analysis, LCA is a model-based technique which allows inferences about a population to be made and helps to summarise data in the form of one latent variable with a number of classes or profiles without significant loss of information (McCutcheon 1987; Magnusson 1998 cited in Beadnell *et al*. 2003; Hagenaars and McCutcheon 2002; Bartholomew *et al*. 2008). Latent class analysis can help to split a heterogeneous sample into classes which are more homogeneous with respect to the latent variable, and within these mutually exclusive and exhaustive classes people will share similar characteristics (Clogg and Goodman 1984; Stuart and Hinde 2009). The main aim of LCA is to determine the smallest number of latent classes that is sufficient to explain relationships between the manifest variables (Magidson and Vermunt 2004). LCA can help identifying groups of women which should be targeted by interventions as they require specific information about HIV.

In contrast to factor analysis or latent trait analysis which provides scores for respondents for each factor extracted, latent class analysis provides a classification of individuals (Muthén and Muthén 2009).

In this analysis LCA was used to empirically investigate patterns of HIV knowledge among women of reproductive age in different cultural and epidemiological contexts.

If there are four manifest or observed variables (A, B, C and D), then latent class model can be expressed as

$$\pi_{ijklt} = \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X} \qquad (5.5)$$

where π_{ijklt} is the probability that a person who gave response *i* to item *A*, response *j* to item *B*, response *k* to item *C* and response *l* to item *D* is in latent class *t* of latent variable *X*, π_t^X denotes the probability of being in latent class $t = 1,2,\ldots,T$ of latent variable *X*; $\pi_{it}^{A|X}$ denotes the conditional probability of obtaining the i^{th} response to item *A*, from members of class *t*, $i=1,2,\ldots,I$; and $\pi_{jt}^{B|X}$, $\pi_{kt}^{C|X}$, $\pi_{lt}^{D|X}$, $j=1,2,\ldots,J$, $k=1,2,\ldots,K$, $l=1,2,\ldots L$, denote the corresponding conditional probabilities for items *B*, *C* and *D*, respectively (Magidson and Vermunt 2004).

Latent class model for binary variables have the following assumptions:

"i) The *n* individuals are a random sample from some population and every individual in that population belongs to just one of the *T* latent classes of the latent variable *X*.
ii) The probability of giving a positive response to a particular item is the same for all individuals in the same class but may be different for individuals in different classes.
iii) Given the latent class to which an individual belongs, its responses to different items are conditionally independent." (Bartholomew *et al*. 2008, p.272-273). This assumption is also known as a local independence assumption (Hagenaars and McCutcheon 2002; Magidson and Vermunt 2004).

As mentioned earlier, one of the main tasks of fitting of the latent class model is to determine the number of homogeneous classes which exists in the heterogeneous population with respect to the latent variable. In order to do this the model fit should be assessed.

One of the approaches used to determine the number of classes (or the model fit) is the likelihood ratio statistics (Magidson and Vermunt 2004). A model fits the data well if the value of L^2 (see equation (5.3)) is sufficiently low (Magidson and Vermunt 2004). However, this approach has problems as in latent class models the likelihood ratio statistic does not always conform to the chi-squared distribution and as the result of that the obtained test statistic becomes an unreliable measure (Storr *et al*. 2004; Stuart and Hinde 2009). The following information criteria provide more reliable measures and will be used instead of the likelihood ratio test: the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC) (Dias 2001; Akaike 1974; Schwarz 1978; Bozdogan 1987; Muthén 1998-2004; Singer and Willet 2003; Magidson and Vermunt 2004; Stuart and Hinde 2009). For all these criteria smaller values represent a better fit of the model (Dias 2001; Field 2009). The AIC is the goodness-of-fit statistic corrected for the complexity of the model by taking into account the number of parameters which were estimated (Field 2009). The BIC is similar but more conservative than the AIC (Field 2009). This statistic balances two components of a model, the likelihood value and parsimony (Muthén and Muthén 2000). Mathematical expressions for the different criteria can be found below:

$$AIC = L^2 + 2M \tag{5.6}$$

$$BIC = L^2 + \ln(N)M\ , \tag{5.7}$$

where L^2 is the log-likelihood ratio test statistic, M is the number of independent parameters for estimated model, and N is the sample size (Muthén 1998-2004; Dias 2001; Singer and Willet 2003; Magidson and Vermunt 2004;).

The level of potential classification error or classification quality is also important to consider when deciding on the final model (Muthén and Muthén 2000; Storr *et al*. 2004; Stuart and Hinde 2009). According to Beadnell *et al*. (2003), classification quality is the ability to distinguish membership in the latent class given the model and the data. The higher the average class probabilities the better the ability to accurately classify individuals into their classes (Beadnell *et al*. 2003).

According to Storr *et al*. (2004), model fit can be improved by adding more latent classes but this additional class may make the model less interpretable. Therefore, it is important to use the judgement, the principal of parsimony and the existing literature when deciding on the final model.

The latent class model "uses the observed data to estimate two sets of parameters: the conditional response probabilities and the latent class prevalences" (Stuart and Hinde 2009, p.3). These parameters can help identification of latent classes.

Once the decision about number of classes is taken, individuals are allocated to the appropriate latent classes on the basis of their responses. This can be done with the help of estimated posterior probabilities. The posterior probability is the probability of an individual being in class t given a specific response pattern of an individual, where t is a latent class of a latent variable X and $t=1,\ldots,T$ (Bartholomew *et al*. 2008).

Posterior probabilities are estimated and then observations are assigned to a class for which the posterior probability is the highest (Magidson and Vermunt 2004). Once the posterior probabilities are estimated, every individual in the dataset can be assigned to a particular class using M-Plus (Muthén and Muthén 2009) and this new variable which contains classes to which individuals were allocated can be used for further analysis as a response or as an explanatory variable.

The created class variables are used as response variables in this paper in order to identify determinants of HIV knowledge and then to compare HIV knowledge across times in China and to compare the levels of HIV knowledge in China with levels in other countries among different groups of women. Binary or multinomial logistic regressions are used to model HIV knowledge as the class variables are nominal[39]. Different types of logistic regression modelling, model selection process, interaction and clustering effects are discussed in sections 3.2.2 and 4.3.3-4.3.6.

5.3 Results

Demographic and other characteristics of women together with their HIV knowledge in seven separate datasets used for the analysis in this paper are presented in Appendix F, Table F.1.

5.3.1 Exploratory data analysis

The Chinese context

Tables 5.1 and 5.2 show all possible patterns of answers to the correct and incorrect routes questions in the separate China datasets and also in the pooled dataset. As there

[39] "Variables having categories without a natural ordering are called nominal" Agresti (2002, p.2).

are four questions about correct routes of HIV transmission, the total number of possible patterns is 16 (2^4). The patterns are presented in the table by the percentages of people who belong to every group starting from the largest group in the pooled dataset. In Table 5.1 and similar following tables, "1" means that the question is answered correctly and "0" otherwise. Table 5.1 shows that in the pooled dataset the largest group contains respondents who answered correctly all four questions about correct routes of HIV transmission (70.90%). Table 5.1 also shows that this pattern of response is the most sizable in all three separate datasets and the percentage increases over time. This suggests a high level of knowledge of correct routes of HIV transmission among respondents in China between 1997 and 2005. The group in which respondents answered incorrectly to all the questions about correct routes of HIV transmission in the pooled dataset represents 4.42%. This suggests that there still were people who completely lacked knowledge of the correct routes of HIV transmission in China between 1997 and 2005. The other two relatively large groups in the pooled dataset are women who answered correctly to three routes and incorrectly to the MTCT question (6.21%) or to the question about needles (4.70%).

Table 5.1: Patterns of answers to correct routes of HIV transmission questions in separate China datasets and in the pooled dataset.

Blood transfusion	MTCT	Sex	Needles	China 1997 %	China 2001 %	China 2005 %	Pooled dataset %
1	1	1	1	51.27	75.67	80.39	70.90
1	0	1	1	6.15	6.76	4.18	6.21
1	1	1	0	8.96	2.99	5.22	4.70
0	0	0	0	10.67	2.75	1.99	4.42
0	0	1	0	7.11	1.09	0.73	2.40
0	1	1	1	2.04	2.64	1.39	2.30
1	0	1	0	3.34	1.12	1.65	1.71
0	1	1	0	5.65	1.28	0.54	2.15
1	1	0	1	1.09	2.05	1.49	1.74
1	0	0	1	0.90	1.13	0.59	1.00
1	0	0	0	0.68	0.49	0.56	0.54
0	0	1	1	0.61	0.87	0.48	0.75
1	1	0	0	0.56	0.25	0.41	0.35
0	1	0	1	0.16	0.37	0.23	0.30
0	1	0	0	0.67	0.28	0.01	0.32
0	0	0	1	0.15	0.26	0.13	0.21
Total sample size				9653	26138	6837	42628

Table 5.2 shows all possible patterns of answers to the three questions about incorrect routes of HIV transmission (8 patterns) in the separate China datasets and in the pooled dataset. The two dominant groups in the pooled dataset are those who either answered

correctly to all three questions (25.23%) or who answered incorrectly to all 3 questions (31.31%) or in other words people with perfect knowledge of incorrect routes of HIV transmission and people with no knowledge about incorrect routes of HIV transmission. The following two patterns are a group who knows about handshake but does not know about the other two routes (17.68%) and a group who knows about handshake and sharing utensils but does not know about kissing (14.17%).

Table 5.2 also shows all possible patterns of answers to the three questions about the incorrect routes of HIV transmission at three different points in time in China. The percentage of people having perfect knowledge about incorrect routes of HIV transmission increases over time and the percentage of women without knowledge about incorrect routes of HIV transmission decreases over time.

Tables 5.1 and 5.2 suggest an improvement of knowledge about correct and incorrect routes of HIV transmission over time in China. They also show that the level of knowledge about the correct routes of HIV transmission is higher than about the incorrect routes of HIV transmission and this result is expected as the HIV epidemic in China is still at the early stage and, therefore, more efforts were so far concentrated on the improvement of HIV prevention knowledge than on the anti-discrimination knowledge.

Table 5.2: Patterns of answers to incorrect routes of HIV transmission questions in separate China datasets (1997-2005) and in the pooled dataset.

Handshake	Utensils	Kissing	China 1997 %	China 2001 %	China 2005 %	Pooled dataset %
0	0	0	43.60	30.95	15.31	31.31
1	1	1	15.90	23.18	46.22	25.23
1	0	0	15.95	19.89	11.67	17.68
1	1	0	11.98	14.27	16.89	14.17
1	0	1	5.58	7.10	4.91	6.41
0	1	0	5.60	3.53	1.97	3.75
0	0	1	0.77	0.65	1.59	0.83
0	1	1	0.61	0.42	1.42	0.63
Total sample size			9653	26138	6837	42628

Table 5.3 shows selected patterns of answers to both correct and incorrect routes of HIV transmission questions together in the separate and the pooled datasets in China. All possible combinations of answers can be presented in 128 patterns (2^7) and it is not practical to present percentages for every group. Therefore, only six selected patterns

which are considered to be of interest are presented in the table. The largest group in the pooled dataset among selected patters on responses represents the pattern in which respondents answer correctly to all seven questions (20.39%).

The second largest group represents the pattern in which respondents answered correctly to all questions about correct routes of HIV transmission but incorrectly to all incorrect routes of HIV transmission questions (17.50%). This group contain some respondents who answered "yes" to every question[40] which can be either their genuine belief of the best possible answer and genuine knowledge or some of them might just be answering "yes" to all questions (this effect is called acquiescence or agreeing-response bias (see section 4.4.3 for details)). The sub-group of women who answered "yes" to all questions can be identified and will be referred to as "yes-women". It is important to investigate percentages of women who agree with all questions. Table 5.4 shows percentages of women in the group who answered "yes" to all questions among all women who were aware of HIV and answered all seven questions about components of HIV knowledge. Table 5.4 shows that high proportions of women were answering "yes" to all questions especially in 1997 and 2001. It possibly could be argued that it is a cultural trait in China to answer "yes" to all questions. However, as this proportion is considerably lower in 2005, it suggests that the majority of women in this group were providing genuine answers. The level of knowledge about incorrect routes of HIV transmission was lower at the beginning of the epidemic and the level of fear about HIV was higher. All these can explain the large proportion of women who belong to this group in 1997 and 2001.

Women who answered incorrectly to all seven questions in the pooled dataset represent 4.17% of the sample. These women do not have any knowledge about HIV.

Table 5.3 also shows that the percentages of respondents with perfect knowledge of both correct and incorrect routes of HIV transmission increase over time in China and the percentage of women without any knowledge about HIV transmission decrease with time.

[40] Other women in this group will have combinations of "yes" and "do not know" answers as for incorrect routes questions the incorrect answers are "yes" and "do not know" (see section 2.3 for details).

Table 5.3: Selected patterns of answers to all HIV knowledge questions in separate China datasets (1997-2005) and the pooled dataset.

BT	MTCT	S	N	H	U	K	China 1997 %	China 2001 %	China 2005 %	Pooled dataset %
1	1	1	1	1	1	1	9.45	19.26	40.18	20.39
1[41]	1	1	1	0	0	0	16.87	20.10	8.50	17.50
1	1	1	1	1	0	0	10.91	16.12	9.55	13.89
1	1	1	1	1	1	0	8.52	11.80	14.79	11.53
0	0	0	0	0	0	0	10.14	2.60	1.74	4.17
1	0	1	1	1	1	1	1.23	1.48	1.46	1.42
0	0	0	0	1	1	1	0.16	0.04	0.13	0.08
Total sample size							9653	26138	6837	42628

Note: BT stands for blood transfusion, MTCT for mother-to-child transmission, S for sexual transmission, N for needles, H for handshake, U for utensils, K for kissing.

Table 5.4: Percentages of women who answered "yes" to all questions among those who answered to all seven constructs of HIV knowledge.

	China 1997	China 2001	China 2005
"yes-women"	12.8%	14.4%	3.7%
Total sample size	9653	26138	6837

The five country context

Tables 5.5, 5.6 and 5.7 show patterns of answers to questions on knowledge about ways to prevent HIV, HIV misconception knowledge and then selected patterns of combination of both types of knowledge in separate countries and in the pooled dataset for the five countries.

Table 5.5 shows that 47.91% of all women in five countries have perfect knowledge about ways to prevent HIV, whereas 5.11% has no knowledge about HIV prevention.

When countries are examined separately, Table 5.5 suggests that the percentage of women with the perfect knowledge about HIV prevention is the highest in the Ukraine and the lowest in Malawi. The percentage of women without any knowledge about HIV prevention is the highest in India and the lowest in Kenya and the Ukraine.

[41] This combination of answers suggests that some respondents in this group answered "yes" to all questions

Table 5.5: Patterns of answers to correct routes of HIV transmission questions in five countries' separate datasets and in the pooled dataset.

MTCT	OP	A	C	China 2005 %	Kenya 2003 %	Malawi 2004 %	India 2006 %	Ukraine 2007 %	Polled dataset %
1	1	1	1	61.90	60.22	39.08	45.10	72.75	47.91
1	1	1	0	4.33	20.84	12.93	8.90	2.18	9.32
1	0	0	0	2.97	0.52	6.68	7.78	1.64	6.65
1	1	0	1	12.78	3.37	4.47	6.06	6.64	6.17
0	1	1	1	3.57	3.64	6.23	5.36	5.75	5.26
0	0	0	0	3.10	0.19	2.17	6.36	0.48	5.11
1	1	0	0	2.40	2.17	4.81	4.90	1.13	4.38
1	0	1	1	2.37	2.66	6.54	3.16	3.75	3.43
1	0	1	0	1.24	1.85	4.78	2.62	1.18	2.61
0	1	1	0	1.23	2.56	2.70	2.31	0.21	2.18
1	0	0	1	1.70	0.52	3.34	1.86	2.20	1.93
0	1	0	0	0.35	0.44	1.27	2.10	0.18	1.72
0	1	0	1	0.88	0.29	1.03	1.34	0.63	1.18
0	0	1	0	0.51	0.36	1.76	0.85	0.24	0.85
0	0	1	1	0.26	0.31	1.53	0.68	0.57	0.71
0	0	0	1	0.41	0.07	0.66	0.62	0.45	0.57
Total sample size				6837	6877	11002	88223	6638	119577

Note: MTCT stands for mother-to-child transmission, OP stands for one partner, A for abstinence and C for condom.

Table 5.6 shows that 44.94% of the pooled sample answered correctly to all three questions about HIV misconceptions and 8.52% answered incorrectly to all these questions.

Table 5.6 also shows that the percentage of women with perfect knowledge about HIV misconceptions is the highest in Kenya and Malawi when compared to other four countries and the lowest in China. The percentage of women with no knowledge about HIV misconceptions is the highest in India and the lowest in Malawi.

Table 5.6: Patterns of answers to incorrect routes of HIV transmission questions in five countries' separate datasets and in the pooled dataset.

Utensils	Healthy-looking	Mosquito bite	China 2005 %	Kenya 2003 %	Malawi 2004 %	India 2006 %	Ukraine 2007 %	Polled dataset %
1	1	1	31.21	60.84	53.46	43.28	50.66	44.94
1	1	0	30.44	16.27	20.04	10.91	11.90	13.23
1	0	1	2.24	4.28	9.90	13.50	11.34	11.87
0	1	0	22.16	7.24	5.77	8.53	9.04	9.01
0	0	0	4.26	2.05	1.76	10.43	5.38	8.52
1	0	0	2.62	2.02	2.90	5.80	3.27	4.99
0	1	1	6.13	6.22	5.04	4.17	5.62	4.56
0	0	1	0.95	1.08	1.13	3.39	2.79	2.87
Total sample size			6837	6877	11002	88223	6638	119577

Table 5.7 shows that 27.91% of women in the pooled dataset have perfect knowledge about ways to prevent HIV as well as HIV misconception knowledge, and only 2.54% of women do not have any knowledge about these components of HIV knowledge. It also shows that 3.32% of women answered correctly to all questions apart from the utensils and mosquito bite questions. In this group some of women answered "yes" to all questions and, therefore, might not be giving genuine answers but might be agreeing with all questions and can be defined as "yes-women". Table 5.8 shows that the proportions of women who answered "yes" to all questions are relatively low in China and the Ukraine but slightly higher in other countries. This suggests that in China the majority of women in the group who answered "yes" to all ways to prevent HIV and the healthy-looking person question and "no" to the two main questions about HIV misconceptions had a mixture of "yes" and "do not know" answers and they were not simply agreeing while providing their answers.

Table 5.7 shows that the percentage of women with perfect HIV knowledge is the highest in the Ukraine and Kenya and the lowest in China and Malawi. The percentage of women without any knowledge about HIV is the highest in India and the lowest in Kenya.

Table 5.7: Selected patterns of answers to all HIV knowledge questions in five countries' separate datasets and in the pooled dataset.

MTCT	OP	A	C	U	H	M	China 2005 %	Kenya 2003 %	Malawi 2004 %	India 2006 %	Ukraine 2007 %	Polled dataset %
1	1	1	1	1	1	1	20.71	39.84	21.51	27.34	41.17	27.91
1	1	1	1	1	1	0	21.85	9.28	9.10	5.33	7.80	6.98
1	1	1	1	1	0	1	1.23	2.09	2.27	5.03	8.41	4.57
1	1	1	0	1	1	1	0.82	12.13	6.97	3.05	0.65	3.67
1[42]	1	1	1	0	1	0	11.28	3.74	2.71	2.60	5.26	3.32
1	1	0	1	1	1	1	5.35	1.85	2.18	3.03	3.22	3.03
0	0	0	0	0	0	0	0.94	0.06	0.12	3.33	0.18	2.54
0	1	1	1	1	1	1	0.83	1.80	2.89	2.10	1.58	2.05
1	0	0	0	1	1	1	0.39	0.17	4.25	1.68	0.29	1.68
0	1	1	1	1	0	1	0.09	0.26	0.93	1.20	0.95	1.04
0	0	0	0	1	0	1	0.04	0.03	0.70	0.87	0.03	0.71
Total sample size							6837	6877	11002	88223	6638	119577

Note: MTCT stands for mother-to-child transmission, OP stands for one partner, A for abstinence and C for condom, U for utensils, H for healthy-looking person question and M for mosquito question.

[42] This combination of answers suggests that some of the respondent in this group answered "yes" to all questions

Table 5.8: Percentages of women who answered "yes" to all questions among those who answered to all seven constructs of HIV knowledge in the five countries.

	China 2005	India 2006	Kenya 2003	Malawi 2004	Ukraine 2007
"yes-women"	0.3	1.7	2.1	1.8	0.9
Total sample size	6837	88232	6877	11002	6638

5.3.2 Latent trait analysis

LTA in a form of exploratory factor analysis (EFA) is conducted in order to extract HIV knowledge factor(s) and to understand if HIV knowledge is a uni-dimensional or a multi-dimensional construct in the Chinese context and in the five country context. It also helps to identify the relative importance of components of HIV knowledge in the extracted factors.

5.3.2.1 Dimensionality of HIV knowledge

The Chinese context

In China seven questions discussed in section 2.3 are used in order to construct HIV knowledge factors. Three models are fitted to the data (the one-factor model, two-factor model and three-factor model). The three-factor solution model fits the data well as the p-value associated with L^2 statistics is 0.109 and suggests that there is no significant difference between this model and the saturated model (Table 5.9). However, the three-factor solution produced one of the factors with a factor loading larger than 1 (1.115) (Table 5.10) and a negative residual variance which might suggest the over-extraction of factors and, therefore, the solution with one less factor should be accepted[43]. Therefore, the best solution in the Chinese context is the two-factor solution (Table 5.10). The change in deviance suggests that the two-factor model fits the data significantly better than the one-factor model (Table 5.9). Despite the fact that the model does not fit the data well, this model simplifies the solution from having seven HIV knowledge variables to having two HIV knowledge factors.

Table 5.9: Model fit statistics in the Chinese context (LTA).

Model	L^2 value	Degrees of freedom	p-value	Change in L^2	Change in df	p-value
One-factor solution	23210.434	14	0.000			

[43] Linda Muthén's answer from forum - http://www.statmodel.com/discussion/messages/8/178.html?1251818383 [Accessed 25 August 2010]

Two-factor solution	698.758	8	0.000	22511.676	6	0.000
Three-factor solution	6.064	3	0.109	692.694	5	0.000

One factor in the one-factor solution represents general HIV knowledge[44]. In the two-factor solution the first factor represents good knowledge of correct routes of HIV transmission whereas the second factor represents good knowledge of incorrect routes of HIV transmission. Therefore, it can be concluded, that in the Chinese context HIV knowledge is a two-dimensional construct. The literature discussed in section 5.1.1 suggests that the same factors were extracted in other contexts. The exploratory factor analysis function in M-Plus allows extracting scores for both factor.

Table 5.10: Factor loadings for different solutions in the Chinese context (LTA)[45].

	One-factor solution	Two-factor solution		Three-factor solution		
Blood transfusion	0.888	0.873	0.300	0.789	0.408	0.289
MTCT	0.705	0.798	0.136	0.605	0.488	0.111
Needle sharing	0.794	0.883	0.136	0.816	0.392	0.114
Sexual intercourse	0.781	0.783	0.273	0.282	1.115	0.174
Handshaking	0.845	0.353	0.854	0.281	0.259	0.841
Sharing utensils	0.733	0.201	0.779	0.131	0.201	0.769
Kissing	0.760	0.094	0.889	0.073	0.112	0.883

Dimensionality of HIV knowledge at different points in time in China

Investigation of dimensionality of HIV knowledge at different points in time in China can help exploring the evolution of HIV knowledge in the Chinese context and also can assess whether the factor extraction procedure in the pooled dataset is dominated by the large sample size of the China 2001 survey.

Detailed results from the analysis of separate points in time in China can be found in Appendix I, Tables I.1-I.6. The analysis of the separate datasets suggest that at every point in time the two-factor solution is extracted and they represent knowledge about the correct and incorrect routes of HIV transmission dimensions. In 1997 and 2005 the two-factor solutions are the best one[46]. In 2001 the three-factor solution can also be

[44] Hereafter the definitions of factors depend on the groups of questions with larger magnitudes of factor loadings.
[45] Varimax orthogonal rotation procedure is used.
[46] In China 1997 three-factor solution produced one of the factors with a factor loading greater than 1 (5.518) which suggest the over-extraction of factors (see Table I.2 in Appendix I).

obtained but the third factor is separating the knowledge about transmission of HIV through sexual intercourse from the knowledge about other correct routes of HIV transmission. The one factor solution at every point of time represents general HIV knowledge. These results suggest that the extraction of factors in pooled dataset reflects the dimensionality of HIV knowledge well at every point in time in the Chinese context and the process is not dominated by the large sample size of the 2001 survey.

The five country context

In a cross-country comparison the seven questions discussed in section 2.3 are used in order to construct factor(s) which would help to answer a question if HIV knowledge is a uni-dimensional or a multi-dimensional construct.

Three models are fitted to the data (the one-factor model, two-factor model and three-factor model). It is not possible to fit four-factor model because of the M-Plus rule for the maximum number of factors that can be extracted[47]. The calculation of the maximum number of factors that can be extracted in the study context can be found in Appendix J.

None of the three models fit data in the pooled dataset well as the p-value associated with L^2 statistics are 0.000 and these results suggest that there was a significant difference between every model and the saturated model (Table 5.11). However, all these models simplify the solution from having seven HIV knowledge variables to having one, two or three components of HIV knowledge. The change in deviance suggested that two-factor model fits the data significantly better than the one-factor model and the three-factor model fits the data significantly better than the two-factor model (Table 5.11). Therefore, it can be concluded that in the five country context the three-factor solution is the best one.

Table 5.11: Model fit statistics in the five country context (LTA).

Model	L^2 value	Degrees of freedom	p-value	Deviance of L^2	Difference in degrees of freedom	p-value
One-factor solution	21899.169	14	0.000			
Two-factor solution	4195.605	8	0.000	17703.564	6	0.000
Three-factor solution	63.425	3	0.000	4132.18	5	0.000

[47] http://www.statmodel.com/download/Topic%201.pdf [Accessed 25 August 2010] – see slide 104.

The factor from the one factor solution represents general HIV knowledge. In the two factor solution the first factor represents the knowledge about ways to prevent HIV together with the knowledge about the healthy-looking person question and the second factor represented knowledge about two main misconceptions (utensils and mosquito bites). It is interesting that this solution included one of the misconception knowledge questions into the first factor as it might suggest that this component of HIV knowledge is wrongly placed within the misconception knowledge questions but should be included into the ways to prevent HIV knowledge questions instead.

The best solution extracts three factors (Table 5.12) which can be defined as:
1. self-protection knowledge (one partner, abstinence and condom);
2. medical knowledge (MTCT and healthy-looking person);
3. misconception knowledge (utensils and mosquito).

The three-factor solution separates MTCT and healthy-looking person from the HIV self-protection knowledge factor in the two-factor solution into a new third factor.

Table 5.12: Factor loadings for different solutions in the five country context (LTA)[48].

	One-factor solution	Two-factor solution		Three-factor solution		
MTCT	0.546	0.476	0.273	0.326	0.547	0.189
One partner	0.780	0.799	0.204	0.784	0.189	0.201
Abstinence	0.752	0.779	0.188	0.775	0.162	0.188
Condom	0.768	0.728	0.291	0.701	0.213	0.280
Healthy-looking person	0.553	0.455	0.314	0.285	0.631	0.239
Sharing utensils	0.661	0.250	0.856	0.219	0.138	0.914
Mosquito bites	0.539	0.153	0.695	0.159	0.107	0.644

Dimensionality of HIV knowledge in different country contexts

Investigation of dimensionality of HIV knowledge in different countries can help exploring the evolution of HIV knowledge in the world and also can assess whether factor extraction procedure in the pooled dataset was dominated by the large sample size of the India 2006 survey.

48 Varimax orthogonal rotation procedure is used.

Investigation of separate countries suggests that in India and Malawi the best solutions are two-factor solutions[49] whereas in China, the Ukraine and Kenya the best solutions are three-factor solutions.

Detailed results for the analysis of separate countries can be found in Appendix I, Tables I.7-I.16. In all five countries the one factor solutions represent different aspects of overall knowledge (for example, overall knowledge in India, self-protection knowledge in Malawi, overall knowledge but without knowledge about mosquito bites and healthy-looking person in China). Therefore, a factor which represents overall knowledge in the pooled dataset covers well all the different combinations which are observed in separate countries. All five two-factor solutions have a factor which represents knowledge about two main misconceptions (utensils and mosquito bites) and the second factor represents self-protection knowledge in Malawi and the Ukraine, knowledge about ways to prevent HIV in India and in China and knowledge about ways to prevent HIV together with the knowledge about a healthy-looking person question in Kenya. Therefore, in the pooled dataset two-factor solution fits all countries well. All available three-factor solutions (China, the Ukraine and Kenya) have the same factor which represents knowledge about two main misconceptions (utensils and mosquito bites). The second factor in the three countries is knowledge about self-protection. The third factor represents knowledge about MTCT in China, about healthy-looking person in the Ukraine and about both questions in Kenya. Therefore, the three-factor solution in the pooled dataset represents all countries well for which three-factor solution is available. This analysis suggests that when the data is pooled and factors are extracted not much information about particular countries is lost in the process and the process of factor extraction is not dominated by the large sample size in India 2006 survey.

As factor loadings can be interpreted as weights (Bartholomew *et al*. 2008), we can assess the relative importance of different component of overall HIV knowledge in the one factor solution in different cultural contexts. For example, in India all components are relatively important, in China everything but healthy-looking person and mosquito bites, in Malawi all self-protection questions are of higher importance than the other components of knowledge, whereas in Kenya we observe a very different picture,

[49] In India 2006 and Malawi 2004 three-factor solutions produced one of the factors with a factor loading greater than 1 (3.535 and 1.064 respectively) which suggest the over-extraction of factors (see Tables I.12 and I.16 in Appendix I).

MTCT and all three misconception questions are of higher importance when one factor is extracted. In order to make more general statements about the evolution of HIV epidemics worldwide, it would be useful to have more countries included in the analysis as, for example, even countries with generalised HIV epidemics (Kenya and Malawi) show different patterns of component importance. It is also plausible that it is not possible to clearly define the evolution of HIV knowledge in the world by dividing countries into two groups: with generalised and with non-generalised HIV epidemics. The reason for this might be that the countries within these two groups have different trajectories which depend on a variety of factors.

The results of this part of the analysis are important as they suggest that HIV knowledge is not a uni-dimensional but multi-dimensional construct in both the Chinese context and the five country context. The results of this exploratory analysis suggest that the approach towards splitting HIV knowledge questions in the Chinese context into correct and incorrect routes used in Chapter 4 is justified by the existence of these two clear-cut dimensions in HIV knowledge. However, the split between the two groups is less straightforward in the five country context. Therefore, different split of the questions to the one used in Chapter 4 might be considered when simple score measures are constructed for separate dimensions of HIV knowledge in the five country context.

5.3.2.2 Modelling factor scores

Once the factor scores were extracted, an attempt was made to use them for further modelling in order to identify the predictors of different knowledge dimensions and in order to further assess the evolution of HIV knowledge in the Chinese context (1997-2005) and to compare HIV knowledge levels in the five country context.

The extracted factor scores are continuous and none of them are normally distributed[50] and, therefore, it is difficult to model them using linear regression models. Linear regression models were fitted to the two pooled datasets but the main assumptions of the modelling were found to be violated (residuals were not normally distributed and did not have a constant variance) and, therefore, the results of the models are not presented and interpreted here.

[50] Histograms of factor scores for the Chinese context and the five country context can be found in Appendix K (Figures K.1-K.9).

The main interest of the paper is to model and to compare levels of general HIV knowledge in China across time and in the five countries. Therefore, one factor solution would be helpful when the main interest of the analysis is addressed. As mentioned earlier, the distributions of factor scores in both contexts are not normal, therefore, in order to model factor scores, continuous variables can be categorised and then categorical variables can be easily modelled using partial proportional odds model or multinomial logistic regression.

In order to decide how to categorise factor score variables, first the one-factor solution scores are compared to Score Three (combined HIV knowledge in the Chinese and the five country contexts) discussed in section 2.3 and section 4.2. Tables L.1- L.2 in Appendix L show that the factor scores resemble closely the distribution of Score Three, with some rare exceptions in both study contexts. These similarities of the two distributions are expected as the scores are obtained on the basis of the factor loadings attached to every component of HIV knowledge. Therefore, if we decide to categorise the factor score distribution, one of the ways of categorisation would be to resemble the groupings of Score Three from Chapter 4 of the thesis. Similar to Score Three, grouping of factor scores can provide meaningful interpretation only for extreme groups. However, the difference between factor scores approach and simple score approach is that for the middle groups even the precise number of correct answers is not available. The only available information is the approximate place of a respondent or a group on the continuum between no knowledge and perfect knowledge.

Another way of categorisation of factor scores is categorising a continuous variable by using quartiles. Magadi and Agwanda (2010) used this approach to categorise an index of HIV/AIDS awareness which was derived using Principal Components Analysis. This approach can be very useful in some research situations. It is possible to create this variable in the Chinese context. However, it is not possible to use this approach in the five country context as the same factor score ended up in different groups. In the one factor solution for the five country context, the quartiles are -0.540 (25%), 0.041 (50%), 0.788 (75%) and 0.788 (100%). This result suggests that some people with the score 0.788 will end up in the 3rd group (50-74.99%) but some in the 4th group (75-100%). The categorised factor scores variable is similar to Score Four discussed in Chapter 4 but with different rules applied to categorisation. As it is easier to derive Score Four than a factor score variable and as categorised factor scores are not able to provide more information, the simple score approach can be preferable to the latent trait approach in some research situations.

Modelling of categorised factor scores for the two-factor solution can also be replaced by the modelling of Score One and Score Two discussed in Chapter 4. However, in the five country context latent trait analysis suggested different split of components between two types of knowledge (see section 5.3.2.1). Therefore, if factor scores extracted from two-factor solution are modelled the results are expected to be different due to non-comparability of the measures used for two different approaches (simple score approach and latent variable approach).

Given the similarities of the categorised factor scores and Score Three and Score Four used in Chapter 4, categorised factor scores are not modelled here. The results of modelling of Score Three and Score Four can be found in section 4.4.2.2.

However, not all parts of the latent trait analysis can be easily replaced by the simple score approach. Exploratory analysis of factor loadings can provide some useful insights into the differences in HIV knowledge in different cultural and epidemiological contexts as factor loadings can be interpreted as weights and provide information about relative importance of different components of HIV knowledge at different points in time and in different cultural and epidemiological contexts as well as about dimensionality of HIV knowledge in different contexts.

5.3.3 Latent class analysis

Latent class analysis is conducted using seven questions in the Chinese context discussed in section 2.3 in order to determine the main classes of respondents in the pooled sample which are identified on the basis of respondents' HIV knowledge. LCA can help splitting the dataset into the classes according to respondents' response patterns.

A one-class model is fitted first to data and then classes are added to the model one by one in both the Chinese and the five country context. Additional class improves model fit but not necessarily interpretability of the model. The main aim of the study is to develop a comprehensive but interpretable measure of HIV knowledge and, therefore, solutions with many classes (above 5) are not considered as they would be difficult to interpret.

Figure 5.1 shows sample split between different groups (A-E) in the Chinese context for the 2-5 class solutions. Figure 5.2 shows sample split between different groups (A-F) in the five country context for 2-5 class solutions.

Model fit statistics (AIC, BIC and classification quality) are used to determine the final model. Parsimony is considered to be of high importance when a decision about number of classes is made. Models with lower AIC and BIC values are preferred as discussed in the methodology section of this chapter.

The Chinese context

If the three criteria are taken into consideration (AIC, BIC and classification quality) then the model which fits the Chinese data best is the three-class model (AIC and BIC are lower in comparison to the two-class solution) and classification quality is the highest for three-class solution than for any other solutions studied (Table 5.13). Therefore, the three-class solution is selected as the best solution in the Chinese context and will be interpreted and used for further analysis.

Table 5.13: Model fit statistics in the Chinese context (LCA).

Number of classes	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	Classification quality (%)
1			
2	268534	268663	77.1
3	249637	249836	80.9
4	246547	246816	80.5
5	245498	245836	77.6

This solution extracts the following classes on the basis of conditional response probabilities[51] (Table 5.14):

1. high overall knowledge (A),
2. low overall knowledge (B),
3. high level of knowledge of correct routes and low level of knowledge of incorrect routes (some women in this group answered "yes" to all questions) (C) (see Figure 5.1 for details).

[51] 0.500 is the cut-off point between low level of knowledge and high level of knowledge.

Table 5.14: Conditional response probabilities in the Chinese context (LCA).

Manifest variable	Latent Class One (A: High overall knowledge)	Latent Class Two (B: Low overall knowledge)	Latent Class Three (C: High level of knowledge about correct routes and low level of knowledge about incorrect routes)
BLOOD			
Other	0.035	0.832	0.051
Correct	0.965	0.168	0.949
MTCT			
Other	0.101	0.782	0.095
Correct	0.899	0.218	0.905
NEEDLES			
Other	0.089	0.885	0.066
Correct	0.911	0.115	0.934
SEX			
Other	0.030	0.554	0.034
Correct	0.970	0.446	0.966
HANDSHAKE			
Other	0.018	0.883	0.614
Correct	0.982	0.117	0.386
UTENSILS			
Other	0.190	0.899	0.886
Correct	0.810	0.101	0.114
KISS			
Other	0.327	0.937	0.977
Correct	0.673	0.063	0.023

Classes in separate Chinese datasets

Investigation of the classes of HIV knowledge at different points in time in China can help exploring the evolution of HIV knowledge in the Chinese context and also can assess whether the class extraction procedure in the pooled dataset is dominated by the large sample size of the China 2001 survey.

For the separate Chinese datasets the three-class solutions have the same classes as the three-class solution in the pooled dataset interpreted in this paper (Appendix M, Tables M.1-M.6). However, the proportions of women in the three classes are different at different points in time. It can be concluded that extraction of classes in the pooled dataset represents well different points in time included in the analysis.

The five country context

In the five country context the substantial drop in AIC and BIC statistics is observed between two- and three-class solutions and then between three- and four-class solutions (Table 5.15). Between four- and five-class solutions the drop in AIC and BIC statistics is not as sizable as between previous solutions (Table 5.15). Classification quality is the lowest for the four-class solution and it is also slightly lower than for the five-class solution. However, as AIC and BIC do not change much between the four- and five-class solutions and, as the final model should be more parsimonious and interpretable, the four-class solution is considered to be the best solution and will be interpreted here and used for further analysis.

Table 5.15: Model fit statistics in the five country context (LCA).

Number of classes	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	Percentage of sample correctly classified
1			
2	875453	875599	77.5
3	864876	865099	70.9
4	856255	856556	68.7
5	854396	854774	71.2

In the four-class solution the following classes are extracted on the basis of conditional response probabilities (Table 5.16):

1. high overall knowledge (A),
2. low overall knowledge (D),
3. high level of knowledge about misconceptions and MTCT and low level of knowledge of the remaining ways to prevent HIV (E),
4. high level of knowledge about ways to prevent HIV and about the healthy-looking person question and low level of knowledge about remaining misconceptions (some of them answered “yes” to all questions) (C) (see Figure 5.2 for details).

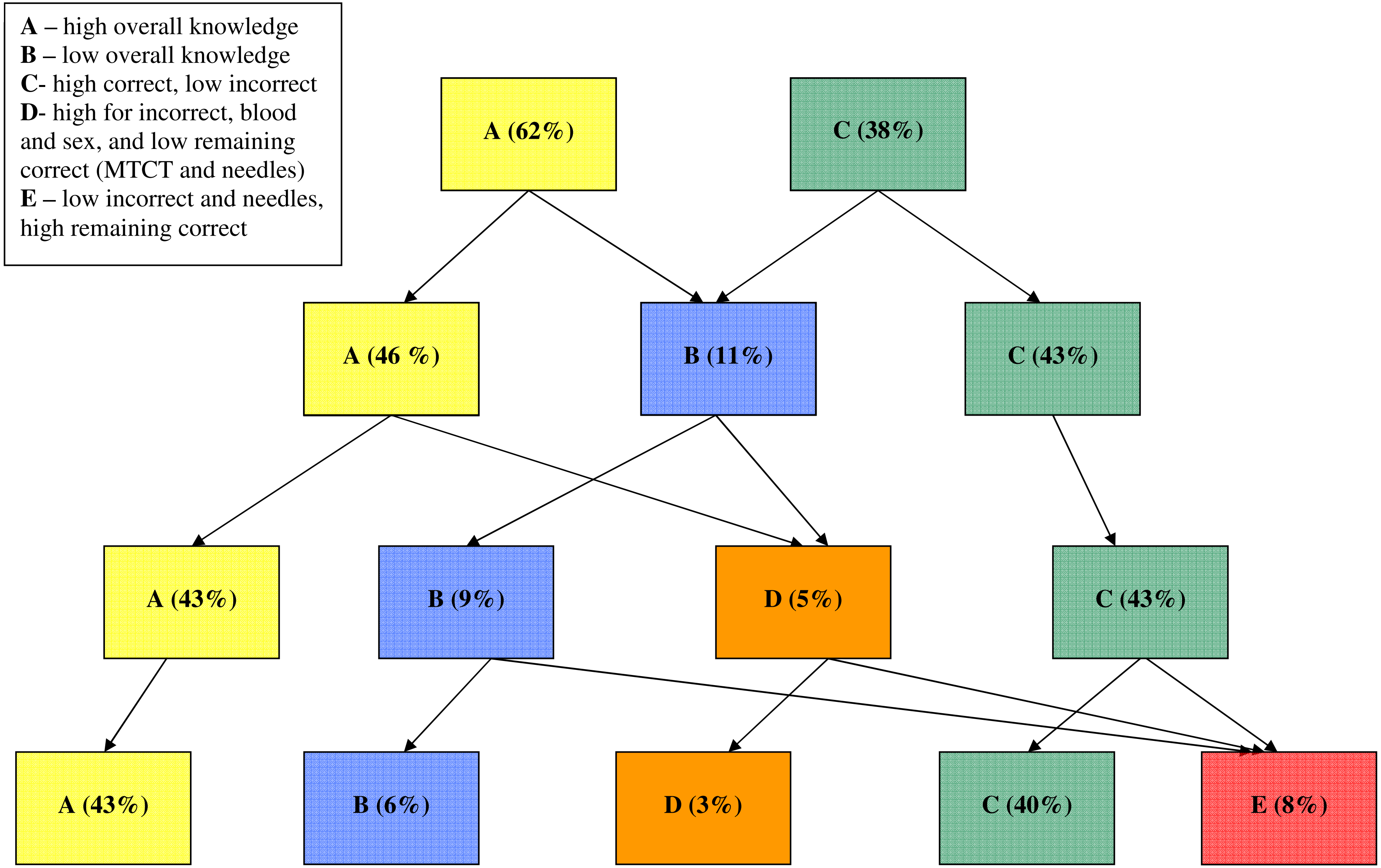


Figure 5.1: Sample split between different groups of HIV knowledge (A-E) in China for 2-5 class solutions (classification of individuals based on their most likely latent class membership).

Table 5.16: Conditional response probabilities in the five country context (LCA).

Manifest variable	Latent Class One (A: High overall knowledge)	Latent Class Two (D: Low overall knowledge)	Latent Class Three (E: High level of knowledge about misconceptions and MTCT, low level of knowledge about remaining ways to prevent HIV)	Latent Class Four (C: High level of knowledge about ways to prevent HIV and healthy-looking person question, low level of knowledge about remaining misconceptions)
MTCT				
other	0.080	0.563	0.281	0.181
correct	0.920	0.437	0.719	0.819
ONEPARTNER				
other	0.044	0.834	0.578	0.121
correct	0.956	0.166	0.422	0.879
ABSTINENCE				
other	0.087	0.926	0.677	0.174
correct	0.913	0.074	0.323	0.826
CONDOM				
other	0.101	0.961	0.724	0.337
correct	0.899	0.039	0.276	0.663
HEALTHY				
other	0.147	0.745	0.410	0.344
correct	0.853	0.255	0.590	0.656
UTENSILS				
other	0.038	0.898	0.133	0.648
correct	0.962	0.102	0.867	0.352
MOSQUITO				
other	0.167	0.881	0.242	0.750
correct	0.833	0.119	0.758	0.250

Classes in separate datasets in the five countries

Investigation of dimensionality of HIV knowledge in different countries might help exploring the evolution of HIV knowledge in the world.

Detailed results of the latent class analysis for the separate datasets can be found in Appendix M, Tables M.7-M.16. In all five countries classes A and C are found. Class E is found in China, India and Malawi and variations of class E are found in other countries. Variations of class D are also found in all five countries: the differences in this class are that in China women had a high level of knowledge about healthy-looking person, in the Ukraine about MTCT and condom, in Kenya about one partner and abstinence, in Malawi about abstinence, utensils and mosquito bites question, and in

India the same class as class D in the pooled dataset exists. As similar classes (with some variations) exist in all separate contexts, it can be concluded that the four classes identified in the pooled dataset reflect classes existing in separate contexts quite well and can be used for further analysis. Analysis of existing classes in separate datasets suggest that in Kenya the level of knowledge about HIV is the highest when compared with the other four countries used for the analysis as it has two classes of women with good overall level of knowledge about all seven questions, whereas India is the country with the worst composition of classes from the point of view of knowledge about seven questions about HIV.

Once the number of classes is determined in both contexts, each woman is allocated to a specific class on the basis of a posterior probability using M-Plus and the variable which contains class identifiers for each respondent are used for further analysis.

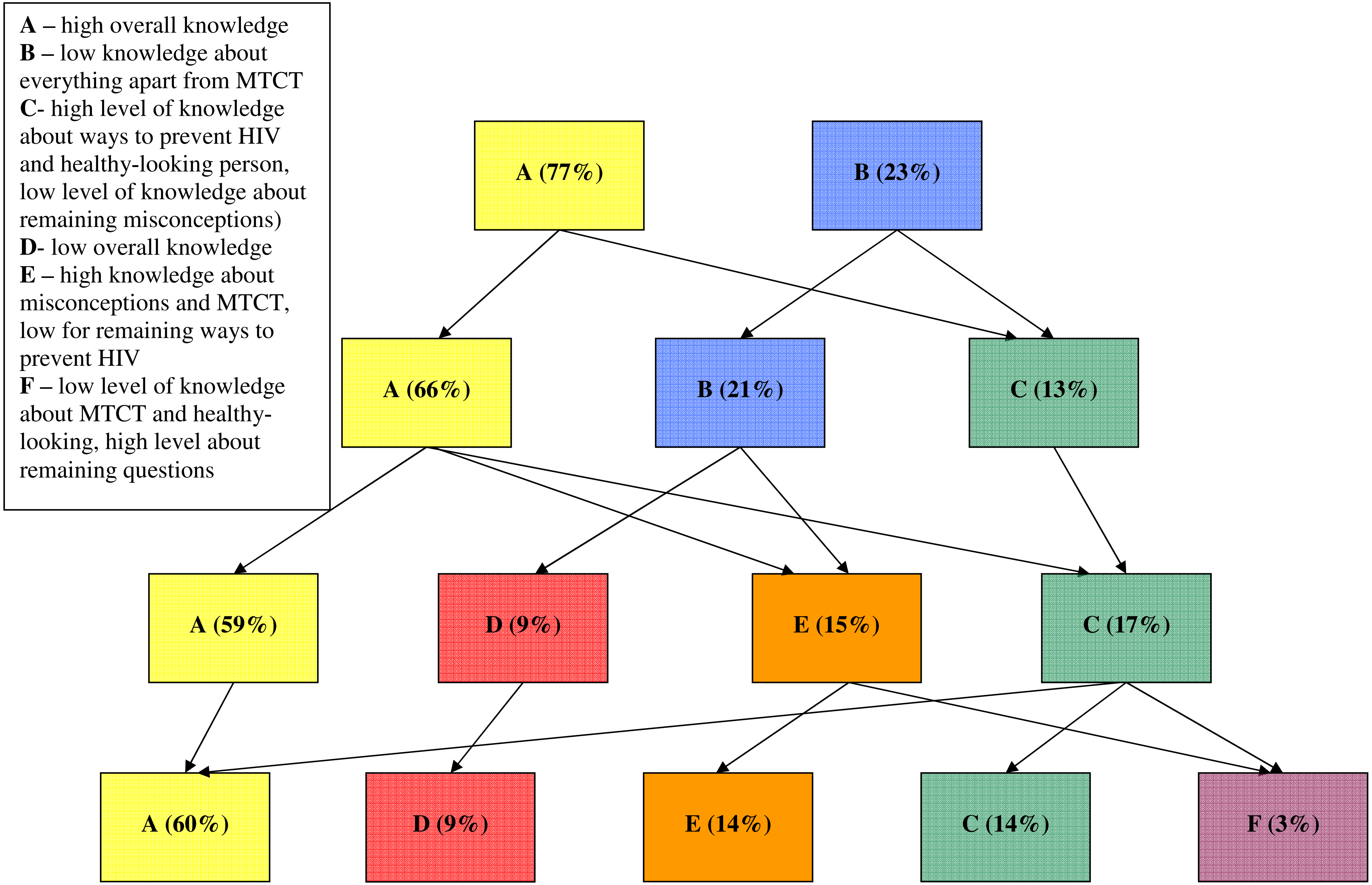


Figure 5.2: Sample split between different groups (A-F) in five countries for 2-5 class solutions (classification of individuals based on their most likely latent class membership).

5.3.3.1 The Chinese context: Three-class solution

Once the number of classes is determined and each woman is allocated to a specific class on the basis of a posterior probability, it becomes possible to identify characteristics of women who belong to different classes. Table 5.17 shows characteristics of women in China (1997-2005) who belong to three different classes of HIV knowledge.

Table 5.17: Characteristics of women in China (1997-2005) by different classes (%).

Covariate	Class One (A: High overall knowledge)	Class Two (B: High overall knowledge)	Class Three (C: High level of knowledge about correct routes and low level of knowledge about incorrect routes)
	45.9	**11.0**	**43.1**
Year of survey			
China 1997	32.8	25.0	42.2
China 2001	44.7	7.3	48.0
China 2005	69.0	5.2	25.8
Residence			
Urban	60.2	4.7	35.1
Rural	38.8	14.1	47.1
Ethnicity			
Han	45.8	10.9	43.3
Minority	46.7	12.0	41.3
Age group			
15-19	48.5	11.0	40.5
20-29	49.1	10.0	40.9
30-39	45.4	10.3	44.3
40-49	41.4	13.4	45.2
Education			
No education	19.1	28.0	52.8
Primary	31.3	17.8	50.9
Secondary	51.3	7.4	41.3
Higher	77.1	1.0	21.9
Marital status			
Widowed or divorced	46.2	9.3	44.5
Married or remarried	44.6	11.3	44.1
Never married	52.0	9.6	38.4

Class One (high overall knowledge): The proportions of women in this class increase over time and this class is most prevalent in 2005. This result suggests that the level of knowledge about HIV improves in China over time. Women from urban areas and with the highest educational levels, those who are 15-29 years old as well as those who are never married were more likely to be in this class.

Class Two (low overall knowledge): This class is most prevalent in 1997 and the proportions of women in this class decrease over time. This result also suggests that the level of knowledge about HIV improves in China over time. Women from rural areas and with no education, those who are 40-49 years old as well as those who are either married or remarried are more likely to be in this class.

Class Three (high level of knowledge about correct routes and low level of knowledge about incorrect routes): This class is more prevalent in 1997 and 2001 but the proportion of women in this class reduced in 2005. High proportions of women in this class in 1997 and 2001 can be explained by the low level of knowledge about incorrect routes of HIV transmission as well as fear of HIV at the beginning of HIV epidemic. As mentioned earlier, some women in this group are so called "yes-women" who agree with every question (see section 5.3.1). Women who belong to this class are more likely to be from rural areas, with either no or primary education and they are more likely to be 30-49 years old and married, remarried, widowed or divorced.

Table 5.17 show that there is almost no difference by ethnicity in memberships of any of the three classes.

Figure 5.3 and Table 5.17 show that in China 1997 and 2001 the most prevalent class is Class Three, whereas in 2005 it is Class One. Prevalence of Class One increases over time, whereas prevalence of Class Two decreases over time. All of these suggest that with time HIV knowledge in general and the knowledge of incorrect routes of HIV transmission in particular improves in China.

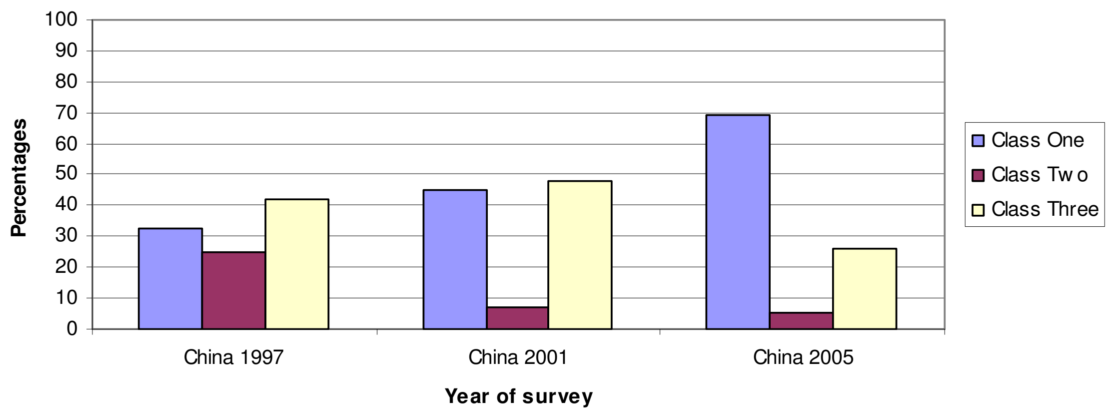


Figure 5.3: Prevalence of different classes at different points in time in China.

Table 5.17 also shows that rural women are more likely to be in Class Two and Three than in Class One, whereas urban women are more likely to be in Class One. A higher proportion of women with higher levels of education belongs to Class One than to any other class, whereas a higher proportion of women with lower levels of education belongs to Class Two and Class Three. Younger women are more likely to be in Class One whereas older women are more likely to be in Class Two and Class Three.

5.3.3.2 The five country context: Four-class solution

Table 5.18 shows characteristics of women in the five countries who are identified to belong to four different classes on the basis of their responses to HIV knowledge questions.

Table 5.18: Characteristics of women in the five country context by different classes (%).

Covariate	Class One (A: High overall knowledge)	Class Two (D: Low overall knowledge)	Class Three (E: High level of knowledge about misconceptions and MTCT, low level of knowledge about remaining ways to prevent HIV)	Class Four (C: High level of knowledge about ways to prevent HIV and healthy-looking person, low level of knowledge about remaining misconceptions)
	59.5	**8.6**	**15.0**	**17.0**
Country				
China 2005	62.6	5.2	5.9	26.3
Ukraine 2007	74.9	1.8	5.4	17.9
India 2006	56.7	10.9	15.9	16.5
Kenya 2003	77.1	0.7	5.4	16.8
Malawi 2004	59.7	1.4	24.6	14.3
Residence				
Urban	65.6	7.1	13.5	13.8
Rural	53.4	10.1	16.4	20.1
Age group				
15-19	56.8	8.6	20.1	14.5
20-29	63.0	7.5	14.1	15.4
30-39	59.8	8.8	13.3	18.1
40-49	54.8	10.5	13.9	20.8
Education				
No education	33.2	20.8	19.8	26.2
Primary	51.2	10.5	17.2	21.2
Secondary	63.3	6.4	14.8	15.5
Higher	84.6	1.0	7.6	6.8
Marital status				
Widowed or	55.6	9.3	15.8	19.3

divorced				
Married or remarried	58.4	9.2	13.8	18.6
Never married	62.9	6.9	17.8	12.4

Class One (high overall knowledge): Women from Kenya and the Ukraine are more likely to be in this class when compared with other countries. Women from urban areas, of age 20-39, with a higher educational level as well as those who are never married are also more likely to be in this class.

Class Two (low overall knowledge): This class is most prevalent in India, among rural women of older age, with a lower educational level who are married, remarried, widowed or divorced.

Class Three (high level of knowledge about misconceptions and about MTCT and low level of knowledge of remaining ways to prevent HIV): This class is more prevalent in Malawi and India. Women from rural areas, of younger ages, with lower educational levels and those who are never married are more likely to be in this class.

Class Four (high level of knowledge about ways to prevent HIV and about healthy-looking person question, and low level of knowledge about remaining misconceptions): This class is more prevalent in China than in other countries. Women from rural areas, with lower levels of education, older and married, remarried, widowed or divorced are more likely to be in this class. Women who answered "yes" to all questions belong to this class (see section 5.3.1).

Figure 5.4 and Table 5.18 show that in all countries Class One is the most prevalent and Class Two is the least prevalent. In Malawi, Kenya and the Ukraine, Class Two is almost non-existent. It is surprising that Class Three is more prevalent in Malawi than in any other countries. It suggests that a relatively large proportion of women in Malawi has good knowledge about misconceptions and MTCT but low knowledge about other ways to prevent HIV. This finding is unexpected due to the stage of HIV epidemic in Malawi.

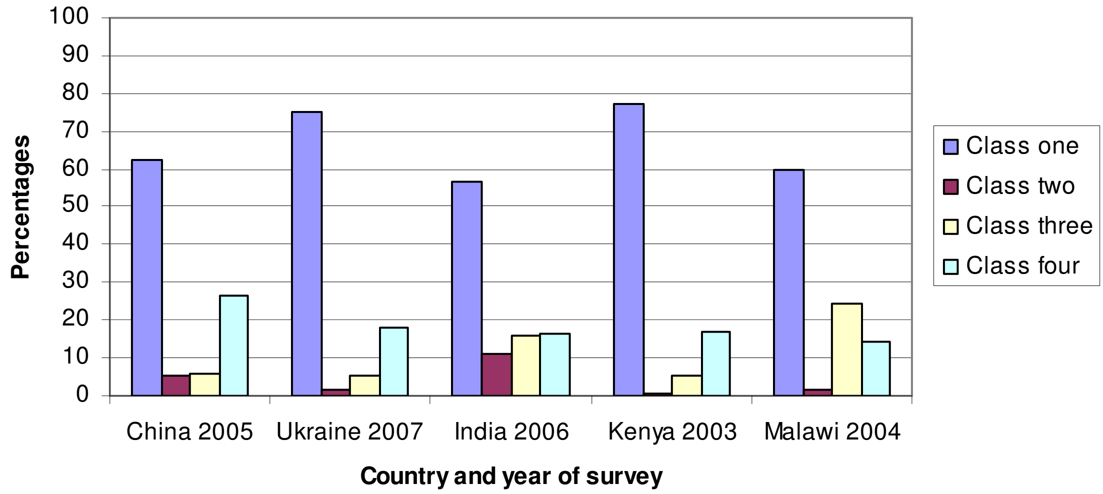


Figure 5.4: Prevalence of different classes in five different countries.

5.3.4 Multinomial logistic regression

Two multinomial logistic regression models are fitted, one to the China pooled dataset and the second one to the five country pooled dataset. In the Chinese context the response variable with the latent class variable with three categories as in this context the best solution is the three-class solution and in the five country context the response variable has four categories as in this context the best solution was the four-class solution. There is no natural ordering of categories. Therefore, multinomial logistic regression is suitable to model HIV knowledge as a latent class variable. Methodological details about multinomial logistic regression can be found in section 4.3.3.3.

A forward stepwise model selection process (see section 4.3.4) is employed for the multinomial logistic regressions. Potential interaction effects (see section 4.3.5) are investigated and are included into the final models if they improve their fits. Likelihood ratio tests are automatically conducted by SPSS in order to test significance of model terms and to inform decision about the inclusion of main effects or interactions into the final model.

Standard errors are corrected using STATA to control for potential effects of clustering (see section 4.3.6).

The Chinese context

The results of the modelling of HIV knowledge as a latent class variable in China found the following main effects and interactions to be significant: age group, residence, ethnicity, education, marital status, year of survey and interactions between age group and country, age group and education, age group and residence, country and education, country and ethnicity, country and residence, education and ethnicity, education and residence, and ethnicity and residence. Table N.1 in Appendix N presents results of the modelling. Table N.1 shows that once the standard errors are corrected and robust standard errors are obtained, the results of the modelling does not change. Therefore, it can be concluded that there is no need to control for clustering effects in this model.

Predicted probabilities are calculated for the same groups of women as in the analysis in Chapter 4 with potentially the best HIV knowledge (urban residence, 20-29 age group, higher educational level, Han ethnicity, married) and potentially the worst HIV knowledge (rural residence, 40-49 age group, no education, from ethnic minority group, widowed or divorced).

Figures 5.5 and 5.6 present predicted probabilities for different classes of women with potentially the best and potentially the worst knowledge in China respectively.

Figure 5.5 shows that for women with potentially the best knowledge the probability of being in Class One (high level of overall HIV knowledge) is the highest in China at every point in time. The probability of being in Class Two (low overall knowledge) decreases with time and is already low in 1997 and almost non-existent in 2001 and 2005 among women with potentially the best knowledge about HIV in China. The probability of being in Class Three (high level of knowledge of correct routes and low level of knowledge of incorrect routes) decreases in China with time as expected, due to increase in level of knowledge about incorrect routes of HIV transmission.

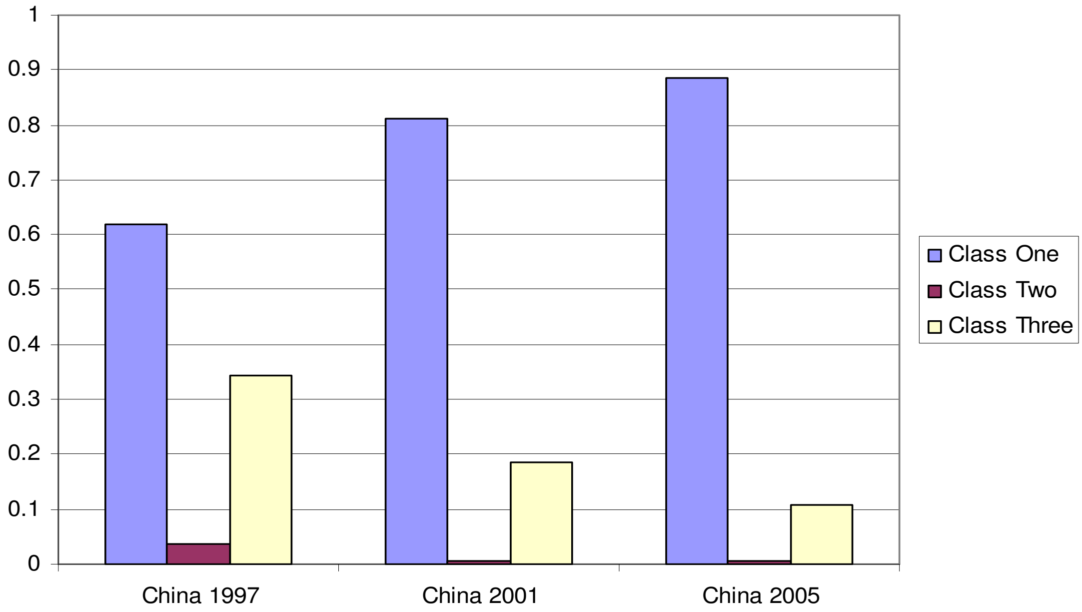


Figure 5.5: Predicted probabilities of belonging to three different classes for women with potentially the best knowledge in China.

Figure 5.6 shows that the probability of being in Class One (high overall knowledge) increases with time for women with potentially the worst HIV knowledge. These results suggest that the level of knowledge increases among women even with potentially the worst HIV knowledge in China. However, the probability of being in this class is the lowest for this group of women at every point in time when compared to the probabilities of being in other classes. Figure 5.6 also shows that the probability of being in Class Two (low overall knowledge) reduces substantially between 1997 and 2001, but this probability increases slightly in 2005. This trend is slightly worrying as in 2005 only best performing counties were put forward for participation of surveys and this result might suggest that the situation in other parts of China which were not included in the survey might be even worse. Figure 5.6 also shows that the probability of being in Class Three (high level of knowledge of correct routes and low level of knowledge of incorrect routes) is higher in 2001 and 2005 in comparison with 1997 and in 2001 the probability of being in the Class Three is the highest when compared with probabilities of being in other classes. However, despite lacking adequate knowledge in this group of women, larger proportions of women now belong to Class Three than to Class Two in 2001 and 2005 when compared to 1997 and, therefore, they at least have better knowledge about correct routes of HIV transmission.

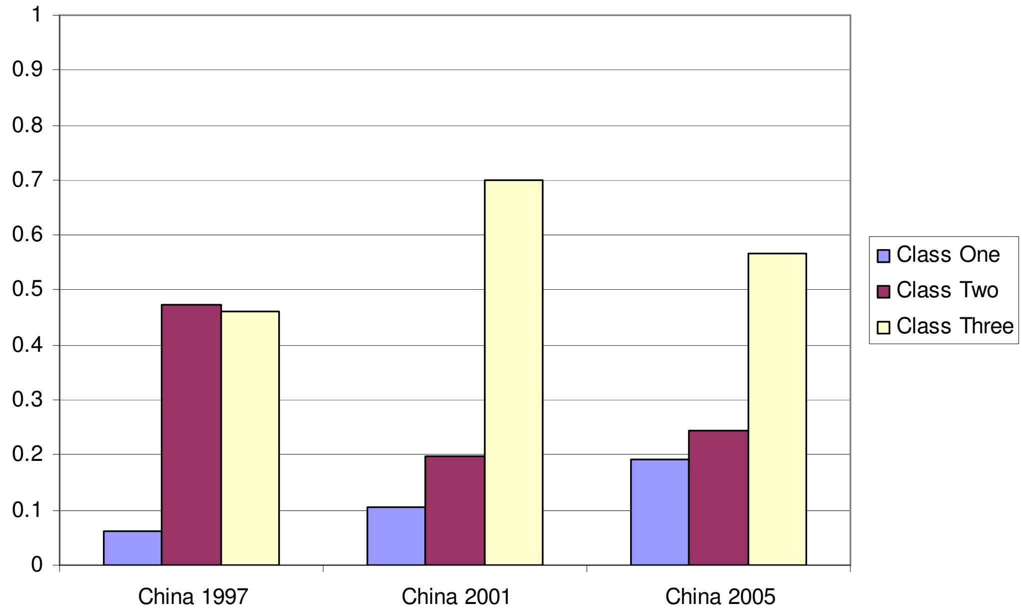


Figure 5.6: Predicted probabilities of belonging to three different classes for women with potentially the worst knowledge in China.

The results of the analysis suggest that HIV knowledge has improved among different groups of women in China over time. However, despite the increased probabilities of having higher knowledge over time in China in different groups, the gap in knowledge between different groups is still large, and the group of women with potentially the worst HIV knowledge is still lacking adequate overall knowledge about HIV in 2005 and, therefore, this group should be targeted for HIV knowledge interventions and campaigns.

The five country context

The results of the modelling of HIV knowledge as a latent class variable in the five country context found the following main effects and interactions to be significant: age group, residence, education, marital status, country and interactions between age and country, age and education, country and marital status, country and residence, education and residence, and marital status and residence. Table N.2 in Appendix N presents the results of the modelling. The same as in the Chinese context, correction of standard errors does not change the results of the analysis. Therefore, it can be concluded that there is no need to control for clustering effects in this model either.

Figures 5.7 and 5.8 present predicted probabilities for different classes of women with potentially the best (urban residence, 20-29 age group, higher educational level, never married) and potentially the worst (rural residence, 40-49 age group, no education, married or remarried) knowledge of both correct and incorrect routes of HIV transmission in the five countries.

Figure 5.7 suggests that for women with potentially the best knowledge, the probability of being in Class One (high overall knowledge) is the highest in all countries. The probability of being in Class Two (low overall knowledge) is very low in China, India and the Ukraine and non-existent in Kenya and Malawi. The difference in probabilities between Class One and other classes is substantial among women with potentially the best knowledge about HIV in all countries.

Figure 5.7 also shows that even among women with potentially the best knowledge, the probability of being in Class Three (high level of knowledge about misconceptions and MTCT and low level of knowledge about remaining ways to prevent HIV) in Malawi is the highest when compared with other countries with generalised and non-generalised epidemics. This finding is not expected as Malawi is a country where HIV epidemics is generalised and, therefore, it was expected that the level of knowledge about ways to prevent HIV is high and definitely higher than in countries where HIV epidemics are still not generalised. This finding might suggest the need for sustained interventions and campaigns even in countries where HIV epidemics are generalised.

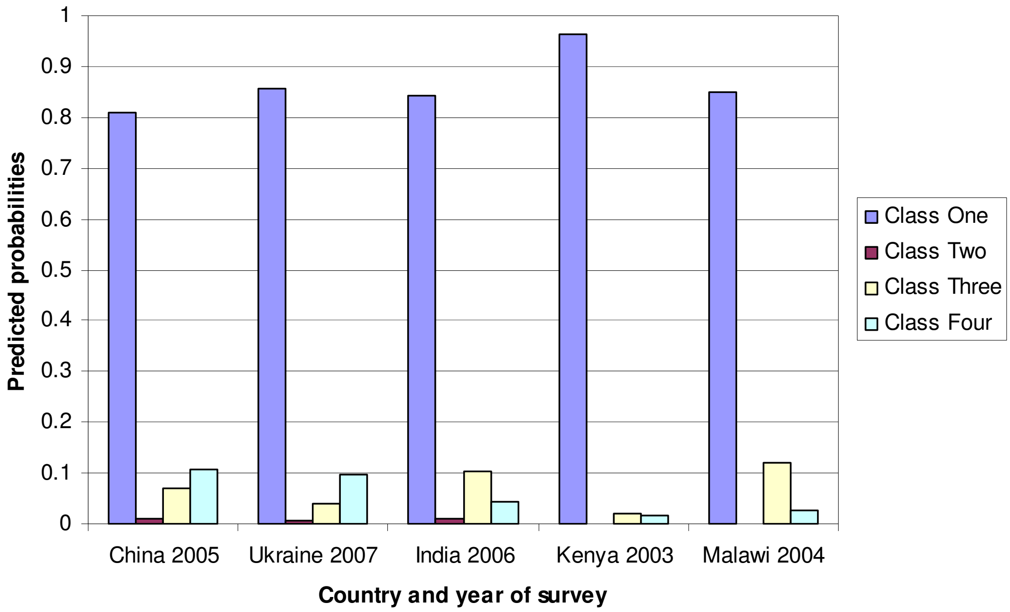


Figure 5.7: Predicted probabilities of belonging to different classes for women with potentially the best knowledge in the five countries.

Figure 5.8 shows that for women with potentially the worst knowledge about HIV in all countries but India, the most prevalent classes are Class One (high overall knowledge) and Class Four (high level of knowledge about ways to prevent HIV and knowledge about healthy-looking person and low knowledge about other misconceptions). In India the most prevalent classes are Class Four and Class Two (low overall knowledge). Figure 5.8 shows that the probability of being in Class Four is the highest for countries with non-generalised HIV epidemics (China, the Ukraine and India), whereas the probability of being in Class One is the highest for countries with generalised HIV epidemics (Kenya and Malawi) when compared with other classes. This result is expected as these two groups of countries are at different stages of HIV epidemics.

Figure 5.8 also shows that the probability of being in Class Three is the highest in Malawi among this group of women too when compared with other countries. This finding suggests the lack of knowledge about ways to prevent HIV in a country where HIV epidemic is generalised and this finding is unexpected.

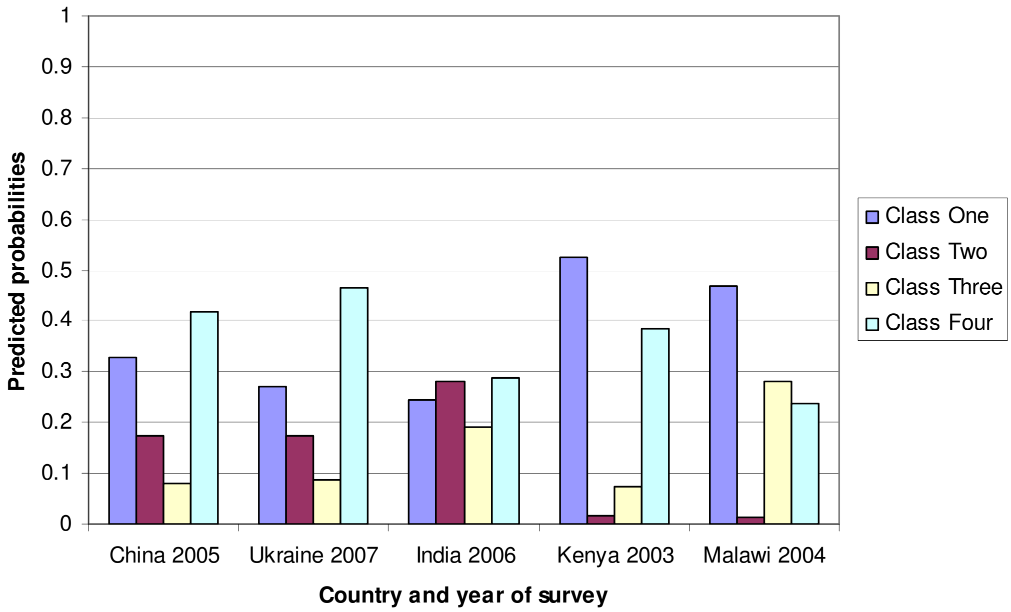


Figure 5.8: Predicted probabilities of belonging to different classes for women with potentially the worst knowledge in the five countries.

This graph suggest that among women with potentially the worst HIV knowledge there are still women without an adequate level of HIV knowledge and these groups should be targeted for educational campaigns in all countries and especially in countries with non-generalised epidemics but also sustained interventions should be in place in countries with mature HIV epidemics.

The results of the analysis suggest that the results for China among women with potentially the best knowledge are comparable to all other countries included in the analysis, whereas for women with potentially the worst knowledge, the level of knowledge in China is comparable to countries where HIV epidemic is non-generalised but better than that in India. The level of knowledge about ways to prevent HIV is higher than the level of knowledge about misconceptions among women in the latter group.

5.3.5 Latent measures of HIV knowledge: Their advantages and limitations

As mentioned earlier, different approaches exist to measuring HIV knowledge. This paper focused on latent variable approach to measuring HIV knowledge. Two different latent variable measures of HIV knowledge were discussed in this paper: a latent trait measure and a latent class measure. In this section these two latent variable measures, their advantages and limitations are discussed in detail. In the next section these measures are compared with simple score measures used in Chapter 4 of the thesis. As HIV knowledge cannot be measured directly, researchers always face the choice between different measures they can use in their analysis. Assessment of different measures of HIV knowledge available can help researchers to decide more easily which measure is most suitable in their study context. Different aspects of measure suitability can be taken into account when a decision about a usage of a certain measure is taken. The summary of all measures used in Chapter 4 and Chapter 5 is presented in Appendix O. All these measures can be used not only within the context of measures of HIV knowledge but also within contexts of any other health-related knowledge measures which also cannot be measured directly and have to be derived on the basis of different components. Examples of other health-related knowledge measures can be found in section 4.4.3.

In section 4.4.3 each HIV knowledge measure was assessed from the point of view of its so-called qualitative and quantitative meanings. The quantitative or ordinal meaning of a simple score measure was defined as a number of components of a measure known by respondents in each group. Or in other words, this meaning of a measure can help to identify a place of a respondent on a continuous spectrum of HIV knowledge between no knowledge and perfect knowledge. The qualitative or nominal meaning of a simple score measure was defined as a specificity of HIV knowledge of people who belong to a certain group, i.e. which specific components of HIV knowledge are known by respondents in this group. Or in other words, the nominal meaning of a measure can help to identify specific patterns of HIV knowledge which exist in a population. Ideally a perfect measure of HIV knowledge should have a precise qualitative and a precise quantitative meaning. However, in some specific research contexts only one of the two requirements or even a part of one requirement is sufficient. In this paper the HIV knowledge measures discussed in this chapter are assessed in the same way as it was carried out in section 4.4.3 for the simple score measures.

The latent trait measure of HIV knowledge measures knowledge on a continuous scale. This measure does not have precise qualitative or quantitative meaning. However, it places respondents on the scale of HIV knowledge between no knowledge and perfect knowledge.

The latent trait measure can be categorised and the distribution of this measure resembles the distributions of the simple score measures discussed in Chapter 4. This categorisation can help obtaining quantitative meaning for the extreme categories if categorised in similar way to Score Three. If the continuous measure is categorised in a similar way to Score Four, used for analysis in Chapter 4, then, similarly to Score Four, the qualitative meanings of this measure is not precise either for all categories.

If dimensionality is taken into account when factor scores are produced, then two (or more) sets of factors scores can be extracted for each respondent. These two measures' distributions might resemble Score One and Score Two used for the analysis in Chapter 4 as the extracted dimensions represent knowledge about correct and knowledge about incorrect routes of HIV transmission in China.

The main limitation of latent trait measure is that not every statistical software package has facilities for derivation of this measure. Therefore, as long as the dimensionality of HIV knowledge or relative importance of separate components is not of the main interest, a simple score measure could be a good substitution for the latent trait measures of HIV knowledge when it is used in regression analysis.

A latent class measure provides information about the main classes or segments of HIV knowledge observed in a population. This measure does not have a clear quantitative meaning as it is not clear how many correct answers respondents in each class obtained (this number might differ from respondent to respondent (Tables 5.19-5.20)). However, this measure has a very clear qualitative meaning for each category (class) as it provides information about the specific type of HIV knowledge for members of each class. Furthermore, some of the classes are impossible to identify using simple score measures. For example, simple score measures are not able to provide information about Class Three in the Chinese context (high level of knowledge about correct routes of HIV transmission and low level of knowledge about incorrect routes) or about Class Three (high level of knowledge about misconceptions and MTCT, low level of knowledge about remaining ways to prevent HIV) and Class Four (high level of knowledge about ways to prevent HIV and healthy-looking person question, low level of knowledge about remaining misconceptions) in the five country context. And this information might be very useful in some research situations and possibly for interventions. Latent class measures can be useful for splitting populations into segment on the basis of their specific knowledge and this information can be helpful when designing necessary interventions.

Table 5.19: Cross-tabulation between variables obtained through the simple score (Score Three) and latent variable (LCA) approaches to measuring HIV knowledge in the Chinese context.

	Class One	Class Two	Class Three
No knowledge	0	1778	0
One route	0	954	0
Two routes	0	1425	0
Three routes	0	501	2301
Four routes	634	30	9031
Five routes	1819	0	7053
Six routes	8409	0	0
Perfect knowledge	8693	0	0

Table 5.20: Cross-tabulation between variables obtained through the simple score (Score Three) and latent variable (LCA) approaches to measuring HIV knowledge in the five country context.

	Class One	Class Two	Class Three	Class Four
No knowledge	0	3034	0	0
One route	0	3828	0	0
Two routes	0	3421	2084	657
Three routes	0	0	5002	4272
Four routes	0	0	7079	7353
Five routes	9350	0	3722	8018
Six routes	28385	0	0	0
Perfect knowledge	33372	0	0	0

The results of the analyses of this section and of the section 4.4.3 suggest that no measure can provide both a clear qualitative and clear quantitative meanings. However, in some specific research contexts only one of the two requirements or even a part of one requirement is sufficient.

5.4 Comparison of the Latent Class Variable and Simple Score Variable Approaches to Modelling HIV Knowledge

In order to obtain a comprehensive understanding of the evolution of HIV knowledge in China two approaches were used to measure HIV knowledge: a simple score variable approach and latent class variable approach. The simple score variable approach enables us to obtain detailed information about two groups: no overall knowledge and perfect overall knowledge. For other groups, information about specific routes is not available. For example, a person with the knowledge of just one route of HIV transmission can have knowledge about one correct route or one incorrect route and the variable would not provide information about which route is known by the respondent. This group of respondents (with one correct answer) combines seven different possibilities of knowledge. It would be difficult to use the results of this analysis for specific interventions as it would be impossible to identify the specific information needs of specific groups in a population. The simple score variable approach does not allow for identification of those who, for example, belong to Class Three in China or Class Three and Class Four in the five country context. However, simple score measures can provide useful information about the location of a specific group of respondents on HIV knowledge spectrum between no knowledge and perfect knowledge. For many research projects and general knowledge interventions this information can be of the main interest. It can be concluded that this approach to measuring HIV knowledge provides useful insights into understanding of the evolution of HIV knowledge in China.

Latent class analysis was used to assess the overall knowledge, but it was possible to identify specific problems with the knowledge about ways to prevent HIV in Malawi when compared with other countries and to obtain an important conclusion about the need for sustainable interventions even in contexts where HIV epidemics are generalised. It was impossible to obtain this result when a simple score of combined knowledge (Score Three) was analysed. Separation of two types of knowledge into two simple score variables (knowledge of correct routes (Score One) and incorrect routes (Score Two) of HIV transmission) was required to identify this potential problem. It can be concluded that it might be important to study overall knowledge and then to study specific types of HIV knowledge separately when a simple score approach is employed in order not to miss some important information.

The latent class variable approach provides information about predominant classes which exist in the sample and can help to identify specific knowledge patterns existing in populations and, therefore, to address specific information needs within each group. Variables obtained through latent class analysis are more suitable when the overall level of HIV knowledge needs to be assessed and when the purpose of analysis is to help designing interventions or to create more homogenous groups on the basis of their knowledge about HIV in the heterogeneous population. Once the needs of every specific group are identified, these groups can be targeted for specially designed educational interventions and campaigns. However, this approach to measuring HIV knowledge does not provide information about the number of known components which might be of interest in some research situations. This approach also would not allow us to precisely place groups of women on the HIV knowledge continuum on the basis of their HIV knowledge. When this specific information is of interest for designs of interventions or assessment of progress of educational and information campaigns, this approach to measuring HIV knowledge is not appropriate.

The latent class variable will often have a smaller number of categories than the simple score response variable (in this study latent class variables have 3-4 categories whereas comparable simple score variable had 8 categories) and, therefore, the model will usually have less parameters to estimate.

It can be concluded that both approaches provide complimentary information as they split populations in different ways and both are useful when the evolution of HIV knowledge is studied. The results from the two analyses confirm the same conclusions about patterns of HIV knowledge among different groups of women but the level of details which can be drawn from the analysis differs by the approach to measuring of HIV knowledge. The reason for the similarity of results about patterns of HIV knowledge among two groups of women (women with potentially the best and potentially the worst knowledge) is that the main focus of the analyses is on presenting general patterns in differences of levels of HIV knowledge observed for the two specific profiles of women existing in the populations. If the main interest of the analysis is to identify the effects of separate factors on HIV knowledge or the magnitudes of effects in different contexts, with different splits of the populations into groups in the two approaches, it is possible that the results would be different. Therefore, these two measures could be used as complementary if detailed analysis of HIV knowledge is required or one more appropriate measure can be selected and used.

The results of the analyses suggest that simple score measures produce ordinal variables and have clear quantitative or ordinal meaning of the categories, whereas latent class measures produce nominal variables and have clear qualitative or nominal meaning of the categories. Both measures split populations into different groups. Different types of measures can be useful in different types of research situations. Therefore, it is important for researchers to have clear objectives and then to decide which measure is sufficient and will provide them with the information needed. Table O.1 in Appendix O can be useful for the decision-making process.

5.5 Limitations of This Study

The current analysis has a number of limitations. The main limitations of the study which are related to the samples used for the analysis, availability of variables and comparability across time and across countries are the same as in section 4.5 as the same samples and the same variables were used for the analysis in this chapter as in Chapter 4.

The pooled datasets represent another limitation as this approach might mask some peculiarities of different years of surveys or of separate countries. This problem can be especially pronounced when latent classes and factor scores are derived as at this stage the model does not control for differences in years of survey or in countries. However, this approach enables comparisons over time and between countries which are the main aims of the paper. Investigation of different years of survey or of countries separately will enable examination of the peculiarities of different years of surveys or countries if this is an interest for the research.

The main limitation of the latent variable approach is the fact that only specialised statistical software packages allow latent trait analysis (M-Plus, GENLAT, GLLAMM, Amos and others) or latent class analysis (M-Plus, Latent Gold, LEM and others).

5.6 Conclusions

This paper studied the evolution of HIV knowledge in China using latent variable approach to measuring HIV knowledge. It also compared the simple score and latent variable approaches to measuring HIV knowledge.

The descriptive stage of analysis suggests that latent trait analysis and latent class analysis can be used as two complementary techniques as they help to obtain answers to different questions: latent trait analysis helps to determine dimensionality of a latent concept as well as to identify relative importance of different components of HIV knowledge concept, whereas latent class analysis helps to identify prevailing sample response patterns and to classify individuals into the classes based on these patterns.

The latent trait analysis shows that HIV knowledge is a multi-dimensional construct in both the Chinese context and the five country context. In the Chinese context, HIV knowledge is best defined as a two-dimensional construct (knowledge about correct routes of HIV transmission and knowledge about incorrect routes), and in the five country context HIV knowledge is best defined as a three-dimensional construct (self-protection knowledge, medical knowledge and knowledge about misconceptions).

The results of the analysis suggest that the descriptive part of the latent trait analysis proved to be useful. However, the results also suggest that when variables which are derived using latent trait analysis are compared to simple score measures, they resemble their distributions and, therefore, the analysis of simple score measures can on some occasions replace modelling of latent trait measures and would produce similar results. Simple score measures could be preferred in some situations for derivation of variables as it is easier to derive them and there is no need to have an access to a specialised software package in order to extract factor scores as all conventional software packages can produce simple score measures easily.

The results of the latent class analysis show that in the Chinese context it is best to split the population into three main classes and in the five country context the population can be best split into four main classes on the basis of their responses to HIV knowledge questions.

The two-factor solution of the latent trait analysis and the four-class solution of the latent class analysis suggest that in the five country context, caution should be taken when placing the question about healthy-looking person in the misconception knowledge dimension when separate simple scores for knowledge about ways to prevent HIV and knowledge about HIV misconceptions are derived and used in Chapter 4. Therefore, it can be concluded that investigation of the dimensionality of a concept can provide some useful insights into the inclusion of variables in measures of specific types (dimensions) of knowledge for the simple score approach. Therefore, simple score and latent variable approaches can be used as complimentary.

The results of the modelling of latent classes show that the level of knowledge in China increases with time for different groups of women. As expected, women in Kenya and Malawi had better overall HIV knowledge than women in India, China and the Ukraine. However, in Malawi a large proportion of women still belong to a group with a low level of knowledge about some ways to prevent HIV but a high level of knowledge about misconceptions and this result was unexpected. Unfortunately, this study cannot identify if the levels of knowledge about some ways to prevent HIV was always low in this country or if they decreased due to the lack of appropriate interventions. This result suggests the need for appropriate and sustained interventions even in the countries where HIV epidemics are generalised.

The results of the analysis also suggest that in China the levels of knowledge about HIV among women with potentially the best knowledge are comparable to those in all other countries, whereas for women with potentially the worst knowledge, the levels of knowledge in China are comparable to countries where epidemics are non-generalised but they are slightly higher than those in India. In Kenya and Malawi among women with potentially the worst knowledge, the most prevalent class is Class One (high overall knowledge), whereas in China, India and the Ukraine it is Class Four (high knowledge about ways to prevent HIV and about healthy-looking person and low knowledge about other misconceptions), and this result is expected due to the differences in the stages of HIV epidemics between the two groups of countries.

The results show that in all contexts there are still groups of women without an adequate level of HIV knowledge and these groups should be targeted for educational interventions and campaigns. There is also still a large gap in knowledge between different groups of women in China.

The overall results of the analysis in this chapter suggest that China has succeeded in improving the HIV knowledge of women over time but there is still room for further improvement. The results for the 2005 survey cannot be generalised to the whole China due to the survey representativeness level. However, it can be argued that if the right interventions are in place in the whole country, similar results could be observed in other parts of China.

The analyses conducted in Chapter 4 and Chapter 5 allowed the comparison between results obtained through modelling of HIV knowledge as simple score variables and as latent class variables. Both approaches are suitable for analysis of the evolution of HIV knowledge. Two different approaches to measuring HIV knowledge split samples differently: a simple score approach produces ordinal variables with clear quantitative meanings of categories, whereas a latent variable approach produces nominal variable with clear qualitative meanings of categories. Different details about HIV knowledge in different contexts can be obtained through the analysis of the variables obtained through two different approaches. It can be concluded that both approaches to measuring HIV knowledge are suitable and useful when the evolution of HIV knowledge is studied. The results obtained from the analysis of the evolution of HIV knowledge are similar to the results obtained in Chapter 4. The reason for this is that the main focus of the analyses is on presenting general patterns in differences of levels of HIV knowledge observed for the two specific profiles of women existing in the populations. If the main interest of the analysis is to identify the effects of separate factors on HIV knowledge or the magnitudes of effects in different contexts, with different splits of the populations into groups in the two approaches, it is possible that the results would be different. However, this was not of an interest for this study. The latent class variable approach can be more suitable than the simple score approach when the purpose of analysis is to help designing specific interventions or to split the population into identifiable more homogeneous groups representing specific response patterns which would further be used for targeted interventions. Latent class analysis also can help in identifying groups of people which would not be easily identified through a simple score approach. The latent class approach cannot provide information about the specific number of correct responses individual in each group gave. The simple score approach, on the contrary, provides clear numeric meaning to the categories and is useful when this is of main interest. It can also place groups of women on the continuum between no knowledge and perfect knowledge as seen in Chapter 4 and this is useful when the evolution of HIV knowledge is studied and when general HIV knowledge interventions are designed.

The results of the analysis suggest that when the simple score approach is employed, it is important to study different types (dimensions) of knowledge separately in order not to miss some important results by analysing only the overall knowledge.

Despite the limitations of different approaches, it can be concluded that both approaches provide useful insights into the understanding of the evolution of HIV knowledge and these two approaches can be treated as complementary in different research situations as they add unique insights into understanding of the phenomenon under the study. However, when the evolution of HIV, which is studied through dynamics of changes in general patterns of HIV knowledge among women with specific characteristics, is the main interest of the study, both approaches can be used interchangeably as they produce similar results.

The current analysis shows that China has succeeded in improving HIV knowledge among women between 1997 and 2005. In order to assess successes of overall HIV prevention in China, other aspects of HIV prevention should be examined. However, as mentioned earlier, this is outside of the remit of this research.

Box 5.1: Key findings in Paper Three

- China has succeeded in improving women's knowledge of HIV over time.
- HIV knowledge increases in all groups of women. However, there is still gap between groups so more efforts are needed to improve level of HIV knowledge among all women in China.
- The levels of knowledge among women with potentially the best knowledge in China are comparable to the levels in countries with generalised and non-generalised HIV epidemics.
- The levels of knowledge among women with potentially the worst knowledge in China are comparable to the levels in countries with non-generalised HIV epidemics but not yet with generalised HIV epidemic (comparable to those in the Ukraine but slightly higher than those in India).

- As expected levels of combined knowledge are higher in Malawi and Kenya than in countries with non-generalised epidemic.
- An unexpectedly high proportion of women in Malawi has a lower level of knowledge about some ways to prevent HIV.
- Among women with potentially the worst knowledge in countries with non-generalised epidemics including China the most prevalent class is the class with high level of knowledge about ways to prevent HIV and low level of knowledge about misconceptions, whereas in countries with generalised epidemics the most prevalent class is the class with high overall knowledge.
- There is a need to improve levels of knowledge among all groups of women even in countries where epidemics are generalised and the levels of knowledge are expected to be high.
- Latent trait analysis is especially useful at a descriptive stage to identify dimensionality and relative importance of different knowledge components.
- Latent class analysis is useful in splitting the data into specific population groups which can be targeted with specific interventions.
- The simple score and the latent variable approaches split study samples into different groups and, therefore, provide specific information about different groups of women and unique insights into the topic.
- The simple score approach produces ordinal variables with clear quantitative meanings of the categories and is useful when there is a need to assess the relative position of different groups of women on a continuum of HIV knowledge between no knowledge and perfect knowledge, whereas the latent variable approach produces nominal variable with clear qualitative meanings of the categories and is useful when there is a need to identify specific segments of population on the basis of their specific HIV knowledge patterns for further interventions.
- Both simple score and latent variables approaches to measuring HIV knowledge are useful when the evolution of HIV knowledge is assessed and produce similar results.

Chapter 6: Conclusions

This chapter of the thesis summarises and brings together the results and findings from all three papers to make overall conclusions with reference to the two main research questions of the thesis. This chapter also discusses policy recommendations and presents suggestions for further work.

6.1 Summary of Results and Conclusions

HIV awareness and knowledge are important pre-requisites of HIV risk reduction behaviour and, therefore, they are important components of any successful HIV prevention strategy. Therefore, a systematic investigation of the dynamics of HIV awareness and knowledge in different cultural and epidemiological contexts is crucially important.

China is the most populous country in the world. It has successfully managed to control the spread of HIV epidemics within the last decade as evident in recent HIV estimates which show that the numbers of people affected by HIV did not reach the projected ten-twenty millions of people. It is important to understand successes and failures in different countries' experiences, including Chinese experience, in the area of HIV prevention in order to be able to effectively control HIV epidemic in the world. This thesis focused its attention on two major components of HIV prevention – HIV awareness and knowledge.

This thesis had two main aims: substantive and methodological. The substantive aim was to assess HIV awareness and knowledge in China across time and to compare HIV knowledge in China to that in India, Kenya, Malawi and the Ukraine in order to answer the first research question whether China has succeeded in improving women's HIV awareness and knowledge over time. The methodological aim was to develop and compare simple score and latent variable measures of HIV knowledge which relates to the second research question of the thesis: how can we better measure HIV knowledge in diverse cultural and epidemiological contexts?

In order to achieve these two main aims and to answer the two main research questions this thesis was developed in three distinct but related papers. The main results of each paper were discussed in detail in relevant chapters and are summarised and presented in Boxes 3.1, 4.1 and 5.1.

A variety of data sources and methods were employed to fully address the two main research questions of the thesis. The following datasets were used for the analysis in the thesis: the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001, the UNFPA Reproductive Health and Family Planning Surveys 2003 and 2005, India DHS 2006, Kenya DHS 2003, Malawi DHS 2004 and Ukraine DHS 2007. In order to examine HIV awareness and knowledge in China, the following methodologies were employed: regression decomposition analysis, simple score approach to measuring HIV knowledge and latent variable approach (latent trait analysis and latent class analysis) to measuring HIV knowledge. In order to model HIV awareness and knowledge in the Chinese and in the five country contexts binary logistic regression, partial proportional odds modelling and multinomial logistic regression were used.

This thesis addressed some of the gaps which were identified in the existing literature and applied several innovative approaches to analysing the existing data. The main contributions of this study are:

1. An assessment of HIV awareness in China over time;
2. An assessment of HIV knowledge in China over time and a comparison of HIV knowledge in China with that in other countries such as India, Kenya, Malawi and the Ukraine;
3. A methodological comparison of simple score and latent variable approaches to measuring HIV knowledge.

In order to address the first research question, levels of HIV awareness and HIV knowledge of correct and incorrect routes of HIV transmission were assessed in China over time (between 1997 and 2005). Levels of knowledge about ways to prevent HIV and knowledge about misconceptions about HIV were also assessed in the five countries and levels in China were compared with those in India, Kenya, Malawi and the Ukraine.

These analyses provided a clearer understanding of the evolution of HIV awareness and knowledge in China over time. They also enabled assessment of relative success of China through comparison of the levels of HIV knowledge in China to those in other countries in the world (India, Kenya, Malawi and the Ukraine). This thesis also addressed issues of dimensionality of HIV knowledge in different cultural and epidemiological contexts as well as the identification of the main patterns of HIV knowledge which exist in different contexts.

The results from the three papers showed that China has succeeded in improving women's HIV awareness and knowledge between 1997 and 2005. The increases in HIV awareness and knowledge are observed in different groups of women in China over time. The main contribution to the changes in HIV awareness in China is attributed to the changes in environment, such as political commitment, effective interventions and campaigns. Changes in population structure also contributed to the changes in HIV awareness, but this contribution was smaller when compared with the contributions from the change in the environment. The level of knowledge about correct routes of HIV transmission and ways to prevent HIV was found to be higher than the level of knowledge about incorrect routes or misconceptions in China. This finding was expected and can be explained by the stage of HIV epidemic in China and by the usual temporal order of the main interventions directed at improvement of HIV knowledge in different countries. At the beginning of HIV epidemic, the focus of HIV knowledge interventions was usually on the HIV prevention knowledge and particularly the knowledge of correct routes of HIV transmission. As the epidemic progresses, specific interventions aimed at reducing stigma and discrimination against people living with HIV were introduced. This led to an improvement in the level of knowledge about misconceptions or incorrect routes of HIV transmission. However, despite the increases in HIV awareness and knowledge among different groups of women in China, there are still groups of women who lack adequate HIV awareness and knowledge and gap in HIV awareness and knowledge still exist between different groups of women. The groups of women with inadequate HIV awareness and knowledge should be targeted in specifically designed interventions and campaigns in order to ensure that HIV awareness and knowledge become more homogenous across the country. The levels of HIV knowledge about ways to prevent HIV in the 30 selected counties of China were found comparable to those in all other countries included in the analyses, including countries where HIV epidemics are generalised (Kenya and Malawi). Surprisingly, the level of knowledge about some ways to prevent HIV was found to be lower in Malawi than in all other countries included in the analyses. This finding was unexpected and suggests the need for sustained efforts in HIV knowledge improvement even in the contexts where HIV epidemics are mature. The levels of knowledge about misconceptions about HIV in the 30 counties in China were found to be comparable to those in other countries where the HIV epidemics are still not generalised but yet not to those in countries where the epidemics are generalised. For some groups of women, these levels were found to be comparable in the 30 counties of China to those in the Ukraine but were found to be higher in China than in India. All findings mentioned above suggest the need for specifically designed targeted interventions in order to address specific needs of certain groups of women in different cultural and epidemiological contexts effectively. These efforts might help to spend limited resources more efficiently.

In order to address the second main research question, simple score and latent variable approaches were used and compared. The results of the comparison suggest that both approaches proved to be useful for the understanding of the evolution of HIV knowledge in China. The simple score approach produced easily derived ordinal measures of HIV knowledge with clear quantitative meanings of categories (these variables provided information about number of questions answered correctly by each respondent), whereas the latent variable approach produced nominal measures with clear qualitative meanings of categories (these variables provided definitions for each main segment in the sample based on the HIV knowledge of the respondents in this segment). Therefore, the two approaches split the samples into different groups and one of the two approaches might be more suitable in different research situations. The simple score approach was found useful when there is a need to assess the relative position of different groups of women on a continuum of HIV knowledge between no knowledge and perfect knowledge. The latent class approach is useful when there is a need to identify specific segments of population on the basis of their specific HIV knowledge for further targeted interventions. The main results obtained from both approaches in relation to the evolution of HIV knowledge in China were similar. This suggests that these two approaches can be used interchangeably when the evolution of HIV knowledge is studied. However, each approach also additionally provided unique insights into the issues of interest. For example, the results of the analysis suggested that latent trait analysis was useful when dimensionality of a concept is of an interest or relative importance of different components is of an interest. This specific information cannot be obtained by applying simple score approach to measuring HIV knowledge.

The application of both methods in the thesis enabled comparison between the two approaches and summary of both methods became available as a result of the analysis. This summary can be helpful for the researchers who are involved in projects which focus on any types of health-related knowledge which cannot be collected as a single variable and, therefore, measures of knowledge need to be derived from different knowledge components.

As mentioned earlier, the main results of the thesis suggest that China has succeeded in improving women's HIV awareness and knowledge over time. If the achievements in other dimensions of HIV prevention are equally successful, China might be considered a success story in controlling HIV epidemic and might provide useful lessons for other countries in the world. However, in order to fully understand the overall successes and failures of HIV prevention in China, other dimensions and aspects of HIV prevention should be considered in order to be able to answer the question how the HIV epidemic in China became controlled (if it did). However, those questions are outside of the remit of this study and further work is required in order to answer this question with confidence.

It could be speculated that increased HIV awareness and knowledge among different groups of women in China might have contributed to adoption of protective behaviours such as avoidance of illegal blood donation, increased rate of condom use and others which in turn might have decreased the speed of spread of HIV infection across the country. However, in order to establish these links, further work needs to be conducted. Avenues for further work are discussed in section 6.3 of this chapter.

6.2 Policy Recommendations

The results of the analyses suggested that despite the fact that HIV awareness and knowledge increased among different groups of women there are still women without adequate levels of HIV awareness and knowledge in China. Therefore, specific groups should be targeted for interventions. The following groups of women still lack HIV awareness and adequate levels of different types of HIV knowledge in China: women with no or primary education, women living in rural areas, women from ethnic minorities as well as women living in Western and Central parts of China. These findings are in agreement with targets for interventions for CP5. The results of the analyses suggest that improvements in the levels of HIV awareness and knowledge are observed over time but more efforts are needed in order to ensure high level of HIV awareness and knowledge across the whole China. China has a specific feature of its HIV epidemic which distinguishes it from other countries in the world: until recently HIV predominantly affected people living in rural areas. Therefore, people in rural areas should be specifically targeted by effective information interventions as the levels of HIV awareness and knowledge is lower in rural areas.

The result of the analysis in the five country context suggested that in Malawi (a country where HIV epidemic is mature and generalised) the level of knowledge about some of the ways to prevent HIV is still low. Unfortunately, the current research could not identify if this level was always low or it reduced recently. This result suggest that sustainability and continuity of intervention efforts is required (and not only in area of HIV awareness and knowledge but also in all other aspects of HIV prevention) to ensure that HIV knowledge does not decrease with time and that new cohorts of women are equipped with necessary HIV awareness and knowledge. Furthermore, these sustained efforts should be in place in all countries irrespective of the level of the HIV epidemic.

Findings from the descriptive analysis suggest that the emphasis should be placed on the messages that HIV is a sexually transmitted infection in China.

6.3 Suggestions for Further Work

There are number of ways how the current analyses can be expanded. New available datasets (Malawi DHS 2010, Kenya DHS 2008-2009 and the Chinese 2006 dataset) can be included in the analysis to obtain more up-to-date information about advances with HIV awareness and knowledge in China and in other cultural and epidemiological contexts. If the Chinese Family Planning and Reproductive Health 2006 Survey becomes publicly available, the same analysis can be repeated with this nationally representative dataset. This would help to obtain results which could be generalised to the entire population of China. Analysis of Malawi DHS 2000 dataset would help to address the question about the levels of knowledge about ways to prevent HIV and to assess whether these levels were similar in 2000 to those in 2004 and were always lower than expected or if they decreased with time due to inadequate efforts of interventions and campaigns.

More countries could be included in the analysis to observe evolution of the HIV epidemic in the world and to answer the question if the assumptions used for the analyses in this thesis are justified.

The analyses presented in the thesis assessed different measures of HIV knowledge as response variables. Further analysis could help the investigation of different approaches to measuring HIV knowledge in the explanatory variable context.

HIV knowledge measures derived in this thesis can be used in the structural equation modelling framework. This approach can help investigating links within the Knowledge, Attitude, Behaviour framework and to study the relationships between different components of the framework in different cultural contexts. In the Chinese context the following dataset can be used for the analysis: the AIDS-Related Knowledge, Attitudes, Behaviour, and Practices: A Survey of 6 Chinese Cities (2008) which is not a nationally representative survey but a rare survey which collected not only HIV awareness and knowledge data but also behavioural data for different groups of people (males and females, migrant workers, youth, blue-collar workers, white-collar workers) (CHAMP 2008).

As mentioned earlier, young people's (15-24 years of age males and females) knowledge of HIV indicator (13) is one of the UNGASS indicators for Millennium Development Goal target 6.A (to halt and begin to reverse the spread of HIV/AIDS by 2015) (UNAIDS 2010). Therefore, it is important to conduct comprehensive analysis of HIV knowledge among young people. It would be useful to repeat the analysis in the five country context for just young women and to reduce the number of components of HIV knowledge to the ones which are considered to be important for the Millennium Development Goal targets.

It would be also important to conduct a study which would collect information about all HIV awareness and knowledge-related interventions which took place in China before 2005. The results of this study might help to isolate specific effects of the interventions, including UNFPA interventions, which were implemented in China between 2003 and 2005 and to assess the effectiveness of these interventions.

All these studies would help advance understanding of different aspects of HIV prevention strategies and might also help improve effectiveness of HIV prevention efforts worldwide.

Appendix A

Tables A.1 and A.2 present the wording of questions from the questionnaires which were used for the calculation of different HIV knowledge measures in both study contexts in Papers Two and Three. These tables show the availability of comparable variables in different surveys used for the analysis for both study contexts.

Table A.1: Wording of questions for variables included in the calculation of different HIV knowledge measures in China.

	China 1997	China 2001	China 2005
Blood transfusion	Can HIV be transmitted through blood transfusion?	Can HIV be transmitted through blood transfusion?	Can HIV be transmitted through blood transfusion?
Sharing needles	Can HIV be transmitted through injections?	Can HIV be transmitted through contaminated needles?	Can HIV be transmitted through contaminated needles?
MTCT	Can HIV be transmitted from the infected mother to her child?	Can HIV be transmitted from the infected mother to her child?	Can HIV be transmitted from the infected mother to her child?
Sexual transmission	Can having multiple sex partners transmit HIV?	Can having multiple sex partners transmit HIV?	Can HIV be transmitted through sexual intercourse?
Handshaking	Can HIV be transmitted through a handshake?	Can HIV be transmitted through a handshake?	Can HIV be transmitted through a handshake?
Sharing food or utensils	Can HIV be transmitted through sharing food and utensils with an HIV infected person?	Can HIV be transmitted through sharing food and utensils with an HIV infected person?	Can HIV be transmitted through sharing food and utensils with an HIV infected person?
Kissing	Can HIV be transmitted through kissing?	Can HIV be transmitted through kissing?	Can HIV be transmitted through kissing?

Sources: the China National Population and Reproductive Health Survey 1997, the China National Family Planning and Reproductive Health Survey 2001, and the UNFPA Reproductive Health and Family Planning Surveys 2005 questionnaires.

Table A.2: Wording of questions for variables included in the calculation of different HIV knowledge measures in the five countries.

	China 2005	India 2006	Kenya 2003	Malawi 2004	Ukraine 2007
One partner	Would any of the following help reduce the risk of getting HIV/AIDS? Limit sex to single partner	In your opinion, can people reduce their chances of getting HIV/AIDS by having just one uninfected sex partner who has no other sex partners?	Can people reduce their chances of getting the AIDS virus by having just one sex partner who has no other partners?	Can people reduce their chances of getting the AIDS virus by having just one sex partner who has no other partners?	Can people reduce their chances of getting the AIDS virus by having just one uninfected sex partner who has no other sex partners?

Condom	Would any of the following help reduce the risk of getting HIV/AIDS? Condom	In your opinion, can people reduce their chances of getting HIV/AIDS by using a condom every time they have sex?	Can people reduce their chances of getting the AIDS virus by using a condom every time they have sex?	Can people reduce their chances of getting the AIDS virus by using a condom every time they have sex?	Can people reduce their chances of getting the AIDS virus by using a condom every time they have sex?
Abstinence	Would any of the following help reduce the risk of getting HIV/AIDS? Abstain from sex	In your opinion, can people reduce their chance of getting HIV/AIDS by abstaining from sexual intercourse?	Can people reduce their chance of getting the AIDS virus by not having sex at all?	Can people reduce their chance of getting the AIDS virus by not having sex at all?	Can people reduce their chance of getting the AIDS virus by not having sexual intercourse at all?
MTCT	Which of the following ways is possible for HIV/AIDS transmission? From mother to baby	Can HIV be transmitted from a mother to her baby?	Can the virus that causes AIDS be transmitted from a mother to a child?	Can the virus that causes AIDS be transmitted from a mother to a child?	Can the virus that causes AIDS be transmitted from a mother to her baby: During pregnancy? During delivery? By breastfeeding?
Healthy-looking person	Can you tell if a healthy-looking person has the HIV virus?	Is it possible for a healthy-looking person to have HIV/AIDS?	Is it possible for a healthy-looking person to have the AIDS virus?	Is it possible for a healthy-looking person to have the AIDS virus?	Is it possible for a healthy-looking person to have the AIDS virus?
Mosquito bites	Which of the following ways is possible for HIV/AIDS transmission? Mosquito bite	In your opinion, can people get HIV/AIDS from mosquito bites?	Can people get the AIDS virus from mosquito or other insect bites?	Can people get the AIDS virus from mosquito bites?	Can people get the AIDS virus from mosquito bites?
Sharing food	Which of the following ways is possible for HIV/AIDS transmission? Sharing utensils with HIV positive person	In your opinion, can people get HIV/AIDS by sharing food with a person who has AIDS?	Can people get the AIDS virus by sharing utensils with a person who has AIDS?	Can people get the AIDS virus by sharing food with a person who has AIDS?	Can people get the AIDS virus by sharing food and utensils with a person who has AIDS virus?

Sources: the UNFPA Reproductive Health and Family Planning Surveys 2005, India DHS 2006, Kenya DHS 2003, Malawi DHS 2004 and Ukraine DHS 2007 questionnaires.

Appendix B

Table B.1 presents demographic and other characteristics of women in four Chinese datasets used for the analysis in Paper One together with HIV awareness by these selected characteristics.

Table B.1: HIV awareness by selected characteristics of women, China, 1997-2005[52].

	% distribution by variables				% of those who are aware of HIV within groups			
Variable	**1997**	**2001**	**2003**	**2005**	**1997**	**2001**	**2003**	**2005**
Total sample size	15213	39586	8400	7356				
Residence								
Rural	76.7	74.6	71.4	68.5	55.0	65.7	75.9	91.8
Urban	23.3	25.4	28.6	31.5	91.4	93.4	93.7	95.5
Age group								
<20	12.0	10.6	6.9	6.5	68.7	75.4	83.7	93.8
20-29	32.6	25.0	22.0	21.5	70.1	79.0	84.6	96.2
30-39	30.3	38.0	42.4	39.7	64.1	74.0	81.3	93.7
40-49	25.0	26.4	28.7	32.3	51.5	63.7	77.6	89.6
Ethnicity								
Non-Han	9.3	9.6	14.0	13.6	42.8	50.7	67.3	89.2
Han	90.7	90.4	86.0	86.4	65.6	75.0	83.5	93.5
Education								
No education	21.4	16.6	9.2	4.3	28.4	36.5	45.8	70.6
Primary	22.9	28.6	28.3	24.3	54.2	62.1	71.2	85.4
Junior secondary	32.2	36.1	41.6	45.0	79.4	85.2	88.2	96.0
Senior secondary and above	16.6	18.7	21.0	26.4	94.3	97.1	96.0	98.7
Marital status								
Never married	20.4	16.3	11.6	10.5	73.2	80.0	84.7	95.6
Married or remarried[53]	78.1	82.0	86.8	87.5	61.3	71.3	80.7	92.8
Divorced or widowed	1.4	1.7	1.5	1.9	64.3	69.3	82.7	88.4
Region								
Eastern			37.6	39.4			88.4	94.5
Central			30.4	28.1			82.6	92.2
Western			32.0	32.6			72.5	92.2
Occupation								

[52] Weights are applied for calculations of percentages for 2003 and 2005 surveys.

[53] This variable under category married or remarried, also include living or not living together, and cohabiting.

Agricultural work	44.1	40.0		72.7	89.6

Non-agricultural manual work	18.5	16.2	85.1	95.5
Non-agricultural intellectual work	9.6	12.2	97.9	98.8
Housework and others	21.0	22.7	84.5	92.8
In school and out of work	6.9	9.0	91.2	96.7
Radio				
Regularly	14.7	11.1	91.8	96.0
Occasionally	32.5	26.9	88.2	95.4
Never	52.8	62.0	74.1	91.4
TV				
Regularly	78.1	85.4	86.1	94.4
Occasionally	19.5	13.6	67.8	84.9
Never	2.5	1.0	33.2	76.0
Newspaper				
Regularly	24.4	25.0	95.0	99.1
Occasionally	34.1	36.2	90.7	95.9
Never	41.6	38.8	81.2	86.7

Appendix C

Tables C.1 and C.2 present results from the detailed decomposition analysis for two sets of datasets used for the analysis in Paper One.

Table C.1: Detailed decomposition analysis (1997 and 2001).

Total

	1997 βs	2001 βs	change due to βs
1997 xs	0.635	0.700	0.065
2001 xs	0.660	0.727	**0.068**
change due to xs	**0.025**	0.027	

Residence - urban

1997 xs	0.914	0.932	0.018
2001 xs	0.915	0.934	**0.019**
change due to xs	**0.001**	0.002	

Residence - rural

1997 xs	0.550	0.629	0.079
2001 xs	0.572	0.657	**0.084**
change due to xs	**0.023**	0.028	

Nationality - han

1997 xs	0.656	0.724	0.068
2001 xs	0.680	0.750	**0.070**
change due to xs	**0.024**	0.027	

Nationality - non-han

1997 xs	0.428	0.466	0.038
2001 xs	0.465	0.507	**0.042**
change due to xs	**0.037**	0.041	

Education - no education

1997 xs	0.284	0.364	0.080
2001 xs	0.282	0.365	**0.083**
change due to xs	**-0.002**	0.002	

Education - primary

1997 xs	0.542	0.625	0.083
2001 xs	0.531	0.621	**0.090**
change due to xs	**-0.011**	-0.004	

Education - junior secondary

1997 xs	0.794	0.852	0.058
2001 xs	0.789	0.852	**0.062**
change due to xs	**-0.005**	0.000	

Education - senior secondary and above

1997 xs	0.943	0.972	0.029
2001 xs	0.942	0.971	**0.029**
change due to xs	**-0.001**	-0.001	

Age group - younger than 20

1997 xs	0.687	0.715	0.028
2001 xs	0.725	0.754	**0.029**
change due to xs	**0.038**	0.040	

Age group - 20-29

1997 xs	0.701	0.753	0.052
2001 xs	0.741	0.790	**0.049**
change due to xs	**0.040**	0.037	

Age group - 30-39

1997 xs	0.641	0.725	0.084
2001 xs	0.655	0.740	**0.085**
change due to xs	**0.014**	0.015	

Age group - 40-49

1997 xs	0.515	0.591	0.077
2001 xs	0.562	0.637	**0.075**
change due to xs	**0.047**	0.046	

Marital status - never married

1997 xs	0.732	0.764	0.031
2001 xs	0.768	0.800	**0.032**
change due to xs	**0.036**	0.037	

Marital status – married or remarried			
1997 xs	0.613	0.685	0.072
2001 xs	0.639	0.713	**0.074**
change due to xs	**0.026**	0.028	

Marital status – divorced or widowed			
1997 xs	0.643	0.721	0.078
2001 xs	0.607	0.692	**0.085**
change due to xs	**-0.038**	-0.029	

Table C.2: Detailed decomposition analysis (2003 and 2005).

Total

	2003 βs	2005 βs	change due to βs
2003 xs	0.812	0.912	0.100
2005 xs	0.837	0.929	**0.092**
change due to xs	**0.025**	0.017	

Residence - urban			
2003 xs	0.937	0.958	0.021
2005 xs	0.931	0.955	**0.024**
change due to xs	**-0.006**	-0.003	

Residence - rural			
2003 xs	0.759	0.892	0.133
2005 xs	0.796	0.918	**0.122**
change due to xs	**0.037**	0.026	

Nationality - han			
2003 xs	0.835	0.920	0.085
2005 xs	0.855	0.935	**0.080**
change due to xs	**0.020**	0.015	

Nationality - non-han			
2003 xs	0.673	0.860	0.187
2005 xs	0.729	0.892	**0.163**
change due to xs	**0.056**	0.032	

Education - no education

2003 xs	0.458	0.688	0.230
2005 xs	0.489	0.706	**0.217**
change due to xs	**0.031**	0.018	

Education - primary

2003 xs	0.712	0.855	0.143
2005 xs	0.708	0.854	**0.146**
change due to xs	**-0.004**	-0.001	

Education - junior secondary

2003 xs	0.882	0.961	0.079
2005 xs	0.874	0.960	**0.086**
change due to xs	**-0.008**	-0.001	

Education - senior secondary and above

2003 xs	0.960	0.987	0.027
2005 xs	0.959	0.987	**0.028**
change due to xs	**-0.001**	0.000	

Age group - younger than 20

2003 xs	0.837	0.938	0.101
2005 xs	0.830	0.938	**0.108**
change due to xs	**-0.007**	0.000	

Age group - 20-29

2003 xs	0.846	0.950	0.104
2005 xs	0.868	0.962	**0.094**
change due to xs	**0.022**	0.012	

Age group - 30-39

2003 xs	0.813	0.918	0.105
2005 xs	0.844	0.937	**0.093**
change due to xs	**0.031**	0.019	

Age group - 40-49

2003 xs	0.776	0.865	0.089
2005 xs	0.809	0.896	**0.087**
change due to xs	**0.033**	0.031	

Marital status - never married

2003 xs	0.847	0.955	0.108
2005 xs	0.851	0.956	**0.105**
change due to xs	**0.004**	0.001	

Marital status – married or remarried

2003 xs	0.807	0.907	0.100
2005 xs	0.836	0.928	**0.092**
change due to xs	**0.029**	0.021	

Marital status – divorced or widowed

2003 xs	0.827	0.868	0.041
2005 xs	0.836	0.884	**0.048**
change due to xs	**0.009**	0.016	

Region -Eastern

2003 xs	0.884	0.937	0.053
2005 xs	0.885	0.944	**0.059**
change due to xs	**0.001**	0.007	

Region-Central

2003 xs	0.826	0.914	0.088
2005 xs	0.839	0.922	**0.083**
change due to xs	**0.013**	0.008	

Region-Western

2003 xs	0.725	0.884	0.159
2005 xs	0.791	0.922	**0.131**
change due to xs	**0.066**	0.038	

Occupation – Agriculture

2003 xs	0.727	0.872	0.145
2005 xs	0.750	0.895	**0.145**
change due to xs	**0.023**	0.023	

Occupation – Non-agricultural manual work

2003 xs	0.851	0.944	0.093
2005 xs	0.877	0.955	**0.078**
change due to xs	**0.026**	0.011	

Occupation – Non-agricultural intellectual work

2003 xs	0.979	0.989	0.010
2005 xs	0.976	0.988	**0.012**
change due to xs	-0.003	-0.001	

Occupation – Housework and Others

2003 xs	0.845	0.914	0.069
2005 xs	0.869	0.928	**0.059**
change due to xs	**0.024**	0.014	

Occupation – In school of Out of work

2003 xs	0.912	0.968	0.056
2005 xs	0.906	0.967	**0.061**
change due to xs	**-0.006**	-0.001	

Radio – Regularly

2003 xs	0.918	0.957	0.039
2005 xs	0.928	0.960	**0.032**
change due to xs	**0.010**	0.003	

Radio – Occasionally

2003 xs	0.882	0.928	0.046
2005 xs	0.918	0.954	**0.036**
change due to xs	**0.036**	0.026	

Radio – Never

2003 xs	0.741	0.889	0.148
2005 xs	0.789	0.914	**0.125**
change due to xs	**0.048**	0.025	

TV – Regularly

2003 xs	0.861	0.941	0.080
2005 xs	0.860	0.944	**0.084**
change due to xs	**-0.001**	0.003	

TV – Occasionally

2003 xs	0.678	0.823	0.145
2005 xs	0.725	0.848	**0.123**
change due to xs	**0.047**	0.025	

TV – Never

2003 xs	0.332	0.692	0.360
2005 xs	0.414	0.760	**0.346**
change due to xs	**0.082**	0.068	

Newspaper – Regularly			
2003 xs	0.950	0.990	0.040
2005 xs	0.953	0.991	**0.038**
change due to xs	**0.003**	0.001	

Newspaper – Occasionally			
2003 xs	0.907	0.952	0.045
2005 xs	0.917	0.959	**0.042**
change due to xs	**0.010**	0.007	

Newspaper – Never			
2003 xs	0.654	0.833	0.179
2005 xs	0.699	0.867	**0.168**
change due to xs	**0.045**	0.034	

Appendix D

This appendix presents the codes for the bootstrap procedures for the estimation of confidence intervals used in Paper One.

Code for the bootstrap procedure for estimation of confidence intervals (1997 and 2001)

```
set more off
clear
set mem 600m
set matsize 800
use "\\cc-arran\users\om206\My Documents\PhDwork\Data China-DATASETS
USED FOR ANALYSIS\Decomposition analysis\mergeddata.dta", clear
gen idobs=_n
char nationality [omit] 2
char maritalst [omit] 5
save "C:\Documents and Settings\om206\BOOTSTRAP\mergeddatac.dta",
replace

forvalues j=1/1000{
set more off
use "C:\Documents and Settings\om206\BOOTSTRAP\mergeddatac.dta", clear
bsample
quietly xi: logit HIV i.residence i.agegroup i.nationality i.education
i.maritalst if year==1997
predict pred`j'
drop _I*
egen totmean=mean(pred`j') if year==2001
sort ageg
by ageg: egen age_m1=mean(pred`j') if ageg==1 & year==2001
sort ageg
by ageg: egen age_m2=mean(pred`j') if ageg==2 & year==2001
sort ageg
by ageg: egen age_m3=mean(pred`j') if ageg==3 & year==2001
sort ageg
by ageg: egen age_m4=mean(pred`j') if ageg==4 & year==2001
sort resi
by resi: egen res_m1=mean(pred`j') if resi==1 & year==2001
sort resi
by resi: egen res_m2=mean(pred`j') if resi==2 & year==2001
sort nati
by nati: egen nat_m1=mean(pred`j') if nati==1 & year==2001
sort nati
by nati: egen nat_m2=mean(pred`j') if nati==2 & year==2001
sort educat
by educat: egen educ_m1=mean(pred`j') if educat==1 & year==2001
sort educat
by educat: egen educ_m2=mean(pred`j') if educat==2 & year==2001
sort educat
by educat: egen educ_m3=mean(pred`j') if educat==3 & year==2001
sort educat
by educat: egen educ_m4=mean(pred`j') if educat==4 & year==2001
```

```
sort marital
by marital: egen marit_m1=mean(pred`j') if marital==1 & year==2001
sort marital
by marital: egen marit_m2=mean(pred`j') if marital==2 & year==2001
sort marital
by marital: egen marit_m3=mean(pred`j') if marital==3 & year==2001

quietly xi: logit HIV i.residence i.agegroup i.nationality i.education
i.maritalst if year==2001
predict predtwo`j'
drop _I*
egen totmeantwo=mean(predtwo`j') if year==2001
sort ageg
by ageg: egen age_m1two=mean(predtwo`j') if ageg==1 & year==2001
sort ageg
by ageg: egen age_m2two=mean(predtwo`j') if ageg==2 & year==2001
sort ageg
by ageg: egen age_m3two=mean(predtwo`j') if ageg==3 & year==2001
sort ageg
by ageg: egen age_m4two=mean(predtwo`j') if ageg==4 & year==2001
sort resi
by resi: egen res_m1two=mean(predtwo`j') if resi==1 & year==2001
sort resi
by resi: egen res_m2two=mean(predtwo`j') if resi==2 & year==2001
sort nati
by nati: egen nat_m1two=mean(predtwo`j') if nati==1 & year==2001
sort nati
by nati: egen nat_m2two=mean(predtwo`j') if nati==2 & year==2001
sort educat
by educat: egen educ_m1two=mean(predtwo`j') if educat==1 & year==2001
sort educat
by educat: egen educ_m2two=mean(predtwo`j') if educat==2 & year==2001
sort educat
by educat: egen educ_m3two=mean(predtwo`j') if educat==3 & year==2001
sort educat
by educat: egen educ_m4two=mean(predtwo`j') if educat==4 & year==2001
sort marital
by marital: egen marit_m1two=mean(predtwo`j') if marital==1 &
year==2001
sort marital
by marital: egen marit_m2two=mean(predtwo`j') if marital==2 &
year==2001
sort marital
by marital: egen marit_m3two=mean(predtwo`j') if marital==3 &
year==2001

drop  HIV residence agegroup nationality education maritalst year
idobs pred* predtwo*
collapse *
save "C:\Documents and Settings\om206\BOOTSTRAP\mean`j'.dta", replace
}

clear
set more off
use "C:\Documents and Settings\om206\BOOTSTRAP\mean1.dta", clear
forvalues j=2/1000{
append using "C:\Documents and Settings\om206\BOOTSTRAP\mean`j'.dta",
nolabel
}

outsheet using "C:\Documents and
Settings\om206\BOOTSTRAP\meanFINAL.csv", comma replace
```

Code for the bootstrap procedure for estimation of confidence intervals (2003 and 2005)

```
set more off
clear
set mem 600m
set matsize 800
use "C:\Documents and
Settings\om206\Desktop\Decomposition\decompositionUNFPAwithextravars-
strata.dta", clear
gen idobs=_n
char region [omit] 3
char nationality [omit] 2
char residence [omit]2
char agegroups [omit] 4
save "C:\Documents and
Settings\om206\BOOTSTRAP\mergeddataUNFPAonestrataextra.dta", replace
forvalues j=1/1000{
set more off
use "C:\Documents and
Settings\om206\BOOTSTRAP\mergeddataUNFPAonestrataextra.dta", clear
bsample, strata (eastur eastrur centralur centralrur westur westrur)
quietly xi: logit HIVnew i.region i.residence i.agegroups
i.nationality i.education i.marstat i.occupation i.radio i.tv
i.newspaper if year==2003
predict pred`j'
drop _I*
egen totmean=mean(pred`j') if year==2005
sort region
by region: egen reg_m1=mean(pred`j') if region==1 & year==2005
sort region
by region: egen reg_m2=mean(pred`j') if region==2 & year==2005
sort region
by region: egen reg_m3=mean(pred`j') if region==3 & year==2005
sort ageg
by ageg: egen age_m1=mean(pred`j') if ageg==1 & year==2005
sort ageg
by ageg: egen age_m2=mean(pred`j') if ageg==2 & year==2005
sort ageg
by ageg: egen age_m3=mean(pred`j') if ageg==3 & year==2005
sort ageg
by ageg: egen age_m4=mean(pred`j') if ageg==4 & year==2005
sort resi
by resi: egen res_m1=mean(pred`j') if resi==1 & year==2005
sort resi
by resi: egen res_m2=mean(pred`j') if resi==2 & year==2005
sort nati
by nati: egen nat_m1=mean(pred`j') if nati==1 & year==2005
sort nati
by nati: egen nat_m2=mean(pred`j') if nati==2 & year==2005
sort educat
by educat: egen educ_m1=mean(pred`j') if educat==1 & year==2005
sort educat
by educat: egen educ_m2=mean(pred`j') if educat==2 & year==2005
sort educat
by educat: egen educ_m3=mean(pred`j') if educat==3 & year==2005
sort educat
by educat: egen educ_m4=mean(pred`j') if educat==4 & year==2005
sort marstat
by marstat: egen marit_m1=mean(pred`j') if marstat==1 & year==2005
```

```
sort marstat
by marstat: egen marit_m2=mean(pred`j') if marstat==2 & year==2005
sort marstat
by marstat: egen marit_m3=mean(pred`j') if marstat==3 & year==2005
sort occupation
by occupation: egen occup_m1=mean(pred`j') if occupation==1 &
year==2005
sort occupation
by occupation: egen occup_m2=mean(pred`j') if occupation==2 &
year==2005
sort occupation
by occupation: egen occup_m3=mean(pred`j') if occupation==3 &
year==2005
sort occupation
by occupation: egen occup_m4=mean(pred`j') if occupation==4 &
year==2005
sort occupation
by occupation: egen occup_m5=mean(pred`j') if occupation==5 &
year==2005
sort radio
by radio:egen radio_m1=mean(pred`j') if radio==1 & year==2005
sort radio
by radio:egen radio_m2=mean(pred`j') if radio==2 & year==2005
sort radio
by radio:egen radio_m3=mean(pred`j') if radio==3 & year==2005
sort tv
by tv:egen tv_m1=mean(pred`j') if tv==1 & year==2005
sort tv
by tv:egen tv_m2=mean(pred`j') if tv==2 & year==2005
sort tv
by tv:egen tv_m3=mean(pred`j') if tv==3 & year==2005
sort newspaper
by newspaper:egen news_m1=mean(pred`j') if newspaper==1 & year==2005
sort newspaper
by newspaper:egen news_m2=mean(pred`j') if newspaper==2 & year==2005
sort newspaper
by newspaper:egen news_m3=mean(pred`j') if newspaper==3 & year==2005


quietly xi: logit HIVnew i.region i.residence i.agegroups
i.nationality i.education i.marstat i.occupation i.radio i.tv
i.newspaper if year==2005
predict predtwo`j'
drop _I*
egen totmeantwo=mean(predtwo`j') if year==2005
sort region
by region: egen reg_m1two=mean(predtwo`j') if region==1 & year==2005
sort region
by region: egen reg_m2two=mean(predtwo`j') if region==2 & year==2005
sort region
by region: egen reg_m3two=mean(predtwo`j') if region==3 & year==2005
sort ageg
by ageg: egen age_m1two=mean(predtwo`j') if ageg==1 & year==2005
sort ageg
by ageg: egen age_m2two=mean(predtwo`j') if ageg==2 & year==2005
sort ageg
by ageg: egen age_m3two=mean(predtwo`j') if ageg==3 & year==2005
sort ageg
by ageg: egen age_m4two=mean(predtwo`j') if ageg==4 & year==2005
sort resi
by resi: egen res_m1two=mean(predtwo`j') if resi==1 & year==2005
sort resi
by resi: egen res_m2two=mean(predtwo`j') if resi==2 & year==2005
sort nati
```

```
by nati: egen nat_m1two=mean(predtwo`j') if nati==1 & year==2005
sort nati
by nati: egen nat_m2two=mean(predtwo`j') if nati==2 & year==2005
sort educat
by educat: egen educ_m1two=mean(predtwo`j') if educat==1 & year==2005
sort educat
by educat: egen educ_m2two=mean(predtwo`j') if educat==2 & year==2005
sort educat
by educat: egen educ_m3two=mean(predtwo`j') if educat==3 & year==2005
sort educat
by educat: egen educ_m4two=mean(predtwo`j') if educat==4 & year==2005
sort marstat
by marstat: egen marit_m1two=mean(predtwo`j') if marstat==1 &
year==2005
sort marstat
by marstat: egen marit_m2two=mean(predtwo`j') if marstat==2 &
year==2005
sort marstat
by marstat: egen marit_m3two=mean(predtwo`j') if marstat==3 &
year==2005
sort occupation
by occupation: egen occup_m1two=mean(predtwo`j') if occupation==1 &
year==2005
sort occupation
by occupation: egen occup_m2two=mean(predtwo`j') if occupation==2 &
year==2005
sort occupation
by occupation: egen occup_m3two=mean(predtwo`j') if occupation==3 &
year==2005
sort occupation
by occupation: egen occup_m4two=mean(predtwo`j') if occupation==4 &
year==2005
sort occupation
by occupation: egen occup_m5two=mean(predtwo`j') if occupation==5 &
year==2005
sort radio
by radio:egen radio_m1two=mean(predtwo`j') if radio==1 & year==2005
sort radio
by radio:egen radio_m2two=mean(predtwo`j') if radio==2 & year==2005
sort radio
by radio:egen radio_m3two=mean(predtwo`j') if radio==3 & year==2005
sort tv
by tv:egen tv_m1two=mean(predtwo`j') if tv==1 & year==2005
sort tv
by tv:egen tv_m2two=mean(predtwo`j') if tv==2 & year==2005
sort tv
by tv:egen tv_m3two=mean(predtwo`j') if tv==3 & year==2005
sort newspaper
by newspaper:egen news_m1two=mean(predtwo`j') if newspaper==1 &
year==2005
sort newspaper
by newspaper:egen news_m2two=mean(predtwo`j') if newspaper==2 &
year==2005
sort newspaper
by newspaper:egen news_m3two=mean(predtwo`j') if newspaper==3 &
year==2005

drop  HIV region residence agegroups nationality education marstat tv
radio occupation newspaper year eastur eastrur centralur centralrur
westur westrur idobs pred* predtwo*
collapse *
save "C:\Documents and
Settings\om206\BOOTSTRAP\meanUNFPAonestratacombextra`j'.dta", replace
}
```

```
clear
set more off
use "C:\Documents and
Settings\om206\BOOTSTRAP\meanUNFPAonestratacombextra1.dta", clear
forvalues j=2/1000{
append using "C:\Documents and
Settings\om206\BOOTSTRAP\meanUNFPAonestratacombextra`j'.dta", nolabel
}

outsheet using "C:\Documents and
Settings\om206\BOOTSTRAP\meanUNFPAonestratacombextraFINAL.csv", comma
replace
```

Appendix E

Table E.1 presents constructed 95% confidence intervals for the average predicted probabilities of HIV awareness estimated by the bootstrap procedures in Paper One.

Table E.1: 95% confidence intervals estimated for the average predicted probabilities of HIV awareness for different groups of women.

	95% CI for the average predicted probabilities of HIV awareness			
	2001 sample, 1997 equation	**2001 sample, 2001 equation**	**2005 sample, 2003 equation**	**2005 sample, 2005 equation**
Total	(0.652, 0.667)	(0.722, 0.731)	(0.829, 0.846)	(0.923, 0.935)
Residence				
Urban	(0.906, 0.924)	(0.928, 0.938)	(0.920, 0.941)	(0.946, 0.963)
Rural	(0.564, 0.581)	(0.651, 0.662)	(0.784, 0.807)	(0.910, 0.926)
Nationality				
Han	(0.673, 0.688)	(0.746, 0.755)	(0.846, 0.864)	(0.929, 0.941)
Ethnic minorities	(0.442, 0.491)	(0.491, 0.522)	(0.704, 0.753)	(0.872, 0.910)
Education				
No education	(0.267, 0.298)	(0.354, 0.377)	(0.453, 0.529)	(0.652, 0.757)
Primary	(0.515, 0.545)	(0.612, 0.629)	(0.689, 0.728)	(0.838, 0.870)
Junior secondary	(0.778, 0.800)	(0.846, 0.857)	(0.862, 0.885)	(0.953, 0.967)
Senior secondary and above	(0.933, 0.951)	(0.968, 0.975)	(0.950, 0.969)	(0.982, 0.991)
Age group				
< 20	(0.708, 0.744)	(0.741, 0.767)	(0.796, 0.862)	(0.913, 0.960)
20-29	(0.731, 0.753)	(0.782, 0.798)	(0.852, 0.883)	(0.952, 0.972)
30-39	(0.642, 0.668)	(0.733, 0.747)	(0.832, 0.856)	(0.929, 0.946)
40-49	(0.547, 0.576)	(0.628, 0.646)	(0.792, 0.826)	(0.884, 0.908)
Marital status				
Never married	(0.755, 0.782)	(0.790, 0.809)	(0.826, 0.876)	(0.940, 0.970)
Married or remarried	(0.630, 0.647)	(0.708, 0.718)	(0.827, 0.846)	(0.922, 0.934)
Divorced or widowed	(0.551, 0.668)	(0.659, 0.728)	(0.767, 0.893)	(0.823, 0.932)
Region				
Eastern			(0.873, 0.896)	(0.935, 0.954)
Central			(0.823, 0.854)	(0.911, 0.932)
Western			(0.774, 0.806)	(0.912, 0.932)
Occupation				
Agriculture			(0.734, 0.767)	(0.885, 0.906)
Non-agricultural manual work			(0.862, 0.894)	(0.943, 0.966)
Non-agricultural intellectual work			(0.964, 0.987)	(0.981, 0.994)
Housework and others			(0.853, 0.883)	(0.916, 0.941)
In school or out of job			(0.881, 0.928)	(0.952, 0.981)
Radio				
Regularly			(0.912, 0.942)	(0.945, 0.974)
Occasionally			(0.908, 0.927)	(0.944, 0.962)

Never	(0.777, 0.801)	(0.906, 0.922)
TV		
Regularly	(0.851, 0.870)	(0.939, 0.950)
Occasionally	(0.702, 0.749)	(0.826, 0.871)
Never	(0.332, 0.493)	(0.655, 0.855)
Newspaper		
Regularly	(0.943, 0.961)	(0.986, 0.995)
Occasionally	(0.907, 0.927)	(0.951, 0.966)
Never	(0.682, 0.716)	(0.855, 0.879)

Appendix F

Table F.1 presents demographic and other characteristics of respondents together with their HIV knowledge in seven separate datasets which are used for the analysis in Papers Two and Three. These seven datasets only include women who reported being aware of HIV.

Table F.1: Characteristics of respondents in seven datasets.

Variables	China 1997, frequency (percentage)	China 2001, frequency (percentage)	China 2005, frequency (percentage)	Kenya 2003, frequency (percentage)	Malawi 2004, frequency (percentage)	India 2006, frequency (percentage)	Ukraine 2007, frequency (percentage)
Sample size	9,653	26,138	6,837	6,877	11,002	88,223	6,638
Area of residence							
Rural	6412 (66.4)	17408 (66.6)	4667 (68.3)	4421 (64.3)	9413 (85.6)	39174 (44.4)	2442 (36.8)
Urban	3241 (33.6)	8730 (33.4)	2170 (31.7)	2456 (35.7)	1589 (14.4)	49049 (55.6)	4196 (63.2)
Age groups							
15-19	1252(13.0)	2862 (10.9)	410 (6.0)	1432 (20.8)	2236 (20.3)	17589 (19.9)	782 (11.8)
20-29	3482 (36.1)	7293 (27.9)	1524 (22.3)	2671 (38.8)	4685 (42.6)	32342 (36.7)	1887 (28.4)
30-39	2959 (30.7)	10169 (38.9)	2796 (40.9)	1717 (25.0)	2476 (22.5)	23389 (26.5)	2004 (30.2)
40-49	1960 (20.3)	5814 (22.2)	2107 (30.8)	1057 (15.4)	1605 (14.6)	14903 (16.9)	1965 (29.6)
Education							
No education	922 (9.6)	1984 (7.6)	218 (3.2)	653 (9.5)	2411 (21.9)	14719 (16.7)	2 (0.0)
Primary	2464 (25.5)	6197 (23.7)	1581 (23.1)	3756 (54.6)	6920 (62.9)	11887 (13.5)	2 (0.0)
Secondary	5718 (59.2)	16109 (61.6)	4585 (67.1)	1891 (27.5)	1603 (14.6)	48713 (55.2)	2821 (42.5)
Higher	549 (5.7)	1848 (7.1)	453 (6.6)	577 (8.4)	68 (0.6)	12904 (14.6)	3813 (57.4)
Marital status							
Never married	1973 (20.4)	4737 (18.1)	659 (9.6)	2088 (30.4)	1777 (16.2)	25524 (28.9)	1459 (22.0)
Married or remarried	7543 (78.1)	21000 (80.3)	6051 (88.5)	4399 (64.0)	8257 (75.0)	59968 (68.0)	4259 (64.2)
Widowed or divorced	137 (1.4)	401 (1.5)	127 (1.9)	390 (5.7)	968 (8.8)	2731 (3.1)	920 (13.9)
Sexual intercourse							
Correct	8217 (85.1)	24158 (92.4)	6467 (94.6)				
Incorrect	1436 (14.9)	1980 (7.6)	370 (5.4)				
Sharing needles							
Correct	6020 (62.4)	23460 (89.8)	6077 (88.9)				
Incorrect	3633 (37.6)	2678 (10.2)	760 (11.1)				
Blood transfusion							
Correct	7042 (73.0)	23643 (90.5)	6460 (94.5)				
Incorrect	2611 (27.0)	2495 (9.5)	377 (5.5)				
MTCT							

Correct	6795 (70.4)	22354 (85.5)	6132 (89.7)	6337 (92.1)	9093 (82.6)	70905 (80.4)	6072 (91.5)
Incorrect	2858 (29.6)	3784 (14.5)	705 (10.3)	540 (7.9)	1909 (17.4)	17318 (19.6)	566 (8.5)
Knowledge that people can reduce chances of getting HIV/AIDS by having just **one partner**							
Correct			5978 (87.4)	6431 (93.5)	7979 (72.5)	67107 (76.1)	5940 (89.5)
Incorrect			859 (12.6)	446 (6.5)	3023 (27.5)	21116 (23.9)	698 (10.5)
Knowledge that people can reduce chances of getting HIV/AIDS by **using a condom** every time they have sex							
Correct			5734 (83.9)	4888 (71.1)	6919 (62.9)	56619 (64.2)	6157 (92.8)
Incorrect			1103 (16.1)	1989 (28.9)	4083 (37.1)	31604 (35.8)	481 (7.2)
Knowledge that people can reduce chance of getting HIV/AIDS by **abstaining** from sexual intercourse							
Correct			5156 (75.4)	6356 (92.4)	8313 (75.6)	60855 (69.0)	5751 (86.6)
Incorrect			1681 (24.6)	521 (7.6)	2689 (24.4)	27368 (31.0)	887 (13.4)
Handshaking							
Correct	4770 (49.4)	16845 (64.4)	5449 (79.7)				
Incorrect	4883 (50.6)	9293 (35.6)	1388 (20.3)				
Sharing utensils or eating together with HIV+ person							
Correct	3291 (34.1)	10823 (41.4)	4547 (66.5)	5736 (83.4)	9495 (86.3)	64830 (73.5)	5123 (77.2)
Incorrect	6362 (65.9)	15315 (58.6)	2290 (33.5)	1141 (16.6)	1507 (13.7)	23393 (26.5)	1515 (22.8)
Kissing							
Correct	2207 (22.9)	8196 (31.4)	3702 (54.1)				
Incorrect	7446 (77.1)	17942 (68.6)	3135 (45.9)				
Mosquito bite							
Correct			2771 (40.5)	4980 (72.4)	7649 (69.5)	56751 (64.3)	4674 (70.4)

Incorrect	4066 (59.5)	1897 (27.6)	3353 (30.5)	31472 (35.7)	1964 (29.6)
Knowledge that a **healthy-looking person** can have AIDS?					
Correct	6149 (89.9)	6229 (90.6)	9276 (84.3)	59008 (66.9)	5126 (77.2)
Incorrect	688 (10.1)	648 (9.4)	1726 (15.7)	29215 (33.1)	1512 (22.8)
Macro-level variables (for cross-country comparison)					
HIV awareness (%)	93.19	98.25	98.72	71.06	98.36
Total HIV prevalence (%)	0.1	6.7	11.8	0.3	1.6
Type of HIV epidemic	Non-generalised	Generalised	Generalised	Non-generalised	Non-generalised
Female literacy[54] (%)	78.8	79.7	49.8	58.3	99.6

54 http://www.nationmaster.com/graph/edu_lit_fem-education-literacy-female for the year 2003 [Accessed 20 November 2010]

Appendix G

Tables G.1-G.8 present results of the partial proportional odds models fitted to both study contexts' pooled datasets from Paper Two.

Table G.1: Results of the partial proportional odds model for Score One in the Chinese context.

Variable	No knowledge β (SE)	One routes β (SE)	Two routes β (SE)	Three routes β (SE)
Intercept	1.581 (0.145)	0.890 (0.140)	0.351 (0.137)	-0.296 (0.136)
Age group				
<20 (ref)	0.000	0.000	0.000	0.000
20-29	0.189 (0.096)	Same	Same	Same
30-39	0.217 (0.101)	Same	Same	Same
40-49	0.179 (0.104)	Same	Same	Same
Education				
No education (ref)	0.000	0.000	0.000	0.000
Primary	0.485 (0.068)	0.454 (0.056)	0.368 (0.048)	0.290 (0.044)
Secondary	1.379 (0.071)	1.291 (0.057)	1.110 (0.049)	0.834 (0.043)
Higher	2.503 (0.287)	2.733 (0.226)	2.446 (0.149)	1.550 (0.079)
Year of survey				
1997 (ref)	0.000	0.000	0.000	0.000
2001	1.789 (0.162)	1.697 (0.122)	1.310 (0.095)	0.782 (0.079)
2005	1.604 (0.276)	1.444 (0.207)	1.134 (0.164)	0.528 (0.136)
Residence				
Urban (ref)	0.000	0.000	0.000	0.000
Rural	-0.547 (0.080)	Same	Same	Same
Marital status				
Widowed or divorced (ref)	0.000	0.000	0.000	0.000
Married or remarried	0.074 (0.092)	Same	Same	Same
Never married	-0.092 (0.103)	Same	Same	Same
Ethnicity				
Han (ref)	0.000	0.000	0.000	0.000
Minority	-0.175 (0.041)	Same	Same	Same
Age*year of survey				
20-29*2001	0.202 (0.079)	Same	Same	Same
20-29*2005	0.360 (0.146)	Same	Same	Same
30-39*2001	0.084 (0.119)	0.123 (0.100)	0.279 (0.087)	0.368 (0.080)
30-39*2005	0.465 (0.139)	Same	Same	Same
40-49*2001	-0.149 (0.125)	-0.022 (0.107)	0.190 (0.094)	0.318 (0.086)
40-49*2005	-0.347 (0.215)	-0.100 (0.183)	0.122 (0.156)	0.185 (0.142)
Age*residence				
20-29*rural	-0.108 (0.087)	Same	Same	Same
30-39*rural	-0.328 (0.084)	Same	Same	Same
40-49*rural	-0.322 (0.088)	Same	Same	Same
Year of survey*residence				

2001*rural	-0.407 (0.143)	-0.228 (0.104)	-0.084 (0.075)	0.121 (0.057)
2005*rural	0.044 (0.145)	0.296 (0.179)	0.369 (0.125)	0.730 (0.082)

Ref – reference category.
Note: Reference category of the response variable is perfect knowledge (four routes).

Table G.2: Results of the partial proportional odds model for Score Two in the Chinese context.

Variable	No knowledge β (SE)	One route β (SE)	Two routes β (SE)
Intercept	-0.389 (0.565)	-1.339 (0.566)	-2.363 (0.569)
Age group			
<20 (ref)	0.000	0.000	0.000
20-29	-0.146 (0.469)	Same	Same
30-39	-0.327 (0.475)	Same	Same
40-49	-0.175 (0.475)	Same	Same
Education			
No education (ref)	0.000	0.000	0.000
Primary	0.216 (0.605)	0.780 (0.606)	0.812 (0.631)
Secondary	1.245 (0.571)	1.080 (0.572)	1.105 (0.575)
Higher	1.523 (0.654)	Same	Same
Year of survey			
1997 (ref)	0.000	0.000	0.000
2001	0.530 (0.125)	0.339 (0.125)	0.302 (0.126)
2005	0.550 (0.198)	Same	Same
Residence			
Urban (ref)	0.000	0.000	0.000
Rural	0.128 (0.154)	Same	Same
Marital status			
Widowed or divorced (ref)	0.000	0.000	0.000
Married or remarried	-0.186 (0.288)	Same	Same
Never married	-0.086 (0.445)	Same	Same
Ethnicity			
Han (ref)	0.000	0.000	0.000
Minority	-0.371 (0.081)	Same	Same
Age*year of survey			
20-29*2001	0.156 (0.073)	Same	Same
20-29*2005	0.261 (0.123)	Same	Same
30-39*2001	0.174 (0.078)	0.239 (0.077)	0.123 (0.081)
30-39*2005	0.551 (0.131)	0.584 (0.124)	0.386 (0.124)
40-49*2001	0.111 (0.084)	Same	Same
40-49*2005	0.399 (0.127)	Same	Same
Age*education			
20-29*primary	0.176 (0.488)	-0.373 (0.488)	-0.463 (0.514)
20-29*secondary	0.090 (0.466)	Same	Same
20-29*higher	0.198 (0.497)	Same	Same
30-39*primary	0.230 (0.495)	-0.373 (0.496)	-0.389 (0.525)
30-39*secondary	0.033 (0.473)	Same	Same
30-39*higher	0.279 (0.511)	Same	Same
40-49*primary	0.166 (0.496)	-0.364 (0.497)	-0.471 (0.527)
40-49*secondary	-0.174 (0.474)	-0.031 (0.475)	-0.130 (0.476)
40-49*higher	-0.237 (0.516)	Same	Same
Age*residence			
20-29*rural	0.224 (0.074)	Same	Same
30-39*rural	0.013 (0.078)	0.124 (0.074)	0.243 (0.077)

40-49*rural	-0.148 (0.085)	-0.060 (0.082)	0.081 (0.087)
Year of survey*education			
2001*primary	0.164 (0.096)	Same	Same
2001*secondary	0.245 (0.094)	Same	Same
2001*higher	0.539 (0.189)	0.383 (0.144)	0.604 (0.144)
2005*primary	0.017 (0.161)	0.254 (0.160)	0.263 (0.166)
2005*secondary	0.591 (0.154)	Same	Same
2005*higher	0.600 (0.206)	Same	Same
Year of survey*ethnicity			
2001*minority	0.407 (0.093)	Same	Same
2005*minority	0.212 (0.105)	Same	Same
Year of survey*residence			
2001*rural	-0.380 (0.054)	Same	Same
2005*rural	0.345 (0.083)	0.188 (0.075)	0.282 (0.075)
Education*marital status			
Primary*married	0.194 (0.330)	Same	Same
Primary*never married	0.671 (0.489)	0.129 (0.489)	-0.167 (0.510)
Secondary*married	0.315 (0.302)	Same	Same
Secondary*never married	0.481 (0.456)	0.394 (0.456)	0.327 (0.456)
Higher*married	1.056 (0.413)	0.564 (0.400)	0.192 (0.397)
Higher*never married	0.810 (0.551)	0.516 (0.535)	0.231 (0.531)
Education*residence			
Primary*rural	-0.296 (0.146)	Same	Same
Secondary*rural	-0.799 (0.138)	-0.653 (0.138)	-0.704 (0.140)
Higher*rural	-0.567 (0.190)	Same	Same

Ref – reference category.

Note: Reference category of the response variable is perfect knowledge (three routes).

Table G.3: Results of the partial proportional odds model for Score Three in the Chinese context.

Variable	No knowledge β (SE)	One route β (SE)	Two routes β (SE)	Three routes β (SE)	Four routes β (SE)	Five routes β (SE)	Six routes β (SE)
Intercept	1.335 (0.241)	1.065 (0.204)	0.586 (0.167)	0.139 (0.143)	-0.820 (0.123)	-1.853 (0.119)	-2.965 (0.129)
Age group							
<20 (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20-29	0.862 (0.263)	0.440 (0.209)	0.385 (0.160)	0.254 (0.128)	-0.023 (0.098)	-0.028 (0.089)	-0.037 (0.095)
30-39	0.460 (0.255)	0.267 (0.208)	0.303 (0.163)	0.165 (0.131)	-0.171 (0.102)	-0.153 (0.094)	-0.223 (0.102)
40-49	0.445 (0.258)	0.070 (0.208)	0.181 (0.165)	0.133 (0.134)	-0.189 (0.105)	-0.121 (0.098)	-0.183 (0.107)
Education							
No education (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Primary	0.457 (0.038)	Same	Same	Same	Same	Same	Same
Secondary	1.371 (0.063)	1.310 (0.054)	1.277 (0.048)	1.218 (0.042)	1.164 (0.039)	1.198 (0.042)	1.208 (0.049)
Higher	2.945 (0.342)	3.085 (0.296)	2.885 (0.212)	2.536 (0.136)	2.403 (0.080)	2.137 (0.062)	1.948 (0.064)

Year of survey							
1997 (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2001	1.562	1.595	1.524	1.290	0.867	0.739	0.815
	(0.091)	(0.085)	(0.081)	(0.077)	(0.072)	(0.071)	(0.074)
2005	1.504	1.256	1.186	0.883	1.231	1.288	1.254
	(0.289)	(0.210)	(0.176)	(0.143)	(0.126)	(0.118)	(0.119)
Residence							
Urban (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rural	-0.283	-0.348	-0.369	-0.562	-0.735	-0.623	-0.793
		(0.169)	(0.130)	(0.105)	(0.084)	(0.077)	(0.089)
Marital status							
Widowed or divorced (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Married or remarried	0.102 (0.070)	Same	Same	Same	Same	Same	Same
Never married	0.077	0.057	-0.029	0.029	0.290	0.219	0.211
	(0.149)	(0.125)	(0.108)	(0.096)	(0.085)	(0.083)	(0.086)
Ethnicity							
Han (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minority	-0.170	-0.412	-0.381	-0.325	-0.377	-0.296	-0.248
	(0.121)	(0.098)	(0.089)	(0.083)	(0.084)	(0.091)	(0.105)
Age*year of survey							
20-29*2001	0.192 (0.072)	Same	Same	Same	Same	Same	Same
20-29*2005	0.290 (0.120)	Same	Same	Same	Same	Same	Same
30-39*2001	0.207 (0.072)	Same	Same	Same	Same	Same	Same
30-39*2005	0.458 (0.117)	Same	Same	Same	Same	Same	Same
40-49*2001	0.077 (0.078)	Same	Same	Same	Same	Same	Same
40-49*2005	0.225 (0.121)	Same	Same	Same	Same	Same	Same
Age*residence							
20-29*rural	-0.710	-0.431	-0.318	-0.070	0.218	0.136	0.215
	(0.267)	(0.210)	(0.160)	(0.124)	(0.091)	(0.080)	(0.093)
30-39*rural	-0.564	-0.552	-0.541	-0.268	0.031	0.061	0.190
	(0.238)	(0.195)	(0.150)	(0.118)	(0.087)	(0.078)	(0.091)
40-49*rural	-0.707	-0.419	-0.492	-0.265	0.006	0.019	0.201
	(0.242)	(0.195)	(0.152)	(0.121)	(0.091)	(0.083)	(0.099)
Year of survey*ethnicity							
2001*minority	-0.122	0.246	0.317	0.215	0.369	0.296	0.201
	(0.185)	(0.156)	(0.133)	(0.111)	(0.100)	(0.106)	(0.125)
2005*minority	0.078 (0.109)	Same	Same	Same	Same	Same	Same
Year of survey*residence							
2001*rural	-0.257 (0.048)	Same	Same	Same	Same	Same	Same
2005*rural	0.008	0.332	0.440	0.601	0.339	0.346	0.577
	(0.287)	(0.201)	(0.159)	(0.114)	(0.084)	(0.073)	(0.074)

Ref – reference category.

Note: Reference category of the response variable is perfect knowledge (seven routes).

Table G.4: Results of the partial proportional odds model for Score Four in the Chinese context.

Variable	Poorest knowledge β (SE)	Bad knowledge β (SE)	OK knowledge β (SE)
Intercept	1.931 (0.341)	0.742 (0.338)	-1.101 (0.339)
Age group			
<20 (ref)	0.000	0.000	0.000
20-29	-0.351 (0.375)	-0.381 (0.358)	-0.827 (0.355)
30-39	-0.566 (0.359)	-0.465 (0.347)	-0.927 (0.345)
40-49	-0.552 (0.356)	-0.286 (0.347)	-0.707 (0.345)
Education			
No education (ref)	0.000	0.000	0.000
Primary	0.053 (0.342)	Same	Same
Secondary	0.674 (0.332)	Same	Same
Higher	1.497 (0.378)	Same	Same
Year of survey			
1997 (ref)	0.000	0.000	0.000
2001	1.599 (0.088)	1.287 (0.080)	0.732 (0.074)
2005	1.220 (0.214)	0.856 (0.149)	1.263 (0.126)
Residence			
Urban (ref)	0.000	0.000	0.000
Rural	-0.627 (0.075)	Same	Same
Ethnicity			
Han (ref)	0.000	0.000	0.000
Minority	-0.155 (0.036)	Same	Same
Age*year of survey			
20-29*2001	0.221 (0.076)	Same	Same
20-29*2005	0.328 (0.133)	Same	Same
30-39*2001	0.262 (0.077)	Same	Same
30-39*2005	0.551 (0.128)	Same	Same
40-49*2001	0.100 (0.082)	Same	Same
40-49*2005	0.224 (0.132)	Same	Same
Age*education			
20-29*primary	0.466 (0.358)	Same	Same
20-29*secondary	0.688 (0.348)	Same	Same
20-29*higher	0.842 (0.399)	Same	Same
30-39*primary	0.458 (0.348)	Same	Same
30-39*secondary	0.591 (0.338)	Same	Same
30-39*higher	2.947 (1.073)	1.837 (0.500)	0.870 (0.391)
40-49*primary	0.373 (0.347)	Same	Same
40-49*secondary	0.467 (0.337)	Same	Same
40-49*higher	0.365 (0.397)	Same	Same
Age*residence			
20-29*rural	-0.206 (0.147)	0.003 (0.102)	0.174 (0.080)
30-39*rural	-0.258 (0.125)	-0.149 (0.093)	0.100 (0.078)
40-49*rural	-0.251 (0.125)	-0.254 (0.098)	0.007 (0.084)
Year of survey*residence			
2001*rural	-0.271 (0.052)	Same	Same
2005*rural	0.335 (0.202)	0.585	0.324 (0.075)

Ref – reference category.

Note: Reference category of the response variable is best knowledge (6-7 routes).

Table G.5: Results of the partial proportional odds model for Score One in the five country context.

Variable	No knowledge β (SE)	One routes β (SE)	Two routes β (SE)	Three routes β (SE)
Intercept	1.603 (0.185)	0.641 (0.145)	-0.170 (0.137)	-1.345 (0.136)
Age group				
<20 (ref)	0.000	0.000	0.000	0.000
20-29	1.667 (0.240)	1.242 (0.155)	1.094 (0.135)	0.885 (0.127)
30-39	1.055 (0.122)	Same	Same	Same
40-49	0.905 (0.124)	Same	Same	Same
Education				
No education (ref)	0.000	0.000	0.000	0.000
Primary	0.500 (0.108)	Same	Same	Same
Secondary	1.082 (0.100)	Same	Same	Same
Higher	2.842 (0.200)	1.995 (0.157)	1.791 (0.150)	1.551 (0.146)
Country				
China 2005 (ref)	0.000	0.000	0.000	0.000
Ukraine 2007	1.625 (0.227)	1.077 (0.147)	0.907 (0.131)	0.926 (0.127)
India 2006	-0.025 (0.161)	-0.120 (0.113)	-0.027 (0.103)	0.180 (0.100)
Kenya 2003	4.182 (0.359)	2.967 (0.173)	1.681 (0.128)	0.977 (0.121)
Malawi 2004	1.420 (0.161)	0.482 (0.127)	0.090 (0.119)	-0.066 (0.119)
Residence				
Urban (ref)	0.000	0.000	0.000	0.000
Rural	-0.352 (0.151)	-0.114 (0.098)	-0.011 (0.087)	-0.043 (0.083)
Marital status				
Widowed or divorced (ref)	0.000	0.000	0.000	0.000
Married or remarried	0.080 (0.065)	Same	Same	Same
Never married	-0.328 (0.106)	-0.192 (0.097)	-0.122 (0.096)	-0.081 (0.097)
Age*country				
20-29*Ukraine 2007	-0.776 (0.141)	Same	Same	Same
20-29*India 2006	-1.307 (0.231)	-0.890 (0.140)	-0.790 (0.117)	-0.627 (0.109)
20-29*Kenya 2003	-1.262 (0.645)	-1.144 (0.246)	-0.582 (0.155)	-0.488 (0.130)
20-29*Malawi 2004	-1.081 (0.275)	-0.885 (0.158)	-0.726 (0.132)	-0.411 (0.123)
30-39*Ukraine 2007	-0.862 (0.136)	Same	Same	Same
30-39*India 2006	-0.757 (0.102)	Same	Same	Same
30-39*Kenya 2003	-0.680 (0.126)	Same	Same	Same
30-39*Malawi 2004	-0.121 (0.218)	-0.522 (0.134)	-0.630 (0.122)	-0.603 (0.121)
40-49*Ukraine 2007	-0.808 (0.138)	Same	Same	Same
40-49*India 2006	-0.661 (0.104)	Same	Same	Same
40-49*Kenya 2003	-0.707 (0.132)	Same	Same	Same
40-49*Malawi 2004	-0.466 (0.122)	Same	Same	Same
Age*education				
20-29*primary	-0.066 (0.097)	-0.145 (0.085)	-0.162 (0.082)	-0.053 (0.082)
20-29*secondary	-0.091 (0.071)	Same	Same	Same
20-29*higher	0.137 (0.094)	Same	Same	Same
30-39*primary	-0.076 (0.083)	Same	Same	Same
30-39*secondary	0.193 (0.097)	0.073 (0.080)	-0.011 (0.075)	-0.098 (0.074)
30-39*higher	0.083 (0.102)	Same	Same	Same
40-49*primary	-0.128 (0.085)	Same	Same	Same
40-49*secondary	0.101 (0.104)	-0.030 (0.084)	-0.108 (0.079)	-0.139 (0.077)
40-49*higher	0.015 (0.108)	Same	Same	Same
Country*residence				
Ukraine 2007*rural	-0.247 (0.079)	Same	Same	Same
India 2006*rural	0.148 (0.139)	-0.166 (0.077)	-0.305 (0.062)	-0.333 (0.056)
Kenya 2003*rural	-0.225 (0.076)	Same	Same	Same
Malawi 2004*rural	0.134 (0.075)	Same	Same	Same

Education*marital status				
Primary*married	-0.001 (0.073)	Same	Same	Same
Primary*never married	-0.159 (0.107)	Same	Same	Same
Secondary*married	0.287 (0.090)	0.092 (0.077)	0.035 (0.074)	-0.022 (0.073)
Secondary*never married	0.444 (0.114)	0.099 (0.104)	-0.013 (0.103)	-0.046 (0.103)
Higher*married	0.321 (0.231)	0.383 (0.137)	0.148 (0.119)	0.082 (0.109)
Higher*never married	-0.096 (0.135)	Same	Same	Same
Education*residence				
Primary*rural	0.060 (0.039)	Same	Same	Same
Secondary*rural	-0.074 (0.061)	0.031 (0.042)	0.050 (0.037)	0.091 (0.037)
Higher*rural	0.069 (0.051)	Same	Same	Same
Marital status*residence				
Married*rural	-0.006 (0.058)	Same	Same	Same
Never married*rural	0.279 (0.085)	0.186 (0.069)	0.119 (0.064)	0.087 (0.063)

Ref – reference category.

Note: Reference category of the response variable is perfect knowledge (four routes).

Table G.6: Results of the partial proportional odds model for Score Two in the five country context.

Variable	No knowledge β (SE)	One route β (SE)	Two routes β (SE)
Intercept	1.965 (0.187)	-0.394 (0.178)	-2.074 (0.178)
Age group			
<20 (ref)	0.000	0.000	0.000
20-29	0.100 (0.020)	Same	Same
30-39	0.105 (0.022)	Same	Same
40-49	0.005 (0.024)	Same	Same
Education			
No education (ref)	0.000	0.000	0.000
Primary	0.359 (0.086)	0.275 (0.080)	0.215 (0.081)
Secondary	1.299 (0.082)	1.239 (0.077)	1.045 (0.077)
Higher	2.840 (0.144)	2.413 (0.124)	1.922 (0.121)
Country			
China 2005 (ref)	0.000	0.000	0.000
Ukraine 2007	-0.977 (0.211)	-0.009 (0.187)	0.567 (0.184)
India 2006	-0.697 (0.180)	0.438 (0.169)	0.926 (0.170)
Kenya 2003	2.047 (0.199)	Same	Same
Malawi 2004	1.605 (0.280)	2.207 (0.202)	2.182 (0.189)
Residence			
Urban (ref)	0.000	0.000	0.000
Rural	-0.007 (0.058)	Same	Same
Marital status			
Widowed or divorced (ref)	0.000	0.000	0.000
Married or remarried	0.140 (0.177)	Same	Same
Never married	-0.602 (0.265)	-0.181 (0.205)	0.090 (0.202)
Country*marital status			
Ukraine 2007*married	-0.199 (0.186)	Same	Same
Ukraine 2007*never married	0.617 (0.296)	0.093 (0.214)	-0.212 (0.207)
India 2006*married	-0.137 (0.170)	Same	Same
India 2006*never married	0.680 (0.256)	0.233 (0.191)	-0.073 (0.188)

Kenya 2003*married	-0.433 (0.230)	0.021 (0.203)	0.081 (0.199)
Kenya 2003*never married	0.421 (0.319)	0.474 (0.228)	0.184 (0.217)
Malawi 2004*married	-0.360 (0.183)	Same	Same
Malawi 2004*never married	0.460 (0.343)	0.018 (0.221)	-0.321 (0.206)
Country*residence			
Ukraine 2007*rural	-0.296 (0.133)	-0.360 (0.084)	-0.562 (0.076)
India 2006*rural	-0.300 (0.061)	-0.425 (0.053)	-0.608 (0.053)
Kenya 2003*rural	-0.754 (0.076)	Same	Same
Malawi 2004*rural	0.218 (0.226)	-0.052 (0.109)	-0.318 (0.079)
Education*marital status			
Primary*married	0.033 (0.076)	Same	Same
Primary*never married	0.018 (0.099)	Same	Same
Secondary*married	0.012 (0.075)	Same	Same
Secondary*never married	0.235 (0.094)	Same	Same
Higher*married	0.158 (0.121)	Same	Same
Higher*never married	0.194 (0.135)	Same	Same
Education*residence			
Primary*rural	0.008 (0.059)	0.108 (0.044)	0.192 (0.047)
Secondary*rural	-0.036 (0.051)	-0.015 (0.037)	0.198 (0.039)
Higher*rural	-0.231 (0.146)	0.093 (0.075)	0.333 (0.055)

Ref – reference category.
Note: Reference category of the response variable is perfect knowledge (three routes).

Table G.7: Results of the partial proportional odds model for Score Three in the five country context.

Variable	No knowledge β (SE)	One route β (SE)	Two routes β (SE)	Three routes β (SE)	Four routes β (SE)	Five routes β (SE)	Six routes β (SE)
Intercept	3.235 (0.155)	2.121 (0.106)	1.401 (0.091)	0.719 (0.083)	-0.024 (0.078)	-1.176 (0.076)	-2.766 (0.079)
Age group							
<20 (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20-29	0.269 (0.058)	Same	Same	Same	Same	Same	Same
30-39	0.278 (0.057)	Same	Same	Same	Same	Same	Same
40-49	0.205 (0.058)	Same	Same	Same	Same	Same	Same
Education							
No education (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Primary	0.352 (0.065)	Same	Same	Same	Same	Same	Same
Secondary	1.787 (0.090)	1.692 (0.071)	1.594 (0.064)	1.439 (0.060)	1.366 (0.059)	1.331 (0.060)	1.354 (0.062)
Higher	3.462 (0.226)	3.360 (0.150)	2.930 (0.107)	2.670 (0.091)	2.314 (0.083)	2.100 (0.081)	1.918 (0.082)
Country							
China 2005 (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ukraine 2007	0.843 (0.526)	0.295 (0.230)	0.061 (0.139)	-0.109 (0.093)	-0.198 (0.069)	0.098 (0.056)	0.583 (0.057)
India 2006	-1.063	-0.767	-0.710	-0.618	-0.484	0.005	0.612

	(0.140)	(0.085)	(0.068)	(0.056)	(0.048)	(0.043)	(0.047)
Kenya 2003	3.817	2.987	2.535	2.106	1.661	1.597	1.650
	(0.519)	(0.206)	(0.130)	(0.091)	(0.071)	(0.062)	(0.061)
Malawi 2004	2.379	1.589	1.007	0.808	0.358	0.490	0.776
	(0.721)	(0.281)	(0.163)	(0.111)	(0.079)	(0.067)	(0.074)
Residence							
Urban (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rural	-0.138	Same	Same	Same	Same	Same	Same
	(0.057)						
Marital status							
Widowed or divorced (ref)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Married or remarried	0.073	0.049	0.103	0.116	0.101	0.042	0.025
	(0.052)	(0.039)	(0.032)	(0.029)	(0.028)	(0.027)	(0.028)
Never married	-0.002	Same	Same	Same	Same	Same	Same
	(0.031)						
Age*education							
20-29*primary	0.186	0.131	0.056	0.028	0.010	0.062	0.165
	(0.102)	(0.082)	(0.074)	(0.070)	(0.068)	(0.068)	(0.070)
20-29*secondary	-0.164	-0.105	-0.164	-0.134	-0.121	-0.117	-0.180
	(0.095)	(0.075)	(0.067)	(0.063)	(0.062)	(0.061)	(0.062)
20-29*higher	0.150	Same	Same	Same	Same	Same	Same
	(0.082)						
30-39*primary	0.082	Same	Same	Same	Same	Same	Same
	(0.066)						
30-39*secondary	-0.181	-0.104	-0.073)	-0.037	-0.091	-0.105	0.182
	(0.106)	(0.080)		(0.065)	(0.062)	(0.061)	(0.062)
30-39*higher	0.178	Same	Same	Same	Same	Same	Same
	(0.084)						
40-49*primary	0.015	Same	Same	Same	Same	Same	Same
	(0.069)						
40-49*secondary	-0.200	Same	Same	Same	Same	Same	Same
	(0.063)						
40-49*higher	0.121	Same	Same	Same	Same	Same	Same
	(0.087)						
Country*residence							
Ukraine 2007*rural	-0.721	-0.035	-0.084	-0.157	-0.359	-0.502	-0.520
	(0.626)	(0.296)	(0.180)	(0.120)	(0.089)	(0.076)	(0.077)
India 2006*rural	-0.112	-0.054	-0.201	-0.254	-0.326	-0.468	-0.542
	(0.072)	(0.062)	(0.058)	(0.055)	(0.052)	(0.050)	(0.053)
Kenya 2003*rural	-0.556	Same	Same	Same	Same	Same	Same
	(0.069)						
Malawi 2004*rural	0.544	0.912	0.484	0.111	0.128	0.026	-0.087
	(0.772)	(0.306)	(0.170)	(0.115)	(0.084)	(0.075)	(0.0815)
Education*residence							
Primary*rural	0.139	0.052	0.085	0.093	0.136	0.180	0.211
	(0.071)	(0.054)	(0.046)	(0.042)	(0.041)	(0.042)	(0.049)
Secondary*rural	0.037	-0.113	-0.010	0.064	0.137	0.227	0.277
	(0.084)	(0.057)	(0.045)	(0.038)	(0.036)	(0.037)	(0.044)
Higher*rural	0.446	-0.118	0.069	0.035	0.169	0.343	0.399
	(0.501)	(0.255)	(0.151)	(0.099)	(0.069)	(0.056)	(0.057)

Ref – reference category.

Note: Reference category of the response variable is perfect knowledge (seven routes).

Table G.8: Results of the partial proportional odds model for Score Four in the five country context.

Variable	Poorest knowledge β (SE)	Bad knowledge β (SE)	OK knowledge β (SE)
Intercept	2.344 (0.351)	0.168 (0.293)	-1.354 (0.274)
Age group			
<20 (ref)	0.000	0.000	0.000
20-29	0.632 (0.409)	1.038 (0.304)	0.502 (0.276)
30-39	0.298 (0.371)	1.002 (0.294)	0.505 (0.271)
40-49	-0.055 (0.352)	0.681 (0.290)	0.396 (0.271)
Education			
No education (ref)	0.000	0.000	0.000
Primary	0.417 (0.113)	Same	Same
Secondary	1.642 (0.116)	1.348 (0.109)	1.303 (0.108)
Higher	3.174 (0.226)	2.595 (0.170)	2.144 (0.159)
Country			
China 2005 (ref)	0.000	0.000	0.000
Ukraine 2007	0.713 (0.128)	Same	Same
India 2006	-0.193 (0.171)	0.051 (0.112)	0.471 (0.104)
Kenya 2003	3.802 (0.289)	2.635 (0.148)	2.059 (0.129)
Malawi 2004	1.807 (0.302)	1.266 (0.150)	0.909 (0.123)
Residence			
Urban (ref)	0.000	0.000	0.000
Rural	-0.106 (0.085)	Same	Same
Marital status			
Widowed or divorced (ref)	0.000	0.000	0.000
Married or remarried	-0.670 (0.314)	0.040 (0.268)	-0.296 (0.248)
Never married	-0.926 (0.311)	-0.126 (0.273)	-0.350 (0.256)
Age*country			
20-29*Ukraine 2007	-0.666 (0.142)	Same	Same
20-29*India 2006	-1.024 (0.210)	-0.712 (0.122)	-0.498 (0.111)
20-29*Kenya 2003	-1.509 (0.419)	-0.485 (0.189)	-0.389 (0.136)
20-29*Malawi 2004	-0.424 (0.124)	Same	Same
30-39*Ukraine 2007	-0.320 (0.385)	-0.978 (0.173)	-0.781 (0.138)
30-39*India 2006	-0.834 (0.191)	-0.768 (0.117)	-0.532 (0.106)
30-39*Kenya 2003	-0.584 (0.135)	Same	Same
30-39*Malawi 2004	-0.425 (0.122)	Same	Same
40-49*Ukraine 2007	-0.739 (0.307)	-0.932 (0.169)	-0.570 (0.141)
40-49*India 2006	-0.562 (0.190)	-0.658 (0.119)	-0.445 (0.108)
40-49*Kenya 2003	-0.664 (0.140)	Same	Same
40-49*Malawi 2004	-0.364 (0.127)	Same	Same
Age*education			
20-29*primary	-0.045 (0.084)	Same	Same
20-29*secondary	-0.070 (0.076)	Same	Same
20-29*higher	0.160 (0.106)	Same	Same
30-39*primary	-0.041 (0.087)	Same	Same
30-39*secondary	-0.024 (0.098)	0.037 (0.083)	-0.085 (0.081)
30-39*higher	0.088 (0.118)	Same	Same
40-49*primary	-0.096 (0.088)	Same	Same
40-49*secondary	0.001 (0.106)	-0.112 (0.086)	-0.202 (0.084)
40-49*higher	-0.051 (0.123)	Same	Same
Age*marital status			
20-29*married	0.552 (0.358)	-0.112 (0.278)	0.283 (0.248)
20-29*never married	0.783 (0.356)	0.057 (0.277)	0.228 (0.248)
30-39*married	0.718 (0.323)	0.005 (0.267)	0.355 (0.243)

30-39*never married	0.680 (0.350)	0.072 (0.282)	0.442 (0.252)
40-49*married	0.739 (0.301)	0.172 (0.261)	0.337 (0.242)
40-49*never married	0.269 (0.262)	Same	Same
Country*residence			
Ukraine 2007*rural	0.101 (0.254)	-0.178 (0.115)	-0.513 (0.079)
India 2006*rural	0.085 (0.081)	-0.272 (0.063)	-0.494 (0.055)
Kenya 2003*rural	-0.526 (0.084)	Same	Same
Malawi 2004*rural	1.067 (0.311)	0.067 (0.120)	-0.003 (0.079)
Education*marital status			
Primary*married	0.156 (0.090)	0.029 (0.080)	0.041 (0.078)
Primary*never married	0.002 (0.132)	-0.130 (0.116)	-0.007 (0.115)
Secondary*married	0.010 (0.076)	Same	Same
Secondary*never married	0.086 (0.106)	Same	Same
Higher*married	0.594 (0.249)	0.255 (0.138)	0.095 (0.115)
Higher*never married	-0.037 (0.143)	Same	Same
Education*residence			
Primary*rural	-0.036 (0.066)	0.097 (0.047)	0.167 (0.046)
Secondary*rural	-0.186 (0.060)	0.053 (0.040)	0.210 (0.040)
Higher*rural	-0.244 (0.255)	0.018 (0.099)	0.326 (0.058)
Marital status*residence			
Married*rural	-0.180 (0.078)	-0.031 (0.064)	-0.013 (0.061)
Never married*rural	0.053 (0.064)	Same	Same

Ref – reference category.

Note: Reference category of the response variable is best knowledge (6-7 routes).

Appendix H

Table H.1 presents the summary of advantages and limitations of four simple score measures used for the analysis in Paper Two.

Table H.1: Summary of simple score measures used for the analysis in Paper Two.

	Score One: Correct routes or prevention knowledge	Score Two: Incorrect routes or misconceptions	Score Three: Combined knowledge	Score Four: Combined knowledge grouped
Number of categories	5	4	8	4 (depends on the decision)
Meaning of categories	0 – no knowledge, 4 – perfect knowledge, other categories represent number of routes correctly answered without specification of which routes are known, just provides numerical value of correctly answered routes.	0 – no knowledge, 3 – perfect knowledge, other categories represent number of routes correctly answered without specification of which routes are known, just provides numerical value of correctly answered routes.	0 – no knowledge, 7 – perfect knowledge, other categories represent number of routes correctly answered without specification of which routes are known, just provides numerical value of correctly answered routes.	Arbitrary split of the answers into 4 groups, every next group represents better HIV knowledge level than the previous one without providing any details about which routes are known and which are not known.
Quantitative meaning	For every group.	For every group.	For every group.	Roughly for every group.
Qualitative meaning	Only for 2 extreme groups.	Only for 2 extreme groups.	Only for 2 extreme groups.	No groups.
Ways of modelling (as a response variable)	Ordinal, partial proportional odds, multinomial.	Ordinal, partial proportional odds, multinomial.	Ordinal, partial proportional odds, multinomial or linear regression if scores are normally distributed.	Ordinal, partial proportional odds, multinomial.
Difficulty of deriving	Easy	Easy	Easy	Easy
Limitations	No qualitative details about categories which are not extreme. Does not take into account weights for each contributing route.	No qualitative details about categories which are not extreme. Does not take into account weights for each contributing route.	No qualitative details about categories which are not extreme. Does not take into account weights for each contributing route Mask those who answered "yes" to all HIV knowledge questions.	No qualitative details about any of the categories apart from continuum from the poorest to the best knowledge.
Summary of measure	Good quantitative details for each category and qualitative details for 2 extreme categories.	Good quantitative details for each category and qualitative details for 2 extreme categories.	Good quantitative details for each category and qualitative details for 2 extreme categories.	Rough quantitative detail and continuum from the poorest to the best knowledge.

Appendix I

Tables I.1-I.16 present results of latent trait analysis for the separate datasets used for the analysis in both study contexts in Paper Three. Orthogonal rotation of factors and Varimax procedure are used.

Table I.1: Model fit statistics in China 1997 (LTA).

Model	L^2 value	Degrees of freedom	p-value	Change in L^2	Change in df	p-value
One-factor solution	7181.749	14	0.000			
Two-factor solution	487.498	8	0.000	6694.251	6	0.000
Three-factor solution	9.920	3	0.019	477.578	5	0.000

Table I.2: Factor loadings for different solutions in China 1997 (LTA).

	One-factor solution	Two-factor solution		Three-factor solution		
Blood transfusion	0.941	0.921	0.245	0.135	0.934	0.260
MTCT	0.751	0.817	0.103	0.140	0.755	0.097
Needle sharing	0.845	0.905	0.093	0.126	0.897	0.098
Sexual intercourse	0.840	0.802	0.325	5.518	0.013	0.008
Handshaking	0.809	0.388	0.854	0.104	0.373	0.856
Sharing utensils	0.639	0.226	0.716	0.082	0.202	0.715
Kissing	0.686	0.032	0.931	0.055	0.023	0.928

Table I.3: Model fit statistics in China 2001 (LTA).

Model	L^2 value	Degrees of freedom	p-value	Change in L^2	Change in df	p-value
One-factor solution	13051.888	14	0.000			
Two-factor solution	262.221	8	0.000	12789.667	6	0.000
Three-factor solution	14.004	3	0.003	248.217	5	0.000

Table I.4: Factor loadings for different solutions in China 2001 (LTA).

	One-factor solution	Two-factor solution		Three-factor solution		
Blood transfusion	0.818	0.826	0.300	0.765	0.298	0.331
MTCT	0.603	0.762	0.090	0.571	0.080	0.485
Needle sharing	0.735	0.873	0.105	0.854	0.090	0.320
Sexual intercourse	0.729	0.787	0.214	0.455	0.179	0.843
Handshaking	0.855	0.288	0.878	0.232	0.871	0.193
Sharing utensils	0.719	0.154	0.761	0.097	0.758	0.150
Kissing	0.784	0.067	0.888	0.074	0.889	0.037

Table I.5: Model fit statistics in China 2005 – the Chinese context (LTA).

Model	L^2 value	Degrees of freedom	p-value	Change in L^2	Change in df	p-value
One-factor solution	1941.978	14	0.000			
Two-factor solution	12.867	8	0.1165	1929.111	6	0.000
Three-factor solution	2.289	3	0.515	10.578	5	0.060

Table I.6: Factor loadings for different solutions in China 2005 – the Chinese context (LTA).

	One-factor solution	Two-factor solution		Three-factor solution		
Blood transfusion	0.802	0.817	0.295	0.811	0.299	0.097
MTCT	0.746	0.804	0.259	0.827	0.253	-0.140
Needle sharing	0.665	0.791	0.160	0.785	0.165	0.099
Sexual intercourse	0.792	0.777	0.334	0.771	0.337	0.017
Handshaking	0.861	0.400	0.793	0.387	0.820	0.184
Sharing utensils	0.849	0.232	0.886	0.238	0.863	-0.021
Kissing	0.716	0.146	0.777	0.142	0.793	-0.165

Table I.7: Model fit statistics in China 2005 – the five country context (LTA).

Model	L^2 value	Degrees of freedom	p-value	Deviance of L^2	Difference in degrees of freedom	p-value
One factor solution	748.743	14	0.000			
Two factor solution	37.470	8	0.000	711.273	6	0.000
Three factor solution	13.263	3	0.004	24.207	5	0.000

Table I.8: Factor loadings for different solutions in China 2005 – the five country context (LTA).

	One-factor solution	Two-factor solution		Three-factor solution		
MTCT	0.664	0.574	0.331	0.491	0.577	0.161
One partner	0.888	0.880	0.204	0.888	0.158	0.170
Abstinence	0.575	0.707	-0.088	0.707	0.038	-0.098
Condom	0.838	0.786	0.275	0.754	0.281	0.196
Healthy-looking person	0.369	0.300	0.217	0.264	0.245	0.138
Sharing utensils	0.549	0.304	0.720	0.294	0.340	0.544
Mosquito bites	0.344	0.078	0.600	0.093	0.050	0.764

Table I.9: Model fit statistics in Ukraine 2007 (LTA).

Model	L^2 value	Degrees of freedom	p-value	Deviance of L^2	Difference in degrees of freedom	p-value
One factor solution	973.648	14	0.000			
Two factor solution	158.634	8	0.000	815.014	6	0.000
Three factor solution	5.842	3	0.120	152.792	5	0.000

Table I.10: Factor loadings for different solutions in Ukraine 2007(LTA).

	One-factor solution	Two-factor solution		Three-factor solution		
MTCT	0.410	0.194	0.382	0.193	0.332	0.265
One partner	0.685	0.813	0.135	0.822	0.107	0.113
Abstinence	0.561	0.714	0.063	0.709	-0.037	0.105
Condom	0.744	0.753	0.262	0.743	0.065	0.274
Healthy-looking person	0.300	0.071	0.344	0.046	0.931	0.124
Sharing utensils	0.671	0.139	0.873	0.142	0.213	0.754
Mosquito bites	0.665	0.238	0.657	0.206	0.079	0.765

Table I.11: Model fit statistics in India 2006 (LTA).

Model	L^2 value	Degrees of freedom	p-value	Deviance of L^2	Difference in degrees of freedom	p-value
One factor solution	15792.29	14	0.000			
Two factor solution	3618.927	8	0.000	12173.363	6	0.000
Three factor solution	49.006	3	0.000	3569.921	5	0.000

Table I.12: Factor loadings for different solutions in India 2006(LTA).

	One-factor solution	Two-factor solution		Three-factor solution		
MTCT	0.544	0.472	0.277	0.137	0.422	0.234
One partner	0.792	0.798	0.241	0.114	0.794	0.230
Abstinence	0.767	0.794	0.211	0.100	0.805	0.198
Condom	0.803	0.734	0.353	0.116	0.734	0.341
Healthy-looking person	0.562	0.454	0.333	3.535	0.018	0.014
Sharing utensils	0.713	0.298	0.824	0.107	0.306	0.816
Mosquito bites	0.608	0.193	0.744	0.085	0.202	0.740

Table I.13: Model fit statistics in Kenya 2003 (LTA).

Model	L^2 value	Degrees of freedom	p-value	Deviance of L^2	Difference in degrees of freedom	p-value
One factor solution	264.471	14	0.000			
Two factor solution	33.785	8	0.000	230.686	6	0.000
Three factor solution	0.466	3	0.926	33.319	5	0.000

Table I.14: Factor loadings for different solutions in Kenya 2003(LTA).

	One-factor solution	Two-factor solution		Three-factor solution		
MTCT	0.519	0.450	0.303	0.830	0.087	0.164
One partner	0.344	0.540	0.026	0.160	0.570	0.030
Abstinence	0.359	0.477	0.093	0.118	0.533	0.106
Condom	0.358	0.391	0.164	0.221	0.323	0.144
Healthy-looking person	0.574	0.520	0.328	0.413	0.323	0.269
Sharing utensils	0.636	0.083	0.824	0.143	0.040	0.878
Mosquito bites	0.566	0.138	0.573	0.158	0.112	0.525

Table I.15: Model fit statistics in Malawi 2004 (LTA).

Model	L^2 value	Degrees of freedom	p-value	Deviance of L^2	Difference in degrees of freedom	p-value
One factor solution	1070.392	14	0.000			
Two factor solution	352.672	8	0.000	717.72	6	0.000
Three factor solution	3.225	3	0.358	349.447	5	0.000

Table I.16: Factor loadings for different solutions in Malawi 2004(LTA).

	One-factor solution	Two-factor solution		Three-factor solution		
MTCT	0.252	0.281	0.239	1.064	0.056	0.050
One partner	0.658	0.656	-0.030	0.105	0.647	-0.019
Abstinence	0.752	0.744	-0.071	0.063	0.780	-0.034
Condom	0.601	0.598	-0.041	0.114	0.582	-0.035
Healthy-looking person	0.298	0.332	0.255	0.327	0.214	0.080
Sharing utensils	-0.097	-0.051	0.750	0.021	-0.010	1.588
Mosquito bites	-0.183	-0.165	0.460	0.002	-0.161	0.247

Appendix J

This appendix presents the calculation of the maximum number of factors that can be extracted in latent trait analysis for both study contexts in Paper Three.

Calculation of Maximum Number of Factors

The maximum number of factors can be estimated as follows:
The number of parameters to be estimated *a* must be less or equal to the number of sample variances/ covariances *b,* where

$$a = pm + m(m+1)/2 + p - m^2$$
$$b = p(p+1)/2,$$

p is number of observed variables, and *m* is number of factors [55].

First we can obtain b as we know number of observed variables (it is seven):
b=7(7+1)/2=28.

As $a \le b$, then number of parameters to be estimated cannot be larger than 28.

$$7m + m(m+1)/2 + 7 - m^2 \le 28$$
$$7m + m^2/2 + m/2 + 7 - m^2 \le 28$$
$$14m/2 + m^2/2 + m/2 - 2m^2/2 \le 21$$
$$15m/2 - m^2/2 \le 21$$
$$15m - m^2 \le 42$$

This inequality is true in the regions $m \le 3.72$ and $m \ge 11.27$. The second region for *m* does not make sense in the context of number of factors which can be extracted (as we have only 7 manifest variables), and therefore, we conclude that maximum number of factors which can be extracted is 3.

---

[55] http://www.statmodel.com/download/Topic%201.pdf [Accessed 25 August 2010] – see slide 104.

Appendix K

Figures K.1- K.9 present all distributions of factor scores which were obtained for both study contexts in Paper Three.

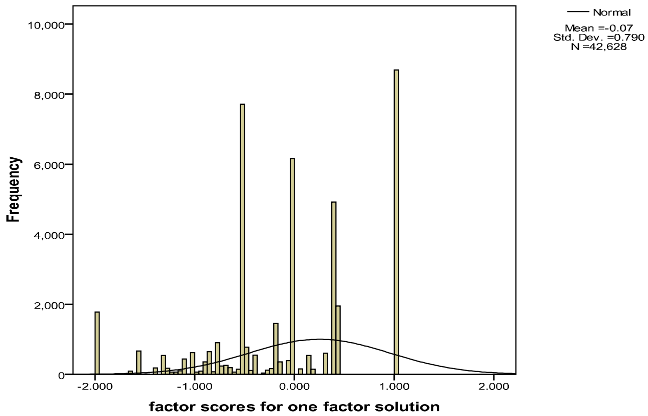


Figure K.1: Histogram of the factor scores for one-factor solution in China.

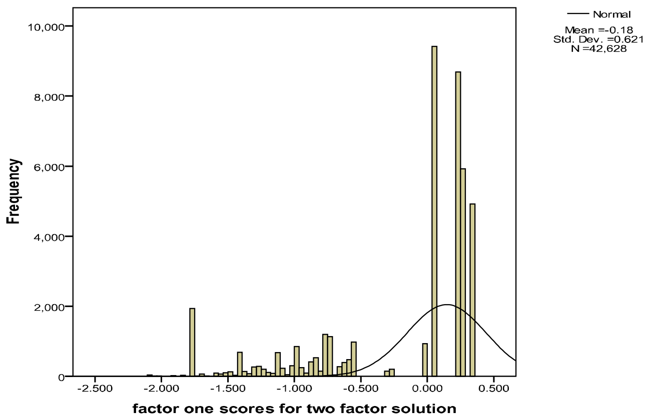


Figure K.2: Histogram of the factor scores for factor one in the two-factor solution in China.

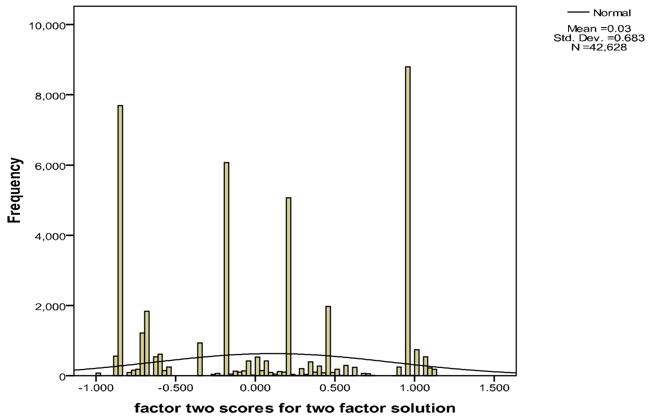


Figure K.3: Histogram of the factor scores for factor two in the two-factor solution in China.

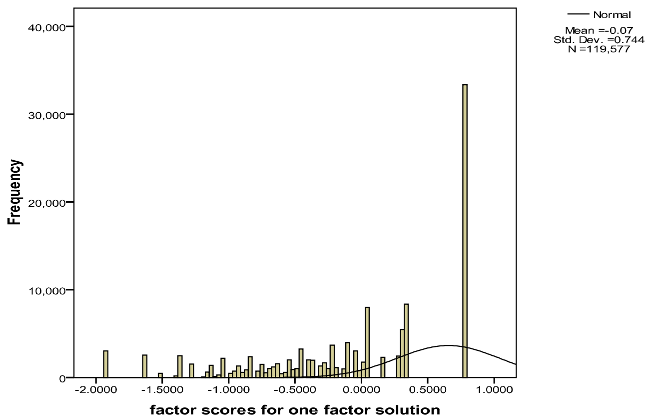


Figure K.4: Histogram of the factor scores for one-factor solution in the five countries.

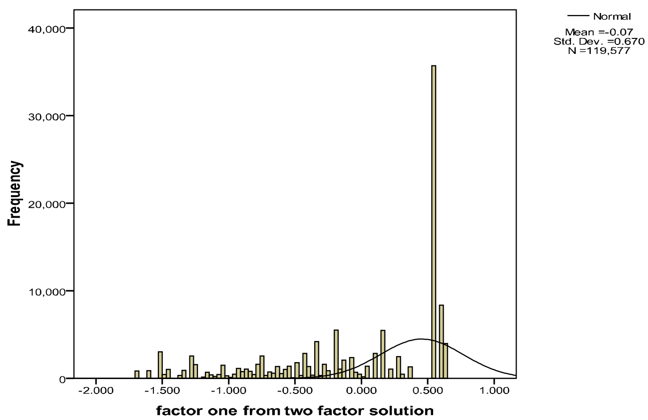


Figure K.5: Histogram of the factor scores for factor one in the two-factor solution in the five countries.

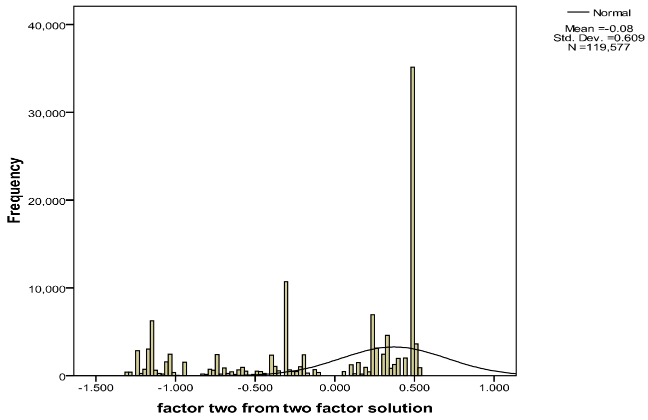


Figure K.6: Histogram of the factor scores for factor two in the two-factor solution in the five countries.

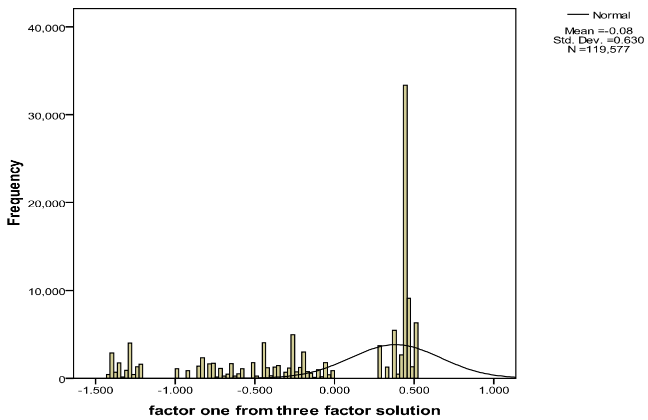


Figure K.7: Histogram of the factor scores for factor one in the three-factor solution in the five countries.

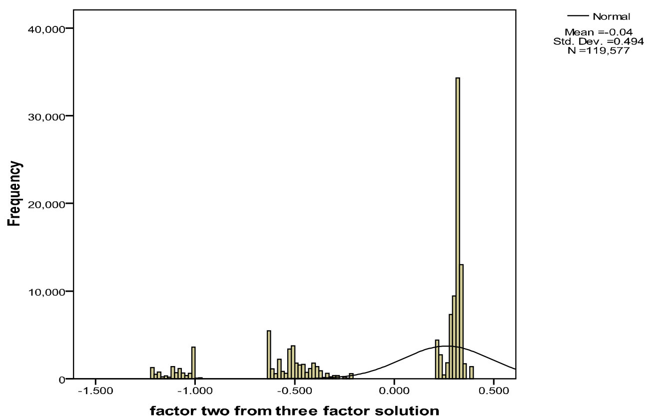


Figure K.8: Histogram of the factor scores for factor two in the three-factor solution in the five countries.

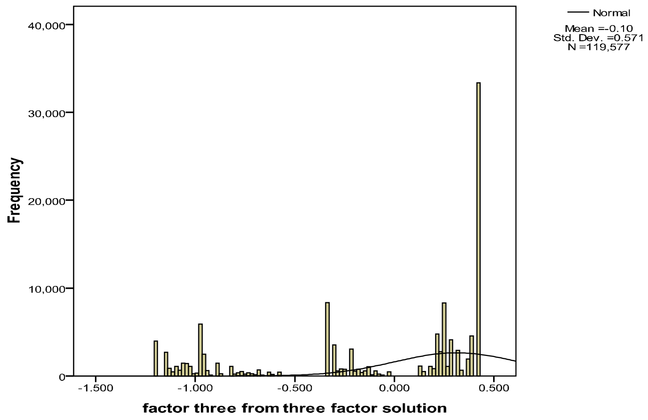


Figure K.9: Histogram of the factor scores for factor three in the three-factor solution in the five countries.

Appendix L

Tables L.1 and L.2 present cross-tabulations of factor scores for one factor solutions from Paper Three and categories of Score Three from Paper Two for both study contexts. These tables demonstrate similarity of factor score distributions to those of the simple scores.

Table L.1: Cross-tabulation of factor scores for one factor solution and categories of Score Three in the Chinese context.

	No knowledge	One route	Two route	Three route	Four route	Five route	Six route	Seven route
-1.979	1778							
-1.650 to -1.377		954						
-1.327 to -1.064			1399					
-1.039				7				
-1.031			7	503				
-1.027				87				
-1.017			19					
-0.974 to -0.775				2013				
-0.768 to -0.759					66			
-0.756 to -0.749				66				
-0.744 to -0.738					22			
-0.728 to -0.725				126				
-0.721 to -0.514					8402			
-0.508						5		
-0.507 to -0.503					49			
-0.497						7		
-0.493 to -0.483					501			
-0.479						110		
-0.467					71	90		
-0.456 to -0.433					63			
-0.422						48		
-0.415					49	31		
-0.382					472			
-0.314 to -0.219						227		
-0.217							93	
-0.207 to -0.015						8354		
0.069 to 0.444							8316	
1.025								8693
Total	1778	954	1425	2802	9695	8872	8409	8693

Table L.2: Cross-tabulation of factor scores for one factor solution and categories of Score Three in the five country context.

	No knowledge	One route	Two route	Three route	Four route	Five route	Six route	Seven route
-1.941	3034							
-1.635 to -1.377		3731						
-1.367 to -1.363			1865					
-1.356		97						
-1.282 to -1.139			3226					
-1.126				402				
-1.104 to -1.046			414					
-1.053 to -1.046				2176				
-0.987 to -0.985			494					
-0.966 to -0.963				174				
-0.956			163	397				
-0.938				3168				
-0.831					2006			
-0.793 to -0.757				1657				
-0.749					183			
-0.748 to -0.742				408				
-0.729					122			
-0.718				85				
-0.709					359			
-0.698				270				
-0.686 to -0.679					616			
-0.678				132				
-0.669 to -0.625					2798			
-0.597				405				
-0.591 to -0.469					4878			
-0.462						982		
-0.461 to -0.458					727			
-0.449						933		
-0.446					300			

-0.407						1807		
-0.385 to -0.372					669			
-0.368						193		
-0.358					1306			
-0.318 to -0.288						1915		
-0.286					468			
-0.274 to -0.102						11460		
-0.047							1784	
-0.036 to 0.007						3800		
0.041 to 0.352							26601	
0.788								33372
Total	3034	3828	6162	9274	14432	21090	28385	33372

Appendix M

Tables M.1-M.16 present results of latent class analysis for the separate datasets used for the analysis in both study contexts in Paper Three.

Table M.1: Model fit statistics in China 1997 (LCA).

Number of classes	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	Classification quality (%)
1			
2	68213.229	68320.854	91.1
3	64422.139	64587.165	80.3
4	63223.632	63446.058	80.1
5	62695.035	62974.861	77.6

Table M.2: Conditional response probabilities in China 1997(LCA).

Manifest variable	Latent Class One	Latent Class Two	Latent Class Three
BLOOD			
Other	0.122	0.926	0.054
Correct	0.878	0.074	0.946
MTCT			
Other	0.215	0.787	0.111
Correct	0.785	0.213	0.889
NEEDLES			
Other	0.272	0.980	0.152
Correct	0.728	0.020	0.848
SEX			
Other	0.045	0.543	0.030
Correct	0.955	0.457	0.970
HANDSHAKE			
Other	0.034	0.955	0.648
Correct	0.966	0.045	0.352
UTENSILS			
Other	0.247	0.931	0.844
Correct	0.753	0.069	0.156
KISS			
Other	0.360	0.973	0.991
Correct	0.640	0.027	0.009

Table M.3: Model fit statistics in China 2001 (LCA).

Number of classes	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	Classification quality (%)
1			
2	153490.486	153613.053	71.9
3	144746.136	144934.073	80.2
4	143520.295	143773.600	78.7
5	143056.158	143374.833	74.5

Table M.4: Conditional response probabilities in China 2001(LCA).

Manifest variable	Latent Class One	Latent Class Two	Latent Class Three
BLOOD			
Other	0.028	0.763	0.054
Correct	0.972	0.237	0.946
MTCT			
Other	0.100	0.773	0.088
Correct	0.900	0.227	0.912
NEEDLES			
Other	0.062	0.775	0.034
Correct	0.938	0.225	0.966
SEX			
Other	0.034	0.603	0.032
Correct	0.966	0.397	0.968
HANDSHAKE			
Other	0.013	0.867	0.614
Correct	0.987	0.133	0.386
UTENSILS			
Other	0.225	0.910	0.894
Correct	0.775	0.090	0.106
KISS			
Other	0.348	0.946	0.982
Correct	0.652	0.054	0.018

Table M.5: Model fit statistics in China 2005 – the Chinese context (LCA).

Number of classes	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	Classification quality (%)
1			
2	34931.104	35033.555	79.3
3	33461.669	33618.762	84.2
4	33183.468	33395.202	83.6
5	33090.874	33357.248	80.6

Table M.6: Conditional response probabilities in China 2005 – the Chinese context (LCA).

Manifest variable	Latent Class One	Latent Class Two	Latent Class Three
BLOOD			
Other	0.603	0.027	0.017
Correct	0.397	0.973	0.983
MTCT			
Other	0.850	0.079	0.046
Correct	0.150	0.921	0.954
NEEDLES			
Other	0.772	0.083	0.063
Correct	0.228	0.917	0.937
SEX			
Other	0.583	0.034	0.015
Correct	0.417	0.966	0.985
HANDSHAKE			
Other	0.811	0.527	0.022
Correct	0.189	0.473	0.978
UTENSILS			

	Other	0.847	0.888	0.073
	Correct	0.153	0.112	0.927
KISS				
	Other	0.868	0.893	0.252
	Correct	0.132	0.107	0.748

Table M.7: Model fit statistics in China 2005 – the five country context (LCA).

Number of classes	Akaike Information Criterion	Bayesian Information Criterion	Percentage of sample correctly classified
1			
2	42477.359	42579.811	86.2
3	42033.683	42190.775	69.6
4	41790.054	42001.787	67.5
5	41704.720	41971.094	69.2

Table M.8: Conditional response probabilities in China 2005 – the five country context (LCA).

Manifest variable	Latent Class One	Latent Class Two	Latent Class Three	Latent Class Four
MTCT				
other	0.133	0.168	0.029	0.575
correct	0.867	0.832	0.971	0.425
ONEPARTNER				
other	0.088	0.485	0.012	0.915
correct	0.912	0.515	0.988	0.085
ABSTINENCE				
other	0.178	0.687	0.155	0.858
correct	0.822	0.313	0.845	0.142
CONDOM				
other	0.175	0.528	0.023	0.939
correct	0.825	0.472	0.977	0.061
HEALTHY				
other	0.142	0.096	0.059	0.303
correct	0.858	0.904	0.941	0.697
UTENSILS				
other	0.747	0.107	0.099	0.920
correct	0.253	0.893	0.901	0.080
MOSQUITO				
other	0.842	0.346	0.468	0.942
correct	0.158	0.654	0.532	0.058

Table M.9: Model fit statistics in Ukraine 2007 (LCA).

Number of classes	Akaike Information Criterion	Bayesian Information Criterion	Percentage of sample correctly classified
1			
2	37302.045	37404.054	60.0
3	36663.721	36820.134	71.6
4	36523.775	36734.592	71.4
5	36442.357	36707.579	74.3

Table M.10: Conditional response probabilities in Ukraine 2007 (LCA).

Manifest variable	Latent Class One	Latent Class Two	Latent Class Three	Latent Class Four
MTCT				
other	0.042	0.325	0.094	0.171
correct	0.958	0.675	0.906	0.829
ONEPARTNER				
other	0.016	0.768	0.549	0.079
correct	0.984	0.232	0.451	0.921
ABSTINENCE				
other	0.053	0.706	0.555	0.106
correct	0.947	0.294	0.445	0.894
CONDOM				
other	0.006	0.755	0.308	0.068
correct	0.994	0.245	0.692	0.932
HEALTHY				
other	0.166	0.468	0.190	0.382
correct	0.834	0.532	0.810	0.618
UTENSILS				
other	0.045	0.835	0.087	0.712
correct	0.955	0.165	0.913	0.288
MOSQUITO				
other	0.124	0.882	0.291	0.694
correct	0.876	0.118	0.709	0.306

Table M.11: Model fit statistics in India 2006 (LCA).

Number of classes	Akaike Information Criterion	Bayesian Information Criterion	Percentage of sample correctly classified
1			
2	660165.563	660306.378	77.5
3	651684.834	651900.749	69.8
4	644739.795	645030.811	69.3
5	643257.936	643624.054	70.3

Table M.12: Conditional response probabilities in India 2006 (LCA).

Manifest variable	Latent Class One	Latent Class Two	Latent Class Three	Latent Class Four
MTCT				
other	0.086	0.565	0.297	0.196
correct	0.914	0.435	0.703	0.804
ONEPARTNER				
other	0.042	0.843	0.581	0.135
correct	0.958	0.157	0.419	0.865
ABSTINENCE				
other	0.090	0.939	0.716	0.193
correct	0.910	0.061	0.284	0.807
CONDOM				
other	0.087	0.967	0.750	0.390
correct	0.913	0.033	0.250	0.610
HEALTHY				
other	0.167	0.765	0.462	0.406
correct	0.833	0.235	0.538	0.594
UTENSILS				
other	0.029	0.899	0.153	0.621
correct	0.971	0.101	0.847	0.379

MOSQUITO				
other	0.132	0.887	0.241	0.740
correct	0.868	0.113	0.759	0.260

Table M.13: Model fit statistics in Kenya 2003 (LCA).

Number of classes	Akaike Information Criterion	Bayesian Information Criterion	Percentage of sample correctly classified
1			
2	36590.621	36693.160	59.2
3	36404.141	36561.367	75.6
4	36350.540	36562.454	57.3
5	36336.435	36603.036	59.6

Table M.14: Conditional response probabilities in Kenya 2003 (LCA).

Manifest variable	Latent Class One	Latent Class Two	Latent Class Three	Latent Class Four
MTCT				
other	0.082	0.065	0.047	0.570
correct	0.918	0.935	0.953	0.430
ONEPARTNER				
other	0.022	0.232	0.021	0.242
correct	0.978	0.768	0.979	0.758
ABSTINENCE				
other	0.057	0.248	0.026	0.257
correct	0.943	0.752	0.974	0.743
CONDOM				
other	0.334	0.460	0.209	0.738
correct	0.666	0.540	0.791	0.262
HEALTHY				
other	0.123	0.134	0.042	0.629
correct	0.877	0.866	0.958	0.371
UTENSILS				
other	0.625	0.113	0.062	0.573
correct	0.375	0.887	0.938	0.427
MOSQUITO				
other	0.794	0.284	0.150	0.631
correct	0.206	0.716	0.850	0.369

Table M.15: Model fit statistics in Malawi 2004 (LCA).

Number of classes	Akaike Information Criterion	Bayesian Information Criterion	Percentage of sample correctly classified
1			
2	79154.185	79263.772	57.9
3	78716.472	78884.506	63.3
4	78478.614	78705.095	69.4
5	78405.607	78690.534	48.9

Table M.16: Conditional response probabilities in Malawi 2004 (LCA).

Manifest variable	Latent Class One	Latent Class Two	Latent Class Three	Latent Class Four
MTCT				
other	0.192	0.097	0.119	0.767

correct	0.808	0.903	0.881	0.233
ONEPARTNER				
other	0.197	0.640	0.118	0.535
correct	0.803	0.360	0.882	0.465
ABSTINENCE				
other	0.145	0.685	0.069	0.465
correct	0.855	0.315	0.931	0.535
CONDOM				
other	0.266	0.710	0.220	0.677
correct	0.734	0.290	0.780	0.323
HEALTHY				
other	0.234	0.157	0.098	0.550
correct	0.766	0.843	0.902	0.450
UTENSILS				
other	1.000	0.060	0.087	0.193
correct	0.000	0.940	0.913	0.807
MOSQUITO				
other	0.940	0.182	0.300	0.281
correct	0.060	0.818	0.700	0.719

Appendix N

Tables N.1 and N.2 present results of multinomial logistic regressions conducted in both study contexts in Paper Three. Two types of standard errors (ordinary standard errors and robust standard errors) are presented in these tables in order to conclude whether there is a need to control for clustering effects in the models or not.

Table N.1: Results of the multinomial logistic regression for the latent class approach in the Chinese context (three-class solution).

Variable	Class One			Class Two		
	β	SE	RSE	β	SE	RSE
Intercept	1.498	0.283***	0.285***	-3.313	1.395*	1.811 NS
Age group						
<20	-0.068	0.276	0.280	0.628	0.691	0.683
20-29	0.463	0.167**	0.166**	-0.434	0.552	0.549
30-39	0.618	0.161***	0.160***	-2.032	0.848*	0.836*
40-49 (ref)	0.000			0.000		
Residence						
Urban	0.039	0.202	0.203	0.591	1.069	1.055
Rural (ref)	0.000			0.000		
Ethnicity						
Han	0.268	0.202	0.204	-0.451	0.814	0.774
Minority (ref)	0.000			0.000		
Education						
No education	-2.288	0.387***	0.390***	2.650	1.409	1.806
Primary	-1.012	0.283***	0.285***	3.038	1.389*	1.794 NS
Secondary	-0.491	0.266	0.267	2.347	1.385	1.788
Higher (ref)	0.000			0.000		
Marital status						
Widowed or divorced	-0.290	0.099**	0.098**	-0.181	0.169	0.169
Married or remarried	-0.154	0.047**	0.047**	-0.218	0.084*	0.084*
Never married (ref)	0.000			0.000		
Year of survey						
1997	-1.688	0.246***	0.247***	1.010	0.831	0.772
2001	-1.147	0.205***	0.206***	-0.854	0.880	0.827
2005 (ref)	0.000			0.000		
Age*year of survey						
15-19*1997	0.302	0.161	0.161	-0.447	0.270	0.268
15-19*2001	0.249	0.142	0.142	-0.102	0.270	0.268
20-29*1997	0.037	0.110	0.110	0.068	0.196	0.195
20-29*2001	0.210	0.092*	0.092*	0.253	0.195	0.195

30-39*1997	-0.210	0.101*	0.101*	0.248	0.156	0.155
30-39*2001	-0.034	0.078	0.078	0.329	0.150*	0.151*
Age*education						
15-19*no education	-0.308	0.543	0.544	-1.202	0.785	0.780
15-19*primary	0.174	0.267	0.269	-0.513	0.652	0.651
15-19*secondary	-0.018	0.236	0.239	-0.781	0.635	0.634
20-29*no education	-0.222	0.213	0.212	0.265	0.541	0.539
20-29*primary	-0.347	0.160*	0.159*	0.128	0.528	0.525
20-29*secondary	-0.256	0.143	0.142	-0.289	0.521	0.519
30-39*no education	-0.494	0.182**	0.182**	1.688	0.842*	0.831*
30-39*primary	-0.459	0.156**	0.155**	1.673	0.839*	0.828*
30-39*secondary	-0.393	0.143**	0.142**	1.447	0.836	0.825
Age*residence						
15-19*urban	0.048	0.095	0.094	0.923	0.171***	0.172***
20-29*urban	-0.211	0.075**	0.075**	0.493	0.142***	0.141***
30-39*urban	-0.142	0.069*	0.069*	0.274	0.127*	0.127*
Year of survey* education						
1997*no education	0.781	0.271**	0.273**	-0.137	0.833	0.772
1997*primary	0.520	0.210*	0.212*	-0.096	0.811	0.749
1997*secondary	0.130	0.193	0.195	0.182	0.806	0.744
2001*no education	0.322	0.241	0.242	0.431	0.881	0.826
2001*primary	0.137	0.185	0.186	0.043	0.861	0.807
2001*secondary	-0.098	0.174	0.175	0.202	0.857	0.802
Year of survey*ethnicity						
1997*Han	0.179	0.137	0.136	0.239	0.190	0.190
2001*Han	-0.206	0.100*	0.100*	0.349	0.189	0.189
Year of survey*residence						
1997*urban	-0.033	0.089	0.089	-0.552	0.172**	0.172**
2001*urban	0.524	0.075***	0.075***	-0.451	0.175*	0.175*
Education*ethnicity						

No education*Han	0.485	0.300	0.302	0.421	0.817	0.771
Primary*Han	-0.302	0.200	0.202	-0.156	0.802	0.754
Secondary*Han	-0.144	0.183	0.185	0.040	0.798	0.751
Education*residence						
No education*urban	-0.384	0.244	0.245	-0.864	1.062	1.014
Primary*urban	-0.225	0.185	0.186	-1.191	1.052	1.006
Secondary*urban	0.113	0.171	0.172	-1.613	1.048	0.999
Ethnicity*residence						
Han*urban	0.197	0.095*	0.096*	0.360	0.189	0.192

Ref – reference category. SE – standard error, RSE – robust standard error.
***-p<0.001, **-p<0.01, *-p<0.05, NS – not significant.
Note: Reference category of the response variable is Class Three.

Table N.2: Results of the multinomial logistic regression for the latent class approach in the five country context (four-class solution).

Variable	Class One			Class Two			Class Three		
	β	SE	RSE	β	SE	RSE	β	SE	RSE
Intercept	2.932	0.160***	0.157***	-4.445	0.449***	0.446***	0.668	0.195**	0.193**
Age group									
<20	0.356	0.187	0.183	1.194	0.476*	0.468*	0.798	0.224***	0.221***
20-29	0.603	0.125***	0.125***	0.583	0.362	0.356	0.523	0.158**	0.157**
30-39	0.574	0.134***	0.133***	0.480	0.395	0.392	0.417	0.170*	0.169*
40-49 (ref)	0.000			0.000			0.000		
Residence									
Urban	-0.081	0.122	0.121	0.900	0.288**	0.290**	0.320	0.146*	0.145*
Rural (ref)	0.000			0.000			0.000		
Education									
No education	-2.334	0.100***	0.099***	1.508	0.236***	0.238***	-0.423	0.136**	0.134**
Primary	-1.837	0.100***	0.099***	1.305	0.239***	0.240***	-0.397	0.138**	0.137**
Secondary	-1.177	0.093***	0.092***	0.695	0.234**	0.238**	-0.253	0.131	0.129

Higher (ref)	0.000			0.000			0.000		
Marital status									
Widowed or divorced	0.307	0.148*	0.144*	0.151	0.430	0.432	0.163	0.164	0.162
Married or remarried	0.076	0.109	0.105	0.077	0.291	0.279	-0.081	0.121	0.119
Never married (ref)	0.000			0.000			0.000		
Country and year of survey									
China 2005	-1.075	0.224***	0.227***	3.203	0.524***	0.507***	-1.093	0.332**	0.338**
Ukraine 2007	-1.158	0.196***	0.196***	2.398	0.545***	0.536***	-1.866	0.296***	0.299***
India 2006	-0.568	0.137***	0.134***	3.024	0.388***	0.383***	-0.258	0.154	0.152
Kenya 2003	-0.022	0.182	0.176	-0.702	0.688	0.624	-1.397	0.268***	0.271***
Malawi 2004 (ref)	0.000			0.000			0.000		
Age*country									
15-19*China	-0.394	0.250	0.253	-1.400	0.525**	0.507**	-0.339	0.358	0.370
15-19*Ukraine	-0.538	0.205**	0.206**	-2.621	0.586***	0.571***	-0.568	0.311	0.314
15-19*India	-0.438	0.130**	0.128**	-0.995	0.348**	0.339**	-0.418	0.147**	0.145**
15-19*Kenya	-0.187	0.172	0.167	-0.496	0.634	0.573	-0.173	0.261	0.264
20-29*China	-0.285	0.122*	0.122*	-1.254	0.349***	0.345***	-0.478	0.199*	0.200*
20-29*Ukraine	-0.515	0.138***	0.139***	-1.258	0.402**	0.400**	-0.562	0.214*	0.215*
20-29*India	-0.301	0.091**	0.091**	-0.596	0.286*	0.283*	-0.309	0.104*	0.105*
20-29*Kenya	0.149	0.128	0.127	-0.107	0.519	0.507	-0.143	0.209	0.210
30-39*China	-0.229	0.116*	0.117*	-0.665	0.339*	0.337*	-0.438	0.175*	0.175*
30-39*Ukraine	-0.509	0.138***	0.138***	-1.059	0.404**	0.399**	-0.359	0.203	0.201
30-39*India	-0.249	0.097*	0.097*	-0.416	0.312	0.311	-0.251	0.111*	0.112*
30-39*Kenya	0.023	0.135	0.135	-0.409	0.584	0.576	-0.123	0.220	0.219
Age*education									
15-19*no education	-0.131	0.166	0.164	-0.085	0.340	0.339	-0.321	0.198	0.195
15-19*primary	-0.147	0.154	0.151	-0.110	0.337	0.336	-0.400	0.191*	0.188*
15-19*secondary	0.249	0.139	0.137	0.068	0.329	0.328	-0.111	0.176	0.172
20-29*no education	-0.313	0.103**	0.102**	-0.061	0.231	0.224	-0.314	0.135*	0.133*
20-29*primary	-0.251	0.102*	0.101*	0.010	0.234	0.227	-0.132	0.137	0.135
20-29*secondary	-0.106	0.091	0.090	0.112	0.227	0.222	-0.171	0.125	0.123
30-39*no education	-0.277	0.106**	0.105**	-0.152	0.249	0.240	-0.103	0.142	0.139

30-39*primary	-0.250	0.107*	0.105*	-0.079	0.253	0.244	-0.187	0.146	0.144
30-39*secondary	-0.122	0.097	0.095	-0.022	0.248	0.241	-0.140	0.136	0.134
Country*marital status									
China* widowed or divorced	-0.357	0.308	0.308	-1.118	0.675	0.663	-0.754	0.492	0.501
China*married or remarried	0.158	0.203	0.206	-1.229	0.447**	0.426**	-0.738	0.300*	0.308*
Ukraine*widowed or divorced	-0.371	0.212	0.213	-0.321	0.587	0.588	-0.085	0.314	0.318
Ukraine*married or remarried	-0.050	0.161	0.160	-0.517	0.445	0.436	0.040	0.244	0.248
India*widowed or divorced	-0.609	0.164***	0.161***	-0.209	0.438	0.440	-0.424	0.183*	0.182*
India*married or remarried	-0.274	0.114*	0.110*	-0.186	0.294	0.282	-0.323	0.127*	0.125*
Kenya*widowed or divorced	-0.770	0.220***	0.214***	-1.134	1.186	1.137	-0.666	0.336*	0.337*
Kenya*married or remarried	-0.335	0.150*	0.144*	0.379	0.527	0.449	-0.408	0.210	0.213
Country*residence									
China*urban	-0.079	0.112	0.113	-1.421	0.267***	0.268***	-0.347	0.161*	0.160*
Ukraine*urban	0.413	0.119**	0.119**	-0.790	0.298**	0.298**	0.091	0.166	0.165
India*urban	0.371	0.094***	0.094***	-0.918	0.220***	0.222***	-0.096	0.105	0.105
Kenya*urban	0.550	0.120***	0.119***	-0.362	0.392	0.385	0.306	0.166	0.167
Education*residence									
No education* urban	0.272	0.085**	0.085**	0.161	0.190	0.188	-0.014	0.110	0.109
Primary*urban	0.071	0.085	0.084	0.118	0.192	0.190	-0.067	0.111	0.110
Secondary* urban	0.128	0.075	0.075	0.109	0.186	0.184	0.055	0.100	0.099
Marital status*residence									

Widowed or divorced*urban	0.040	0.092	0.092	-0.128	0.132	0.132	-0.184	0.115	0.115
Married or remarried*urban	-0.001	0.044	0.043	-0.201	0.066**	0.066**	-0.123	0.053*	0.053*

Ref – reference category. SE – standard error, RSE – robust standard error.
***-p<0.001, **-p<0.01, *-p<0.05
Note: Reference category of the response variable is Class Four.

Appendix O

Table O.1 presents a summary of different measures of HIV knowledge which were derived using simple score and latent variables approaches to measuring HIV knowledge in Papers Two and Three.

Table O.1: Summary of different approaches to measuring HIV knowledge.

	Simple Score	Simple Score Grouped	Latent Trait	Latent Trait Grouped	Latent Class
Number of categories	8 or (5 for correct and 4 for incorrect).	Depends on the decision (4 groups created in Paper Two).	Continuous score.	Depends on the decision.	2-5 (depends on the decision).
Meaning of categories	0 – no knowledge, 7 – perfect knowledge, other categories represent number of routes correctly answered without specification of which routes are known, just provides numerical value of correctly answered routes.	Arbitrary split of the answers into 4 groups, every next group represents better HIV knowledge level than the previous one without providing any details about which routes are known and which are not known.	Score represents a level of HIV knowledge with weights applied to knowledge of every separate routes of transmission.	Scores can be split, for example, into 8 groups similarly to Simple Score or by 4 quartiles with first group representing people with the poorest knowledge and the top group – with the best knowledge, it is not possible to identify which routes are known.	Groups represent people with high probabilities of obtaining specific patterns in their answers but do not say how many correct answers people provided in each group.
Quantitative meaning	For every group.	Roughly for every group.	No groups, but it places people on continuum between no knowledge and perfect knowledge.	No groups, but it places people on continuum between no knowledge and perfect knowledge.	No groups.
Qualitative meaning	Only for 2 extreme groups.	No groups.	Continuous measure of HIV knowledge. No qualitative meaning.	Based on continuous measure, no qualitative meaning.	Provide most probable patterns of answers in each group.
Ways of modelling	Ordinal, partial proportional odds, multinomial or linear regression if scores are normally distributed.	Ordinal, partial proportional odds, multinomial.	Linear regression if scores are normally distributed.	Ordinal, partial proportional odds, multinomial.	Multinomial.

Difficulty of obtaining variable	Easy.	Easy.	Difficult (specialised software is required).	Easy once latent trait continuous variable is obtained.	Difficult (specialised software is required).
One country at a time VS pooled	Same (no problem with pooling).	Same (no problem with pooling).	Different (when groups are extracted in pooled dataset, we might be loosing a bit of detailed information about specific countries but it enables comparison across countries).	Different (when groups are extracted in pooled dataset, we might be loosing a bit of detailed information about specific countries but it enables comparison across countries).	Different (when groups are extracted in pooled dataset, we might be loosing a bit of detailed information about specific countries but it enables comparison across countries).
Limitations	No qualitative details about categories which are not extreme; Does not take into account weights for each contributing route.	No qualitative details about any of the categories apart from continuum from the poorest to the best knowledge.	No well-defined qualitative meanings can be attached to scores. No quantitative details either.	No qualitative details about any of the categories. No quantitative details either.	No quantitative details about number of routes known.
Summary of measure	Good quantitative details and qualitative details for two extreme categories.	Rough quantitative detail and continuum from the poorest to the best knowledge.	No quantitative details, continuum from the poorest to the best knowledge. Important when dimensionality of a concept or relative importance of different components are investigated.	No quantitative details, continuum from the poorest to the best knowledge.	No quantitative details, good quantitative details about specific patterns of responses.

References

Adams, V., Erwin, K. and Le, P.V. (2009). "Public health works: Blood donation in urban China." Social Science & Medicine **68**(3): 410-418.

Agresti, A. (1996). An introduction to categorical data analysis (1st edition). New York: John Wiley & Sons.

Agresti, A. (2002). Categorical data analysis (2nd edition). New York: Wiley.

Agresti, A. and Finlay, B. (1997). Statistical method for the social sciences (3rd edition). Upper Saddle River, New Jersey: Prentice Hall.

Ajzen, I. (1991). "The theory of planned behavior." Organizational Behavior and Human Decision Processes **50**(2): 179-211.

Ajzen, I. and Fishbein, M. (1980). Understanding attitudes and predicting social behaviour (1st edition). Eglewood Cliffs, NJ: Prentice-Hall.

Akaike, H. (1974). "A new look at statistical model identification." IEEE Transactions on Automatic Control **AC-19**: 716-723.

Ambati, B.K., Ambati, J. and Rao, A.M. (1997). "Dynamics of knowledge and attitudes about AIDS among the educated in southern India." AIDS Care **9**(3): 319-330.

Ananth, C.V. and Kleinbaum, D.G. (1997). "Regression models for ordinal responses: a review of methods and applications." International Journal of Epidemiology **26**(6): 1323-1333.

Anderson, A.F., Qingsi, Z., Hua, X. and Jianfeng, B. (2003). "China's floating population and the potential for HIV transmission: a social-behavioural perspective." AIDS Care **15**(2): 177-185.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory (1st edition). Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1990). "Perceived self-efficacy in the exercise of control over AIDS infection." Evaluation and Program Planning **13**(1): 9-17.

Barden-O'Fallon, J.L., deGraft-Johnson, J., Bisika, T., Sulzbach, S., Benson, A. and Tsui, A.O. (2004). "Factors associated with HIV/AIDS knowledge and risk perception in rural Malawi." AIDS and Behavior **8**(2): 131-140.

Bartholomew, D.J., Steele, F., Moustaki, I and Galbraith, J. (2008). Analysis of multivariate social science data (2nd edition). Boca Raton: Chapman & Hall.

Beadnell, B., Baker, S., Knox, K., Stielstra, S., Morrison, D.M., DeGooyer, E., Wickizer, L., Doyle, A. and Oxford, M. (2003). "The influence of psychosocial difficulties on women s attrition in an HIV/STD prevention program." AIDS Care **15**(6): 807-820.

Becker, M.H. (1974). "The health belief model and personal health behaviour." Health Education Monographs **2**(4): 324-508.

Benotsch, E.G., Pinkerton, S.D., Dyatlov, R.V., DiFranceisco, W., Smirnova, T.S., Swain, G. R., Dudko, V.Y. and Kozlov, A.P. (2004). "A comparison of HIV/AIDS knowledge and attitudes of STD clinic clients in St. Petersburg, Russia and Milwaukee, WI, USA." Journal of Community Health **29**(6): 451-465.

Bentler, P.M., Jackson, D.N. and Messick, S. (1971). "Identification of content and style: A two-dimensional interpretation of acquiescence." Psychological Bulletin **76**(3): 186-204).

Billiet, J.B. and Davidov, E. (2008). "Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design." Sociological Methods and Research **36** (4): 542-562.

Blinder, A.S. (1973). "Wage discrimination: Reduced form and structural estimates." Journal of Human Resources **8**(4): 436-455.

Bongaarts, J. (1996). "Global trends in AIDS mortality." Population and Development Review **22**(1): 21-45.

Bozdogan, H. (1987). "Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions." Psychometrika **52**: 345-370.

Braithwaite, K. and Thomas, V.G. (2001). "HIV/AIDS knowledge, attitudes, and risk-behaviors among African-American and Caribbean college women." International Journal for the Advancement of Counselling **23**(2): 115-129.

Brant, R. (1990). "Assessing proportionality in the proportional odds model for ordinal logistic regression." Biometrics **46**(4): 1171-1178.

Brown, W.J. and Bocarnea, M.C. (1998). Assessing AIDS-related concern, beliefs, and communication behaviour, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 310-312. SAGE publications.

Bulik, C.M., Sullivan, P.F. and Kendler, K.S. (2000). "An empirical study of the classification of eating disorders." American Journal of Psychiatry **157**(6): 886-895.

Burkholder, G. J., Harlow, L.L. and Washkwich, J.L. (1999). "Social stigma, HIV/AIDS knowledge, and sexual risk." Journal of Applied Biobehavioral Research **4**(1): 27-44.

Buszin, J., Nieto-Andrade, B., Rivas, J. and Longfield, K. (2009). "Multiple partnerships and HIV among the Garifuna minority population in Belize." Detroit, Michigan, USA.

Buvé, A., Caraël, M., Hayes, R.J., Auvert, B., Ferry, B., Robinson, N.J., Anagonou, S., Kanhonou, L., Laourou, M., Abega, S., Akam, E., Zekeng, L., Chege, J., Kahindo, M., Rutenberg, N., Kaona, F., Musonda, R., Sukwa, T., Morison, L., Weiss, H.A. and Laga, M. (2001). "The multicentre study on factors determining the differential spread of HIV in four African cities: Summary and conclusions." AIDS **15**(supplement 4): S127-S131.

Cai, G., Moji, K., Honda, S., Wu, X. and Zhang, K. (2007). "Inequality and unwillingness to care for people living with HIV/AIDS: A survey of medical professionals in Southeast China." AIDS Patient Care and STDs **21**(8): 593-601.

Cai, Y., Hong, H., Shi, R., Ye, X., Xu, G., Li, S. and Shen, L. (2008). "Long-term follow-up study on peer-led school-based HIV/AIDS prevention among youths in Shanghai." International Journal of STD and AIDS **19**(12): 848.

Canudas Romo, V. (2003). Decomposition methods in demography, University of Groningen.

Cao, H., Wang, D., Qin, X. and Hu, Z. (2006). "The evaluation of health education on STD/AIDS knowledge among market vendors in Hefei city." Chinese Journal of Disease Control and Prevention **10**(2): 132-135. (in Chinese)

Carey, M.P., Morrison-Beedy, D. and Johnson, B.T. (1997). "The HIV-knowledge questionnaire: Development and evaluation of a reliable, valid, and practical self-administered questionnaire." AIDS and Behavior **1**(1): 61-74.

Carey, M.P., Morrison-Beedy, D. and Johnson, B.T. (1998). The HIV-knowledge questionnaire, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 313-315. SAGE publications.

Catania, J.A., Kegeles, S.M. and Coates, T.J. (1990). "Towards an understanding of risk behavior: An AIDS risk reduction model (ARRM)." Health Education & Behavior **17**(1): 53-72.

CHAMP. (2008). "AIDS-related knowledge, attitudes, behaviour, and practices: A survey of 6 Chinese cities: Summary report of CHAMP 2008 KAB/P." China HIV/AIDS Media Partnership, UNAIDS, GBC Initiative. http://www.unaids.org.cn/uploadfiles/20080925150557.pdf [Accessed 20 April 2009].

Charasse-Pouélé, C. and Fournier, M. (2006). "Health disparities between racial groups in South Africa: A decomposition analysis." Social Science & Medicine **62**(11): 2897-2914.

Chen, J., Chen, S. and Minjia, K. (2003a). Who had correct information and knowledge about HIV/AIDS in China, in Chen, S., Zhang, S., Hsia, J. and Mo, L. (eds.), Alarming sign (1st edition), pp. 87-113. Beijing: Chinese Population. [*sic*] (in Chinese)

Chen, J., Choe, M.K., Chen, S. and Zhang, S. (2007). "The effects of individual-and community-level knowledge, beliefs, and fear on stigmatization of people living with HIV/AIDS in China." AIDS Care **19**(5): 666-673.

Chen, S., Zhang, S., Mo, L. and Yang, S. (2003b). HIV/AIDS in China: survey provides guidelines for improving awareness, in Chen, S., Zhang, S., Hsia, J. and Mo, L. (eds.), Alarming sign (1st edition), pp. 76-86. Beijing: Chinese Population. [*sic*] (in Chinese)

Chen, S., Zhang, S., Mo, L. and Yang, S. (2003c). Results of baseline survey for IEC in HIV/AIDS prevention project, in Chen, S., Zhang, S., Hsia, J. and Mo, L. (eds.), Alarming sign (1st edition), pp. 48-60. Beijing: Chinese Population. (in Chinese)

Chen, X., Li, X., Stanton, B., Mao, R., Sun, Z., Zhang, H., Qu, M., Wang, J. and Thomas, R. (2004). "Patterns of cigarette smoking among students from 19 colleges and universities in Jiangsu Province, China: A latent class analysis." Drug and Alcohol Dependence **76**(2): 153-163.

Cheng, Y., Guo, X., Li, Y., Huang, N., Wang, X., Wu, J. and Ru, X. (2005). Survey on AIDS related knowledge, attitudes and practice among rural married women in China. 7th International Congress on AIDS in Asia and the Pacific (CAAP). Japan, Kabo.

Chou, I. (2004). "China's new AIDS policy faces great wall of scepticism." Nature Medicine **10**(1): 4.

Chu, T.X. and Levy, J.A. (2005). "Injection drug use and HIV/AIDS transmission in China." Cell Research **15**(11-12): 865-869.

Clogg, C.C. and Goodman, L.A. (1984). "Latent structure analysis of a set of multidimensional contingency tables." Journal of the American Statistical Association **79**(388): 762-771.

Cloud, J. and Vaughan, G.M. (1970). "Using balanced scales to control acquiescence." Sociometry **33**: 193-212.

Coale, A.J. (1992). "Age of entry into marriage and the date of the initiation of voluntary birth control." Demography **29**(3): 333-341.

Coates, T.J., Richter, L. and Caceres, C. (2008). "Behavioural strategies to reduce HIV transmission: How to make them work better." The Lancet **372**(9639): 669-684.

Collins, J.M. (2009). "Differential effects of education by gender and age cohort on engagement in risky and protective sexual behaviors in 7 sub-Saharan African countries." Population Association of America Conference. Detroit, Michigan, USA.

CPDRC and S3RI. (2004). China/UNFPA Reproductive Health/Family Planning CPR/98/P01. China Population Development Research Centre & Opportunities & Choices Reproductive Health Research Programme, Southampton Statistical Sciences Research Institute, University of Southampton, UK. http://www.s3ri.soton.ac.uk/projects/proj-unfpa.php [Accessed 19 August 2009].

Davis, C., Tang, C.S.K., Chan, S.F.F. and Noel, B. (1999). "The development and validation of the international AIDS questionnaire-Chinese version (IAQ-C)." Educational and Psychological Measurement **59**(3): 481.

Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S. L. (1998). Handbook of sexuality-related measures (1st edition). SAGE publications.

De Vaus, D.A. (2002). Surveys in social research (5^{th} edition). London: Routledge.

Dias, J.G. (2001). Components of knowledge on AIDS in Brazil: Identifying information needs using a segmented approach. Population Research Centre Working Paper. University of Groningen: Population Research Centre.

Ding, X., Wang, D., Yang, M. and Xu, S. (2006). "A high risk behaviors analysis and KABP investigation on STD/AIDS among 489 female attendants of the entertainment establishments in Chongqing." Chinese Journal of Disease Control and Prevention **10**(4): 365-368. [*sic*] (in Chinese)

Do, M. and Guend, H. (2009). "HIV-related stigma and HIV testing: A cross-country comparison in Vietnam, Tanzania, and Cote d'Ivoire." XXVI IUSSP Conference, Marrakech, Morocco.

Dong, Q., Li, C. and Zhao, G. (2003). "Status of knowledge about STD/HIV and related factors among mobile people in Shanghai Pudong new district." Journal of Fudan University (Medical Sciences) **30**(4): 325-329. (in Chinese)

Dummer, T.J.B. and Cook, I.G. (2007). "Exploring China's rural health crisis: Processes and policy implications." Health Policy **83**: 1-16.

Eberstadt, N. (2002). "The future of AIDS." Foreign Affairs **81** (6): 22-45.

Efron, B. and Tibshirani, R.J. (1993). An introduction to the bootstrap (1^{st} edition). London: Chapman and Hall.

Ellis, J.A., Scurrah, K.J., Cobb, J.E., Zaloumis, S.G., Duncan, A.E. and Harrap, S.B. (2007). "Baldness and the androgen receptor: the AR polyglycine repeat polymorphism does not confer susceptibility to androgenetic alopecia." Human Genetics **121**: 451-457.

Everitt, B.S. and Dunn, G. (1991). Applied multivariate data analysis (1^{st} edition). London: Edward Arnold.

Ezedinachi, E.N.U., Ross, M.W., Meremiku, M., Essien, E.J., Edem, C.B., Ekure, E. and Ita, O. (2002). "The impact of an intervention to change health workers' HIV/AIDS attitudes and knowledge in Nigeria: A controlled trial." Public Health **116**(2): 106-112.

Fang, J. and Kaufman, J. (2008). "Reproductive health in China: Improve the means to the end." The Lancet **372**: 1619-1620.

Feldman, B.S., Kark, J.D., Zarka, S., Ankol, O., Letyagina, V., and Shtarkshall, R.A. (2011). "Behavioral surveillance of knowledge about HIV/AIDS transmission and perceived need for additional knowledge in a national sample of young Israeli men and women between 1993 and 2005." AIDS and Behavior **15**: 193-203.

Ferguson, E., Cox, T., Irving, K., Leiter, M. and Farnsworth, B. (1995). "A measure of knowledge and confidence in relation to HIV and AIDS: Reliability and validity." AIDS Care **7**(2): 211-228.

Field, A. (2009). Discovering statistics using SPSS (3^{rd} edition). Los Angeles: SAGE.

Fisher, J.D. and Fisher, W.A. (1992). "Changing AIDS-risk behavior." Psychological Bulletin **111**(3): 455-474.

Fisher, J.D., Fisher, W.A., Bryan, A.D. and Misovich, S.J. (2002). "Information-motivation-behavioral skills model-based HIV risk behavior change intervention for inner-city high school youth." Health Psychology **21**(2): 177-186.

Fisher, J.D., Fisher, W.A., Misovich, S.J., Kimble, D.L. and Malloy, T.E. (1996). "Changing AIDS risk behavior: Effects of an intervention emphasizing AIDS risk reduction information, motivation, and behavioral skills in a college student population." Health Psychology **15**(2): 114-123.

Fitzgerald, K., Chakraborty, J., Shah, T., Khuder, S. and Duggan, J. (2003). "HIV/AIDS knowledge among female migrant farm workers in the midwest." Journal of Immigrant Health **5**(1): 29-36.

Ford, K., Wirawan, D.N., Reed, B.D., Muliawan, P. and Sutarga, M. (2000). "AIDS and STD knowledge, condom use and HIV/STD infection among female sex workers in Bali, Indonesia." AIDS Care **12**(5): 523-534.

Fotso, J.-C., Ezeh, A., Madise, N., Ziraba, A. and Ogollah, R. (2009). "What does access to maternal care mean among the urban poor? Factors associated with use of appropriate maternal health services in the slum settlements of Nairobi, Kenya." Maternal and Child Health Journal **13**(1): 130-137.

Fox, J. (2002). An R and S-Plus companion to applied regression, Sage Pubs.

Gao, Y., Lu, Z.Z., Shi, R., Sun, X.Y. and Cai, Y. (2001). "AIDS and sex education for young people in China." Reproduction, Fertility and Development **13**: 729-737.

Genberg, B.L., Hlavka, Z., Konda, K.A., Maman, S., Chariyalertsak, S., Chingono, A., Mbwambo, J., Modiba, P., Van Rooyen, H. and Celentano, D.D. (2009). "A comparison of HIV/AIDS-related stigma in four countries: Negative attitudes and perceived acts of discrimination towards people living with HIV/AIDS." Social Science & Medicine **68**(12): 2279-2287.

Genberg, B.L., Kawichai, S., Chingono, A., Sendah, M., Chariyalertsak, S., Konda, K.A. and Celentano, D.D. (2008). "Assessing HIV/AIDS stigma and discrimination in developing countries." AIDS and Behavior **12**(5): 772-780.

Gill, B., Chang, J. and Palmer, S. (2002). "China's HIV crisis." Foreign Affairs **81**(2): 96.

Gill, B. and Okie, S. (2007). "China and HIV - A window of opportunity." The New England Journal of Medicine **356**(18): 1801-1806.

Gomulka, J. and Stern, N. (1990). "The employment of married women in the United Kingdom 1970-83." Economica **57**(226): 171-199.

González Álvarez, M.L. and Clavero Barranquero, A. (2009). "Inequalities in health care utilization in Spain due to double insurance coverage: An Oaxaca-Ransom decomposition." Social Science & Medicine **69**(5), 793-801.

Goodman, L.A. (1974a). “The analysis of systems of qualitative variables when some of the variables are unobservable. Part I. A modified latent structure approach.” American Journal of Sociology **79**(5): 1179-1259.

Goodman, L.A. (1974b). “Exploratory latent structure analysis using both identifiable and unidentifiable models.” Biometrika **61**(2): 215-231.

Gregson, S., Waddell, H. and Chandiwana, S. (2001). “School education and HIV control in Sub-Saharan Africa: From discord to harmony?” Journal of International Development **13**(4): 467-485.

Gu, J. and Renwick, N. (2008). “China's fight against HIV/AIDS.” Journal of Contemporary China **17**(54): 85-106.

Guo, J. and Xu, Q. (2006). “Analysis on knowledge and influence factors towards AIDS among mobile workers.” Chinese Journal of Public Health **22**(11): 1291-1292. (in Chinese)

Gupta, G.R., Parkhurst, J.O., Ogden, J.A., Aggleton, P. and Mahal, A. (2008). “Structural approaches to HIV prevention.” The Lancet **372**(9640): 764-775.

Hagenaars, J.A. and McCutcheon, A.L. (2002). Applied latent class analysis (1st edition). Cambridge: Cambridge University Press.

Hamilton, D.L. (1968). “Personality attributes associated with extreme response style.” Psychological Bulletin **69**: 192-203.

Harkness, J. (1999). “In pursuit of quality: Issues for cross-national survey research.” International Journal of Social Research Methodology **2**(2): 125-140.

Harkness, J.A., Van de Vijver, F.J.R. and Mohler, P.P. (2003). Cross-cultural survey methods (1st edition). Hoboken, N.J: Wiley.

He, J.L. and Rehnstrom, J. (2005). “United Nations system efforts to support the response to AIDS in China.” Cell Research **15**(11-12): 908-913.

He, N. and Detels, R. (2005). “The HIV epidemic in China: History, response, and challenge.” Cell Research **15**(11-12): 825-832.

Helweg-Larsen, M. and Collins, B.E. (1997). “A social psychological perspective on the role of knowledge about AIDS in AIDS prevention.” Current Directions in Psychological Science **6**(2): 23-26.

Hesketh, T., Zhang, J. and Qiang, D.J. (2005). “HIV knowledge and risk behaviour of female sex workers in Yunnan Province, China: Potential as bridging groups to the general population.” AIDS Care **17**(8): 958-966.

Holtzman, D., Chen, S., Zhang, S., Hsia, J., Rubinson, R., Bao, F. Y., Mo, L. and McQueen, D.V. (2003). “Current HIV/AIDS-related knowledge, attitudes, and practices among the general population in China: Implications for action.” AIDS Science **3**(1): 1-6.

Hoyle, R.H. and Duvall, J.L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis, in Kaplan, D. (ed.), The SAGE handbook of quantitative methodology for social sciences (1st edition), pp. 301-316. Thousand Oaks: SAGE publications.

Hu, Z., Liu, H., Li, X., Stanton, B. and Chen, X. (2006). "HIV-related sexual behaviour among migrants and non-migrants in a rural area of China: Role of rural-to-urban migration." Public Health **120**(4): 339-345.

Huang, Y., Henderson, G.E., Pan, S. and Cohen, M.S. (2004). "HIV/AIDS risk among brothel-based female sex workers in China: Assessing the terms, content, and knowledge of sex work." Sexually Transmitted Diseases **31**(11): 695-700.

Ingham, R. (1995). AIDS: Knowledge, awareness and attitudes, in Cleland, J. and Ferry, B. (eds.), Sexual behaviour and AIDS in the developing world (1st edition), pp. 43-74. Taylor & Francis.

Intelligence Community Assessment (ICA). (2002). The next wave of HIV/AIDS: Nigeria, Ethiopia, Russia, India, and China. http://www.integration-projects.org/keydocs/ICA_HIV-AIDS_CIA-unclassified.pdf [Accessed 20 November 2007].

International Conference on Population and Development (ICPD). (1994). Summary of the programme of action. http://www.un.org/ecosocdev/geninfo/populatin/icpd.htm#chapter8 [Accessed 2 December 2008].

International Institute for Population Sciences (IIPS) and Macro International. (2007). National Family Health Survey (NFHS-3), 2005–06: India: Volume I. Mumbai: IIPS. http://pdf.usaid.gov/pdf_docs/PNADK385.pdf [Accessed 20 March 2011].

Jacobson, J.O., Robinson, P. and Bluthenthal, R.N. (2007). "A multilevel decomposition approach to estimate the role of program location and neighborhood disadvantage in racial disparities in alcohol treatment completion." Social Science & Medicine **64**(2), 462-476.

Janz, N.K. and Becker, M.H. (1984). "The health belief model: A decade later." Health Education & Behavior **11**(1): 1-47.

Jaworski, B.C. and Carey, M.P. (2007). "Development and psychometric evaluation of a self-administered questionnaire to measure knowledge of sexually transmitted diseases." AIDS and Behavior **11**(4): 557-574.

Jejeebhoy, S.J. and Sathar, Z.A. (2001). "Women's autonomy in India and Pakistan: The influence of religion and region." Population and Development Review **27**(4): 687-712.

Jiang, Z. (2000). Data collection of 1997 National Population and Reproductive Health Survey. China Population Publishing House. (in Chinese)

Kattumuri, R. (2003). "One-and-a-half decades of HIV/AIDS in Tamil Nadu: How much do patients know now?" International Journal of STD & AIDS **14**(8): 552-559.

Koch, P.B. and Singer, M.D. (1998). HIV/AIDS knowledge and attitudes scales for teachers, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 317-320. SAGE publications.

Koopman, C. and Reid, H. (1998). Assessment of knowledge and beliefs about HIV/AIDS among adolescents, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 321-324. SAGE publications.

Kutcher, N. (2003). "To speak the unspeakable: AIDS, culture, and the rule of law in China." Syracuse Journal of International Law and Commerce **30**(2): 271-286.

Lagarde, E., Pison, G. and Enel, C. (1997). "Improvement in AIDS knowledge, perceptions and risk behaviours over a short period in a rural community of Senegal." International journal of STD & AIDS **8**(11): 681.

Lancet (2010). "HIV: The fight is far from over (editorial)." The Lancet **376**: 1874.

Lau, J.T.F., Tsui, H.Y., Siah, P.C. and Zhang, K.L. (2002). "A study on female sex workers in southern China (Shenzhen): HIV-related knowledge, condom use and STD history." AIDS Care **14**(2): 219-233.

Lazarsfeld, P.F. and Henry, N.W. (1968). Latent structure analysis (1st edition). New York: Houghton Mifflin Boston.

Lee, L. (2004). "The current state of public health in China." Annual Review of Public Health **25**: 327-339.

Li, B., Guo, S., Zhao, F., Liu, H., Wu, K., Brown, J., Padmadas, S., Stones, W., Falkingham, J. and Chambers, R. (2004a). UNFPA/CHINA Reproductive Health/Family Planning Project – CPR/03/P01Baseline Survey, Technical Report. China Population & Development Research Centre, National Centre for Women and Children Health, Chinese Centre for Disease Control and Prevention, and Southampton Statistical Sciences Research Institute, UK. http://www.s3ri.soton.ac.uk/projects/proj-unfpa.php [Accessed 19 August 2009].

Li, B., Guo, S., Zhao, F., Liu, H., Wu, K., Padmadas, S., Brown, J., Stones, W. and Falkingham, J. (2004b*).* UNFPA/CHINA Reproductive Health/Family Planning Project-CPR/03/P01Baseline Survey, Key findings. China Population & Development Research Centre, National Centre for Women and Children Health, Chinese Centre for Disease Control and Prevention, and Southampton Statistical Sciences Research Institute, UK. http://www.s3ri.soton.ac.uk/projects/proj-unfpa.php [Accessed 19 August 2009].

Li, B., Qi, J., Gao, W., Wu, J., Wu, Q., Brown, J.J. and Padmadas, S.S. (2008). Baseline survey technical report: Focus on young people, gender and HIV/AIDS, UNFPA/CHINA Sixth Country Programme (2006-2010). China Population & Development Research Centre, National Centre for Women and Children Health, Chinese Centre for Disease Control and Prevention and Southampton Statistical Sciences Research Institute, UK.

Li, H., Kuo, N.T., Liu, H., Korhonen, C., Pond, E., Guo, H., Smith, L., Xue, H. and Sun, J. (2010). "From spectators to implementers: Civil society organizations involved in AIDS programmes in China." International Journal of Epidemiology **39**: ii65-ii71.

Li, L., Morrow, M. and Kermode, M. (2007). "Vulnerable but feeling safe: HIV risk among male rural-to-urban migrant workers in Chengdu, China." AIDS Care **19**(10): 1288-1295.

Li, X., Lin, C., Gao, Z., Stanton, B., Fang, X., Yin, Q. and Wu, Y. (2004c). "HIV/AIDS knowledge and implications for health promotion programs among Chinese college students: Geographic, gender and age differences." Health Promotion International **19**(3): 345-356.

Little, R.J.A. and Rubin, D.B. (2002). Statistical analysis with missing data (2nd edition). New Jersey: John Wiley and Sons, Inc.

Long, J.S. and Freese, J. (2006). Regression models for categorical dependent variables using Stata (2nd edition). TX: STATA Press Publication.

Lou, C.Q., Zhao, Q., Gao, E.S. and Shah, I.H. (2006). "Can the Internet be used effectively to provide sex education to young people in China?" Journal of Adolescent Health **39**(5): 720-728.

Loue, S., Cooper, M. and Fiedler, J. (2003). "HIV knowledge among a sample of Puerto Rican and Mexican men and women." Journal of Immigrant Health **5**(2): 59-65.

Loue, S., Cooper, M., Traore, F. and Fiedler, J. (2004). "Locus of control and HIV risk among a sample of Mexican and Puerto Rican women." Journal of Immigrant Health **6**(4): 155-165.

Luquis, R.R. and Koch, P.B. (1998). HIV/AIDS knowledge and attitudes scales for hispanics, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 325-328. SAGE publications.

Madise, N.J., Matthews, Z. and Margetts, B. (1999). "Heterogeneity of child nutrition status between households: A comparison of six Sub-Saharan African countries." Population Studies **53**(3): 331-343.

Magadi, M.A. and Agwanda, A.O. (2010). "Investigating the association between HIV/AIDS and recent fertility patterns in Kenya." Social Science and Medicine **71**(2): 335-344.

Magidson, J. and Vermunt, J.K. (2004). Latent class models, in Kaplan, D. (ed.), The SAGE handbook of quantitative methodology for social sciences (1st edition), pp. 175-198. Thousand Oaks: SAGE publications.

Magnusson, D. (1998). The logic and implications of a person-oriented approach, in Cairns, R.B., Bergman, L.R. and Kagan, J. (eds.), Methods and models for studying the individual (1st edition), pp. 33-64. Thousands Oaks: SAGE publications.

Manchester, T. (2002). Attitudes towards HIV/AIDS in China: Research on public knowledge, attitudes and behaviour in cities and towns. http://www.kaisernetwork.org/health_cast/uploaded_files/futures_group_09.24.02.pdf [Accessed 20 September 2008]

MAP. (2005). Sex work and HIV/AIDS in Asia. The Monitoring AIDS Pandemic Network (The Map Network). http://www.mapnetwork.org/docs/MAP_SW%20in%20Asia%20Final%2004July05_en.pdf [Accessed 20 November 2007].

Mauldin, W.P. (1967). "Measurement and evaluation of national family planning programs." Demography **4**(1): 71-80.

McCullagh, P. (1998). "The proportional-odds model." Encyclopedia of Biostatistics **5**: 3560-3563.

McCutcheon, A.L. (1987). Latent class analysis (1st edition). Newbury Park: SAGE publications.

Mehta, C. and Patel, N. (1998). StatXact 4 for Windows. Cambridge: CYTEL Software Corporation.

Merli, M.G., and Hertog, S. (2010). "Masculine sex ratios, population age structure and the potential spread of HIV in China." Demographic Research **22**: 63-94.

Merson, M., O'Malley, D., Serwadda, D. and Apisuk, C. (2008). "The history and challenge of HIV prevention." The Lancet **372**(9637): 475-488.

Misovich, S.J., Fisher, W.A. and Fisher, J.D. (1998). A measure of AIDS prevention information, motivation, behavioral skills, and behaviour, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 328-337. SAGE publications.

Monga, N., Rehm, J., Fischer, B., Brissette, S., Bruneau, J., El-Guebaly, N., Noël, L., Tyndall, M., Wild, C. and Leri, F. (2007). "Using latent class analysis (LCA) to analyze patterns of drug use in a population of illegal opioid users." Drug and alcohol dependence **88**(1): 1-8.

Moors, G. (2004). "Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behaviour." European Sociological Review **20**(4): 303-320.

Muthén, B.O. (1998-2004). Mplus Technical Appendices. Los Angeles, CA: Muthén & Muthén. http://www.statmodel.com/download/techappen.pdf [Accessed 20 December 2010].

Muthén, B.O. and Muthén, L.K. (2000). "Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes." Alcoholism: Clinical and Experimental Research **24**(6): 882-891.

Muthén, L.K. and Muthén, B.O. (2009). Mplus user's guide (5th edition). Los Angeles, CA: Muthén and Muthén.

Nielsen, H.S. (1998). “Discrimination and detailed decomposition in a logit model.” Economics Letters **61**(1): 115-120.

Nutbeam, D. and Blakey, V. (1990). “The concept of health promotion and AIDS prevention. A comprehensive and integrated basis for action in the 1990s.” Health Promotion International **5**(3): 233-242.

Nyamathi, A.M., Stein, J.A. and Swanson, J.M. (2000). “Personal, cognitive, behavioral, and demographic predictors of HIV testing and STDs in homeless women.” Journal of Behavioral Medicine **23**(2): 123-147.

Oaxaca, R. (1973). “Male-female wage differentials in urban labor markets.” International Economic Review **14**(3): 693-709.

Oaxaca, R.L. and Ransom, M.R. (1999). “Identification in detailed wage decompositions.” Review of Economics and Statistics **81**(1): 154-157.

Padian, N.S., Buvé, A., Balkus, J., Serwadda, D. and Cates, W. (2008). “Biomedical interventions to prevent HIV infection: evidence, challenges, and way forward.” The Lancet 372(9638): 585-599.

Pan, G., Zhang, H. and Wang, Q. (2003). Collection of data of 2001 National Family Planning and Reproductive Health Survey. Volume 1. China Population Publishing House. (in Chinese)

Park, A. (2003). “China’s secret plague.” Time Asia, 15.12.2003.

Pei, Y. and Wang, G. (2007). “Investigation on knowledge and perception of AIDS/STDs among female employee in entertainment establishments.” Chinese Journal of Public Health **23**(2): 178-179. [*sic*] (in Chinese)

Peterson, B. and Harrell, F.E. Jr. (1990). “Partial proportional odds models for ordinal response variables.” Applied Statistics **39**(2): 205-217.

Ping, Y., Chuanwen, D. and Hualong, L. (2006). “Survey on the knowledge of HIV/AIDS in criminal suspect.” Tropical Medicine **6**(3): 328-330. [*sic*] (in Chinese)

Pinkerton, S.D., Dyatlov, R.V., DiFranceisco, W., Benotsch, E.G., Smirnova, T.S., Dudko, V.Y., Belyanin, D.V. and Kozlov, A. (2003). “HIV/AIDS knowledge and attitudes of STD clinic attendees in St. Petersburg, Russia.” AIDS and Behavior **7**(3): 221-228.

Pleck, J.H. (1998). AIDS-phobia scale, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 341-342. SAGE publications.

Qi, G., Zhou, X. and Yan, H. (2007). “Baseline survey on STD/AIDS knowledge, attitude and practice among female sex workers.” Chinese Journal of Public Health **23**(7): 860-862. (in Chinese)

Renwick, N. (2002). "The 'nameless fever': The HIV/AIDS pandemic and China's women." Third World Quarterly **23**(2): 377-393.

Ross, M.W. and Kelaher, M. (1993). Knowledge, attitudes and behaviour in heterosexual men and women: the research evidence, in Sherr, L. (ed.), AIDS in the heterosexual population (1st edition), pp. 253-262. Harwood academic publishers.

Schwarz, G. (1978). "Estimating the dimension of a model." The annals of statistics **6**(2): 461-464.

Shapiro, R.L. (2002). "Drawing lines in the sand: the boundaries of the HIV pandemic in perspective." Social Science & Medicine **55**: 2189-2191.

Shen, J. and Yu, D.B. (2005). "Governmental policies on HIV infection in China." Cell Research **15**(11-12): 903-907.

Shrum, J., Turner, N. and Bruce, K. (1998). AIDS attitude scale, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 346-348. SAGE publications.

Singer, J.D. and Willett, J.B. (2003). Applied longitudinal data analysis: Modelling change and event occurrence (1st edition). New York: Oxford University Press.

Singh, S. and Samara, R. (1996). "Early marriage among women in developing countries." International Family Planning Perspectives **22**(4): 148-175.

Snell, W.E., Finney, P.D. and Godwin, L.J. (1998). The stereotypes about AIDS scale, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 354-358. SAGE publications.

Snelling, D., Omariba, D.W.R., Hong, S., Georgiades, K., Racine, Y. and Boyle, M.H. (2007). "HIV/AIDS knowledge, women's education, epidemic severity and protective sexual behaviour in low-and middle-income countries." Journal of Biosocial Science **39**(03): 421-442.

Soon, J.-J. (2010). "The determinants of students' return intentions: A partial proportional odds model." Journal of Choice Modelling **3**(2): 89-112.

SPSS. (2006). SPSS Base 15.0 User's Guide. http://sw.cs.uoguelph.ca/dsoft/SPSS/SPSS15Manuals/SPSS%20Base%20User's%20Guide%2015.0.pdf [Accessed 19 August 2009].

SPSS. (2007a). SPSS Base 16.0 User's Guide. www.hanken.fi/student/media/3618/spssbaseusersguide160.pdf [Accessed 19 August 2009].

SPSS. (2007b). SPSS Regression Models 16.0. www.hanken.fi/student/media/3618/spssregressionmodels160.pdf [Accessed 20 August 2009].

STATA. 2005. Statistical Software: Release 9.0. College Station, TX: Stata Corporation.

State Council. (2004). A joint assessment of HIV/AIDS prevention, treatment and care in China. State Council AIDS Working Committee Office and UN Theme Group on HIV/AIDS in China. http://www.casy.org/engdocs/JAREng04.pdf [Accessed 20 November 2007].

Stein, J.A. and Li, L. (2008). "Measuring HIV-related stigma among Chinese service providers: confirmatory factor analysis of a multidimensional scale." AIDS and Behavior **12**(5): 789-795.

Stephenson, R. (2009). "A community perspective on young people's knowledge of HIV/AIDS in three African countries." AIDS Care **21**(3): 378-383.

Stewart Williams, J.A. (2009). "Using non-linear decomposition to explain the discriminatory effects of male–female differentials in access to care: A cardiac rehabilitation case study." Social Science and Medicine **69**(7), 1072-1079.

Storr, C.L., Reboussin, B.A. and Anthony, J.C. (2005). "The Fagerström test for nicotine dependence: a comparison of standard scoring and latent class analysis approaches." Drug and Alcohol Dependence **80**(2): 241-250.

Storr, C. L., Zhou, H., Liang, K.Y. and Anthony, J.C. (2004). "Empirically derived latent classes of tobacco dependence syndromes observed in recent-onset tobacco smokers: epidemiological evidence from a national probability sample survey." Nicotine and Tobacco Research **6**(3): 533-545.

Stuart, B. and Hinde, A. (2009). "Identifying individuals engaging in risky sexual behaviour for chlamydia infection in the UK: A latent class approach." Journal of Biosocial Science **42**(01): 27-42.

Sullivan, P.F., Kessler, R.C. and Kendler, K.S. (1998). "Latent class analysis of lifetime depressive symptoms in the national comorbidity survey." American Journal of Psychiatry **155**(10): 1398-1406.

Tan, X. (2008). "Results of a questionnaire survey among Chinese students on knowledge of and attitudes about HIV/AIDS." Medical Education **42**(2): 227.

Tan, X., Pan, J., Zhou, D., Wang, C, and Xie, C. (2007). "HIV/AIDS knowledge, attitudes and behaviors assessment of Chinese students: A questionnaire study." International Journal of Environmental Research and Public Health **4**(3): 248-253.

Tan, X., Pan, J., Zhou, D., Xie, C., Wen, X. and Hong, Y. (2006). "HIV/AIDS knowledge, attitudes and behaviors among undergraduate students in China." Journal of US-China Medical Science **3**(5): 39-46.

Thompson, A. (2003). HIV/AIDS epidemic in China spreads into the general population. Population Reference Bureau. http://www.prb.org/Articles/2003/HIVAIDSEpidemicinChinaSpreadsIntotheGeneralPopulation.aspx [Accessed 25 November 2007].

Thompson, A. (2005). China confronts HIV/AIDS. Population Reference Bureau.

http://www.casy.org/engdocs/ChinaConfrontsHIVAIDS.pdf [Accessed 23 November 2007].

Tian, L., Tang, S., Cao, W., Zhang, K., Li, V. and Detels, R. (2007). "Evaluation of a web-based intervention for improving HIV/AIDS knowledge in rural Yunnan, China." AIDS **21**(Supplement 8): S137-42.

Tucker, J.D., Henderson, G.E., Wang, T.F., Huang, Y.Y., Parish, W., Pan, S.M., Chen, X.S. and Cohen, M.S. (2005). "Surplus men, sex work, and the spread of HIV in China." AIDS **19**(6): 539-547.

UNAIDS. (2002). HIV/AIDS: China's titanic peril. UN Theme Group on HIV/AIDS in China. http://www.youandaids.org/unfiles/chinastitanicperillast.pdf [Accessed 15 November 2007].

UNAIDS. (2007). 2007 AIDS epidemic update. http://data.unaids.org/pub/EPISlides/2007/2007_epiupdate_en.pdf [Accessed 28 November 2007].

UNAIDS. (2008a). Report on the Global AIDS epidemic – Executive summary. http://data.unaids.org/pub/GlobalReport/2008/JC1511_GR08_ExecutiveSummary_en.pdf [Accessed 30 July 2009].

UNAIDS. (2008b). Q&A on HIV/AIDS estimates: Understanding the latest estimates of the 2008 report on the global AIDS epidemic. http://data.unaids.org/pub/InformationNote/2008/QA_HIVestimates_methodologybackgrounder_en.pdf [Accessed 29 July 2009].

UNAIDS. (2009). "HIV transmission in intimate partner relationships in Asia." UNAIDS website featured story. http://www.unaids.org/en/KnowledgeCentre/Resources/FeatureStories/archive/2009/20090811_Intimate_partners.asp [Accessed 24 August 2009].

UNAIDS. (2010). Global report: UNAIDS report on the global AIDS epidemic - 2010. http://www.unaids.org/documents/20101123_GlobalReport_em.pdf [Accessed 30 November 2010].

UNFPA. (2003). Component project document agreement between the government of the People's Republic of China and United Nations Population Fund. China.

UNGASS. (2008). UNGASS country progress report, P.R. China. http://data.unaids.org/pub/Report/2008/china_2008_country_progress_report_en.pdf. [Accessed 28 July 2009].

Van de Vijver, F.J.R. and Leung, K. (1997). Methods and data analysis for cross-cultural research (1st edition). Thousand Oaks: SAGE publications.

Van den Hoek, Yuliang, F., Duckers, N.H.T.M., Zhiheng, C., Jiangting, F., Lina, Z. and Xiuxing, Z. (2001). "High prevalence of syphilis and other sexually transmitted diseases among sex workers in China: Potential for fast spread of HIV." AIDS **15**(6): 753-759.

Wade, A.S., Enel, C. and Lagarde, E. (2006). “Qualitative changes in AIDS preventative attitudes in a rural Senegalese population.” AIDS Care **18**(5): 514-519.

Wang, B., Li, X., Stanton, B., Yang, H., Fang, X., Zhao, R., Dong, B., Zhou, Y., Liu, W. and Liang, S. (2005). “Vaginal douching, condom use, and sexually transmitted infections among Chinese female sex workers.” Sexually Transmitted Diseases **32**(11): 696-702.

Wang, J., Jiang, B., Siegal, H., Falck, R. and Carlson, R. (2001). “Level of AIDS and HIV knowledge and sexual practices among sexually transmitted disease patients in China.” Sexually Transmitted Diseases **28**(3): 171-175.

Wang, N., Wang, L., Wu, Z., Guo, W., Sun, X., Poundstone, K. and Wang, Y. (2010). “Estimating the number of people living with HIV/AIDS in China: 2003-2009.” International Journal of Epidemiology **39**: ii21-ii28.

Wang, S. and Keats, D. (2005). “Developing an innovative cross-cultural strategy to promote HIV/AIDS prevention in different ethnic cultural groups of China.” AIDS Care **17**(7): 874-891.

Watson, D. (1992). “Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness.” Sociological Methods and Research **21**(1): 52-88.

Williams, R. (2006). “Generalized ordered logit/partial proportional odds models for ordinal dependent variables.” Stata Journal **6**(1): 58-82.

Wong, F. Y., Z. J. Huang, Z.J., He, N., Smith, B.D., Ding, Y., Fu, C. and Young, D. (2008). “HIV risks among gay-and non-gay-identified migrant money boys in Shanghai, China.” AIDS Care **20**(2): 170-180.

World Bank. (1997). Confronting AIDS: A World bank policy research report. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1997/10/01/000009265_3980219162747/Rendered/PDF/multi0page.pdf [Accessed 18 August 2009].

World Health Organization/ Canadian Public Health Association (WHO). (1986). Ottawa charter for health promotion. http://www.who.int/hpr/NPH/docs/ottawa_charter_hp.pdf [Accessed 20 November 2010].

Wu, F.S. (2005). “International non-governmental actors in HIV/AIDS prevention in China.” Cell Research **15**(11-12): 919-922.

Wu, J., Liu, B., Li, W., Li, Y. and Jiang, Z. (2007a). “Knowledge level about HIV/AIDS among reproductive men in five counties in China.” Journal of Reproduction and Contraception **18**(2): 121-126. [*sic*] (in Chinese)

Wu, Y. (2006). “Survey of sexual knowledge, sexual attitudes and sexual behaviors among college students in Guangdong Business Studies University.” Modern Preventive Medicine **33**(4): 563-565. (in Chinese)

Wu, Z., Qi, G., Zeng, Y. and Detels, R. (1999). "Knowledge of HIV/AIDS among health care workers in China." AIDS Education and Prevention **11**(4): 353-363.

Wu, Z., Rou, K. and Cui, H. (2004). "The HIV/AIDS epidemic in China: History, current strategies and future challenges." AIDS Education and Prevention **16** (Supplement A): 7-17.

Wu, Z., Sullivan, S. G., Wang, Y., Rotheram-Borus, M.J., and Detels, R. (2007b). "Evolution of China's response to HIV/AIDS." The Lancet **369**: 679-690.

Wu, Z., Wang, Y., Detels, R. and Rotheram-Borus, M.J. (2010). "China AIDS policy implementation: Reversing the HIV/AIDS epidemic by 2015." International Journal of Epidemiology **39**: ii1-ii3.

Xiaoming, S., Yong, W., Choi, K-H., Lurie, P., and Mandel, J. (2000). "Integrating HIV prevention education into existing family planning services: Results of a controlled trial of a community-level intervention for young adults in rural China." AIDS and Behavior **4**(1): 103-110.

Yang, H., Li, X., Stanton, B., Fang, X., Zhao, R., Dong, B., Liu, W., Liang, S., Zhou, Y. and Hong, Y. (2005). "Condom use among female sex workers in China: Role of gatekeepers." Sexually Transmitted Diseases **32**(9): 572-580.

Yang, X. (2004). "Temporary migration and the spread of STDs/HIV in China: Is there a link?." International Migration Review **38**(1): 212-235.

Yang, X. (2006). "Temporary migration and HIV risk behaviors in China." Environment and Planning A **38**(8): 1527–1543.

Yang, X. and Xia, G. (2006). "Gender, migration, risky sex, and HIV infection in China." Studies in Family Planning **37**(4): 241-250.

Yang, Y., Lu, L., Jia, M., Zhang, S., Zhang, W. and Pu, H. (2007). "Survey of the floating miners witting situation about STD and AIDS in a city of Yunnan province." Occupation and Health **23**(22): 2062-2064. [*sic*] (in Chinese)

Yuan, X.Z., Zi, L.H. and Li, M. (2003). "Questionnaire survey and countermeasure about knowledge relevant to HIV among pregnant." Nursing Journal of Chinese People's Liberation Army **20**(7): 35-36. [*sic*] (in Chinese)

Zhang H.X. (2004). "The gathering storm: AIDS policy in China." Journal of International Development **16**: 1155-1168.

Zhang, K. and Beck, E.J. (1999). "Changing sexual attitudes and behaviour in China: implications for the spread of HIV and other sexually transmitted diseases." AIDS Care **11**(5): 581-589.

Zhang, L.Y., Jejeebhoy, S., Shah, I.H., Zhang, L.H., Hsia, J. and Im-Em, W. (2004). "Access to contraceptive services among unmarried young people in the north-east of China." The European Journal of Contraception & Reproductive Health Care **9**(3): 147-154.

Zhang, Y., Zhang, B. and Wan, B. (2006). "An investigation on the veneral disease/ AIDS knowledge and its influential factors of public entertainment service workers." Modern Preventive Medicine **33**(12): 2274-2276. [*sic*] (in Chinese)

Zhou, J., Sun, X., Mantell, J., Ru, X. and Wen, Y. (2007). "AIDS-related knowledge, attitudes and behavior survey among the "migrant" population in China." Journal of Reproduction and Contraception **18**(2): 155-162.

Zhu, J. and Tan, X. (2006). "Characteristics of the knowledge, attitude and behavior of HIV/AIDS among citizens, road construction workers and public servers in Central China." Journal of US-China Medical Science **3**(6): 52-60. [*sic*]

Zhu, T.F., Wang, C.H., Lin, P. and He, N. (2005). "High risk populations and HIV-1 infection in China." Cell Research **15**(11-12): 852-857.

Zimet, G.D. (1998). Adolescent AIDS knowledge score, in Davis, C.M., Yarber, W.L., Bauserman, R., Schreer, G. and Davis, S.L. (eds.), Handbook of sexuality-related measures (1st edition), pp. 365-366. SAGE publications.

Zuo, D., Wu H and Li, S. (2007). "Investigation on knowledge about sexually transmitted diseases and AIDS among female college students in Wuhan." Chinese Journal of Public Health **23**(9): 1032-1033. (in Chinese)