



**Università
degli Studi
di Palermo**

AREA RICERCA E TRASFERIMENTO TECNOLOGICO
SETTORE DOTTORATI E CONTRATTI PER LA RICERCA
U. O. DOTTORATI DI RICERCA

Health Promotion and Cognitive Sciences
Dipartimento di Scienze Psicologiche, Pedagogiche, dell'Esercizio Fisico e della Formazione
STAT-03/B

Towards a Sustainable Future: Addressing Gender-Based Violence and HIV in Sub-Saharan Africa

IL DOTTORE
MICAELA ARCAIO

IL COORDINATORE
GIUSEPPA CAPPUCCIO

IL TUTOR
ANNA MARIA PARROCO

CICLO XXXVII
ANNO CONSEGUIMENTO TITOLO 2024

Preface

Social sustainability is a broad concept that, as of yet, still has to find an official definition. However, scholars and global institutions alike agree that it involves the well-being of people and populations, so long as policies and actions that could potentially improve one group's conditions do not harm others and nature.

The United Nations' Sustainable Development Goals (SDGs) function as a framework for sustainability, in which its social aspects are crucial in their realisation. This thesis refers to two particular SDGs, "Goal 3 - Good Health and Well-Being", which calls for the health promotion and well-being of all populations, and "Goal 5 - Gender Equality", invoking gender equality and empowerment for all women and girls. Specifically, this thesis aims to explore and identify the social determinants of intimate partner violence (IPV) and HIV in Sub-Saharan Africa, as well as to evaluate the connection between these two phenomena through a spatial lens. The focus on Sub-Saharan Africa is due to the heavy burden of both HIV and intimate partner violence prevalence in the region. These goals are pursued from the point of view to social statistics.

The core of this work is, thus, based on the compilation of four academic published (or soon-to-be published) manuscripts carried out in the course of the PhD Program in Health Promotion and Cognitive Sciences at the Department of Psychology, Educational Science and Human Movement of the University of Palermo. Accordingly, this thesis is structured as follows:

1. A brief introduction, to give the theoretical framework as well as a more detailed description of the contents of this dissertation;

2. Four chapters presenting studies that constitute the main core of this work, all with the focus on Sub-Saharan Africa, given the diffusion of the HIV epidemic and intimate partner violence in the region. These studies are related to different goals concerning: 1) the identification of the determinants of IPV to build a model-based composite indicator based on the ecological model; 2) the analysis of neighbourhood dynamics and their effects on HIV prevalence; 3) the evaluation of the association between HIV prevalence and IPV prevalence taking into account their spatial component; 4) the investigation of gender-based violence among victims of human trafficking from Sub-Saharan Africa from the perspective of operators in the third sector in Palermo (Italy). The four manuscripts¹ here presented are:

- (a) **Arcaio**, M., Parroco, A. M. (2024). A Composite Indicator of Polyvictimisation Through the Lens of the Ecological Model in Sub-Saharan Africa. *Social Indicator Research* 173, 421–438. 77 doi:10.1007/s11205-024-03344-5
- (b) Parroco, A. M., **Arcaio**, M., Mendola, D. (2024). Does neighborhood matter? An analysis of HIV prevalence in Sub-Saharan African countries. *Biodemography and Social Biology*, 69(3), 149–162. doi:10.1080/19485565.2024.2384995
- (c) **Arcaio**, M., Parroco, A. M. (N.D.). Connected threads: Africa, HIV and Intimate Partner Violence using the INLA-SPDE approach. *Springer Proceedings in Mathematics and Statistics: SIS 2023-Statistical Learning, Sustainability and Impact Evaluation*
- (d) **Arcaio**, M., Barbaro, F., Parroco, A.M. (N.D.) “Inhumane”: quantitative textual analysis of a qualitative survey on gender violence among victims of international human trafficking. *Book of Short Papers of the ASA Conference 2024 - Measuring and Interpreting World Changes with Statistics, Data Science and AI*

3. Some conclusions, in which the main findings of the dissertation, as

¹Authors’ contributions can be found at the end of the Preface.

well as more generalised findings taking all the previous chapters into account, are presented.

I would like to mention some personal remarks before leaving readers to the contents of my thesis.

The choice of social sustainability as a guiding framework for my work is due to multiple reasons. First, it is deeply rooted in my academic background, which has consistently focused on the intersections of sociology, anthropology, economics and population studies with sustainable practices. Given my Bachelor’s degree in International Cooperation and Economic Development first, and my Master’s degree in Statistics and Data Science after, my academic journey has provided me with a strong foundation to critically analyse and address the social dimensions of intimate partner violence and HIV in Sub-Saharan Africa, using quantitative and robust methodologies as tools to evaluate these phenomena. This, in conjunction with my further education offered by the researchers and other PhD students in Psychology working within my Department, have granted me a unique point of view in the evaluation and study of intimate partner violence and HIV.

Moreover, my research was supported by National Operational Programme (PON) on Research and Innovation 2014-2020 (scholarship code: DOT1720935 - 4; Project ID: B73D21009040006) of the Italian Ministry of University and Research, which is dedicated to fostering solutions to social challenges as stated by the SDGs, further motivating me to prioritise this perspective in my work. This grant has allowed me to dedicate these last three years to my research project full time, and also provided me with the possibility to be a visiting research student for the WorldPop Group of the School of Geography of the University of Southampton (UK), which gave me the opportunity to create long-standing friendships with authors from different countries.

Additionally, I have presented research outputs in eight international conferences, only one of which is not directly related to my thesis. Presentation 3 in the list below also granted me the *VADISTAT – Per Simona Balbi 2024 Young Award for best presentation*.

1. **Arcaio**, M., Barbaro, F., Parroco, A. M. (2024). “Inhumane”: quanti-

- tative textual analysis of a qualitative survey on gender violence among victims of international human trafficking. ASA 2024 conference – “Measuring and Interpreting World Changes with Statistics, Data Science and AI” (organized by the Italian Association for Applied Statistics – University of Rome “Sapienza”, Italy, September 18-20).
2. **Arcaio**, M., Parroco, A.M., Nnanatu, C.C., (2024). Evaluating the Geographical Patterns of Intimate Partner Violence in Nigeria, Zambia and Zimbabwe using the INLA-SPDE Approach. SIS 2024 – 52nd Scientific Meeting of the Italian Statistical Society (Università degli Studi di Bari, Italy, Aldo Moro – June 17-20).
 3. **Arcaio**, M., Parroco, A.M., Nnanatu, C.C., (2024). Exploring HIV hotspots in Zambia and Zimbabwe: an analysis using the INLA-SPDE approach. Data Science & Social Research, 4th international conference, Naples (Italy) Winner of the VADISTAT – Per Simona Balbi 2024 Young Award for best presentation.
 4. **Arcaio**, M., Parroco, A.M., (2023). Women’s Exposure to HIV in Africa: the Role of Intimate Partner Violence. SIS 2023 – Statistical Learning, Sustainability and Impact Evaluation (organised by the Italian Statistical Society – Università Politecnica delle Marche, Ancona, Italy, June 21-24).
 5. **Arcaio**, M., Parroco, A.M., (2023). A composite indicator for intimate partner violence: a focus on African countries. PopDays 2023 Conference (organized by the Italian Association for Population Studies – Università degli Studi Roma3, Italy, February 01-03).
 6. Parroco, A.M., **Arcaio**, A.M., (2022). Assessing intimate partner violence in African countries through a model-based composite indicator. ASA 2022 Conference (organized by the Italian Association for Applied Statistics – University of Genoa, Italy, September 12-14).
 7. **Arcaio**, M., Mendola, D., Parroco, A.M., (2022). Domestic violence in Africa: a glance through the DHS survey. SIS 2022 – 51st Conference

of the Italian Statistical Society (Università degli Studi della Campania “Luigi Vanvitelli”, Caserta, Italy, June 22-24).

8. **Arcaio**, M., Mendola, D., Parroco, A.M., (2022). HIV Prevalence in some African territories: socio-economic drivers. SIS 2022 – 51st Conference of the Italian Statistical Society (Università degli Studi della Campania “Luigi Vanvitelli”, Caserta, Italy, June 22-24).

All the steps in the process of achieving the research goals have seen my full involvement, from the formulation of research questions, to the data processing necessary to conduct the quantitative analysis, to the interpretations and discussions of the results. To achieve these goals, I have cultivated a deep understanding and expertise in both theoretical concepts and methodological approaches that are, and always will be, at the very core of my work as a researcher.

Authors’ Contributions.

1. *A Composite Indicator of Polyvictimisation Through the Lens of the Ecological Model in Sub-Saharan Africa*. Conceptualization: Arcaio, Parroco; Methodology: Arcaio, Parroco; Data preparation, Software, Analysis, Visualization, Writing—original draft preparation: Arcaio; Writing, review: Arcaio, Parroco.
2. *Does neighbourhood matter? An analysis of HIV prevalence in Sub-Saharan African countries*. Conceptualization: Parroco, Arcaio, Mendola; Methodology: Arcaio, Mendola, Parroco; Data preparation, Software, Analysis, Visualization, writing—original draft preparation: Arcaio; Writing, review: Parroco, Mendola, Arcaio.
3. *Connected threads: Africa, HIV and Intimate Partner Violence using the INLA-SPDE approach*. Conceptualization: Arcaio, Parroco; Methodology: Arcaio, Parroco; Data preparation, Software, Analysis, Visualization, Writing—original draft preparation: Arcaio; Writing, review: Arcaio, Parroco.

4. *“Inhumane”: quantitative textual analysis of a qualitative survey on gender violence among victims of international human trafficking*. Conceptualisation: Arcaio, Barbaro, Parroco; Methodology: Arcaio; Data preparation, software, visualisation: Arcaio; Analysis, Writing-original draft: Barbaro; Writing, review: Arcaio, Barbaro, Parroco.

Acknowledgements

There are several people whose support and mentoring was fundamental not only for me as a researcher but also as a person.

First and foremost, I would like to thank my supervisor, Professor Anna Maria Parroco, for the invaluable lessons and opportunities she has given me over and over throughout these past three years.

I will never thank Professor Daria Mendola enough, not only for the knowledge she has passed onto me, but also for the priceless suggestions and support, and the conversations we have had during the years.

I do not think I would have been here had it not been for Professor Annalisa Busetta. Her passion and devotion to her research showed me a path in my career and personal life I had never imagined, and I would be remiss not to remember that this thesis actually started with her.

Warmest thanks to Professor Mauro Ferrante for the advice and knowledge and conversations, giving me the courage to always do more and the chance to prove myself.

Thank you to all the members of my Department, for the continuous academic stimuli.

I am grateful to Dr Chibuzor Christopher Nnanatu and the whole World-Pop Group of the University of Southampton. You made me feel incredibly welcome and I will always cherish my time in the city.

I would also like to thank Dr Simone Lucido and NEXT - Nuove Energie x il Territorio for the chance to be an intern in your organisation, and whose collaboration is now permanently inked in Chapter 4.

Thanks to the National Operational Programme (PON) on Research and Innovation 2014-2020 (scholarship code: DOT1720935 - 4; Project ID:

B73D21009040006) of the Italian Ministry of University and Research, that granted me the possibility to focus on this research in the past three years.

Last but not least, I thank my family, my Mom, my Dad, my brother, for the infinite support and encouragement.

And to D., who is always in my heart.

Ringraziamenti

Sono diverse le persone il cui supporto e insegnamenti sono stati per me fondamentali, non solo in quanto ricercatrice ma anche come persona.

Innanzitutto, vorrei ringraziare la mia tutor, la Professoressa Anna Maria Parroco, per le inestimabili lezioni e opportunità che mi ha dato in questi ultimi tre anni.

Non ringrazierò mai abbastanza la Professoressa Daria Mendola, non solo per la conoscenza impartitami, ma anche per i suoi preziosissimi consigli e il suo supporto, e le conversazioni durante gli anni.

Non credo sarei qui se non fosse stato per la Professoressa Annalisa Busetta. La sua passione e devozione per la ricerca mi hanno fatto scoprire una strada sia nella mia vita professionale che privata che non immaginavo esisese, e commetterei un errore se non ricordassi che questa tesi è in realtà nata da lavoro iniziato con lei.

I miei più calorosi ringraziamenti vanno al Professor Mauro Ferrante, per i consigli e la sua conoscenza e le conversazioni che mi danno sempre la forza di fare di più e la possibilità di farmi valere.

Un ringraziamento va anche ai componenti del mio Dipartimento, per i continui stimoli accademici.

Sono grata al Dr Chibuzor Christopher Nnanatu e a tutto il WorldPop Group dell'Università di Southampton. Mi avete accolta con molto calore e ricorderò sempre i mesi passati in città.

Vorrei anche ringraziare il Dr Simone Lucido e NEXT - Nuove Energie x il Territorio per avermi dato la possibilità di svolgere un tirocinio presso la vostra organizzazione, e la cui collaborazione resterà perennemente ricordata nel quarto capitolo di questa tesi.

Un ringraziamento anche al Programma Operativo Nazionale (PON) Ricerca e Innovazione 2014-2020 (codice borsa DOT1720935 - 4; CUP: B73D21009040006) del Ministero dell'Università e della Ricerca italiano, che mi ha dato la possibilità di concentrarmi su questa ricerca negli ultimi tre anni.

Infine, vorrei ringraziare la mia famiglia, mia Madre, mio Padre, mio Fratello, per l'infinito supporto e incoraggiamento.

Index

Introduction	15
References	21
1 An IPV Model-Based Composite Indicator	27
1.1 Introduction	28
1.2 Theoretical Background of Intimate Partner Violence	30
1.2.1 Definitions	30
1.2.2 Literature and research hypotheses	31
1.3 Data and Methods	34
1.3.1 Data and Sample Selection	34
1.3.2 Methods	36
1.4 Results	40
1.4.1 Measurement Model Testing	40
1.4.2 Structural Equation Model Testing	40
1.4.3 Hypotheses tests	41
1.4.4 Post-estimation findings	44
1.5 Conclusion	46
References	48
2 HIV Determinants within Neighbourhoods	55
2.1 Introduction	56
2.2 HIV Risk Factors within Neighbourhoods	57
2.3 Data and methods	58
2.3.1 The fractional logistic regression models	61
2.4 Results and Discussion	63

INDEX

2.5	Conclusions	67
	References	71
3	HIV-IPV Association using INLA	75
3.1	Introduction	76
3.2	HIV and Intimate Partner Violence: a Brief Literature Review through the Ecological Model	78
3.3	Data and Methods	80
3.4	Results	82
	3.4.1 Logistic Regression Model	82
	3.4.2 Bayesian hierarchical models.	84
3.5	Conclusions	84
	References	86
4	Human Trafficking Matters by TDM	91
4.1	Introduction	92
4.2	Human Trafficking and Gender-Based Violence	93
4.3	The focus group: the transcript and methods of analysis . . .	95
4.4	Results	97
4.5	Discussion	99
4.6	Conclusions	101
	References	101
	Conclusions	105
	References	114

Introduction

Although an unequivocal definition of social sustainability has not been found yet (Boyer et al., 2016; Policy Department for Economic, Scientific and Quality of Life Policies, 2020; United Nations Global Compact, 2024), scholars and institutions alike agree that it involves the way individuals inhabit a society, compatibly with their needs, betterment, and development, without ostracising others and nature. Boström (2012), for example, summarises some of the substantive aspects of social sustainability, highlighting areas of intervention, such as equality of rights (including human rights), access to social infrastructure and facilities, opportunities for learning, economic and environmental security, health, quality of life, happiness, and well-being. In another work, Hajirasouli and Kumarasuriyar (2016) discuss different levels of impact of social sustainability, all referring to the goals of basic physical well-being, good quality of life, and social equity.

The United Nations' (UN's) Sustainable Development Goals (SDGs) (United Nations, 2024) – and its consequent Global Indicator Framework (United Nations Statistics Division, 2024) – help identify the dimensions of sustainability, dividing this broader concept into 17 goals, including 169 targets and 231 indicators to measure it. While the UN does not explicitly point to some SDGs as exclusively dealing with environmental or social sustainability, the well-being of populations is crucial and central to most of them. These goals range from ending poverty and hunger to promoting health and pushing for education and gender equality, all focusing on people and populations (United Nations, 2024). This framework was also endorsed by the European Union (EU), which developed an EU-specific set of 100 sustainable development indicators (Policy Department for Economic, Scientific and

INTRODUCTION

Quality of Life Policies, 2020).

Social sustainability is also the framework for this thesis. According to the Joint United Nations Programme on HIV/AIDS (UNAIDS, 2024d), several SDGs include actions that could lower the number of new HIV infections worldwide and improve the quality of life for people already living with HIV (UNAIDS, 2024a). Among these, Goal 5 (“Gender Equality”) plays a crucial role. Gender inequality and gender-based violence (GBV) are linked to a higher risk of HIV infections, which have been identified as the leading cause of death among women of reproductive age (15-44 years old) (Dunkle & Jewkes, 2007; Dunkle et al., 2004; Jewkes & Morrell, 2010; Jewkes et al., 2010; Maman et al., 2000; UNAIDS, 2024a, 2024c).

Given this premise, the overarching aim of this thesis is to study intimate partner violence (IPV) and HIV prevalence in East and Southern Africa, as well as their association. This aim was pursued through a structured, four-step process, with each step aligned to a distinct research objective:

1. Studying the determinants of IPV and building a model-based composite indicator based on the theoretical framework of the ecological model;
2. Analysing neighbourhood dynamics and their effects on HIV prevalence using fractional logistic regression models;
3. Assessing the association between HIV prevalence and IPV using an approach that accounts for their spatial component;
4. Investigating gender-based violence among victims of human trafficking from Sub-Saharan Africa from the perspective of operators in the third sector in Palermo (Italy), using quantitative text analysis tools.

In this section, some relevant literature on the association between IPV and HIV will be presented, alongside a brief summary of the published (or soon-to-be published) works that form this thesis, presented in full in Chapters 1-4. In the conclusion, a brief recapitulation of the thesis will be followed by a comprehensive discussion of the results.

A Brief Literature Review Since the outbreak of HIV in East and Southern Africa, several works of literature have aimed to evaluate the connection between HIV prevalence and gender-based violence (GBV), particularly intimate partner violence (IPV) (Case et al., 2012; DiCarlo et al., 2014; Dunkle & Jewkes, 2007; Dunkle et al., 2004; Jewkes & Morrell, 2010; Jewkes et al., 2010; Maman et al., 2000). This stream of literature is driven by the alarming numbers related to the HIV epidemic in the region. Of the 39.9 million people living with HIV worldwide in 2023, nearly 65% are residents of East and Southern Africa, or West and Central Africa (UNAIDS, 2024b). Additionally, the prevalence of IPV in the region is equally staggering: one in three women aged 15-49 years has been abused at least once in her lifetime, making Sub-Saharan Africa one of the regions with the highest IPV prevalence (World Health Organization, 2021).

The literature has mainly explored the relationship between HIV and gender-based violence through ad hoc surveys or qualitative studies (Dunkle & Jewkes, 2007; Dunkle et al., 2004; Jewkes et al., 2010). Several studies based in Sub-Saharan Africa have used the theory of 'hegemonic masculinity' (DiCarlo et al., 2014). According to Connell (2005), hegemonic masculinity describes a societal pattern where stereotypical male characteristics are upheld as the ideal standard of masculinity, explaining how and why men sustain dominant social roles over women and groups perceived as feminine.

Gender inequality and violence against women and girls are thought to increase the risk of HIV infection. A 2001 qualitative study in South Africa found that both physical and sexual violence by partners were linked to higher chances of HIV infection, with women reporting more partner violence in relationships where men exerted greater control (Dunkle et al., 2004). Another South African study confirmed that power imbalances within couples and IPV contribute to the spread of HIV among young women. However, this study did not find a link between HIV infection and rape by someone outside the relationship (Jewkes et al., 2010).

The acknowledgement of hegemonic masculinity as a social norm is one reason why women are considered vulnerable to AIDS, alongside biological factors (Mills et al., 2012). One solution the international community has

INTRODUCTION

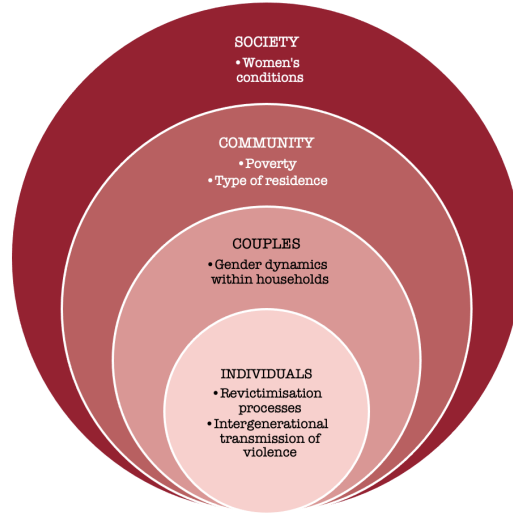
pursued to address this issue is engaging men in prevention and treatment programs, including HIV testing campaigns and antiretroviral therapy coverage (World Health Organization, 2023).

In contrast to hegemonic masculinity is the concept of women's empowerment, which involves enabling women to make life-defining decisions, reshape traditional gender roles, and gain freedom to pursue their goals. Women's empowerment may involve elevating their status through education, skill development, employment opportunities, increased awareness, and self-esteem (Kabeer, 2005). However, not all theories of women's empowerment agree on whether it improves conditions overall, especially regarding IPV prevalence. According to the ameliorative hypothesis, women's empowerment should lower IPV prevalence, as gender equality is expected to improve women's overall conditions (Heirigs & Moore, 2017). Conversely, the backlash hypothesis suggests that if women's empowerment is perceived as a threat to a patriarchal society, it may lead to increased violence (Classen et al., 2005; Levinson, 1989; Meinck et al., 2015).

Chapter 1: An IPV Model-Based Composite Indicator In the first chapter, a model-based composite indicator of polyvictimisation in intimate partner violence (IPV) is constructed using a structural equation model (SEM) (Muthén, 1984; Muthén et al., 1997). Polyvictimisation refers to experiencing multiple types of violence (e.g., physical, emotional, sexual) within a relationship (Coetzee et al., 2017; Finkelhor et al., 2007; Logie et al., 2019; Okumu et al., 2021). This paper builds on previous work (Arcaio et al., 2022a; Parroco & Arcaio, 2023), where revictimisation processes and the intergenerational transmission of violence were identified as the most relevant factors associated with recent experiences of physical, emotional, or sexual abuse by an intimate partner.

These findings were then integrated into the ecological model framework (Bronfenbrenner, 1979; Centers for Disease Control and Prevention, 2015). The ecological model posits that protective and risk factors act at different levels of influence, from societal-level factors to community, couple, and individual-level factors, with each level influencing the others.

Figure 1: The Ecological Model of Intimate Partner Violence



This theoretical model was subsequently used as the basis for constructing a SEM, which allowed for the creation of a composite indicator for polyvictimisation, to identify and quantify the multiple dimensions of violence and how they interact across different levels of influence.

Thus, Chapter 1 provides a detailed explanation of the construction of the composite indicator and its relevance to policy-making, particularly in assessing and mitigating IPV risk factors in vulnerable populations.

Chapter 2: HIV Prevalence Determinants within Neighbourhoods

In this chapter, building from (Arcaio et al., 2022b), neighbourhood dynamics as well as their effects on HIV prevalence in East and Southern Africa are analysed. Using fractional logistic regression models (Papke & Wooldridge, 1996), the analysis explores how neighbourhood-level characteristics influence HIV transmission rates, focusing on how social and cultural factors contribute to the variability in HIV prevalence.

The results indicate that neighbourhoods with higher education tend to have lower HIV prevalence rates. Conversely, areas with high HIV testing coverage are associated with higher HIV prevalence rates. An interesting results concerns the association between higher numbers of children and lower HIV prevalence, probably due to access to treatment therapy once women

enter antenatal care programs.

The findings from this chapter emphasize the critical role of neighbourhood dynamics in shaping HIV prevalence and support the case for comprehensive, localized interventions to combat the spread of HIV in East and Southern Africa.

Chapter 3: Bayesian Hierarchical Models and HIV-IPV Association Building from Arcaio and Parroco (2023) and Parroco et al. (2024), in this chapter the association between HIV prevalence and IPV prevalence is assessed using Bayesian hierarchical models. These models are implemented within the integrated nested Laplace approximation (INLA) framework, incorporating the stochastic partial differential equation (SPDE) approach to account for the spatial correlation between HIV and IPV across Zambia and Zimbabwe (Gómez-Rubio, 2020; Lindgren & Rue, 2015; Moraga, 2023). The choice of these countries is due to their national HIV prevalence: indeed, in both countries one in ten people aged 15-49 years lives with HIV (UNAIDS, 2024b).

The stochastic partial differential equation (SPDE) approach is utilized to capture the spatial dependence between HIV prevalence and IPV prevalence. This method is particularly useful in understanding how the spatial structure influences the joint distribution of HIV and IPV across regions.

Key elements of the model include data obtained from nationally representative surveys - Demographic and Health Survey (DHS) (Croft et al., 2018) data on Zambia and Zimbabwe.

The Bayesian hierarchical model reveals a significant association between HIV prevalence and IPV prevalence. Neighbourhoods with higher IPV rates (particularly, physical and emotional abuse) tend to have correspondingly higher HIV prevalence, even after controlling for socioeconomic and healthcare-related factors.

This finding aligns with previous studies suggesting that IPV is a significant risk factor for HIV transmission, particularly among women in Sub-Saharan Africa (Jewkes & Morrell, 2010; Jewkes et al., 2010). The spatial structure of the data highlights clusters of high HIV and IPV prevalence,

suggesting the need for geographically targeted interventions.

Chapter 4: Human Trafficking Matters by Text Data Mining A focus group discussion on gender-based violence (GBV) among victims of human trafficking from Sub-Saharan Africa is here explored applying text mining techniques. The participants were operators working in the third sector in the Province of Palermo (Italy) and expert researchers in migration studies.

The focus group transcript was processed using text mining techniques to identify recurring themes, sentiments, and key concepts related to human trafficking, as well as the challenges faced by victims and service providers. The quantitative results were further explored through more qualitative techniques to ensure that the nuances of the discussions were captured accurately.

References

- Arcaio, M., Mendola, D., & Parroco, A. M. (2022a). Domestic violence in Africa: A glance through the DHS survey. In A. Balzanella, M. Bini, C. Cavicchia, & R. Verde (Eds.), *Book of the short papers* (pp. 1108–1113). Pearson Italia.
- Arcaio, M., Mendola, D., & Parroco, A. M. (2022b). HIV prevalence in some african territories: Socio-economic drivers. In A. Balzanella, M. Bini, C. Cavicchia, & R. Verde (Eds.), *Book of the short papers* (pp. 1126–1131). Pearson Italia.
- Arcaio, M., & Parroco, A. M. (2023). Women’s exposure to HIV in africa: The role of intimate partner violence. In *Sis 2023 - the book of short papers* (pp. 599–604). Pearson.
- Boström, M. (2012). A missing pillar? challenges in theorizing and practicing social sustainability: Introduction to the special issue. *Sustainability: Science, practice and policy*, 8(1), 3–14.
- Boyer, R. H., Peterson, N. D., Arora, P., & Caldwell, K. (2016). Five approaches to social sustainability and an integrated way forward. *Sustainability*, 8(9), 878.

INTRODUCTION

- Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Harvard University Press.
- Case, K., Ghys, P., Gouws, E., Eaton, J., Borquez, A., Stover, J., Cuchi, P., Abu-Raddad, L., Garnett, G., & Hallett, T. (2012). Understanding the modes of transmission model of new HIV infection and its use in prevention planning. *Bulletin of the World Health Organization*, 831–838.
- Centers for Disease Control and Prevention. (2015). The social-ecological model: A framework for prevention [Accessed 2024-01-30]. <https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html>
- Classen, C. C., Palesh, O. G., & Aggarwal, R. (2005). Sexual revictimization: A review of the empirical literature. *Trauma, Violence, & Abuse*, 6, 103–129.
- Coetzee, J., Gray, G. E., & Jewkes, R. K. (2017). Prevalence and patterns of victimization and polyvictimization among female sex workers in Soweto, a South African township: A cross-sectional, respondent-driven sampling study. *Global Health Action*, 10(1), 1403815.
- Connell, R. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society*, 19(6), 829–859.
- Croft, T. N., Marshall, A. M. J., Allen, C. K., et al. (2018). *Guide to DHS statistics* (tech. rep.). ICF. Rockville, Maryland, USA.
- DiCarlo, A. L., Mantell, J. E., Remien, R. H., Zerbe, A., Morris, D., Pitt, B., Abrams, E. J., & El-Sadr, W. M. (2014). 'men usually say that HIV testing is for women': Gender dynamics and perceptions of HIV testing in Lesotho. *Culture, Health & Sexuality: An International Journal for Research, Intervention and Care*, 16(8).
- Dunkle, K. L., & Jewkes, R. K. (2007). Effective HIV prevention requires gender-transformative work with men. *Sexually Transmitted Infections*, 83, 173–174.
- Dunkle, K. L., Jewkes, R. K., Brown, H. C., Gray, G. E., McIntyre, J. A., & Harlow, S. D. (2004). Gender-based violence, relationship power, and risk of HIV infection in women attending antenatal clinics in south africa. *The Lancet*, 363, 1415–1421.

- Finkelhor, D., Ormrod, R. K., & Turner, H. A. (2007). Poly-victimization: A neglected component in child victimization. *Child Abuse & Neglect*, 31(1), 7–26.
- Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. Chapman & Hall/CRC Press.
- Hajirasouli, A., & Kumarasuriyar, A. (2016). The social dimension of sustainability: Towards some definitions and analysis. *Journal of Social Science for Policy Implications*, 4(2), 23–34.
- Heirigs, M., & Moore, M. (2017). Gender inequality and homicide: A cross-national examination. *International Journal Of Comparative And Applied Criminal Justice*, 42(4), 273–285. <https://doi.org/10.1080/01924036.2017.1322112>
- Jewkes, R. K., Dunkle, K. L., Nduna, M., & Shai, N. (2010). Intimate partner violence, relationship power inequity, and incidence of HIV infection in young women in south africa: A cohort study. *The Lancet*, 376, 41–48.
- Jewkes, R. K., & Morrell, R. (2010). Gender and sexuality: Emerging perspectives from the heterosexual epidemic in South Africa and implications for HIV risk and prevention. *Journal of the International AIDS Society*, 13(6).
- Kabeer, N. (2005). Gender equality and women’s empowerment: A critical analysis of the third millennium development goal 1. *Gender & Development*, 13(1), 13–24.
- Levinson, D. (1989). Family violence in cross-cultural perspective. In *Family violence in cross-cultural perspective* (pp. 435–455). Springer US.
- Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19), 1–25. <https://doi.org/10.18637/jss.v063.i19>
- Logie, C. H., Okumu, M., Mwima, S., Hakiza, R., Irungi, K. P., Kyambadde, P., et al. (2019). Social ecological factors associated with experiencing violence among urban refugee and displaced adolescent girls and young women in informal settlements in Kampala, Uganda: A cross-sectional study. *Conflict and Health*, 13(1), 1–15.

INTRODUCTION

- Maman, S., Campbell, J., Sweat, M. D., & Gielen, A. C. (2000). The intersections of HIV and violence: Directions for future research and interventions. *Social Science & Medicine*, *50*, 459–478.
- Meinck, F., Cluver, L. D., Boyes, M. E., & Mhlongo, E. L. (2015). Risk and protective factors for physical and sexual abuse of children and adolescents in Africa: A review and implications for practice. *Trauma, Violence, & Abuse*, *16*, 81–107.
- Mills, E. J., Beyrer, C., Birungi, J., & Dybul, M. R. (2012). Engaging men in prevention and care for HIV/AIDS in Africa. *PLoS Med*, *9*(2), e1001167.
- Moraga, P. (2023). *Spatial statistics for data science: Theory and practice with R*. Chapman & Hall/CRC Press.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132.
- Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* (tech. rep.). Unpublished technical report.
- Okumu, M., Orwenyo, E., Nyoni, T., Mengo, C., Steiner, J. J., & Tonui, B. C. (2021). Socioeconomic factors and patterns of intimate partner violence among ever-married women in Uganda: Pathways and actions for multicomponent violence prevention strategies. *Journal of Interpersonal Violence*, 08862605211021976.
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Economics*, *11*(6), 619–632.
- Parroco, A. M., & Arcaio, M. (2023). Assessing intimate partner violence in African countries through a model-based composite indicator. In *ASA 2022 data-driven decision making*. Firenze University Press. <https://doi.org/10.36253/979-12-215-0106-3>
- Parroco, A. M., Arcaio, M., & Mendola, D. (2024). Does neighborhood matter? an analysis of HIV prevalence in Sub-Saharan African coun-

- tries. *Biodemography and Social Biology*. <https://doi.org/10.1080/19485565.2024.2384995>
- Policy Department for Economic, Scientific and Quality of Life Policies. (2020). *Social sustainability: Concepts and benchmarks* (tech. rep.). European Union.
- UNAIDS. (2024a). *AIDS and the Sustainable Development Goals* [Accessed: 2024-10-06]. https://www.unaids.org/en/AIDS_SDGs
- UNAIDS. (2024b). Aidsinfo | unaids [Accessed: 2024-10-01]. <http://aidsinfo.unaids.org/>
- UNAIDS. (2024c). *Gender and HIV/AIDS* [Accessed: 2024-10-06]. <https://www.unaids.org/en/topic/gender>
- UNAIDS. (2024d). *Unaid - joint united nations programme on HIV/AIDS* [Accessed: 2024-10-06]. <https://www.unaids.org/en>
- United Nations. (2024). *Sustainable Development Goals* [Accessed: 2024-10-06]. <https://sdgs.un.org/goals>
- United Nations Global Compact. (2024). Social sustainability [Accessed: 2024-10-06]. <https://unglobalcompact.org/what-is-gc/our-work/social>
- United Nations Statistics Division. (2024). *SDG indicators - global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development* [Accessed: 2024-10-06]. <https://unstats.un.org/sdgs/indicators/indicators-list/>
- World Health Organization. (2021). *Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring: Recommendations for a public health approach* [Accessed: 2024-10-06]. <https://www.who.int/publications/i/item/9789240022256>
- World Health Organization. (2023). *Engaging men in HIV testing services to improve uptake and outcomes: Technical brief* [Accessed: 2024-10-06]. <https://www.who.int/publications/i/item/9789240085794>

INTRODUCTION

Chapter 1

An IPV Model-Based Composite Indicator

A Composite Indicator of Polyvictimisation through the lens of the Ecological Model in Sub-Saharan Africa¹

Co-Author: Anna Maria Parroco

Abstract

Intimate partner violence (IPV) is a worldwide issue. One of the latest developments in its theoretical framework deals with the concept of polyvictimisation – the simultaneous occurrence of physical, emotional, and sexual abuse. However, the literature lacks an overall measure of violence for surveys.

The aim of this research is to study IPV within the framework of the ecological model. A model-based composite indicator that takes into account the relationship between domestic abuse and individual characteristics of

¹Copyright © 2024, The Authors. Article published on Social Indicator Research - An International and Interdisciplinary Journal for Quality-of-Life Measurement. This is an open access article distributed under the terms of the Creative Commons CC BY license.

respondents, family dynamics, and community and societal traits is built using survey data.

The data are from the Demographic and Health Survey collected in eleven African countries on women aged 15-49.

The employed structural equation model shows the importance of individual characteristics while community and societal factors are less relevant. The composite indicator is also used for classification and ranking purposes, allowing areas where socio-educational interventions are more urgent to be identified.

Keywords

Intimate partner violence; composite indicator; structural equation model; polyvictimisation; ecological model.

1.1 Introduction

Intimate partner violence (IPV) is a worldwide issue (World Health Organization, 2021). It has been estimated that, globally, 27% of women aged 15-49 have experienced an act of either physical or sexual violence by an intimate partner in their lifetime, and 13% of them have had such an experience in the past year. In general, it is more widespread in what the United Nations classify as “Least Developed Countries”. In Sub-Saharan Africa, for example, 33% of women aged 15-49 have experienced IPV at least once in their lifetime, and 20% have been abused in the last twelve months (World Health Organization, 2021).

In the literature, the common measures used for IPV consider the different kinds of violence independently – giving a measure of either physical, emotional, or sexual violence. At the national level, IPV is mostly measured using indicators that are part of the “Violence against Women and Girls: a Compendium of Monitoring and Evaluation Indicators” (Bloom, 2008). Other psychometric scales are often used in surveys, such as the Safe Dates

Scales, but again they differentiate between physical and psychological abuse (Al-Modallal et al., 2020).

In the last few years, the literature has also started to move towards the study of polyvictimisation, i.e. the simultaneous occurrence of different experiences of abuse, and the mechanisms that facilitate and reinforce each of these events (Coetzee et al., 2017; Finkelhor et al., 2007; Logie et al., 2019; Okumu et al., 2021).

The polyvictimisation of individuals has been evaluated using the Composite Abuse Scale (Revised) – Short Form (Ford-Gilboe et al., 2016). However, the authors themselves currently advise against using this scale in developing countries, as the questionnaire has not been validated in this context - meaning that some items might have to be reworded.

Coetzee et al. (2017), Finkelhor et al. (2007), and Logie et al. (2019) have also proposed measures of polyvictimisation, using an additive approach that combines the three types of domestic abuse, while Okumu et al. (2021) focuses on the effect of socio-economic drivers on abuse.

The aim of this research is, thus, to study polyvictims of IPV in African countries within the framework of the ecological model, via a structural equation model (SEM), that will consider physical, emotional, and sexual violence at the same time – as well as individual characteristics of respondents, family dynamics, and community and societal traits. The structure of the relationship between the dimensions involved is the key tool for the construction of an IPV model-based indicator, which is then proposed also for classification and ranking purposes.

The data in this research are part of the Demographic and Health Survey (Croft et al., 2018). Because of its higher diffusion among women (Centers for Disease Control and Prevention, 2023; World Health Organization, 2012), this work will be focused on partnered women in heterosexual relationships.

In Section 1.2, the theoretical background of the implemented SEM in the framework of the ecological model of IPV, with a focus on Sub-Saharan Africa, is explored; Section 1.3 provides a description of the data and of the statistical model; finally, the results of the model are presented in Section 1.4, followed by the conclusions.

1.2 Theoretical Background of Intimate Partner Violence

1.2.1 Definitions

The Declaration on the Elimination of Violence against Women (United Nations, 1993) defines violence against women as gender-based acts of abuse. In particular, IPV, or domestic abuse, consists of abuse within families, typically towards a partner. It has been defined as a behavioural pattern of control and power, including actions meant to manipulate, threaten, and terrorise a partner.

Intimate partner violence can be perpetrated in different ways. Emotional or psychological abuse refers to verbal abuse and isolation from family and friends, as well as to threats to hurt a partner or their family and pets. Threats of violence via the use of a weapon – be it a knife or a gun – are categorised as a more physical kind of abuse, alongside impeding said partner from seeking for help, damaging property in anger and, of course, physical harm to the abused and their children. Sexual abuse is referred to as the perpetration of non-consensual sexual acts of any kind (United Nations, 2022).

Even though intimate partner violence affects both women and men, it is most commonly experienced by women – with men being the most common perpetrators (Centers for Disease Control and Prevention, 2023; World Health Organization, 2012). Indeed, recent data from the United States Center for Disease Control and Prevention show that while 26% of men experience domestic violence, 41% of women experience this sort of abuse during their lifetime.

Domestic violence has been identified as having a significant effect on women’s health – both physical and mental – as well as on their general well-being and that of their children and families (World Health Organization, 2021). Violence can also culminate in femicide, i.e. the murder of women because of their gender: recent data show that 50,000 femicides happen each year, with 137 women and girls being killed every day by a member of their

family (Center, 2021). Most notably, in South Africa, a woman is killed by a partner every eight hours, but numbers are on the rise in more developed regions as well: in the European Union member States, 980 femicides were estimated in 2018 and 1225 in 2020 (Women Against Violence Europe, 2019, 2021).

1.2.2 Literature and research hypotheses

Domestic abuse knows no boundaries. It is known to affect people of all ages and genders, of any race and religion, of every sexual orientation, of any socio-economic background (United Nations, 2022). To better understand its dynamics, the ecological model is often employed to check for the interplay between personal and contextual characteristics (Gage, 2005; Oyediran & Feyisetan, 2017). This model assumes that drivers operate at different levels of influence that also influence one another: the societal dimension influences factors at the community level which in turn affect relationships between people – who have further determinants of their own (Centers for Disease Control and Prevention, 2015).

On the other hand, the theory of polyvictimisation gives a comprehensive explanation as to why different types of abuse can happen not only simultaneously but also how and why they influence and reinforce each other. While one of the first studies of polyvictimisation concerned childhood abuse (Finkelhor et al., 2007), this theory was quickly adopted in the research on IPV (Coetzee et al., 2017; Logie et al., 2019; Okumu et al., 2021).

The research hypotheses to be verified integrate the framework of the ecological model with the theory of polyvictimisation. Indeed, IPV can be considered as the combination of physical, emotional, and sexual abuse, but its drivers also operate on three different levels: societal, community, and individual. These levels are presented as latent dimensions and they are part of a measurement model within a SEM.

Using a model-based approach to study IPV allows for the evaluation not only of how each kind of violence contributes to violence but also for the consideration of how each level in the ecological model is associated with IPV

itself, thus identifying the relationship between all the dimensions involved.

The societal level. Women’s empowerment should be connected – as stated by the ameliorative hypothesis – to decreased victimisation, given that gender equality is supposed to improve living conditions overall (Heirigs & Moore, 2017). On the other hand, according to the backlash hypothesis, if female empowerment is perceived as a threat, a pre-existing patriarchal society may respond with increased levels of violence (Classen et al., 2005; Levinson, 1989; Meinck et al., 2015). Women’s participation in decisions concerning the household is used here as a proxy of women’s empowerment to check for factors regarding the societal context in which violence is perpetrated. Accordingly, the present work puts forward the following hypothesis:

H1. Empowerment is associated with intimate partner violence.

The community level. Poverty, as well as living in rural areas, has been deemed to facilitate violence (Deyessa et al., 2010; Jeyaseelan et al., 2007). Indeed, in a study by Sedziafa et al. (2017), economic dependency on a partner left women more exposed to sexual violence. Moreover, general acceptance of violence within couples itself facilitates the perpetration of violence (Abrahams et al., 2006; Gage, 2005; Lawoko, 2006; Sunmola et al., 2019). Bamiwuye and Odimegwu (2014) showed that the context plays a role in determining how household wealth is associated with experiences of violence. Logie et al. (2019) show the peculiar connections between household wealth, education, labour participation and IPV: in this particular study, lower wealth is positively associated with violence, with violence less prevalent in richer households.

However, global statistics on IPV show that the socio-economic background of victims and perpetrators does not matter, since it can be found in every stratum of the population (United Nations, 2022). Indeed, other studies show that the relationship between the socio-economic status and IPV is uncertain (Abramsky et al., 2019; Ince-Yenilmez, 2022; Kilgallen et al., 2022).

Here socio-economic deprivation in the household will be used to assess the community-level dynamics of IPV. Despite the contrasting literature on the matter, the hypothesis is formulated as:

H2. Socio-economic deprivation exposes to IPV.

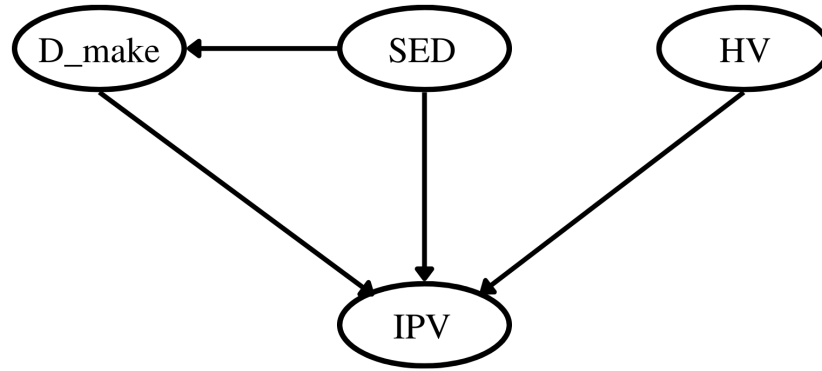
Relationship and individual factors. Gender dynamics within couples as well as disparities in educational attainment – especially if women are more highly educated than men – make women more vulnerable to violence (Deyessa et al., 2010). Individual factors point to both the socio-economic characteristics of the victim and their personal history of abuse. People with a higher education are less likely to either perpetrate or suffer IPV (Abrahams et al., 2006; Kishor & Johnson, 2006). Revictimisation processes and the intergenerational transmission of violence have also both been found to be indicators of domestic violence later in life. The intergenerational transmission of violence – defined as witnessing parental violence – makes women more vulnerable to violence later in life and men much more likely to become perpetrators themselves. This is due to the assimilation processes of acceptability and perpetuation of these same patterns of violence (Jeyaseelan et al., 2007; Kalmuss, 1984; Kishor, 2004). Revictimisation processes also matter: indeed, abused children are more likely to enter into abusive relationships as adults, and women that have been subjected to sexual violence are more vulnerable to further incidences of this in the future (Classen et al., 2005).

In this framework, socio-economic deprivation covers both community and relationship and individual factors; respondents' history of experienced violence represents both relationship and individual factors from the ecological model. Consequently – regarding to the last factors examined – we hypothesise:

H3. Socio-economic deprivation hinders women's empowerment within couples.

H4. A positive association between exposure to violence and current experiences of abuse.

Figure 1.1: Conceptual model



Given the hypotheses formulated above, the conceptual model can be articulated between the four latent dimensions. A graphical representation of this model can be found in Figure 1.1.

1.3 Data and Methods

1.3.1 Data and Sample Selection

Intimate partner violence is investigated using the Demographic and Health Survey (DHS), which was created by the United States Agency for International Development in 1984. The Agency has provided assistance to more than 350 surveys in over 90 Countries, implemented in overlapping five-year phases. The aim of the DHS Program is to improve the collection, analysis, and dissemination of population and health data. Model questionnaires – harmonised across countries and phases of collection - gather data on marriage, fertility, mortality, family planning, reproductive health, nutrition, and HIV/AIDS. The focus of the survey is women of reproductive age (15-49).

Women eligible for an individual interview are identified through the sampled households.

Some DHS surveys also involve a special module on IPV, and in that case, ever-partnered women in heterosexual relationships – already involved in the main survey - are sampled at random to collect information on their experiences of gender-based violence, both with their current partner and potential past instances of abuse (Croft et al., 2018).

All the surveys that carried the domestic violence module in the five years between 2016 and 2020 were chosen for this study. If a country appeared on this list more than once, the latest survey was selected. This work focuses on eleven surveys in Sub-Saharan Africa: Benin 2017-2018, Burundi 2016-2017, Cameroon 2018, Gambia 2019-2020, Liberia 2019-2020, Mali 2018, Nigeria 2018, Rwanda 2019-2020, Sierra Leone 2019, Uganda 2016, Zambia 2018. Though Senegal 2019 and South Africa 2016 were also initially part of this analysis, they were excluded due to the considerable amount of missing data on crucial variables.

So the sample used in this analysis consists of 43,106 currently partnered women, with an average age of 31 years ($SD = 8.12$).

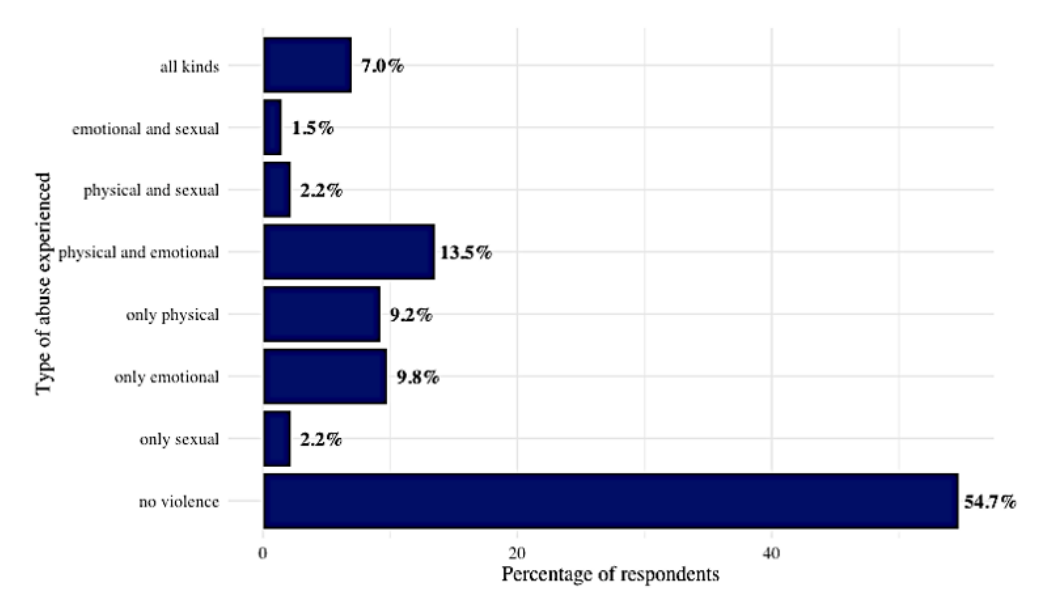
IPV is primarily assessed via three binary indicators referred to as actions perpetrated by the respondent's partner:

- “Physical Violence”, referring to being pushed, shaken, slapped, punched, or threatened at gunpoint;
- “Emotional Violence”, involving humiliation, threats of physical harm or insults;
- “Sexual Violence”, indicating forced sexual acts.

The country prevalence of women who had been subjected to at least an act of domestic abuse in the twelve months prior to the interview varies, from 35.8% in Nigeria to 59.6% in Sierra Leone (see Table 1).

Figure 1.2 shows the percentage of women who have experienced any form of the three types of IPV by their partner in their country of residence. More than 45% of all respondents have experienced at least one form of abuse by

Figure 1.2: Percentage of respondents to have experienced each type of abuse



their partner, and while marital rape seems to be the least common form of violence, 7% of the respondents have been subjected to all three types of violence and 17.2% have suffered two.

1.3.2 Methods

This work estimates a composite indicator of IPV using a Structural Equation Model (Muthén, 1984) , based on four latent variables, as already shown in Figure 1.1.

The Measurement Model

Latent variables in SEMs are specified as equations in the measurement models, here with a reflective approach.

The literature widely discusses some concerns with models including latent variables: first, factor indeterminacy, due to the possibility that individual factor scores may be generated by an infinite number of models (Grice, 2001; Kline, 2023); second, the scale of the latent variables. Setting either

one indicator in each latent dimension or its variance equal to 1 are addressed as possible solutions (Ramlall, 2016); in this work, the first option was chosen and the variable set as constraint is specified for each latent dimension. These are:

- Socio-economic deprivation (SED), which considers both personal and contextual characteristics. The exogenous manifest variables in SED are all dichotomous, for which 1 indicates: low wealth (set as constraint), for those in the two lowest quintiles of wealth; whether the respondent lives in a rural area (versus urban); whether the respondent and her husband have none to primary education. The equation is written as:

$$SED = \lambda_{x_1} Low.wealth + \lambda_{x_2} Rural.area + \lambda_{x_3} Husband.low.edu + \lambda_{x_4} Wife.low.edu + \delta_1^2$$

- History of violence (HV), concerning information about the respondent's (self-reported) past experiences with abuse. Its exogenous manifest variables are: intergenerational transmission of IPV (set as constraint), dichotomous, which takes 1 if the respondent's father used to physically abuse her mother; previous sexual violence, taking 1 if the respondent has ever been sexually abused by anyone other than the current partner; number of abusers in life, a numerical variable used to assess how many (if any) have abused the respondent excluding her current partner; number of control issues exercised by the partner; number of justifications respondents give for physical violence. The following expression is yielded:

$$HV = \lambda_{x_5} Violent.father + \lambda_{x_6} Previous.rape + \lambda_{x_7} No.abusers + \lambda_{x_8} No.Control.issues + \lambda_{x_9} No.Wife.beating + \delta_2$$

² $\lambda_{x_j}, j = 1, \dots, 9$ are the loadings for the exogenous manifest variables; $\delta_j, j = 1, 2$ are the error terms for the exogenous latent variables.

- Empowerment (DMake), which assesses women's participation in decisions concerning their households. All items take value 1 if the respondent is either the main decision maker in each matter or at least has some say in it. Four different endogenous items are included in this dimension: the respondent's health care (set as constraint); large household purchases; visits to friends and family; husband's earnings. This latent variable's equation assumes the form:

$$DMake = \lambda_{y4} Health.care + \lambda_{y5} HH.purchase + \lambda_{y6} Family.visit + \lambda_{y7} Partner.s.earning + \epsilon_2$$

- IPV, that is assessed by an equation taking into account the presence of physical, emotional, and sexual violence by the current partner in the twelve months prior the interview. To account for the scale of the latent variable, physical violence is set to 1. IPV is written as:

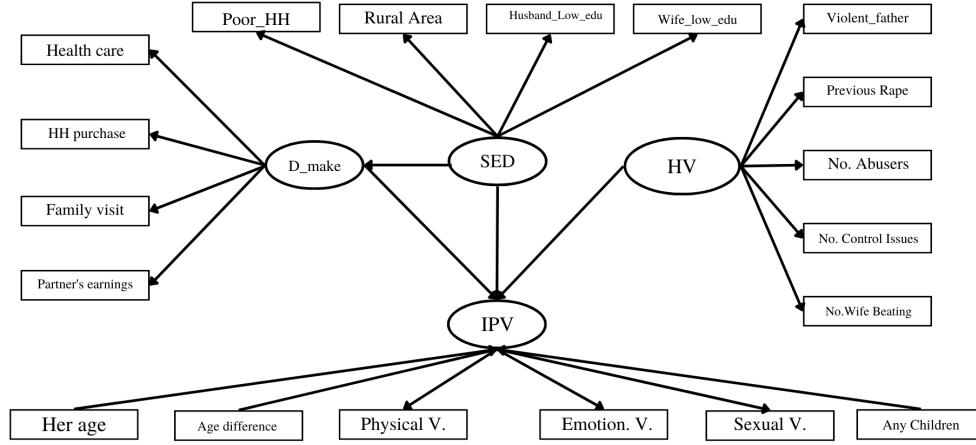
$$IPV = \lambda_{y1} PhysicalV. + \lambda_{y2} Emotional.V. + \lambda_{y3} Sexual.V. + \epsilon_1$$

To check for the internal coherence of the latent variables, each of these models is assessed using both Cronbach's alpha and the related goodness-of-fit statistics - comparative fit indices (CFI), root mean square error of approximation (RMSEA), standardised root mean square residual (SRMR).

Structural model

The structural component of SEMs estimates the association between the latent variables. The structural model equivalent to the conceptual model in Figure 1 consists of two equations. The first one investigates the association between socio-economic deprivation of respondents and women's empowerment within couples (which reflects the hypothesis H3 above formulated). The second equation assesses the effect of their history of violence, socio-economic deprivation, and women's empowerment (to verify hypotheses H1,

Figure 1.3: Path diagram



H2, H4) and other exogenous variables on intimate partner violence, i.e., age of the respondent; age difference between the respondent and her partner (computed as his age minus her age), and whether the respondent has any living children. The equations in the structural model are written as follows:

$$DMake = \gamma_1 SED + \zeta_1^3$$

$$IPV = \gamma_2 SED + \gamma_3 HV + \beta_1 DMake + w_1 Age + w_2 Age.difference + w_3 Any.children + \zeta_2$$

The IPV composite indicator

Figure 1.3 shows a schematic representation of the model.

³ β_1 is the path coefficient for the endogenous latent variable; $\gamma_j, j = 1, \dots, 3$ are the path coefficients for the exogenous latent variables, $w_j, j = 1, \dots, 3$ are the regression coefficients for the exogenous manifest variables, and $\zeta_j, j = 1, 2$ are the error terms.

Since most of the manifest variables in the measurement models are dichotomous, the usual SEM statistical assumption that observations are drawn from a continuous and multivariate normal population is not satisfied. Thus, the diagonal weighted least squares estimator (Muthén et al., 1997) with robust standard errors is chosen. The classic goodness-of-fit statistics of SEMs were then employed (CFI, RMSEA, SRMR).

Factor scores were computed using the empirical Bayes modal (EBM) approach, which is particularly suited for models including categorical variables (Bhaktha & Lechner, 2021; Rabe-Hesketh et al., 2004).

R 4.3.2 (Team, 2023) and the lavaan (version 0.6.16; (Rosseel, 2012)) package were used for the estimation of the model and the computation of the factor scores. QGIS (QGIS.org, 2023) was used to create the maps.

1.4 Results

Socio-economic deprivation, empowerment, and history of violence are three latent variables shaping IPV. Their interrelations are assumed as in Figure 1.3. In this section, the main descriptive statistics of the manifest variables in the model are shown (Table 1.1) and the reliability coefficients on measurement models and goodness of fit statistics relating to the structural model are discussed. Then hypotheses testing and post-estimation findings are presented.

1.4.1 Measurement Model Testing

All the latent variables in the measurement model use a reflective approach. The decision to use this approach is validated by the assessment of the Cronbach's α and/or by the goodness-of-fit statistics of the related CFA models, as in Table 1.2.

1.4.2 Structural Equation Model Testing

Aiming to estimate a structural equation model for IPV polyvictimisation, the first step was to look at the values of goodness-of-fit tests. The Chi-

CHAPTER 1. AN IPV MODEL-BASED COMPOSITE INDICATOR

Table 1.1: Summary statistics of the manifest variables in the SEM model by latent variables and countries

Latent variables	Manifest variables	Benin	Burundi	Cameroon	Gambia	Liberia	Mali	Nigeria	Rwanda	Sierra Leone	Uganda	Zambia
	Number of respondents	3712	6241	3665	1513	1659	2660	7662	1535	3219	5910	5326
	Victims of IPV (%)	39.70	48.50	40.30	39.60	52.80	42.30	35.80	41.10	59.60	55.00	44.90
	Age	31.10	31.40	30.60	31.50	32.70	29.90	31.30	33.50	32.10	30.00	30.80
	s.d.	8.09	7.47	8.24	8.17	8.61	8.37	8.20	7.32	8.15	8.14	8.21
	Age difference w. partner	6.85	5.52	9.03	11.80	6.10	11.10	9.56	4.31	8.83	5.96	5.81
	s.d.	6.46	6.58	7.46	7.92	6.32	7.56	6.75	6.24	7.56	5.96	4.68
IPV	Physical violence (%)	17.78	36.96	31.61	30.10	43.16	32.01	19.68	32.65	47.75	39.45	35.24
	Emotional violence (%)	35.29	22.93	25.67	25.38	39.39	34.36	32.19	29.65	45.65	39.31	27.90
	Sexual violence (%)	8.15	24.43	9.12	5.92	7.04	10.15	7.32	12.30	7.44	21.42	13.44
	Lowest two quintiles of wealth (%)	37.48	40.71	39.92	54.90	59.20	39.21	39.89	41.69	50.67	46.61	46.92
SED	Living in rural areas (%)	58.15	83.93	52.09	49.23	67.52	73.69	59.80	81.41	70.52	81.13	67.30
	Respondent has low education (%)	82.54	87.67	57.80	67.38	74.74	85.39	54.51	77.03	75.63	74.85	62.05
	Partner has low education (%)	72.50	87.21	51.98	63.78	46.57	82.78	44.62	82.92	65.44	63.33	44.98
HV	Witnessing parental violence (%)	10.04	40.94	23.80	13.88	31.21	9.65	11.87	37.50	30.32	38.80	31.03
	Rape by non-partner (%)	3.10	5.30	5.40	3.94	3.16	2.65	2.90	11.78	2.93	4.40	2.63
	Husband has control issues (%)	64.62	31.96	60.63	54.17	79.30	62.45	59.39	50.72	79.03	69.54	69.14
	Physical violence is justified (%)	31.30	60.94	29.20	58.91	44.32	76.04	27.96	51.24	53.04	50.66	50.03
D_Make	Own health (%)	48.46	71.97	56.31	51.43	79.05	19.00	47.36	81.61	43.44	73.76	80.06
	Large household purchases (%)	50.78	68.74	59.86	38.83	80.69	20.10	44.27	77.29	48.46	65.96	65.41
	Visits to family or relatives (%)	68.67	80.96	64.10	54.04	81.66	26.57	61.62	86.91	45.68	72.14	76.44
	How to spend husband's earnings (%)	27.85	63.45	46.14	21.21	74.26	12.38	30.66	75.59	38.64	50.83	64.97

Table 1.2: Assessing the dimensionality of the latent dimensions ($n = 43,106$).

	IPV	SED	HV	D_Make
Cronbach's α	0.632	0.709	0.214	0.709
CFI	1	0.976	0.924	0.999
RMSEA	0	0.142	0.032	0.039
SRMR	0	0.075	0.027	0.019

Square test returned a $p - value \approx 0$; the CFI is equal to 0.952; the RMSEA and the SRMR are respectively equal to 0.052 and 0.060. All the tests point to a good fit of the model and all the coefficients in the model have $p - value \approx 0$.

1.4.3 Hypotheses tests

Table 1.3 and Table 1.4 show results from the SEM model introduced in the Methods section. They will be presented separately for each hypothesis formulated within the framework of the ecological model.

H1. Empowerment is associated with intimate partner violence.

Women's empowerment is assessed in this model via the D_Make latent variable, expressing the participation of respondents in decision-making within their households. All the indicators contributing to this latent variable are significantly associated with D_Make: the highest standardised loading here belongs to "respondent participates in the decision concerning large house-

hold's purchases" ($St.loading = 0.931$).

In the structural model, we test the association between empowerment and polyvictimisation. While the association is significant, the standardized path coefficient is only -0.026, thus showing a negligible association.

H2. Socio-economic deprivation exposes to IPV. The latent variable SED assesses socio-economic deprivation in the respondents' households. It uses household poverty, living in rural areas, and whether the respondent and her husband have none to primary education. The indicator with the highest standardised loading is the one regarding the respondent's education ($St.loading = 0.875$).

The association between SED and IPV is negative. Once again, the standardised path coefficient is still very low (-0.040).

H3. Socio-economic deprivation hinders women's empowerment within couples. The second equation in the structural model tests the effect of socio-economic deprivation within the respondents' households on women's participation in decision-making within those same households. With a standardised coefficient of -0.234, deprivation worsens women's empowerment in their families – where generalised conditions of poverty are present, women lack agency more often.

H4. A positive association between exposure to violence and current experiences of abuse. Exposure to violence is measured in the model as the latent variable HV. The indicator with the highest standardised loading is the one concerning how many control issues the respondent's husband exercises over her ($St.loading = 0.557$). HV has the highest standardised coefficient in the equation in the structural model about IPV: this coefficient shows that one change in the standard deviation of HV causes a change equal to 0.867 in the standard deviation of IPV.

CHAPTER 1. AN IPV MODEL-BASED COMPOSITE INDICATOR

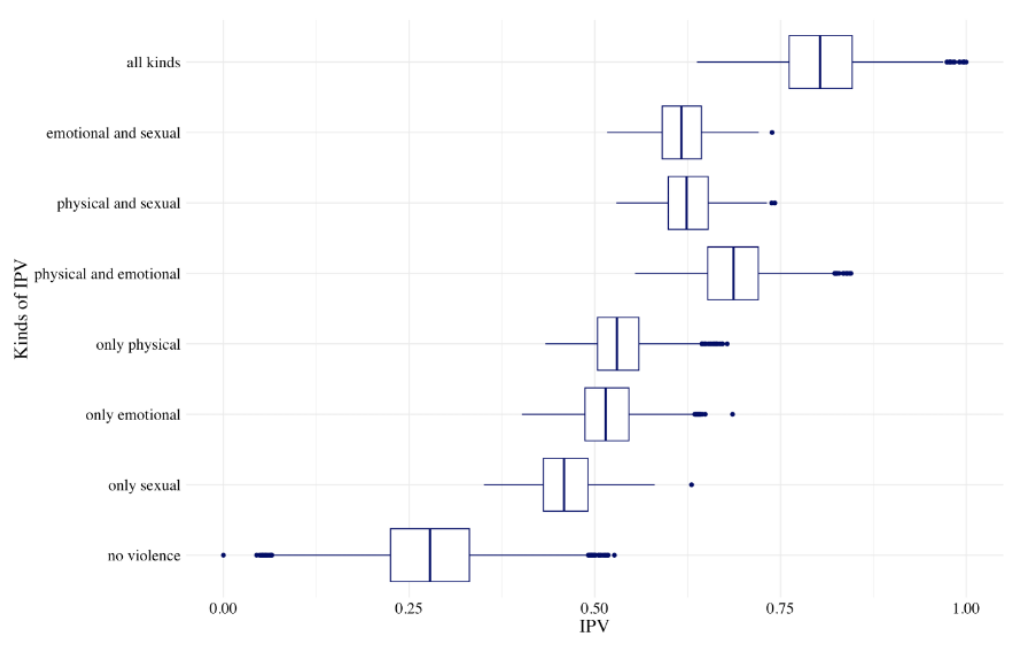
Table 1.3: Results of the overall measurement model ($n = 43,106$) by latent variables. All manifest variables contribute significantly ($p - value < 0.001$) to the latent variable

Latent variables	Manifest variables	Standardised Loading	Standardised Variance
IPV			0.242
	Physical violence	0.852	0.275
	Emotional violence	0.820	0.328
	Sexual violence	0.667	0.555
SED			1
	Lowest two quintiles of wealth	0.772	0.405
	Living in rural areas	0.732	0.464
	Respondent has low education	0.875	0.234
	Partner has low education	0.764	0.416
HV			1
	Witnessed Parental Violence	0.292	0.915
	Rape by non-partner	0.105	0.989
	Number of abusers in life	0.119	0.986
	Number of husband's control issues	0.557	0.690
	Number of justifications for physical violence against wives	0.324	0.895
D_Make			0.945
	Own health	0.912	0.168
	Large household purchases	0.931	0.134
	Visits to family or relatives	0.861	0.259
	How to spend husband's earnings	0.764	0.417

Table 1.4: Results of the structural model (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Hypothesised path	Standardised coefficient	SE
H1. Empowerment is - in one way or another - associated with IPV	-0.026***	0.007
H2. Socio-economic deprivation exposes to violence.	-0.040***	0.013
H3. Socio-economic deprivation hinders women's empowerment within couples.	-0.234***	0.008
H4. A positive association between exposure to violence and current experiences of abuse.	0.867***	0.622

Figure 1.4: Box-plots of respondents' overall IPV with respect to the type of abuse experienced.



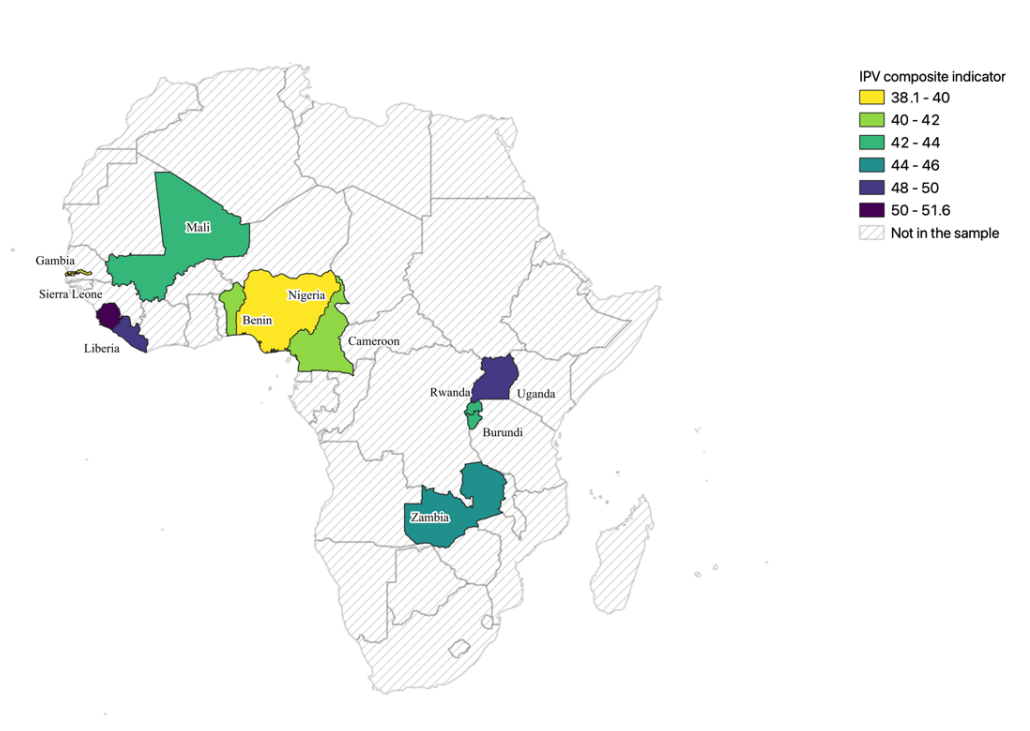
1.4.4 Post-estimation findings

Using the post-estimation results of the model, the factor scores of each statistical unit were computed using the EBM approach. These estimated scores were then normalised, thus the values are presented on a scale going from 0 (minimum levels of IPV) to 100 (maximum levels of IPV).

We check for the different levels of the IPV indicator among victims of the different types of violence. Figure 1.4 shows relevant differences in generalised levels of IPV for respondents who have been victims of all different types of IPV. Indeed, higher levels of IPV correspond to women who have experienced either all kinds of violence or two of them. This comparison – here shown graphically using box-plots – also acts as an internal validity test for the composite indicator proposed.

The IPV distribution for the respondents who declare no violence has a greater range than the others. Indeed, about 25% of non-abused respondents has an IPV value comparable to those who have experienced at least one

Figure 1.5: Map showing the IPV index.



type of violence. This can highlight that the IPV indicator may catch – via the relationships within the SEM’s structural model – a quota of hidden or undeclared violence that is not captured by self-reported answers in the survey.

Finally, the normalised factor scores were also employed to rank the countries in the sample. Using the DHS domestic violence module cluster weights, the weighted country averages were computed to identify which countries are characterised by a higher prevalence of IPV. The results show that Sierra Leone is the country with the highest average value of IPV, while Nigeria is the country with the lowest. All the results are shown in the map in Figure 1.5.

1.5 Conclusion

The aim of this research was to study IPV in African countries within the framework of the ecological model. This was done by building a composite indicator of IPV using survey data taking into account the association between domestic abuse and individual characteristics of respondents, family dynamics, and community and societal traits. A structural equation model was used to create the overall composite indicator of IPV. This indicator showed the great relevance of physical and emotional violence, given the prevalence of these two kinds of abuse among respondents: indeed, almost 20% of the respondents had experienced either one of them, while 13.5% had experienced both.

The model also points to a strong association between IPV and a history of experienced violence by victims. The latter, used to represent the personal dimension within the ecological model framework, is highly associated with the respondents' levels of overall IPV. As it has been acknowledged in the literature (Classen et al., 2005), people who have been victimised during childhood or adolescence are more likely to fall into processes of revictimisation. Additionally, this confirms how the intergenerational transmission of violence – here represented by having witnessed their father use physical violence against their mother – plays a role in whether women will be victims themselves (Jeyaseelan et al., 2007; Kabeer, 2005; Kishor, 2004).

While women's empowerment is generally acknowledged to improve women's conditions overall (Heirigs & Moore, 2017), in this analysis, higher levels of empowerment - defined as women's decision-making power in their households - do not equal high changes in the abuse received. Neither relationship nor community-level characteristics seem to have a definite association with intimate partner violence.

The IPV indicator – through the relationships in the structural model – seems to catch undeclared cases of violence, otherwise unobservable. Moreover, as for the countries considered in this analysis, the data suggest that Sierra Leone, Liberia and Uganda in particular require major interventions to fight this phenomenon. Usual practices of women's education are next

to futile in this context. Rather, men themselves need to be addressed to facilitate a more proactive fight against IPV.

Therefore, this research provides some innovative ideas. In the framework of the literature on intimate partner violence, this work moves to integrate the study of polyvictimisation within the framework of the ecological model of intimate partner violence.

Then, the estimation of a composite indicator of IPV measurable for each statistical unit that does not arise from aggregated data as with most composite measures. Moreover, considering the whole set of dimensions associated with IPV could bring forth cases of abuse that self-reported data fails to account for.

The indicator also provides the possibility of classifying individuals and ranking countries by computing summary statistics of their scores – thus identifying “where” to invest the most to mitigate the phenomenon at hand in countries at greatest risk. Specifically, given that the model uses a set of explanatory variables, it can also guide policy makers toward specific areas of investment as they grapple with intimate partner violence at the community-level.

These aspects are relevant not only in relation to the countries examined in this study but it may also be used to extend this analysis to developed countries as well, where intimate partner violence is just as problematic.

Despite its contribution to the topic, this study is not free from limitations. The ecological model constitutes the framework in which all the variables are set, thus we should consider community-level characteristics and/or local statistics to evaluate empowerment and the socio-economic context of respondents. Instead, these two dimensions are still evaluated on individuals, which could underestimate the role of neighbourhood dynamics and country-level aspects.

Data on intimate partner violence also carry severe limitations. While the DHS is particularly careful in the administration of the domestic violence module – making sure no other person is present in the room when women are interviewed – this kind of data is subject to social desirability, which reduces their general reliability. Here again, it is worth drawing atten-

tion to the methodological issue of factor indeterminacy in SEMs. Although some answers have been proposed, the literature has not yet come to an unequivocal solution (Grice, 2001; Kline, 2023).

But there is also potential for future development. A possible avenue could involve the implementation of structural equation models with spatial data to account for two different elements: second-level predictors, since the data are collected in different countries, each with peculiar characteristics; taking advantage of the ecological model at its fullest, thus, for example, evaluating women’s empowerment within a country and community-level characteristics using macro measures available from different data sources.

References

- Abrahams, N., Jewkes, R. K., Laubscher, R., & Hoffman, M. (2006). Intimate partner violence: Prevalence and risk factors for men in Cape Town, South Africa. *Violence and Victims*, *21*(2), 247–264.
- Abramsky, T., Lees, S., Stöckl, H., Harvey, S., Kapiga, I., Ranganathan, M., Mshana, G., & Kapiga, S. (2019). Women’s income and risk of intimate partner violence: Secondary findings from the MAISHA cluster randomised trial in North-Western Tanzania. *BMC Public Health*, *19*(1), 1108. <https://doi.org/10.1186/s12889-019-7454-1>
- Al-Modallal, H., Mudallal, R., Abujilban, S., Hamaideh, S., & Mrayan, L. (2020). Physical violence in college women: Psychometric evaluation of the safe dates-physical violence victimization scale. *Health Care for Women International*, *41*(8), 949–964.
- Bamiwuye, S. O., & Odimegwu, C. (2014). Spousal violence in Sub-Saharan Africa: Does household poverty-wealth matter? *Reproductive Health*, *11*(1), 1–10.
- Bhaktha, N., & Lechner, C. M. (2021). To score or not to score? a simulation study on the performance of test scores, plausible values, and SEM, in regression with socio-emotional skill or personality scales as predictors. *Frontiers in Psychology*, *12*, 679481.

- Bloom, S. S. (2008). Violence against women and girls: A compendium of monitoring and evaluation indicators. <https://www.measureevaluation.org/publications/pdf/ms-08-30.pdf.html>
- Center, W. (2021). Infographic: A global look at femicide [Accessed 2024-01-30]. <https://www.wilsoncenter.org/article/infographic-global-look-femicide>
- Centers for Disease Control and Prevention. (2015). The social-ecological model: A framework for prevention [Accessed 2024-01-30]. <https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html>
- Centers for Disease Control and Prevention. (2023). Intimate partner violence: Fast facts [Accessed: 2024-10-06]. <https://www.cdc.gov/violenceprevention/intimatepartnerviolence>
- Classen, C. C., Palesh, O. G., & Aggarwal, R. (2005). Sexual revictimization: A review of the empirical literature. *Trauma, Violence, & Abuse*, 6, 103–129.
- Coetzee, J., Gray, G. E., & Jewkes, R. K. (2017). Prevalence and patterns of victimization and polyvictimization among female sex workers in Soweto, a South African township: A cross-sectional, respondent-driven sampling study. *Global Health Action*, 10(1), 1403815.
- Croft, T. N., Marshall, A. M. J., Allen, C. K., et al. (2018). *Guide to DHS statistics* (tech. rep.). ICF. Rockville, Maryland, USA.
- Deyessa, N., Berhane, Y., Ellsberg, M., Emmelin, M., Kullgren, G., & Högborg, U. (2010). Violence against women in relation to literacy and area of residence in Ethiopia. *Global Health Action*, 3(1), 2070.
- Finkelhor, D., Ormrod, R. K., & Turner, H. A. (2007). Poly-victimization: A neglected component in child victimization. *Child Abuse & Neglect*, 31(1), 7–26.
- Ford-Gilboe, M., Wathen, C. N., Varcoe, C., MacMillan, H. L., Scott-Storey, K., Mantler, T., et al. (2016). Development of a brief measure of intimate partner violence experiences: The Composite Abuse Scale (revised)—short form (CASR-SF). *BMJ Open*, 6(12), e012824.
- Gage, A. J. (2005). Women’s experience of intimate partner violence in Haiti. *Social Science & Medicine*, 61(2), 343–364.

- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4), 430–450.
- Heirigs, M., & Moore, M. (2017). Gender inequality and homicide: A cross-national examination. *International Journal Of Comparative And Applied Criminal Justice*, 42(4), 273–285. <https://doi.org/10.1080/01924036.2017.1322112>
- Ince-Yenilmez, M. (2022). The role of socioeconomic factors on women’s risk of being exposed to intimate partner violence. *Journal of Interpersonal Violence*, 37(9-10), NP6084–NP6111. <https://doi.org/10.1177/0886260520966668>
- Jeyaseelan, L., Kumar, S., Neelakantan, N., Peedicayil, A., Pillai, R., & Duvvury, N. (2007). Physical spousal violence against women in India: Some risk factors. *Journal of Biosocial Science*, 39(5), 657–670.
- Kabeer, N. (2005). Gender equality and women’s empowerment: A critical analysis of the third millennium development goal 1. *Gender & Development*, 13(1), 13–24.
- Kalmuss, D. (1984). The intergenerational transmission of marital aggression. *Journal of Marriage and the Family*, 11–19.
- Kilgallen, J. A., Schaffnit, S. B., Kumogola, Y., Galura, A., Urassa, M., & Lawson, D. W. (2022). Positive correlation between women’s status and intimate partner violence suggests violence backlash in Mwanza, Tanzania. *Journal of Interpersonal Violence*, 37(21-22), NP20331–NP20360. <https://doi.org/10.1177/08862605211050095>
- Kishor, S. (2004). Profiling domestic violence: A multi-country study [Accessed 2024-01-30]. <https://dhsprogram.com/pubs/pdf/od31/od31.pdf>
- Kishor, S., & Johnson, K. (2006). Reproductive health and domestic violence: Are the poorest women uniquely disadvantaged? *Demography*, 43(2), 293–307.
- Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford Publications.

- Lawoko, S. (2006). Factors associated with attitudes toward intimate partner violence: A study of women in Zambia. *Violence and Victims*, 21(5), 645–656.
- Levinson, D. (1989). Family violence in cross-cultural perspective. In *Family violence in cross-cultural perspective* (pp. 435–455). Springer US.
- Logie, C. H., Okumu, M., Mwima, S., Hakiza, R., Irungi, K. P., Kyambadde, P., et al. (2019). Social ecological factors associated with experiencing violence among urban refugee and displaced adolescent girls and young women in informal settlements in Kampala, Uganda: A cross-sectional study. *Conflict and Health*, 13(1), 1–15.
- Meinck, F., Cluver, L. D., Boyes, M. E., & Mhlongo, E. L. (2015). Risk and protective factors for physical and sexual abuse of children and adolescents in Africa: A review and implications for practice. *Trauma, Violence, & Abuse*, 16, 81–107.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132.
- Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* (tech. rep.). Unpublished technical report.
- Okumu, M., Orwenyo, E., Nyoni, T., Mengo, C., Steiner, J. J., & Tonui, B. C. (2021). Socioeconomic factors and patterns of intimate partner violence among ever-married women in Uganda: Pathways and actions for multicomponent violence prevention strategies. *Journal of Interpersonal Violence*, 08862605211021976.
- Oyediran, K. A., & Feyisetan, B. (2017). Prevalence and contextual determinants of intimate partner violence in Nigeria. *African Population Studies*, 31(1).
- QGIS.org. (2023). QGIS geographic information system [Retrieved from <http://www.qgis.org>]. <http://www.qgis.org>
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69, 167–190.

- Ramlall, I. (2016). Drawbacks of sem. In *Applied structural equation modelling for researchers and practitioners* (pp. 19–20). Emerald Group Publishing Limited.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <http://www.jstatsoft.org/v48/i02/>
- Sedziafa, A. P., Tenkorang, E. Y., Owusu, A. Y., & Sano, Y. (2017). Women’s experiences of intimate partner economic abuse in the eastern region of ghana. *Journal of Family Issues*, 38(18), 2620–2641.
- Sunmola, A. M., Mayungbo, O. A., Ashefor, G. A., & Morakinyo, L. A. (2019). Does relation between women’s justification of wife beating and intimate partner violence differ in context of husband’s controlling attitudes in Nigeria? *Journal of Family Issues*, 41(1), 85–108.
- Team, R. C. (2023). R: A language and environment for statistical computing. <https://www.R-project.org/>
- United Nations. (1993). Declaration on the elimination of violence against women [Accessed 2024-01-30]. https://www.un.org/en/genocideprevention/documents/atrocities-crimes/Doc.21_declaration%20elimination%20vaw.pdf
- United Nations. (2022). What is domestic abuse? [Accessed 2024-01-30]. <https://www.un.org/en/coronavirus/what-is-domestic-abuse>
- Women Against Violence Europe. (2019). Wave country report 2019 [Accessed 2024-01-30]. <https://wave-network.org/wave-country-report-2019/>
- Women Against Violence Europe. (2021). Wave country report 2021 [Accessed 2024-01-30]. <https://wave-network.org/wave-country-report-2021/>
- World Health Organization. (2012). Understanding and addressing violence against women: Intimate partner violence [Accessed 2024-01-30]. <https://apps.who.int/iris/handle/10665/77432>
- World Health Organization. (2021). *Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring: Recommen-*

dations for a public health approach [Accessed: 2024-10-06]. <https://www.who.int/publications/i/item/9789240022256>

Chapter 2

HIV Prevalence Determinants within Neighbourhoods

Does neighborhood matter? An analysis of HIV prevalence in Sub-Saharan African countries¹

Co-authors: Anna Maria Parroco, Daria Mendola

Abstract

Over two-thirds of the population living with HIV were concentrated in Eastern, Southern, Western, and Central Africa in 2021. This paper employs data from the Demographic and Health Survey to assess the relationship between HIV prevalence and its socio-economic and demographic drivers at the neighbourhood (macro-cluster) level. Additionally, the study examines the existence of differences in such relationships among countries.

The results of the fractional logistic regression models highlight that highly educated neighbourhoods are less likely to be affected by HIV. A greater average number of children, potentially due to programs that promote access to preventive antenatal care and prevention, is associated with

¹ Authors' Original Manuscript (AOM). Article published on Biodemography and Social Biology.

a lower likelihood of residents living with HIV.

Notably, HIV testing coverage is prevalent in neighbourhoods with a high prevalence of HIV. It is also evident that there are notable differences between countries, which demonstrate national context plays a crucial role in the association between education, number of children, testing coverage, and HIV prevalences.

Keywords

HIV; fractional logistic regression model; HIV determinants; neighbourhoods; Demographic and Health Survey.

2.1 Introduction

The diffusion of HIV in Africa is impressive; indeed, the twenty countries with the highest prevalence in the world can be found in the Sub-Saharan region, ranging from 27.9% in Eswatini to 1.9% in Côte d'Ivoire (UNAIDS, 2024c). Although the numbers of new infections and deaths due to HIV are declining, 20.6 million people in Eastern and Southern Africa and 5 million people in Western and Central Africa were living with HIV in 2021 (UNAIDS, 2024c).

HIV is a long-standing epidemiological concern, but it also has a deep effect on the demographic structure of countries (mostly in terms of fertility and mortality), affecting the reproductive behaviours of neighbourhoods and their well-being and hampering their future economic development, particularly among the active age strata of the population. Because of these transversal effects, actions to prevent HIV are included in ten of the seventeen Sustainable Development Goals of the United Nations (UNAIDS, 2024a).

Most studies adopt a national lens to explain HIV prevalence or focus on individuals to assess factors predicting infections. Instead, in this study, a different point of view is adopted by assuming neighbourhoods as the unit of analysis. Neighbourhoods are identified as geographical units, accounting for

physical proximity among individuals. This geographical aggregation allows us to assess how socio-economic and behavioural characteristics that shape the place where people live may affect HIV diffusion (Gillespie et al., 2007).

Hence, using a large dataset from the Demographic and Health Survey covering 11 countries in Sub-Saharan Africa, this paper aims to assess how the spread of HIV is associated with neighbourhood characteristics and to evaluate differences between countries concerning possible relevant covariates that shape HIV diffusion within those neighbourhoods.

The paper is structured as follows: Section 2.2 draws attention to the relevance of the adoption of a neighbourhood lens and introduces the main social factors that the scientific literature has highlighted as characteristics associated with HIV. Section 2.3 describes the data and methods, presents the sample data and introduces the fractional logistic regression model used to estimate the association between HIV and the selected variables. Section 2.4 concerns the results of the analyses. Section 2.5 concludes and acknowledges some limitations of this study.

2.2 HIV Risk Factors within Neighbourhoods

The contextualization of risk factors is relevant to the analysis of disease (Gillespie et al., 2007; Link & Phelan, 1995). In particular, when studying HIV, it has been suggested that neighbourhood or community-specific measures should be used since local hotspots may not be identified when the prevalence decreases at the national level (Cuadros & Abu-Raddad, 2014). The effect of socioeconomic characteristics (such as education and wealth) can vary across countries due to their interaction with local factors. Therefore, these characteristics must be analysed within their specific contexts (Gillespie et al., 2007).

The scientific literature often focused on the association between education and HIV, with its prevalence being greater among less educated people (Fortson, 2008; Hargreaves et al., 2008), and Levinsohn et al. (2013) showed that in South Africa, cohorts that are more affected by unemployment are also more affected by HIV. Furthermore, while HIV is more prevalent in

poorer countries, the richest segments of the population are more impacted than other strata of the population (Mishra et al., 2007), even though the association is not always clear. Fox (2012) also highlights that in richer areas of Sub-Saharan Africa, poorer people are more likely to be infected, while the richest individuals are more affected in poorer regions.

Living in rural areas has also been found to be negatively associated with HIV. Tanser et al. (2009) showed that, in South Africa, rural areas are characterised by a lower prevalence of HIV; however, this may conceal the presence of hotspots due to the lower population density in these same areas.

Research indicates that individuals in stable heterosexual relationships are significantly impacted by HIV (Case et al., 2012), as are separated individuals (Nutor et al., 2020). Additionally, the literature highlights the protective effects of antenatal care and family planning against HIV (Psaros et al., 2015).

2.3 Data and methods

The data used in this analysis were drawn from the Demographic and Health Survey (DHS) Program (USAID, 2022). These are nationally representative household surveys in which harmonised questionnaires collect data on marriage, fertility, mortality, family planning, reproductive health, nutrition, and HIV/AIDS (Croft et al., 2018). In addition to the standard modules, we also relied on the HIV datasets² and the geographic datasets of DHS³.

The sample consisted of 148,620 people aged 15-64 years who were residents of Angola, Burundi, Cameroon, Chad, Guinea, Malawi, Mozambique, Senegal, Sierra Leone, South Africa, and Zimbabwe and were selected due to the availability of both surveys with HIV tests and the geolocalization of survey units in a period ranging from 2015 to 2019. For each country, we choose the most recent survey. Despite accounting for less than 3% of the global

²The HIV datasets contain biomarker data and results of ELISA (enzyme-linked immunosorbent assay) tests to determine respondents' serostatus.

³The geographic datasets contain the GPS coordinates of the centroids of the clusters of households involved in the survey.

population in 2021 (UN 2022), the above-mentioned countries represent almost 29% of all people living with HIV worldwide⁴. According to (UNAIDS, 2024b), the HIV prevalence in adults aged 15-49 years ranges from 0.3% in Senegal to 18.3% in South Africa (see Figure 2.1).

In DHS surveys, households are grouped into so-called “clusters”⁵. Across the eleven selected countries, there are 5,960 clusters with sizes ranging from 2 to 113 individuals. However, to increase the reliability of the aggregated data and maintain the analysis at the subregional level, the original DHS clusters were combined into “macro-clusters” (henceforth referred to as “neighbourhoods” or “territorial units”), which have larger population sizes than DHS clusters. Neighbourhoods do not align with specific administrative divisions but are segments of national territories that encompass geographically proximate individuals. This perspective is valuable for studying the spread of HIV, as residents in these areas are more likely to have sexual contact with one another.

In particular, neighbourhoods were operationalized by using a spatially aware k-means clustering technique (Alexandrov & Kobarg, 2011; QGIS Development Team, 2009, 2020). Hence, clusters are aggregated based on their geographical proximity, which is determined by their centroid’s geographical coordinates⁶ and the constraints of country borders. We chose to have 50 macro-clusters (i.e., neighbourhoods) for each country to allow for a reliable estimation of the country’s effect.

The final sample is composed of 498 macro-clusters⁷ in the eleven countries (average population = 224.09; $sd = 139.53$). One-way ANOVA was performed to evaluate the differences between countries in terms of HIV prevalence at the macro-cluster level, which was statistically significant ($F =$

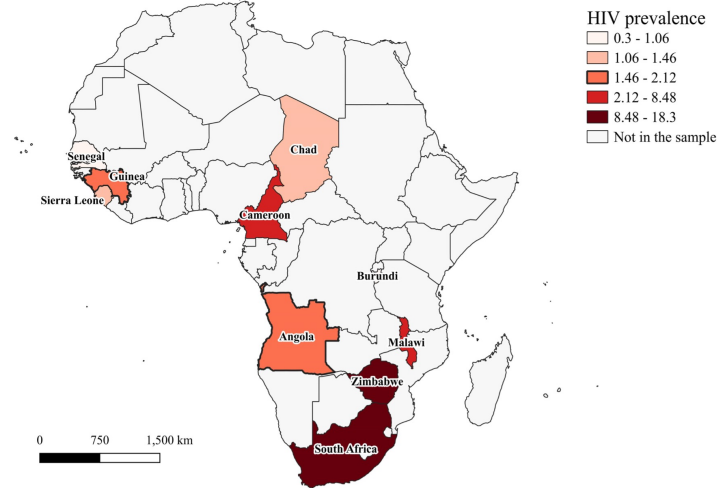
⁴This statistic must be considered as an underestimation since no data about HIV is available in official statistics for Mozambique (Baffes et al., 2022; UNAIDS, 2024c).

⁵Clusters in the DHS are census units. For further details see Croft et al. (2018).

⁶Centroids are defined as single points representing the barycentre of a geographical area. Since, in this case, clusters represent groups of households, their GPS coordinates refer to the location of the centroid of each territorial area (i.e., cluster). Hence, macro-clusters have the barycentre of the clustered geographical coordinates as their mean point.

⁷Macro-clusters with missing data and those in the top and bottom 5% for population size were dropped from the analysis.

Figure 2.1: HIV prevalence in people aged 15-49 years in Africa in 2021



Authors' own elaboration on UNAIDS (2024b) data

98.86; $p - value < 0.001$).

Therefore, the variables of interest for this study are listed in Table 2.2. These variables were measured at the territorial level to assess the meso-level drivers of HIV spread. Numeric variables were averaged, while dichotomous variables were presented as proportions. For ease of interpretation, these variables were grouped into three domains:

- *Socio-economic status of the neighbourhood*: number of years of education; proportion of people in the middle-to-high quintiles of household wealth; employed respondents; proportion of people living in rural areas;
- *Demographic characteristics of the neighbourhood*: age of respondents; number of children of respondents (both ever-born and alive); people who never married; people in monogamous or polygamous relationships;

- *Sexual behaviours*: never in a sexual relationship; age at first intercourse (as before or after sixteen years of age); people who used a condom during their most recent intercourse; people who had been tested at least once for HIV;
- Country dummies.

2.3.1 The fractional logistic regression models

The prevalence of HIV in macro-clusters was the response variable (y) of the statistical model in this study. It ranges from 0-3.73% in Burundi's macro-clusters to 0.19-35.43% in Mozambique's macro-clusters, with a general average of 5.41% ($s.d. = 7.10$). Figure 2.2 shows the territorial distribution of HIV (among people aged 15-49) in each selected Sub-Saharan African country. Since we deal with a limited dependent variable in the interval $[0, 1]$, even if empirically it spans in $[0, 1)$, a fractional logistic regression model (FLRM) (Papke & Wooldridge, 1996) was implemented to evaluate the association of HIV prevalence with the selected covariates. The FLRM was preferred to the Beta regression model because of the presence of several macro-clusters in which the HIV prevalence was zero (see Figure 2.2).

With $\mathbf{X}$ being the matrix of the covariates x_j , and $y \in [0, 1]$, the conditional expected value $E(y|\mathbf{X})$ is thus modelled as a logistic function:

$$E(y | \mathbf{X}) = \frac{\exp(\mathbf{X}'\boldsymbol{\beta})}{[1 + \exp(\mathbf{X}'\boldsymbol{\beta})]}$$

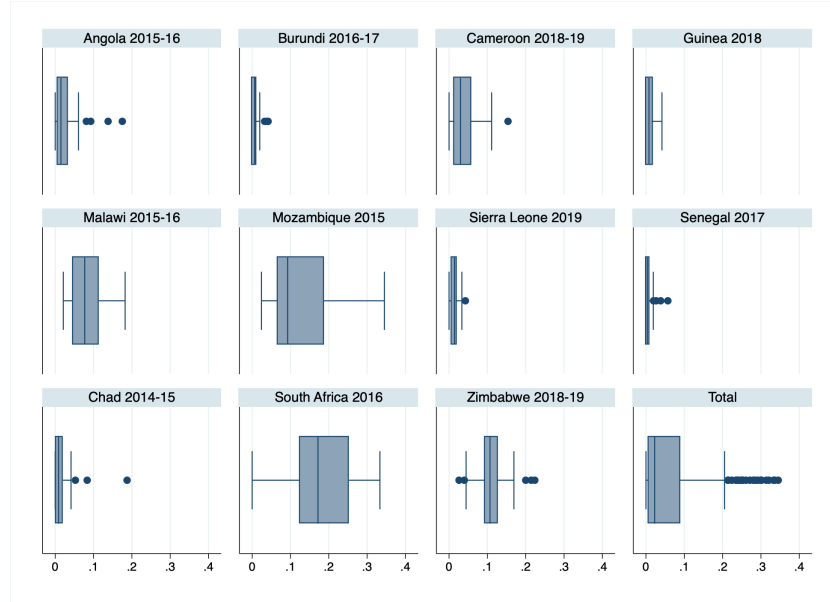
The approach used to obtain robust estimates of $\boldsymbol{\beta}$ is the quasi-maximum likelihood estimation method. This model thus ensures that the predicted values of y are in $[0, 1]$ and that the effect of any x_j on $E(y|X)$ diminishes as $\mathbf{X}\boldsymbol{\beta} \rightarrow \infty$.

Using the logit transformation, the log-likelihood function for this model is:

$$\log L = \sum_{j=1}^N y_j \log \left\{ \frac{\exp(\mathbf{X}'\boldsymbol{\beta})}{[1 + \exp(\mathbf{X}'\boldsymbol{\beta})]} \right\} + (1 - y_j) \log \left\{ 1 - \frac{\exp(\mathbf{X}'\boldsymbol{\beta})}{[1 + \exp(\mathbf{X}'\boldsymbol{\beta})]} \right\}$$

CHAPTER 2. HIV DETERMINANTS WITHIN NEIGHBOURHOODS

Figure 2.2: HIV prevalence in communities of Sub-Saharan Africa in 2015-2019 (people aged 15-49)



Authors' own elaborations on DHS data (most recent survey in each country)

Given that the conditional mean of HIV prevalence at the territorial level can be interpreted as the probability of being HIV positive in that particular area (Papke & Wooldridge, 1996), we used odds ratios to interpret the results. In this case, the estimated odds ratios — while still referring to these purposive territorial units—measure the association between living with HIV and the characteristics of the neighbourhoods in which respondents live.

In addition to the base model (Model 1 in Table 2.1), the inclusion of different interaction terms to detect country differences in the mechanism of diffusion of HIV was also tested. Due to the relatively low number of observations (498 neighbourhoods) in the model, a full model including all these interactions at the same time was not feasible. Therefore, a procedure by block was adopted, interacting only one covariate with the country of residence in each model. Notably, the goodness-of-fit criteria do not suggest the inclusion of interaction terms, even when highly significant, highlighting how the base model has the same predictive value as more complex models.

Nevertheless, to provide more insight into the geographical prevalence of HIV, Table 2.1 presents the three additional models, where the interaction terms are added as mentioned above, namely, between country of residence and education (model 2), number of children (model 3), and HIV testing coverage (model 4)⁸.

Finally, for ease of interpretation, average marginal effects (AME) graphs are presented (Figures 2.3, 2.4, 2.5). A marginal effect is the average effect on the conditional mean of HIV prevalence of a unit change in one of the covariates; the choice, in this case, was to compute the marginal effect at each value of the explanatory variable $\mathbf{x} = \mathbf{x}_i$ and then compute the average (Cameron & Trivedi, 2022).

2.4 Results and Discussion

Table 2.1 reports the results of the estimation of the fractional logistic regression models. As previously described, covariates account for the socio-economic status of the neighbourhood, demographic characteristics and sexual behaviours. As a general result, the theoretical framework is quite robust: the same set of neighbourhood characteristics was significantly associated with HIV in all four models, with the same direction and significance levels⁹.

Overall, significant differences emerge among countries, with most of them having higher odds of HIV infection than Angola (namely, Malawi, Mozambique, South Africa and Zimbabwe).

Focusing on the socio-economic status of the neighbourhood, highly educated neighbourhoods are less vulnerable to HIV, likely due to a greater awareness of the mechanisms of HIV diffusion (see, e.g., Baker et al. (2011), Hargreaves et al. (2008), and Sani et al. (2016)). Model 2 detects potential differences in the effect of education among countries. With the exception of Senegal, the prevalence of HIV in highly educated neighbourhoods increased by 11% (Chad) to 63% (Burundi) compared to those in Angola. For

⁸Angola, having the median HIV prevalence among the analyzed countries, was chosen as the reference in all four models.

⁹Wealth is only significant in Model 2 and Model 3.

the other covariates in this domain, neighbourhoods with higher wealth and labour market attachment do not show significant changes in the odds of living with HIV. This result is not unexpected; indeed, Gillespie et al. (2007) showed that the association between wealth and HIV is particularly context-specific and can be found to be statistically not significant when controlling for other characteristics such as age, education and sexual behaviours. The proportion of people living in rural areas was not significantly associated with living with HIV in this study, unlike the findings of previous literature (Reniers & Watkins, 2010; Tanser et al., 2009).

Analysing the demographic characteristics of the neighbourhoods, age was not significantly associated with increased odds of living with HIV, likely due to substantially low variability resulting from the DHS eligibility criteria, which only include individuals aged 15-49. Living arrangements (being never married, separated or widowed and staying in polygamous marriages) are not statistically significant, even if some of them could have predicted promiscuity. Even polygyny, which in the literature has been shown to be a protective factor against HIV diffusion (Reniers & Watkins, 2010), is not significantly different from other familial ties. The average number of children is a particularly interesting aspect to consider. A greater number of children in the neighbourhood is associated with lower odds of living with HIV, probably due to women accessing prevention education and tests during antenatal care visits, as often stated in the literature (Church & Mayhew, 2009; GNP+, 2013; Knettel et al., 2018; Psaros et al., 2015). Unfortunately, we cannot test this interpretation: cross-sectional data do not permit examination of the effects of policies and interventions, and there is no information in the data indicating whether the policy has been implemented in the neighbourhood. Model 3 shows differences among countries: compared with neighbourhoods in Angola, neighbourhoods with more children have an increase in HIV diffusion of 52% (Burundi) to nine times (South Africa).

Given that this study focused on a sexually transmitted disease, it is worth highlighting that common sexual behaviours in close neighbourhoods are key factors associated with living with HIV. As expected, people who have never engaged in sexual relationships are less likely to live with HIV,

all other things being equal. Moreover, having their first intercourse after 16 years of age was shown to reduce the odds of having a high prevalence in the area. Finally, higher HIV testing coverage is significantly associated with greater HIV diffusion, likely signalling awareness about the spread of the virus. For Model 4, it is notable that there do not seem to be many significant differences between countries: in Sierra Leone, higher testing coverage lowers the odds of living with HIV by almost 65% with respect to Angola. Moreover, greater coverage in Burundi, Malawi, and Mozambique was associated with a higher HIV prevalence than in Angola.

Figures 2.3, 2.4, and 2.5 show the average marginal effect graphs and allow us to check for possible disparities in the effect of education (using post-estimation results from Model 2), number of children (from Model 3), and HIV testing coverage (from Model 4) in Senegal, Angola and South Africa, which are respectively the countries with the lowest, median and highest HIV prevalence among the selected countries. These three variables represent the three domains of covariates introduced in Section 3.3.

Figure 2.3: The effect of the average number of years of education on HIV prevalence (AME and CI at 90%)

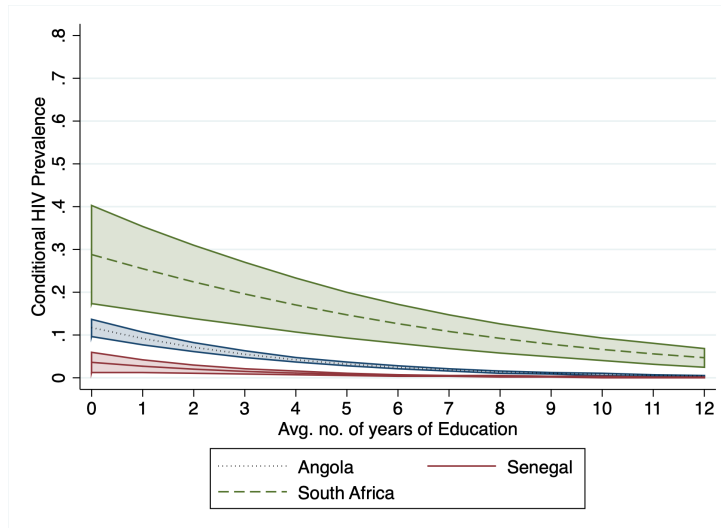


Figure 2.3 shows that when the HIV prevalence is high in a country (such as South Africa), the positive effect of an increase in the average educational

CHAPTER 2. HIV DETERMINANTS WITHIN NEIGHBOURHOODS

Table 2.1: Fractional logit models for HIV prevalence (Odds Ratio estimates)

Domains	Covariates	Base (Model 1)	Model 2	Model 3	Model 4
Socio-economic status	No. years of education (avg.)	0.893***	0.748***	0.903***	0.898***
	People in first quintiles of wealth (prop.)	1.559	1.726**	1.520**	1.539
	Employed (prop.)	0.811	0.710	0.949	0.783
	Living in rural area (prop.)	0.899	0.910	0.930	0.883
Demographic characteristics	Age (avg.)	1.036	1.036	1.047	1.041
	No. children (avg.)	0.729**	0.765*	0.321***	0.732**
	Never in a married (prop.)	9.621**	8.314*	9.554**	8.793*
	Separated (prop.)	11.056	13.915	11.012	9.207
	People in polygamous relationships (prop)	2.100	3.374	2.090	1.921
Sexual behaviors	Never in a sexual relationship (prop.)	0.373	0.338	0.376	0.324
	First intercourse at 16 or later	0.269**	0.234**	0.247***	0.244**
	Condom use at last intercourse	0.854	1.134	1.618	0.819
	HIV testing coverage (prop.)	4.904***	5.830***	3.762***	2.648***
Country dummies	Angola (reference category)				
	Burundi	0.305***	0.036***	0.095***	0.073***
	Cameroon	1.22	0.588	0.064***	1.065
	Chad	0.986	0.346***	0.198***	0.813
	Guinea	0.624***	0.140***	0.090***	0.518***
	Malawi	2.387***	1.059	0.134***	0.620**
	Mozambique	4.627***	1.719	0.600	2.748***
	Senegal	0.288***	0.267***	0.000***	0.305***
	Sierra Leone	0.583***	0.175***	0.059***	0.892
	South Africa	3.699***	3.401***	0.034***	4.928***
	Zimbabwe	2.68***	0.271***	0.510**	3.016*
Interaction education x country (only in Model 2)	Angola x education (reference)				
	Burundi x education		1.634***		
	Cameroon x education		1.168***		
	Guinea x education		1.567***		
	Malawi x education		1.178***		
	Mozambique x education		1.212***		
	Sierra Leone x education		1.310***		
	Senegal x education		0.984		
	Chad x education		1.297***		
	South Africa x education		1.106***		
	Zimbabwe x education		1.449***		
Interaction no. children x country (only in Model 3)	Angola x children (reference)				
	Burundi x children			1.522***	
	Cameroon x children			2.694***	
	Chad x children			1.881***	
	Guinea x children			1.923***	
	Malawi x children			2.643***	
	Mozambique x children			2.071***	
	Senegal x children			8.749***	
	Sierra Leone x children			2.219***	
	South Africa x children			9.088***	
	Zimbabwe x children			1.793***	
Interaction HIV testing coverage x country (only in Model 4)	Angola x HIV testing coverage (reference)				
	Burundi x HIV testing coverage				16.2***
	Cameroon x HIV testing coverage				1.611
	Chad x HIV testing coverage				1.806
	Guinea x HIV testing coverage				1.766
	Malawi x HIV testing coverage				7.949***
	Mozambique x HIV testing coverage				3.193***
	Senegal x HIV testing coverage				0.884
	Sierra Leone x HIV testing coverage				0.353***
	South Africa x HIV testing coverage				0.966
	Zimbabwe x HIV testing coverage				1.229
	Constant	0.013***	0.034**	0.088	0.015***
p-value <0.10 *; p-value <0.05 **; p-value <0.01 ***					
Goodness of Fit and Sample Size					
	AIC	201.783	201.454	200.939	203.684
	BIC	243.889	243.560	243.045	250.000
	Observations	498	498	498	498

level in neighbourhoods is more evident (i.e., the curve rapidly decreases), and there are clear differences among the three countries.

Figure 2.4: The effect of the average number of children on HIV prevalence (AME and CI at 90%)

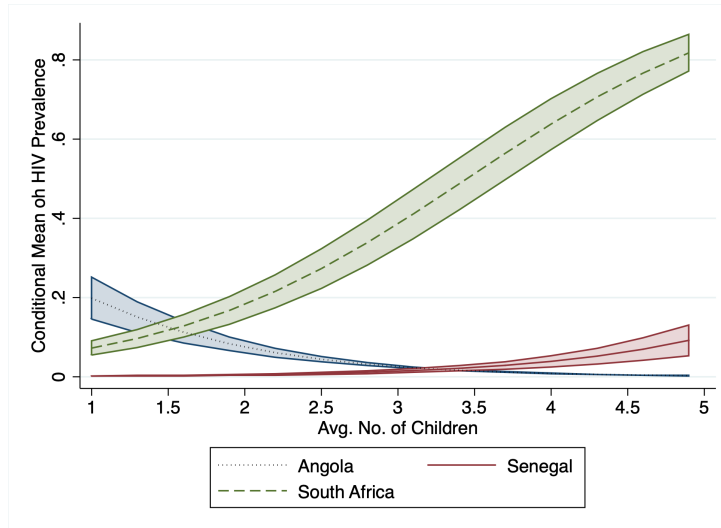


Figure 2.4 shows the association between the number of children and HIV. Indeed, there are some intersections between the curves at specific values. However, there is still a pronounced difference between South Africa and the other two countries.

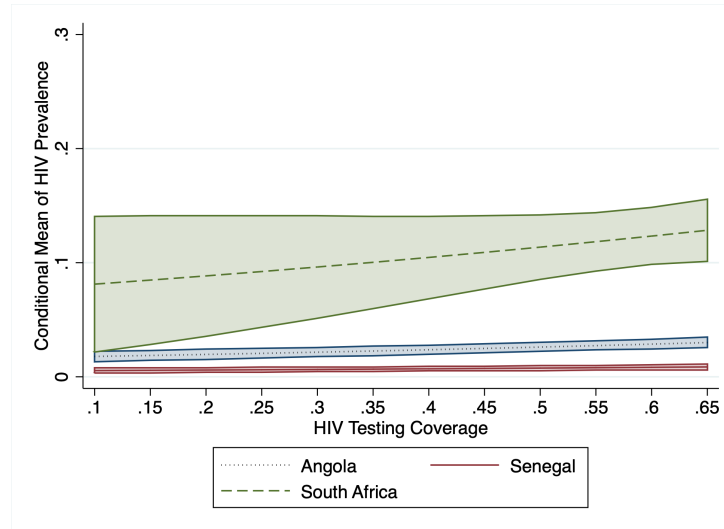
Higher HIV testing coverage increases HIV prevalence, as shown in Figure 2.5. South Africa, once again, has a higher mean prevalence of HIV at the territorial level than the other countries.

It is worth noting that in Senegal, the average effects of education, children, and protective practices on HIV are irrelevant. This is probably due to the small number of people living with HIV which makes the effect statistically undetectable.

2.5 Conclusions

This study aimed to assess the association between neighbourhood characteristics and HIV prevalence in Sub-Saharan African countries. This analysis

Figure 2.5: The effect of the proportion of people who have already tested for HIV on its prevalence (AME and CI at 90%)



Note: The scale of the conditional mean of HIV prevalence was modified for readability.

was carried out for countries that represent more than 29% of people living with HIV worldwide despite accounting for less than 3% of the global population.

The regression analyses revealed that neighbourhoods with higher levels of education exhibit lower HIV prevalence, with significant variations observed across countries. Policies that promote access to treatment therapy – not only during pregnancy but also after pregnancy – could explain why neighbourhoods with a greater number of children experience a lower HIV prevalence. The proportion of people who have never been in a sexual relationship and, in the opposite direction, HIV testing coverage were significantly associated with HIV prevalence, decreasing and increasing its diffusion, respectively. Neighbourhood characteristics thus play a relevant role in the diffusion of HIV.

This study demonstrated the impact of immediate neighbourhood context on HIV transmission, identifying specific neighbourhood characteristics that correlate with higher or lower HIV prevalence among residents. Moreover,

using pooled data over several countries – with different national-level HIV prevalence – emphasised differences between them and differences between the roles that the same set of covariates plays in the diffusion of HIV.

These findings need to be further investigated. Particularly, differences between countries – which are significant even when controlling for factors that the literature has shown to be important for HIV – remain to be explained. Furthermore, this study is not exempt from limitations. First, the data used do not provide information on treatment use or about the availability, proximity and accessibility of health facilities or medical care providers. It could be very useful to have this kind of information in the DHS survey in the future. Moreover, given the geographical lens adopted here, this information is not retrievable, not even recurring to external data sources. Second, the relationships between the variables analysed here cannot be interpreted as causal, given the cross-sectional nature of the data.

While interpreting these results requires caution, the study emphasised the importance of proximity factors in enhancing our understanding of HIV epidemics.

Appendix

Table 2.2: Characteristics of neighborhoods by country -standard deviations in parentheses

Country	Angola	Burundi	Cameroon	Chad	Guinea	Malawi	Mozambique	Sierra Leone	Senegal	South Africa	Zimbabwe
Year of interview	2015-16	2016-17	2018-19	2014-15	2018	2015-16	2015	2019	2017	2016	2018-19
Number of original clusters	621	554	430	623	400	850	306	563	400	669	544
Number of macro-cluster	43	48	44	43	48	48	48	48	44	38	46
Respondents (avg.)	207.58	288.44	195.98	203.93	155.77	278.85	179.48	221.1	220.86	250.37	263.33
HIV prevalence (%)	-149.4	-144.14	-128.19	-140.45	-84.26	-133.99	-120	-131.74	-152.28	-186.29	-123.61
Age (years, avg.)	2.52	0.8	3.54	1.27	1.07	8.19	13.13	1.4	0.57	20.16	10.84
	28.6	29.54	31.95	29.95	30.37	28.47	31.57	31.05	28.73	30.55	29.19
	-1.15	-0.8	-1.51	-1.05	-1.65	-0.76	-1.69	-1.69	-0.82	-1.76	-0.77
Years of education (avg.)	4.67	4.31	5.87	2.04	2.34	6.57	4.33	3.98	3.71	9.69	7.18
	-1.69	-1.19	-2.55	-1.72	-1.5	-1.18	-1.41	-1.53	-1.68	-0.55	-0.74
Number of Children (avg.)	3.18	2.87	3.19	4.05	3.02	2.69	3.23	3.17	2.48	1.57	2.76
	-0.56	-0.48	-0.6	-0.52	-0.66	-0.39	-0.56	-0.6	-0.5	-0.2	-0.32
Middle-to-rich quintiles of wealth (%)	37.92	62.84	50.74	62.27	50.06	64.88	57.24	46	43.47	54.82	59.58
Employed (%)	70.39	83.31	73.77	54.99	73.41	70.58	52.3	78.17	63.05	35.03	58.49
Never had a sexual relationship (%)	8.24	29.53	16.45	18.96	16.88	13	7	10.96	26.75	12.23	13.99
First intercourse under the age of 16 (%)	62.8	12.18	41.47	37.97	38.57	42.75	51.97	45.22	27.6	37.41	44.92
First intercourse at the age of 16 or over (%)	28.95	58.28	42.07	43.06	44.55	44.25	41.03	43.82	45.65	50.35	41.1
People who have. never married (%)	35.98	36.92	34.38	23.16	27.71	28.8	19.47	30.17	36.81	64.54	36.52
People in a monogamous union (%)	24.34	28.56	22.53	23.69	22.42	30.5	30.85	24.27	25.85	16.97	26.37
People in a polygamous union (%)	32.02	29.03	32.54	47.74	47.43	32.39	35.71	41.47	34.07	12.95	28.45
People who used condom during most recent intercourse (%)	10.19	4.05	16.55	2.44	6.75	19.8	10.91	4.18	9.3	48.2	18.01
People who have. tested for HIV at least once (%)	33.39	57.17	57.74	9.22	10.67	78.09	51.06	38.06	32.89	78.97	82.08
People living in rural areas (%)	55.71	86.34	60.09	82.19	80.74	82.88	69.42	76.78	64.69	47.51	62.8

References

- Alexandrov, T., & Kobarg, J. H. (2011). Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics*, 27(13), i230–i238. <https://doi.org/10.1093/bioinformatics/btr246>
- Baffes, J., Hyland, M. C., & Blankespoor, B. (2022). World Bank Open Data [Accessed February 24, 2024]. <https://data.worldbank.org/>
- Baker, D. P., Leon, J., & Collins, J. M. (2011). Facts, attitudes, and health reasoning about HIV and AIDS: Explaining the education effect on condom use among adults in Sub-Saharan Africa. *AIDS and Behavior*, 15(7), 1319–1327.
- Cameron, C., & Trivedi, P. (2022). *Microeconometrics using STATA* (2nd, Vol. 2). Stata Press.
- Case, K., Ghys, P., Gouws, E., Eaton, J., Borquez, A., Stover, J., Cuchi, P., Abu-Raddad, L., Garnett, G., & Hallett, T. (2012). Understanding the modes of transmission model of new HIV infection and its use in prevention planning. *Bulletin of the World Health Organization*, 831–838.
- Church, K., & Mayhew, S. H. (2009). Integration of STI and HIV prevention, care, and treatment into family planning services: A review of the literature. *Journal of Family Studies*, 40(3), 171–186.
- Croft, T. N., Marshall, A. M. J., Allen, C. K., et al. (2018). *Guide to DHS statistics* (tech. rep.). ICF. Rockville, Maryland, USA.
- Cuadros, D. F., & Abu-Raddad, L. J. (2014). Spatial variability in HIV prevalence declines in several countries in Sub-Saharan Africa. *Health & Place*, 28, 45–49.
- Fortson, J. G. (2008). The gradient in Sub-Saharan Africa: Socioeconomic status and HIV/AIDS. *Demography*, 45(2), 303–322.
- Fox, A. M. (2012). The HIV–poverty thesis re-examined: Poverty, wealth or inequality as a social determinant of HIV infection in Sub-Saharan Africa? *Journal of Biosocial Science*, 44(4), 459–480.

- Gillespie, S., Kadiyala, S., & Greener, R. (2007). Is poverty or wealth driving HIV transmission? *AIDS*, *21*, S5–S16.
- GNP+. (2013). Option B+: Understanding the perspective of women living with HIV [Accessed February 24, 2024]. <https://gnpplus.net/resource/option-b-understanding-the-perspective-of-women-living-with-hiv/>
- Hargreaves, J. R., Bonell, C. P., Boler, T., Boccia, D., Birdthistle, I., Fletcher, A., Pronyk, P. M., & Glynn, J. R. (2008). Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-saharan africa. *AIDS*, *22*(3), 403–414.
- Knettel, B. A., Cichowitz, C., Ngocho, J. S., Knippler, E. T., Chumba, L. N., Mmbaga, B. T., & Watt, M. H. (2018). Retention in HIV care during pregnancy and the postpartum period in the option b+ era: A systematic review and meta-analysis of studies in africa. *Journal of Acquired Immune Deficiency Syndromes*, *77*(5), 427–438.
- Levinsohn, J., McLaren, Z. M., Shisana, O., & Zuma, K. (2013). HIV status and labor market participation in south africa. *The Review of Economics and Statistics*, *95*(1), 98–108.
- Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*, 80–94.
- Mishra, V., Bignami-Van Assche, S., Greener, R., Vaessen, M., Hong, R., Ghys, P. D., Boerma, J. T., Van Assche, A., Khan, S., & Rutstein, S. (2007). HIV infection does not disproportionately affect the poorer in Sub-Saharan Africa. *AIDS*, *21*, 17–28.
- Nutor, J. J., Duah, H. O., Agbadi, P., Duodu, P. A., & Gondwe, K. W. (2020). Spatial analysis of factors associated with HIV infection in Malawi: Indicators for effective prevention. *BMC Public Health*, *20*(1), 1–14.
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Economics*, *11*(6), 619–632.
- Psaros, C., Remmert, J. E., Bangsberg, D. R., Safren, S. A., & Smit, J. A. (2015). Adherence to HIV care after pregnancy among women in Sub-

- Saharan Africa: Falling off the cliff of the treatment cascade. *Current HIV/AIDS Reports*, 12(1), 1–5.
- QGIS Development Team. (2009). QGIS geographic information system [computer software manual] [Accessed February 24, 2024]. <http://qgis.osgeo.org>
- QGIS Development Team. (2020). Vector analysis [Accessed February 24, 2024]. https://docs.qgis.org/3.4/en/docs/user_manual/processing_algs/qgis/vectoranalysis.html#k-means-clustering
- Reniers, G., & Watkins, S. (2010). Polygyny and the spread of HIV in Sub-Saharan Africa: A case of benign concurrency. *AIDS*, 24(2), 299–307.
- Sani, A. S., Abraham, C., Denford, S., & Ball, S. (2016). School-based sexual health education interventions to prevent STI/HIV in Sub-Saharan Africa: A systematic review and meta-analysis. *BMC Public Health*, 16(1), 1–26.
- Tanser, F., Barnighausen, T., Cooke, G. S., & Newell, M. L. (2009). Localized spatial clustering of HIV infections in a widely disseminated rural south african epidemic. *International Journal of Epidemiology*, 38(4), 1008–1016.
- UNAIDS. (2024a). *AIDS and the Sustainable Development Goals* [Accessed: 2024-10-06]. https://www.unaids.org/en/AIDS_SDGs
- UNAIDS. (2024b). Aidsinfo | unaids [Accessed: 2024-10-01]. <http://aidsinfo.unaids.org/>
- UNAIDS. (2024c). *Unaid - joint united nations programme on HIV/AIDS* [Accessed: 2024-10-06]. <https://www.unaids.org/en>
- USAID. (2022). Demographic and Health Survey [Accessed February 24, 2024]. <https://dhsprogram.com/>

Chapter 3

Bayesian Hierarchical Models and HIV-IPV Association

Connected threads: Africa, HIV and Intimate Partner Violence using the INLA-SPDE approach¹

Co-author: Anna Maria Parroco

Abstract

This work aims at evaluating the relationship between intimate partner violence (IPV) and HIV using survey data from Zambia and Zimbabwe – specifically, data from the Demographic and Health Survey. Sub-Saharan Africa’s population is strongly affected both by HIV and IPV. Indeed, the literature has paid great attention to the association between these two phenomena and how their contextual characteristics affect their association and their spatial distribution. The sample used in this work includes women living in Zambia and Zimbabwe who were tested for HIV during the survey and who were asked about their experiences with abuse. Two different perspectives of analysis are here considered: first, women as individual respondents, then as

¹Pre-print version.

people being part of a neighbourhood. The estimation results of a logistic regression model and the post estimation results of a Bayesian hierarchical Beta regression model with the integrated nested Laplace approximation (INLA) are presented. HIV is the response variable. Results show that women who are both physically and emotionally abused by their partners are more likely to be affected by HIV, both when considered individually and within their neighbourhood; maps allow to visualise HIV prevalence within regions.

Keywords

HIV; Intimate Partner Violence; Sub-Saharan Africa; Ecological Model; INLA-SPDE

3.1 Introduction

Recent data about Sub-Saharan Africa (SSA) show that this region is extremely affected both by intimate partner violence² (IPV) and HIV. Indeed, in 2021 over two-thirds of people living with HIV were residents in these countries (UNAIDS, 2024b). IPV is here an equally taxing matter, and it has been estimated that one in three women has been abused at least once in her life (World Health Organization, 2021). The relevance of both phenomena has been acknowledged by the international community, which has included actions and targets to improve the living conditions of people living with HIV and to cut down the number of new infections in ten of the seventeen Sustainable Development Goals (SDGs) (UNAIDS, 2024a); among them, given the role that gender relations are attributed to the diffusion of HIV in Sub-Saharan Africa, the target that calls for the termination of violence against women and girls within the fifth SDG (“Gender Equality”) is also relevant for the HIV epidemics in this continent (Nations, 2024).

²The United Nations define intimate partner violence, also referred to as domestic violence, as a behavioural pattern of power and control over an intimate partner (United Nations, 2022).

A systematic review of these matters shows that there are several ways in which violence makes women more vulnerable to this disease: on the one hand, marital rape by an HIV-positive partner is expected to expose them to the virus; wherever abuse is present, women lack agency in asking for safe sexual practices; finally, revictimisation processes put women in vicious cycles of risky sexual practices (Maman et al., 2000). IPV does not only endanger women’s health, but also that of their children: a recent study shows that, in Sub-Saharan Africa, IPV could be responsible for one in eight paediatric HIV infections due to vertical transmission (Kuchukhidze et al., 2024).

This study, building from previous work (Arcaio & Parroco, 2023) , aims at analysing the relationship between IPV and HIV using data from the Demographic and Health Survey (DHS) from Zambia and Zimbabwe, both at the individual level and through the lens of their spatial relationship. The analysis will be carried out using different techniques that account for two specific research questions: a logistic regression model to study the association between these two phenomena at the individual level; and Bayesian hierarchical models within the integrated nested Laplace approximation (INLA) framework with the stochastic partial differential equation (SPDE) approach, to analyse spatial and contextual factors that act at the neighbourhood level, thus highlighting possible HIV hotspots.

As far as we know, in the literature, IPV and HIV are mostly studied via qualitative studies or ad-hoc surveys – or in the clinical environment, which, however, goes beyond the scope of this work. Here the DHS is used, which allows for a broader analysis that includes geostatistical data on both HIV and IPV.

In this work, Section 3.2 reviews the existing literature on HIV and intimate partner violence. Section 3.3 briefly discusses the data and methods used for this analysis. Sections 3.4 and 3.5 respectively contain the results of the analysis and a brief discussion of the results with conclusions.

3.2 HIV and Intimate Partner Violence: a Brief Literature Review through the Ecological Model

The contextualisation of risk factors and the role of social context in mediating the relationship between risk factors and disease have been widely explored in the literature of HIV (Gillespie et al., 2007). For example, gender relations³ seem to be crucial in the increase of HIV prevalence, especially in Sub-Saharan Africa, where the phenomenon is so widespread. Indeed, in 2021, women accounted for 63% of new HIV infections in this region, with a slower decline in their incidence trend than in men's (UNAIDS, 2022c). Moreover, while in the rest of the world 94% of new infections in 2021 were imputable to key populations⁴, in Sub-Saharan Africa, almost half of the new infections are imputable to the general population (UNAIDS, 2022c).

Gender dynamics that privilege men over women, as emphasised by the theory of Hegemonic Masculinity (Connell, 2005), underscore the heightened vulnerability of women and girls to HIV. In Sub-Saharan Africa, young women are three times more likely to live with HIV compared to men of the same age (UNAIDS, 2022b). For example, a cross-sectional study conducted in South Africa in 2001 (Dunkle et al., 2004) shows that both physical and sexual partner violence are associated with increased odds of HIV infection, while a longitudinal study based, once again, in South Africa confirms that power inequity inside couples and intimate partner violence contribute to HIV infection in young women (Jewkes et al., 2010).

The ecological model has been proposed both for HIV and IPV to draw attention to the complex interplay between social and individual characteris-

³“Gender relations” refers to the relationship between men and women, and it reflects their respective roles in power dynamics, in decision making and labour, not only within households but also in the society (USAID, n.d.).

⁴Key populations are considered to be particularly vulnerable to HIV infection. They include: sex workers; clients of sex workers and sex partners of key populations; people who inject drugs; gay men and other men who have sex with men; transgender women. These populations account for less than 5% of the global population, however, they are disproportionately affected by HIV (UNAIDS, 2022a).

tics of both phenomena. The model moves from societal and community-level factors to relationship and individual characteristics that are associated with the presence of HIV or IPV (Centers for Disease Control and Prevention, 2015).

At the societal level, gender inequality and violence against women and girls have been shown to exacerbate the risk of women to be infected by HIV (Dunkle et al., 2004; Jewkes et al., 2010).

At the community level, violence seems to be associated with poverty and living in rural areas (Jeyaseelan et al., 2007). Moreover, the perpetration of violence itself is facilitated by the general acceptance of violence within couples (Arcaio & Parroco, 2024; Gage, 2005). On the contrary, rural areas have been shown to be characterised by a lower prevalence of HIV than urban areas (Nutor et al., 2020). Furthermore, while the highest HIV prevalence rates are found in low-income countries, more often than not, the wealthiest segments of the population are the most affected (Mishra et al., 2007; Parroco et al., 2024). This occurrence is, however, context-specific: the contrary is, indeed, found in rich countries, and it is likely due to several factors such as recurrence to sex workers (Dunkle et al., 2004).

Relationship and individual factors refer to gender relations within couples as well as their socio-economic characteristics. People with a higher education are less likely to either perpetrate or suffer IPV (Abrahams et al., 2006; Kishor & Johnson, 2006) and to be affected by HIV (Hargreaves et al., 2008).

The association between socioeconomic context and gender also plays a key role in the study of HIV. Most of the characteristics of the association between HIV and context here presented, indeed, are amplified in women. For women living in poverty, engaging in sex work may become a necessity, thereby increasing their risk of HIV infection, in particular for those living in substandard housing in urban environments, in case of unemployment, and when they have a history of intimate partner violence (Dunkle et al., 2004; Gillespie et al., 2007).

Table 3.1: Descriptive statistics

	Zambia 2018	Zimbabwe 2015
No. Respondents	5,144	4,011
No. Clusters	545	400
HIV prevalence	13.86%	13.36%
Violence experience		
<i>Only physical</i>	11.22%	10.23%
<i>Only emotional</i>	6.39%	9.70%
<i>Only sexual</i>	2.37%	2.50%
<i>Physical + emotional</i>	12.81%	12.67%
<i>Physical + sexual</i>	2.72%	1.30%
<i>Emotional + sexual</i>	1.53%	2.15%
<i>All three</i>	7.14%	5.78%

3.3 Data and Methods

HIV and intimate partner violence are investigated using data from the Demographic and Health Survey (DHS) programme. The DHS is a nationally representative household survey, employed for the collection and analysis of harmonised demographic and health data (Croft et al., 2018). Women of reproductive age (15-49 years) are the main subject of this survey.

The data used in this work are a merge of the “HIV dataset”, which contains the result of ELISA (enzyme-linked immunosorbent assay) tests to determine respondents’ serostatus, and the domestic violence module within the questionnaire administered to women. To collect data on these matters, in all households in which men are also included in the survey, blood samples are collected from those who volunteer for HIV testing; furthermore, the domestic violence module is administered to randomly selected ever-partnered women in heterosexual relationships already involved in the survey.

In this study, the surveys Zambia 2018 and Zimbabwe 2015 are included. These surveys carry the HIV dataset and the domestic violence module, as well as the geographical information for the clusters. Clusters identify census units of their countries and are comparable to neighbourhoods; the geographical dataset of the DHS provides the GPS coordinates of the clusters’ centroids.

As it was stated before, this analysis aims to evaluate the relationship between HIV and IPV, controlling for individual and contextual characteristics. Considering each respondent as a unit of analysis, the HIV status of respondents, i.e. the response, is modelled using a logistic regression model.

The spatial relationship between HIV and IPV prevalence is modelled using a Bayesian hierarchical model within the INLA framework and the SPDE approach using the same data once aggregated (using sample weights) at the cluster level. One of the assumptions of the Bayesian model is that the diffusion of HIV is spatially continuous and modelled with a Gaussian random field, which is approximated using a finite basis function defined on a triangulated mesh (built on non-convex boundaries for the referred coordinates and by setting the smoothness parameter to 1). Thus, the stochastic partial differential equation approach (SPDE) is implemented⁵ using the R-INLA package (Lindgren & Rue, 2015) to then predict the HIV prevalence in unsampled locations. In this case, the response is assumed to be Beta distributed since it falls in the interval (0,1) once censoring is applied. In the results section, the adjusted posterior odds ratios (AORs), as well as their credibility intervals, are reported for ease of interpretation.

The same set of covariates is used in both models. They were selected among the available variables in the datasets as to explain the phenomenon within the framework of the ecological model, thus they can be split into three categories.

1. *Society*. This category includes the experienced abuse of the respondents and the respondents' employment status to control for gender equality. Violence is measured via dichotomic variables on whether the respondent has been abused either physically, emotionally, or sexually (or any combination thereof⁶) in the 12 months prior the interview. In the individual-level data analysis, a single model is employed, incorpo-

⁵More insight on the models, as well as an explanation of how the INLA approach differs from other spatial analysis methods, can be found in Gómez-Rubio (2020) and Moraga (2023).

⁶In the logistic regression model, proper interaction terms were added to acknowledge the different kinds of violence suffered by the respondents; in the Bayesian models, percentages of the different combinations of abuse at the cluster level were instead computed.

rating a dummy variable for the country of residence to account for this factor. However, in the Bayesian models, where Zambia and Zimbabwe are analysed separately, a random effect is applied to account for the region of residence.

2. *Community*. Here, the type of residence (be it rural or urban) is used as a control variable in the logistic regression model and as a random effect in the Bayesian models. This domain also includes whether the respondent is in the top two quintiles of wealth in her country and whether she justifies that a husband may use physical violence on his wife.
3. *Relationship and individual level characteristics*. These include respondent's age and number of years of education, age at first intercourse, whether the respondent has any children, and whether the respondent's husband has little to no education.

3.4 Results

3.4.1 Logistic Regression Model

The results of the estimation of the logistic regression model can be found in Table 3.2. A likelihood ratio test (LRT) was also employed to check whether the variables indicating violence are relevant in the fit of the model: given that $LRT(7d.f.) = 16.62$ ($p - value = 0.02$), the model that also includes IPV fits significantly better than the model without it. The approximation of the LRT statistic with the Wald test – of the selected model vs. the null model – returns $Wald(17d.f.) = 424.68$ ($p - value \approx 0$), thus pointing to an overall good fit of the model.

Among the different kinds of abuse experienced, only the interaction between physical and emotional violence is positively (and significantly) associated with HIV: indeed, where women are abused both physically and emotionally, they are 76% more likely to be HIV positive than their counterparts. HIV is negatively associated with being employed ($OR = 0.837$,

Table 3.2: Logistic Regression Model Estimates

Domain	Variable	OR	SE
Society	Physical violence	0.890	0.095
	Emotional violence	0.843	0.101
	Sexual violence	0.959	0.202
	Physical&emotional violence	1.760***	0.305
	Physical&sexual violence	1.312	0.417
	Emotional&sexual violence	0.991	0.348
	Physical&emotional&sexual violence	0.600	0.269
	Respondent is employed (ref. Unemployed)	0.837**	0.053
	Country (ref. Zambia)		
	Zimbabwe	1.490***	0.102
Community	Respondent is in the top two quintiles of wealth in her country	1.193 .	0.119
	Respondent lives in a rural area (ref. urban area)	0.621***	0.058
	Respondent justifies physical abuse by her husband	0.958	0.063
Individual and Relationship characteristics	Age	1.059***	0.004
	No. of years of education	1.026*	0.012
	Age at first intercourse	0.911***	0.011
	Respondent has children	0.635**	0.094
	Husband has little to no education	0.813**	0.065
	Constant	0.198***	0.057
*** p <0.01, ** p <0.05, * p <0.1			

$p - value = 0.005$), while the coefficients for the country dummy variable reflect the general condition of their resident populations with respect to HIV prevalence.

As for community characteristics, in this model, women who justify physical violence on women, are not associated with living with HIV. People residing in rural areas are 48% less likely to live with HIV, which is a similar result to what has been shown in the literature (Nutor et al., 2020).

Socio-demographic characteristics of the respondents and their partners also matter. Education seems to play a peculiar role with HIV: indeed, women who have studied longer are less likely to be HIV positive, but a lower education level of their husbands makes them less likely to be HIV positive. This last result, tied with the positive association between high wealth and HIV of respondents, points to a curious effect of context. People who are richer are more likely to get in contact with more sexual partners, which is exacerbated for those male partners with higher educational levels. Poverty and a low education level don't protect from the virus, of course, but they make it difficult for people to have a larger social circle that could expose them to HIV. Age at first intercourse is also significantly associated with HIV: indeed, for each year of delay in age at first sex the odds of being HIV positive decrease by a bit less than 10%.

Table 3.3: Bayesian models: adjusted odds ratios and their credibility interval

Variable	Zambia			Zimbabwe		
	AOR	2.50%	97.50%	AOR	2.50%	97.50%
Physical violence	1.002	0.994	1.009	0.998	0.989	1.006
Emotional violence	0.996	0.986	1.007	0.998	0.989	1.006
Sexual violence	0.981	0.959	1.004	0.996	0.977	1.016
Physical&emotional violence	1.002	0.995	1.010	1.010	1.002	1.018
Physical&sexual violence	1.000	0.984	1.017	0.999	0.973	1.026
Emotional&sexual violence	0.976	0.948	1.005	0.990	0.972	1.009
Physical&emotional&sexual violence	1.008	0.998	1.018	0.995	0.983	1.008
Respondent is employed	0.997	0.993	1.002	0.996	0.986	1.005
Respondent is in the top two quintiles of wealth in her country	1.011	1.007	1.015	1.002	0.998	1.005
Respondent justifies physical violence	1.000	0.995	1.005	0.998	0.992	1.004
Age	1.052	1.019	1.085	1.032	1.000	1.065
No. of years of education	1.025	0.955	1.099	0.912	0.822	1.012
Age at first intercourse	0.816	0.745	0.895	0.997	0.906	1.097
Respondent has children	0.994	0.981	1.007	0.999	0.994	1.004
Husband has little to no education	0.996	0.990	1.002	0.996	0.989	1.002
Constant	1.075	0.139	8.306	0.360	0.060	2.160

3.4.2 Bayesian hierarchical models.

The posterior estimates of the Bayesian hierarchical models are presented in Table 3.3.

The joint physical and emotional abuse is significantly associated with HIV prevalence at the cluster level only in Zimbabwe.

The analysis of the random effects in the models do not show any significant differences between type of residence or among the different regions within either country.

The post-estimation results were then projected on maps to show the regions with higher HIV prevalence in women.

The maps above show the regions with highest HIV prevalence among women. In Zambia, this can be found in the Southern and Western Provinces, as well as in the Central Province. In Zimbabwe, the highest HIV prevalence is in the Matabeleland South Province.

3.5 Conclusions

In this work, the main goal is to analyse the relationship between IPV and HIV using survey data from African countries. This analysis has confirmed

CHAPTER 3. HIV-IPV ASSOCIATION USING INLA

Figure 3.1: Zambia 2018

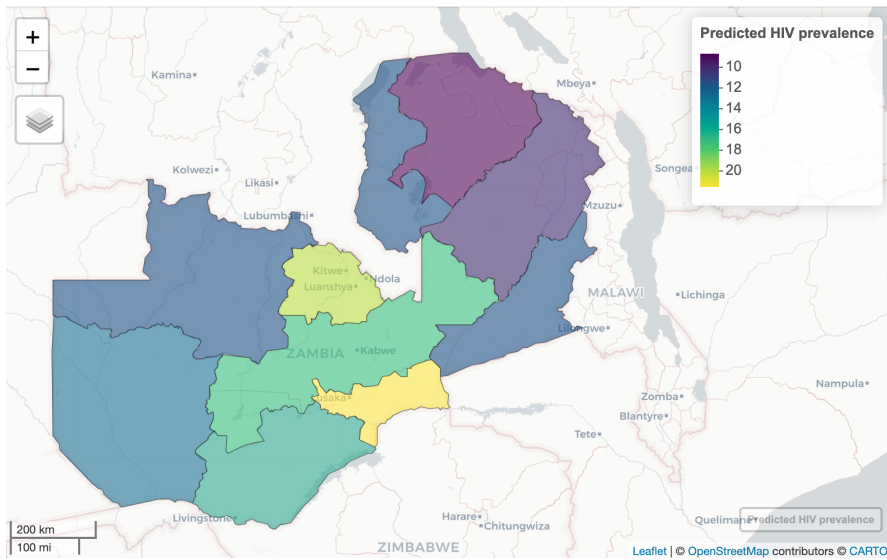
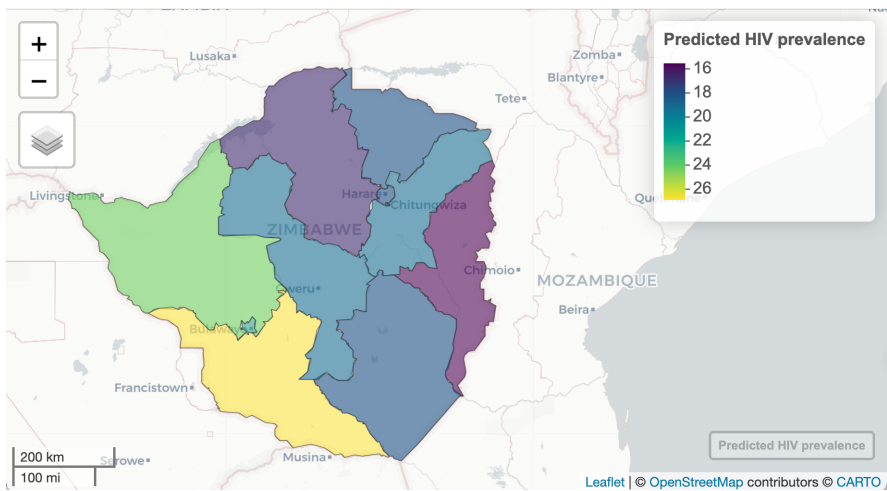


Figure 3.2: Zimbabwe 2015



several findings highlighted in the literature – being abused by the current partner makes women more likely to be HIV-positive, especially when women are abused both physically and emotionally. This becomes especially relevant given that, while the literature on HIV mostly focuses on physical and sexual violence separately, the literature on IPV is progressively moving towards the concept of polyvictimisation (Arcaio & Parroco, 2024; Okumu et al., 2021). Moreover, extending this analysis to a spatial approach gives the opportunity to check for the spatial effect of this association, thus highlighting where it breeds HIV hotspots. Using the Bayesian models allows for a better understanding and interpretation of HIV and IPV than only using classical models could. This proves to be valuable to the international community, given the attention paid to these matters – and to IPV in particular.

This work certainly does not lack limitations. Specifically, given the nature of the data, there is much to be said about the reliability of the responses on items about domestic violence, which are subject to social desirability, thus leading to an overall underestimation of the phenomenon. Another limitation is determined by the fact that both HIV and IPV could be considered ‘rare events’, which turns into low variability in the spatial models used, and this may become an issue in the estimation of the models.

But this work also offers many possibilities for further development. Mainly, fine-tuning the models to achieve better fits and integrating region and national-level covariates. A sensitivity analysis could help assess the robustness of the results. This would allow to further check for the interplay between contextual and individual characteristics of respondents, to exploit the framework of ecological models, often employed in the analysis of HIV and IPV.

References

- Abrahams, N., Jewkes, R. K., Laubscher, R., & Hoffman, M. (2006). Intimate partner violence: Prevalence and risk factors for men in Cape Town, South Africa. *Violence and Victims*, 21(2), 247–264.

- Arcaio, M., & Parroco, A. M. (2023). Women’s exposure to HIV in africa: The role of intimate partner violence. In *Sis 2023 - the book of short papers* (pp. 599–604). Pearson.
- Arcaio, M., & Parroco, A. M. (2024). A composite indicator of polyvictimisation through the lens of the ecological model in Sub-Saharan Africa. *Social Indicator Research*. <https://doi.org/10.1007/s11205-024-03344-5>
- Centers for Disease Control and Prevention. (2015). The social-ecological model: A framework for prevention [Accessed 2024-01-30]. <https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html>
- Connell, R. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society*, 19(6), 829–859.
- Croft, T. N., Marshall, A. M. J., Allen, C. K., et al. (2018). *Guide to DHS statistics* (tech. rep.). ICF. Rockville, Maryland, USA.
- Dunkle, K. L., Jewkes, R. K., Brown, H. C., Gray, G. E., McIntyre, J. A., & Harlow, S. D. (2004). Gender-based violence, relationship power, and risk of HIV infection in women attending antenatal clinics in south africa. *The Lancet*, 363, 1415–1421.
- Gage, A. J. (2005). Women’s experience of intimate partner violence in Haiti. *Social Science & Medicine*, 61(2), 343–364.
- Gillespie, S., Kadiyala, S., & Greener, R. (2007). Is poverty or wealth driving HIV transmission? *AIDS*, 21, S5–S16.
- Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. Chapman & Hall/CRC Press.
- Hargreaves, J. R., Bonell, C. P., Boler, T., Boccia, D., Birdthistle, I., Fletcher, A., Pronyk, P. M., & Glynn, J. R. (2008). Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-saharan africa. *AIDS*, 22(3), 403–414.
- Jewkes, R. K., Dunkle, K. L., Nduna, M., & Shai, N. (2010). Intimate partner violence, relationship power inequity, and incidence of HIV infection in young women in south africa: A cohort study. *The Lancet*, 376, 41–48.

- Jeyaseelan, L., Kumar, S., Neelakantan, N., Peedicayil, A., Pillai, R., & Duvvury, N. (2007). Physical spousal violence against women in India: Some risk factors. *Journal of Biosocial Science*, 39(5), 657–670.
- Kishor, S., & Johnson, K. (2006). Reproductive health and domestic violence: Are the poorest women uniquely disadvantaged? *Demography*, 43(2), 293–307.
- Kuchukhidze, S., Walters, M. K., Panagiotoglou, D., Boily, M. C., Diabaté, S., Russell, W. A., & Maheu-Giroux, M. (2024). The contribution of intimate partner violence to vertical HIV transmission: A modelling analysis of 46 african countries. *The Lancet HIV*, 11(8), e542–e551.
- Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19), 1–25. <https://doi.org/10.18637/jss.v063.i19>
- Maman, S., Campbell, J., Sweat, M. D., & Gielen, A. C. (2000). The intersections of HIV and violence: Directions for future research and interventions. *Social Science & Medicine*, 50, 459–478.
- Mishra, V., Bignami-Van Assche, S., Greener, R., Vaessen, M., Hong, R., Ghys, P. D., Boerma, J. T., Van Assche, A., Khan, S., & Rutstein, S. (2007). HIV infection does not disproportionately affect the poorer in Sub-Saharan Africa. *AIDS*, 21, 17–28.
- Moraga, P. (2023). *Spatial statistics for data science: Theory and practice with R*. Chapman & Hall/CRC Press.
- Nations, U. (2024). Sustainable Development Goals: Goal 5 [Accessed: 2024-10-06]. <https://sdgs.un.org/goals/goal5>
- Nutor, J. J., Duah, H. O., Agbadi, P., Duodu, P. A., & Gondwe, K. W. (2020). Spatial analysis of factors associated with HIV infection in Malawi: Indicators for effective prevention. *BMC Public Health*, 20(1), 1–14.
- Okumu, M., Orwenyo, E., Nyoni, T., Mengo, C., Steiner, J. J., & Tonui, B. C. (2021). Socioeconomic factors and patterns of intimate partner violence among ever-married women in Uganda: Pathways and actions for multicomponent violence prevention strategies. *Journal of Interpersonal Violence*, 08862605211021976.

- Parroco, A. M., Arcaio, M., & Mendola, D. (2024). Does neighborhood matter? an analysis of HIV prevalence in Sub-Saharan African countries. *Biodemography and Social Biology*. <https://doi.org/10.1080/19485565.2024.2384995>
- UNAIDS. (2022a). 2022 global AIDS update summary [Accessed: 2024-10-06]. https://www.unaids.org/sites/default/files/media_asset/2022-global-aids-update-summary_en.pdf
- UNAIDS. (2022b). Dangerous inequalities [Accessed: 2024-10-06]. <https://www.unaids.org/en/resources/documents/2022/dangerous-inequalities>
- UNAIDS. (2022c). In danger: Global AIDS update [Accessed: 2024-10-06]. <https://www.unaids.org/en/resources/documents/2022/in-danger-global-aids-update>
- UNAIDS. (2024a). *AIDS and the Sustainable Development Goals* [Accessed: 2024-10-06]. https://www.unaids.org/en/AIDS_SDGs
- UNAIDS. (2024b). Aidsinfo | unaids [Accessed: 2024-10-01]. <http://aidsinfo.unaids.org/>
- United Nations. (2022). What is domestic abuse? [Accessed 2024-01-30]. <https://www.un.org/en/coronavirus/what-is-domestic-abuse>
- USAID. (n.d.). Demographic and Health Survey [Accessed: 2024-10-06]. https://pdf.usaid.gov/pdf_docs/Pnadm089.pdf
- World Health Organization. (2021). *Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring: Recommendations for a public health approach* [Accessed: 2024-10-06]. <https://www.who.int/publications/i/item/9789240022256>

Chapter 4

Text Mining of A Focus Group on Gender-Based Violence

“Inhumane”: quantitative textual analysis of a qualitative survey on gender violence among victims of international human trafficking¹

Co-authors: Anna Maria Parroco, Federica Barbaro

Abstract

Focus groups are often employed in research because they offer insight on perceptions and experiences from stakeholders. Qualitative interviews also allow researchers to get more in-depth and pragmatic information on the phenomena of interest, thus providing breeding ground for novel interpretations of results coming from solely quantitative analyses.

Furthermore, several computational techniques, including machine learning and generative pre-trained transformers (GPTs), have helped develop quantitative analysis to process massive textual data sets.

In this context, text mining - or text data mining (TDM) - has been

¹Pre-print version.

used to examine and find key themes in group interviews' transcripts to give objectivity to their analysis. Natural language processing (NLP) techniques can be used to first gather information on common themes, as well as to inform on the feelings risen from the text via more advanced sentiment analysis tools.

The aim of this work is to examine the transcript of a focus group about gender violence among victims of international human trafficking, using several text mining techniques. The focus group involved two researchers and four key stakeholders working in the third sector with expertise on the matter – either by working with victims of gender violence, or by working with victims of human trafficking, or both – as well as a moderator. The stakeholders work in the province of Palermo (Italy) and refer to their experiences in this specific territory.

Preliminary results show the reality of human trafficking: the harrowing experience of violence, the consequent loss of identity, coming from systemic and inhumane practices of women's exploitation in the streets. Education and independence, while recognized as essential, are rarely sought after. This also highlights that the naturalization of gender violence makes reporting their perpetrators particularly difficult.

Keywords

Human trafficking; Gender-based violence; Text mining; Focus groups.

4.1 Introduction

Researchers often employ focus groups to gauge more information and insight on the phenomena they are studying (Centers for Disease Control and Prevention, 2018). The qualitative interviews coming from this kind of research offer personal perceptions and experiences from stakeholders, thus providing breeding ground for novel interpretations of results that come from solely quantitative analyses. While usually either audiovisual records or their tran-

scripts are employed to evaluate what is discussed during a focus group, in the last few years, several computational techniques, including machine learning and generative pre-trained transformers (GPTs), have helped develop quantitative analysis methods that process massive textual datasets (Nielbo et al., 2024). Text mining has been used to examine and find key themes in group interviews’ transcripts to give objectivity to their analysis and natural language processing (NLP) techniques are used to first gather information on common themes, as well as to inform on the feelings risen from the text via more advanced sentiment analysis tools (i.e., Payton et al. (2018)).

In this context, the aim of this work is to examine the transcript of a focus group about gender violence among victims of international human trafficking, using two text mining techniques. The focus group involved six key stakeholders working in the third sector with expertise on the matter – either by working with victims of gender violence, or by working with victims of human trafficking, or both – as well as a moderator. The stakeholders work in the province of Palermo (Italy) and refer to their experiences in this specific territory.

4.2 Human Trafficking and Gender-Based Violence

Human trafficking has been defined by the United Nations Office on Drugs and Crime in the Palermo Protocol as “the recruitment, transportation, transfer, harbouring or receipt of people through force, fraud or deception, with the aim of exploiting them for profit” (United Nations Convention against Transnational Organized Crime, 2000) . Violence and fraudulent promises of education and job opportunities are used to force or lure possible victims (United Nations Office on Drugs and Crime, 2024).

No region in the world is exempt from this phenomenon, and people of all sexes and ages have been found to fall victims of this crime. The most significant trans-regional flow – i.e., from one region of the world to another,

in contrast to domestic trafficking that happens within a certain region – concerns African people (United Nations Office on Drugs and Crime, 2022). Moreover, 68% of trafficked individuals are moved by either governance-type crime groups or business-like crime groups: indeed, these kinds of organisations are able to traffic more people than individual traffickers can, as well as to retain them for longer periods of time, likely due to the extent of the violence used to coerce them (United Nations Office on Drugs and Crime, 2022).

The detection of cases of human trafficking more often than not happens because of proactive actions of the victims who manage to exit exploitation and thus self-report to the authorities: 41% of cases arise from victims' actions while only 9% emerge from other institutions such as hospitals, schools or NGOs (United Nations Office on Drugs and Crime, 2022). Self-reporting can, however, be hindered by the perceptions of the victims themselves. Trafficked individuals struggle with identifying themselves as such due to psychological mechanisms such as denial or repression, inability to see themselves as victims, as well as lack of knowledge of their rights and the perception that their exploitation is happening because of their irregular migration status (Bloch & McKay, 2013; Van Meeteren & Hiah, 2020).

The exploitation of victims of human trafficking, thus, assumes several forms: in 2020, 38.8% of the detected cases were victims of forced labour and 38.7% were involved with sexual exploitation (United Nations Office on Drugs and Crime, 2022). Women and girls constitute 60% of all detected victims. This statistic is followed by what can be found in case summaries: women are three times more likely to experience violence than men², which includes both physical and sexual violence (United Nations Office on Drugs and Crime, 2022).

The venues in which this exploitation happens also matter. In the court case summaries for human trafficking in 2012-2022, 11% of victims were reported to work in the street, while 18% were employed in apartments sup-

²Court case summaries report that 31% of men have not experienced violence during their trafficking, whereas 79% of women are reported to have experienced extreme or explicit physical violence or threats and psychological violence (United Nations Office on Drugs and Crime, 2022).

plied by traffickers and 19% in brothels (United Nations Office on Drugs and Crime, 2022).

Given this premise, the existence of a connection between human trafficking, gender-based violence (GBV) and sex work is clear. Sex work is defined as the supply of sexual encounters in exchange of money. Although many scholars debate on whether sex work is the result of women's exploitation or an act of sexual freedom, it has been recognised that there is a structural component to the exploitation and violence experienced by sex workers, with migrants, people living in poverty and women of colour more likely to be abused than their white middle-to-upper class counterparts (Connelly, 2016). Even policies and laws decriminalising prostitution still exclude migrants from their protection. For example, the Prostitution Reform Act issued in New Zealand in 2003 – that aimed to eliminate all laws against consensual sex work, as well as to improve workers' conditions by protecting them from exploitation and by promoting their well-being, health and safety – excluded migrants because it considered them as “seasonal workers” (Armstrong & Abel, 2020).

The three components of GBV – power, privilege and positionality³ – are also shared with human trafficking. Power is exercised through violence by traffickers to force people into sex work; privilege and positionality determine the relationship between trafficker and trafficked (Ali & Rogers, 2023).

4.3 The focus group: the transcript and methods of analysis

The focus group whose transcript was analysed in this study involved seven people: one moderator and six key stakeholders. The stakeholders included two experts in migration studies and agents in the third sector in the province

³In the literature on GBV power is defined as the ability to control whether an individual gets access to resources and opportunities as well as the ability to prevent another individual to make their own decisions. Privilege is due to the position of individuals in hypothetical hierarchies and it's strictly related to positionality, which refers to the placement within relationships themselves (Ali & Rogers, 2023).

of Palermo (Italy): two of them work within an association that provides psycho-social assistance to migrants to promote human rights and to fight GBV amongst victims of sex trafficking; another one works in an association whose aim is to contrast all GBV, and intimate partner violence in particular; the last participant in the focus group is a lawyer who works in an association that offers legal help to migrants.

The information sought during the interviews included: the victims' background, both in terms socio-demographic characteristics and migration history; whether migrant victims of GBV had peculiar characteristics with respect to Italian victims; the profile of traffickers and perpetrators; what compels victims to report the abuse.

The record of the conversation was transcribed so that the textual analysis could be performed. It was kept in Italian to maintain the original language and vernacular used during the focus group. The text ended up being made of 341 entries (i.e., number of times the conversation shifted from one person to another) for a total of 13,266 words.

Prior to the analysis, the text was processed as following: numbers, common and uncommon stop-words, punctuation and extra spaces were removed; then it was transformed to find the stem of the words. After pre-processing, an exploratory word-cloud was created to check for the most common words in the text. Later a sentiment analysis was conducted. In this case, Saif Mohammad's NRC Emotion lexicon was used (Mohammad & Turney, 2013). This lexicon associates a list of words with eight basic emotions – anger, fear, anticipation, trust, surprise, sadness, joy, and disgust – and whether the overall sentiment is positive or negative. The choice for this particular lexicon was due to its availability in Italian.

The text was then evaluated qualitatively in conjunction with the quantitative results to gauge more information about human trafficking from the perspective of operators in the third sector.

The textual data used in this work was examined using R version 4.3.2 (Team, 2023), and the packages tm version 0.7-13 (Feinerer & Hornik, 2024), wordcloud version 2.6 (Fellows, 2018), and syuzhet version 1.0.7 (Jockers, 2015).

4.4 Results

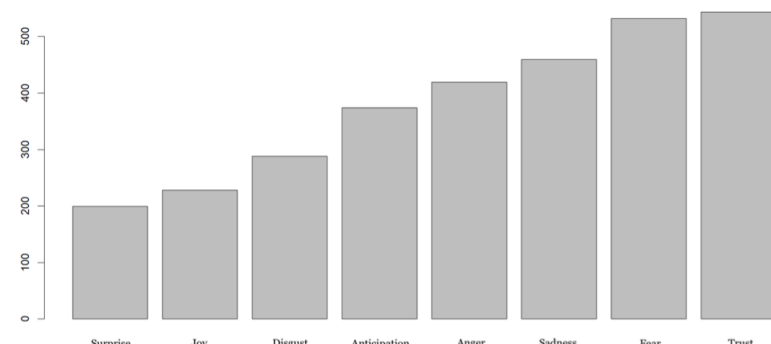
After the pre-processing of the transcript of the record of the focus group, a first exploratory wordcloud⁴ was created (see Figure 4.1). This graph offers a preview of the key concepts connected to the experiences lived by the victims of human trafficking. Here are some of the themes arisen:

- Violence and abuse: words such as *violenza*, *carnefici*, *disumane*, *vittima*, *agghiacciante* (i.e., violence, perpetrator, inhumane, victim, harrowing) highlight what victims of human trafficking live through;
- Women and minors: *donna*, *bambini*, *scuola* (i.e., woman, children, school) point to the peculiar vulnerability of these groups;
- Trafficking and exploitation: *attività*, *denunce*, *rete*, *problema*, *sessuale*, *clienti*, *strada*, *manager*, *sistemica*, *complesso* (i.e, activity, criminal report, network, problem, sexual, clients, street, manager, systemic, complex) all refer to the criminal nature of human trafficking and to the different forms of exploitation victims are subject to;
- Origin and destination: words like *ghanesi*, *cultura*, *mediare*, *italiano*, *nazionalità* (i.e., Ghanaian, culture, mediate, Italian, nationality) indicate how human trafficking is tied to migration flows and culture.

As for the results of the sentiment analysis (see Figure 4.2), they show a significant prevalence of negative emotions, such as sadness, fear, anger and disgust. This reflects the gravity of what victims of human trafficking experience, often including abuse, exploitation and physical and psychological violence.

The fact that trust is the most common sentiment is imputable to two mechanisms, one opposite the other. On the one hand, trust refers to the deception and tricks used to lure victims into trafficking, which leads them to the loss of autonomy and agency; on the other hand, trust is the foundation of the rapport between operators and victims seeking a way out.

⁴The wordcloud in Figure 4.1 is in Italian because it is the language of the text.



4.5 Discussion

The results of the quantitative analysis were then used to support the qualitative text analysis. Several pieces of information on human trafficking were gathered:

- Victims' background. Words in the text (translated): nationality, culture, culturally, educat*, school, minors, woman, Ghanaian, country
 - Socio-demographic characteristics: human trafficking victims are usually from poor and remote villages, mostly presenting a low educational level. Even when trying to get out, education and independence – essential to be free of the exploitation – are rarely sought after. General distrust between different communities – heightened when people of different genders are present – hinders the possibility for women to even learn the language spoken in the place they live in.
 - Migration background: while most victims of human trafficking in the province of Palermo do have a migration history, significative differences are found between ethnicities. Ghanaians, contrary to their Nigerian counterparts, are rarely involved in illicit activities. Most activities involving solicitation are, however, managed by criminal organizations while the “workers” are from Nigeria, Côte d'Ivoire, and Sri Lanka.
- Peculiar characteristics of GBV among migrant women. Words in the text (translated): network, victim, trafficking, children, school, perpetrators, inhumane, systemic, perceptions.
 - The victims of human trafficking discussed in the focus group share a history of migration, which further increases their vulnerability to exploitation by criminal organizations. These crime groups present a highly organized and hierarchical structure, particularly evident when it comes to the Nigerian mob, which presents a pyramid configuration from chiefs to sex workers, managed by

group leaders that also control contact with clients. This organization is a para-statal system, which offers services, assistance and aid – mainly monetarily. The systemic violence as well as women’s exploitation in the street become harrowing experiences that cause the loss of identity in victims.

- Asking for help. Words in the text (translated): resignation, violence, normal, trust, police report, mediation, complex, relationships, sexual.
 - It is usually a mixture of factors that pushes women into asking for help, such as experience of abuse and violence, precarious health conditions, economic and social difficulties, a need for legal support and social protection. “Normal” is quite often used in the text: it sometimes happens that because of cultural reasons, and because of the victims’ lower educational level, they lack awareness about their rights and cognizance about the nature of what they are going through. The very first moment they try to ask for help, the aid relationship might fail: on the one hand, these people come from a complex system, made of abuse and exploitation, so they do not report traffickers (mostly because of fear); on the other hand, because they are already asking for help, building trust with the operators is crucial. But building trust could be hindered both by linguistic and cultural barriers. While the therapeutic value of narration is often employed to assist in case of intimate partner violence, this tool cannot always be used for victims of human trafficking because of the complexity and gravity of the abuse. However, long-term aid measures that focus on the autonomy of people trying to get out, such as job orientation (that also offers psychological and economic independence), can counter the forced subordination to traffickers.

4.6 Conclusions

The aim of this work was to use quantitative text analysis tools to evaluate the record's transcript of a focus group on GBV among victims of sex trafficking. This analysis highlighted how these victims mainly come from low socio-economic environments, heavily characterised by poverty, low education and multi-problematic families. In these contexts, migration is sought after for better living conditions and opportunities. Here it is easy that traffickers often affiliated with highly organised crime groups exploit these disadvantages in their favour.

The focus group showed how human trafficking is a systemic phenomenon, powered by less than effective migration laws and lack of care from state institutions. Indeed, even though the conversation was centred around a criminal act, the lack of mentions to “police”, “institutions” and “justice” – even though a lawyer was part of the group – really stands out. The only time institutions were mentioned at all was to highlight how very little interested in the phenomenon institutions are.

Using advanced methodologies such as text mining lends a hand into understanding and outlining underlying dynamics to complex phenomena discussed in focus groups. Visualising texts through word-clouds and sentiment analysis tools allows for a clearer interpretation of results.

This work, of course, does not lack limitations. The analysis is mainly exploratory and does not deal with more complex text analysis techniques that also study the relationship between words or methods that show how the tone shifts during the whole conversation. These methods could, however, represent a possibility of further development, to better understand the underlying dynamics of human trafficking.

References

Ali, P., & Rogers, M. M. (2023). *Gender-based violence: A comprehensive guide*. Springer.

- Armstrong, L., & Abel, G. (2020). *Sex work and the New Zealand model: Decriminalisation and social change*. Policy Press.
- Bloch, A., & McKay, S. (2013). Hidden dishes—how food gets on to our plates: Undocumented migrants and the restaurant and takeaway sector. *Journal of Workplace Rights*, 17(1).
- Centers for Disease Control and Prevention. (2018). Data collection methods for evaluation: Focus groups. <https://www.cdc.gov/healthyyouth/evaluation/pdf/brief13.pdf>
- Connelly, L. (2016). Debates on prostitution: An introduction to feminist politics and their influence upon international policy and practice. In F. Jacob (Ed.), *Prostitution: A companion of mankind*. Peter Lang.
- Feinerer, I., & Hornik, K. (2024). Tm: Text mining package [R package version 0.7-13]. <https://CRAN.R-project.org/package=tm>
- Fellows, I. (2018). Wordcloud: Word clouds [R package version 2.6]. <https://CRAN.R-project.org/package=wordcloud>
- Jockers, M. L. (2015). Syuzhet: Extract sentiment and plot arcs from text. <https://github.com/mjockers/syuzhet>
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436–465.
- Nielbo, K. L., Karsdorp, F., Wevers, M., Lassche, A., Baglini, R. B., Kestemont, M., & Tahmasebi, N. (2024). Quantitative text analysis. *Nature Reviews Methods Primers*, 4(1).
- Payton, F. C., Yarger, L. K., & Pinter, A. T. (2018). Text mining mental health reports for issues impacting today’s college students: Qualitative study. *JMIR Mental Health*, 5(4).
- Team, R. C. (2023). R: A language and environment for statistical computing. <https://www.R-project.org/>
- United Nations Convention against Transnational Organized Crime. (2000). November 15, 2000. <https://www.unodc.org/documents/treaties/UNTOC/Publications/TOC%20Convention/TOCebook-e.pdf>
- United Nations Office on Drugs and Crime. (2022). Global report on trafficking in persons. https://www.unodc.org/documents/data-and-analysis/glotip/2022/GLOTiP_2022_web.pdf

- United Nations Office on Drugs and Crime. (2024). Human trafficking. <https://www.unodc.org/unodc/en/human-trafficking/human-trafficking.html>
- Van Meeteren, M., & Hiah, J. (2020). Self-identification of victimization of labor trafficking. In *The palgrave international handbook of human trafficking* (pp. 1605–1618). Springer.

Conclusions

This thesis has explored the overarching goal to study the determinants of intimate partner violence (IPV) and HIV prevalence in East and Southern Africa, as well as their association. As shown in Table 4.1, the multiple studies carried out are the result of different research objectives, questions and hypotheses, all pursued and tested by using different techniques.

In the first case study, the main goal was the construction of a model-based composite indicator of polyvictimisation of intimate partner violence (Logie et al., 2019; Okumu et al., 2021), using the ecological model as the theoretical framework for it (Centers for Disease Control and Prevention, 2015). Polyvictimisation is a critical concept in understanding IPV, as it acknowledges the multiple forms of abuse a person may experience simultaneously.

Using data from eleven countries in Sub-Saharan Africa, a structural equation model (Muthén, 1984; Muthén et al., 1997) was then estimated, which is grounded in the ecological model's multidimensionality. By considering protective and risk factors at different levels of influence, the model demonstrates how various societal, community, relationship, and individual factors contribute to the prevalence of polyvictimisation of IPV.

The evidence provided in this first article demonstrate the relevance that women's personal history of violence might have in their future relationships as well - thus, confirming the association between the intergenerational transmission of violence and revictimisation processes and IPV. The consequent composite indicator may help bring to light undetected victims of IPV, thereby it could be a possible solution to the under-reporting of violence experience.

CONCLUSIONS

Hence, since the model uses a set of explanatory variables, it could guide policy makers toward specific areas of intervention by identifying “where” to invest the most to mitigate the phenomenon at hand in countries at greatest risk. For example, the results suggest Sierra Leone, Liberia and Uganda require major interventions to fight this phenomenon.

In Chapter 2, fractional logistic regression models (Papke & Wooldridge, 1996) were used to explore the association between neighbourhood dynamics and HIV prevalence. The Sub-Saharan countries the data were collected in are quite relevant when it comes to the study of HIV prevalence, since more than 29% of people living with HIV worldwide live there despite accounting for less than 3% of the global population (UNAIDS, 2024b).

The models revealed significant variation in HIV prevalence across different neighbourhoods, highlighting the importance of social factors. For example, neighbourhoods with higher levels of education exhibit lower HIV prevalence. Moreover, while no causal effect can be proved, the results show that neighbourhoods with a greater number of children experience a lower HIV prevalence, which could be explained by policies that promote access to treatment therapy – not only during but also after pregnancy. Neighbourhoods with higher HIV testing coverage also present higher HIV prevalence.

These particular results could be of interest to policy makers seeking to create HIV prevention programs in high HIV prevalence countries. Programs promoting access to antiretroviral therapy for people living with HIV or to pre-exposure prophylaxis to people who are at risk of infection could be useful if reducing the new number of HIV infections is the goal. Moreover, high HIV testing coverage rates in high HIV prevalence neighbourhoods could be a sign of social responsibility among people engaging in risky sexual behaviours.

Chapter 3 builds, in some ways, from the results obtained in Chapter 1 and 2. In this work, the goal was to study the relationship between IPV and HIV prevalence through a spatial lens, using the theoretical framework of the ecological model.

Table 4.1: Recall of research objectives, questions, hypotheses, data, methods and results.

Chapter	Objective	Research question	Hypothesis	Data	Methods	Results
2	The construction of a model-based polyvictimisation of intimate partner violence for Sub-Saharan countries within the framework of the ecological model.	What are the factors that shape intimate partner violence when considering societal, community, couples' and individual characteristics as in the framework of the ecological model?	H1. There is an association - whether positive or negative - between women's empowerment and exposition to abuse	Module on domestic violence from the Demographic and Health Survey: Benin 2017-2018, Burundi 2016-2017, Cameroon 2018, Gambia 2019-2020, Liberia 2019-2020, Mali 2018, Nigeria 2018, Rwanda 2019-2020, Sierra Leone 2019, Uganda 2016, Zambia 2018 Married women, age 15-49	The composite indicator was built using a structural equation model based on four latent variables with a reflective approach. Since most of the manifest variables are dichotomous, the diagonal weighted least square estimator with robust standard errors was chosen.	Significant and negative association between empowerment and intimate partner violence: the more empowered women are, the less abused they are.
			H2. Socio-economic deprivation makes women more vulnerable to abuse			Significant and negative association between socio-economic deprivation and intimate partner violence: more isolated and poorer women are exposed to more violence.
			H3. Socio-economic deprivation hinders women's empowerment within couples			Significant and negative association between socio-economic deprivation and women's empowerment: more isolated and poorer women have less power over their own life.
			H4. Exposure to violence through intergenerational transmission of violence and revictimisation processes make women more likely to be abused			Significant and positive association between exposure to violence and current experiences of abuse: women with a history of violence are more likely to be abused by the partner.
3	Assessing how HIV diffusion in Sub-Saharan African is associated with neighbourhood characteristics.	What are the neighbourhood characteristics that facilitate HIV diffusion? Does the national cultural and social context play a role in these associations?	H1. Socio-economic factors of neighbourhoods characterise the presence of HIV	HIV testing results and GIS data from the Demographic and Health Survey: Angola 2015-2016, Burundi 2016-2017, Cameroon 2018-2019, Chad 2014-2015, Guinea 2018, Malawi 2015-2016, Mozambique 2015, Sierra Leone 2019, Senegal 2017, South Africa 2016, Zimbabwe 2018-2019 Neighbourhoods, residents aged 15-64	Operationalisation of neighbourhoods using a spatially aware k-means clustering technique of the clusters of households (based on census units). Variables were then measured at the neighbourhood level. Fractional logistic regression models were implemented to evaluate the association of HIV prevalence with selected covariates.	Highly educated neighbourhoods are less vulnerable to HIV, with noticeable differences between countries; no significant associations were detected for wealth and living in rural areas.
			H2. The demography of neighbourhoods shapes HIV			A higher number of children in a community is associated with lower HIV prevalence - with differences between countries.
			H3. Sexual behaviours in a community determine its HIV prevalence			Higher HIV testing coverage is associated with greater HIV diffusion. Higher proportions of people who have never engaged in sexual relationships and higher proportion of people who delay their first intercourse are associated with lower HIV prevalence.

CONCLUSIONS

Table 4.2: Recall of research objectives, questions, hypotheses, data, methods and results.

Chapter	Objective	Research question	Hypothesis	Data	Methods	Results
4	Evaluating the spatial relationship between intimate partner violence and HIV.	Is intimate partner violence associated with HIV? Does their relationship shape how HIV is spread in a country?	H1. HIV prevalence is associated with intimate partner violence	HIV testing results, module on domestic violence and GIS data from the Demographic and Health Survey: Zambia 2018, Zimbabwe 2015	Logistic regression model; Bayesian hierarchical model with the INLA-SPDE approach with Beta distributed response (after censoring)	Intimate partner violence, in particular the combination of physical and emotional violence, increases the odds for women to be HIV positive.
			H2. The association between HIV and intimate partner violence creates spatial patterns of disease in a country	Married women, aged 15-49; Neighbourhoods, resident married women, aged 15-49		
5	Examining the transcript of a focus group about gender violence among victims of international human trafficking	Are there common themes that can be detected throughout the text? What feelings does the transcript convey?	H1. Common themes and concepts can be identified through the text.	Focus group transcript One moderator, two researchers in migration studies, four key stakeholders from the third sector operating in the province of Palermo (Italy)	Exploratory wordclouds Sentiment analysis	Common themes arisen concern "violence and abuse", "women and minors", "trafficking and exploitation", "origin and destination". The words associated to these themes tell a story about the victims of international human trafficking met by the stakeholders. Prevalence of negative emotions, such as sadness, fear, anger and disgust. Trust is the most common sentiment due to a dual mechanism: trust refers to the deception used to lure victims into trafficking, but it is also the foundation of the rapport between operators and victims seeking a way out.
			H2. The vernacular used during the focus group conveys distinguishable feelings and sentiments.			

In this chapter, there is a focus on Zambia and Zimbabwe because of their similar national HIV prevalence. Two methodologies were employed to assess the relationship between IPV and HIV prevalence: a logistic regression model taking individual women (aged 15-49 years) as statistical units, and Bayesian hierarchical models within the integrated nested Laplace approximation with the stochastic partial differential equation approach (Gómez-Rubio, 2020; Lindgren & Rue, 2015; Moraga, 2023), which assumes neighbourhoods of women as statistical units.

The analysis confirmed that being abused by the current partner makes women more likely to be HIV-positive, especially when women are abused both physically and emotionally, thus also confirming that polyvictimisation exposes to increased risks. The extension of the analysis with a spatial model allows to check whether this association has geographical implications, by identifying whether it produces HIV prevalence hotspots. Maps could represent an invaluable tool for policy-makers of the international community, given the ease of interpretation (if not ease of production).

Finally, Chapter 4 is focused on the investigation of gender-based violence among victims of sex trafficking from the point of view of operators of the third sector. Although the research question guiding the focus group is not directly tied to the main objective of this thesis, the purpose of organizing the focus group was to determine whether the findings and literature concerning women living in Sub-Saharan Africa remain relevant when these women live outside their countries of origin. Because the operators that were part of the focus group work with a very specific category of women coming from Africa, the conversation shifted to the gender-based violence experienced by victims of human trafficking. For this manuscript, a mixed methodology was employed, by using the results of quantitative text analysis methods as a basis for the qualitative evaluation of the transcript.

The focus group discussion provided valuable insights into the experiences of GBV among victims of human trafficking from Sub-Saharan Africa. The findings emphasise the importance of culturally sensitive and comprehensive support services that address the unique needs of trafficking victims. The operators also highlighted the need for improved policies and resources to

CONCLUSIONS

assist victims in their recovery and reintegration as well as more involvement from institutions, and the justice system in particular.

Discussion. In addition to the concluding remarks from each work in this thesis, it is essential to emphasise the collective significance of the contributions as a whole. Thus, some reflections around the determinants of IPV and HIV prevalence, as well as around their association, must be provided.

Overall, this thesis was written keeping social sustainability as its core framework. All the articles that make it provide information that could be useful to policy makers and to organisations - both national and international - that work in Sub-Saharan Africa to favour the implementation process of the United Nations' Sustainable Development Goals (SDGs).

For example, Chapter 2 suggests that, while it is usually practices of women's education and outreach that are used to fight intimate partner violence, empowering women does not necessarily mean that they will not be abused or that they will leave a violent environment. While women's empowerment is an essential target by its own - as stated in Goal 5 "Gender Equality" of the SDGs - it should be pursued alongside other initiatives to effectively address and end intimate partner violence. For example, addressing men directly, by educating them on the effects of violence, could represent a viable solution in a more proactive fight against IPV.

Another result that is somehow in contrast with earlier research, but that could still be of relevance to the international community, concerns the association between IPV and HIV prevalence. In the past, both scholars and international organisations have highlighted the association between these two phenomena. Indeed, the UNAIDS clearly points to the elimination of gender-based violence and intimate partner violence as one of the social sustainability targets that could aid in the reduction of new HIV infections as in the framework of the SDGs (UNAIDS, 2024a). However, the results presented in this work do not tell the same story: while women who declare that their husbands have abused them both physically and emotionally do experience higher likelihood of living with HIV, the association is not as strong as the current literature actually states, and in Zambia the association is not

significant at all. This implies that although a gender perspective should be maintained in HIV prevention programs as to make the proper adjustments when dealing with people of different genders, they should not only address gender dynamics as the sole responsible of HIV diffusion. These programs could, perhaps, focus on the diffusion of Pre-Exposure Prophylaxis (PrEP), which has been shown to reduce the risk of infection of HIV from intercourse by 99% (Centers for Disease Control and Prevention, 2024).

This suggestion could be used in conjunction with methodologies such as those presented in Chapter 3, as to quickly identify areas in higher need of intervention.

Some considerations... and where to go from them. Aside from the results presented in the different works in this thesis - and their consequent policy implications - some considerations about the data used need to be made.

All the data used in this thesis, with the exception of those used in Chapter 4, are part of the Demographic and Health Survey (DHS) Program (Croft et al., 2018). While the accuracy of the data on HIV is guaranteed by the fact that the HIV of respondents is assessed using clinical instruments, the data on intimate partner violence is still subject to the usual issues of social desirability and other forms of self-report bias connected to this kind of data.

Although a Demographic and Health Survey protocol ensures that interviewers are certain that safety measures are in place before collecting data on domestic violence — such as confirming that no male figure is present during the interview — social desirability bias from respondents may still compromise the reliability of the data themselves, potentially leading to an underestimation of the issue. This, however, may also reflect a broader difficulty for women in openly discussing their experiences with abuse: indeed, under-reporting may happen for many reasons, including cultural contexts where some types of violence perpetrated by intimate partners is viewed as normal, when a woman fears retaliation upon disclosure, or where the level of stigma around such violence in the given society is high (Bloom, 2008). This reporting bias is not fixable and it represents a common flaw in domestic

CONCLUSIONS

violence data.

Another consideration to be made is about the sample design of the main survey. The Demographic and Health Survey is currently at its eighth wave of data collection, and usually multiple surveys are available for countries, collected in various years. And while longitudinal data collection across respondents is not envisaged, the general sample design adopted - which is usually based on a multiple stage sample, with some form of cluster sampling at the first stage - does not allow to use these smaller territorial units as a base for longitudinal-type of analyses that use clusters as statistical units either. The availability of any kind of longitudinal data could allow to derive causal evaluations between phenomena, which is not possible to do with cross-sectional data. Thus, the Demographic and Health Survey program could take into consideration the application of sample schemes based on the same, or at the very least comparable, clusters of households in the following data collection waves.

Limits and Future developments. Although it is a crucial step in a larger ongoing process, this thesis is not necessarily the final and complete outcome of this research project, nor is it exempt from limitations. It is certainly a moment of progress, but it leaves space for further exploration.

For example, a limitation of the analyses in the core chapters of this thesis is that all data, including information about respondents' partners, is collected solely from the respondents rather than from the partners themselves. However, the DHS gives the possibility of merging data to recreate households' structures, with particular attention to couples and children that form nuclear family (not necessarily representing households as a whole) (Croft et al., 2018). Thus, a first possibility of future development could be integrating couple's data in the analyses. This possibility was not explored in my research, but modelling IPV and or HIV by also controlling for characteristics directly observed on the respondents' partners could provide novel results.

Another limit is given by the restricted use of the ecological model dimensions only to variables directly available from the Demographic and Health

Survey. Integrating the data with external data sources could, indeed, boost the results of the analyses, by exploiting the ecological model to its fullest potential.

Moreover, the analysis of the determinants of IPV presented in Chapter 2 will also be developed by integrating these results with the spatial dimension of violence. By doing so, it will be possible to identify specific areas where intimate partner violence is more prevalent, or what could be termed "hotspots" in epidemiological terms. These hotspots can provide critical insights into localised patterns of violence, helping to uncover regions or communities that are particularly vulnerable.

Such spatial mapping of IPV could be extremely valuable for stakeholders, including policymakers, local authorities, and organizations working to combat domestic violence. By pinpointing areas with higher levels of violence, resources can be more effectively allocated, and intervention programs can be better targeted to the regions most in need. Furthermore, this information could enhance the integration of domestic violence prevention into existing community initiatives, making it a key component of broader social and health programs already active in the area.

Another possibility worth mentioning is the mixed methodology employed in Chapter 4, where text data mining served as the foundation for a qualitative evaluation of a focus group transcript. This integration of quantitative techniques with qualitative analysis offers a robust framework for gaining deeper insights from textual data. By identifying patterns and trends through computational analysis, followed by a more nuanced, context-rich qualitative interpretation, researchers can uncover both broad trends and subtle, individualised responses within focus group discussions. This kind of mixed methodology holds great potential for psychologists, sociologists, and other qualitative researchers who rely heavily on focus groups and interviews, as it allows them to handle texts without sacrificing the depth and richness of qualitative insights.

References

- Bloom, S. S. (2008). Violence against women and girls: A compendium of monitoring and evaluation indicators. <https://www.measureevaluation.org/publications/pdf/ms-08-30.pdf.html>
- Centers for Disease Control and Prevention. (2015). The social-ecological model: A framework for prevention [Accessed 2024-01-30]. <https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html>
- Centers for Disease Control and Prevention. (2024). Pre-Exposure Prophylaxis (PrEP) [Accessed: 2024-10-30]. <https://www.cdc.gov/hiv/risk/prep/index.html>
- Croft, T. N., Marshall, A. M. J., Allen, C. K., et al. (2018). *Guide to DHS statistics* (tech. rep.). ICF. Rockville, Maryland, USA.
- Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. Chapman & Hall/CRC Press.
- Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19), 1–25. <https://doi.org/10.18637/jss.v063.i19>
- Logie, C. H., Okumu, M., Mwima, S., Hakiza, R., Irungi, K. P., Kyambadde, P., et al. (2019). Social ecological factors associated with experiencing violence among urban refugee and displaced adolescent girls and young women in informal settlements in Kampala, Uganda: A cross-sectional study. *Conflict and Health*, 13(1), 1–15.
- Moraga, P. (2023). *Spatial statistics for data science: Theory and practice with R*. Chapman & Hall/CRC Press.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132.
- Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* (tech. rep.). Unpublished technical report.

- Okumu, M., Orwenyo, E., Nyoni, T., Mengo, C., Steiner, J. J., & Tonui, B. C. (2021). Socioeconomic factors and patterns of intimate partner violence among ever-married women in Uganda: Pathways and actions for multicomponent violence prevention strategies. *Journal of Interpersonal Violence*, 08862605211021976.
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Economics*, 11(6), 619–632.
- UNAIDS. (2024a). *AIDS and the Sustainable Development Goals* [Accessed: 2024-10-06]. https://www.unaids.org/en/AIDS_SDGs
- UNAIDS. (2024b). Aidsinfo | unaids [Accessed: 2024-10-01]. <http://aidsinfo.unaids.org/>