THESIS

Evaluating the impact of a sport-for-development intervention on the physical and mental health of young adolescents in Gulu, Uganda - a post-conflict setting within a low-income country

Justin Richards

University College, University of Oxford

Submitted for the degree of Doctor of Philosophy (Medical Sciences) – Michaelmas term, 2011

TABLE OF CONTENTS

PAGE

ABSTRACT 1

ACKNOWLEDGEMENTS 2

DECLARATION OF AUTHORSHIP 4

GLOSSARY: Abbreviations and Acronyms 5

CHAPTER 1: Introduction 8

PART I – PROBLEM DEFINITION AND SOLUTION GENERATION

CHAPTER 2: Background and Rationale for Research 16

CHAPTER 3: Case-Study Site – Gulu, Uganda 42

PART II – RESEARCH OBJECTIVES

CHAPTER 4: Research Hypothesis and Questions 54

PART III – PRETESTING METHODS AND MATERIALS

CHAPTER 5: Feasibility and Reliability of Adapted Protocols to Measure the Physical Fitness, Mental Health and Physical Activity of Young Adolescents in Gulu, Uganda 61

PART IV – COMMUNITY NEEDS ANALYSIS

CHAPTER 6: Baseline Measurements for Physical Fitness and Mental Health of Young Adolescents in Gulu, Uganda 106

PART V – INNOVATION TESTING

CHAPTER 7: Reach of a Sport-for-Development Intervention for Young Adolescents in Gulu, Uganda 158

CHAPTER 8: Impact of a Sport-for-Development Intervention on the Physical Fitness, Mental Health and Physical Activity of Young Adolescents in Gulu, Uganda 179

PART VI – RESEARCH FINDINGS AND IMPLICATIONS

CHAPTER 9: Final Discussion and Conclusions 237

REFERENCES (Thesis) 276

PAGE

APPENDICES

A1: Methods for Systematic Literature Review **291**

A2: Location of Study **299**

A3: BMI-for-Age and Height-for-Age Measurement Protocol **303**

A4: Standing Broad Jump Measurement Protocol **306**

A5: Multi-Stage Fitness Test Measurement Protocol **308**

A6: APAI Questionnaire Measurement Protocol **311**

A7: Accelerometry Measurement Protocol **313**

A8: Reliability of Physical Fitness Outcomes (Assumptions) **316**

A9: Reliability of Mental Health Outcomes (Assumptions) ..... **-22**

A10: Reliability of Mental Health Outcomes (Results) **329**

A11: Comparison between Stratified Levels for Age, Division and School at M1 .. **333**

A12: Comparison to Normative Data at M1 (Assumptions) **349**

A13: History of Abduction at M1 (Assumptions) **355**

A14: Relationship between Outcome Variables at M1 (Plots) **366**

A15: Relationship between Outcome Variables at M1 (Assumptions) **376**

A16: Gum Marom Kids League Model (Intervention) **392**

A17: Reach of the Intervention (Assumptions) **410**

A18: Impact of the Intervention – Within-Group Differences (Assumptions) **421**

A19: Impact of the Intervention – Between-Group Differences (Assumptions) **450**

A20: Accelerometry – MVPA Between-Group Differences (Assumptions) **465**

A21: Accelerometry – VPA Between-Group Differences (Assumptions) **468**

A22: "Heat Maps" for 10-Minute Bouts of MVPA **471**

A23: "Heat Maps" for 10-Minute Bouts of VPA **489**

A24: Publication and Conference Plan **494**

A25: Published Articles **497**

A26: Accepted Conference Abstracts **507**

REFERENCES (Appendices) **509**

Evaluating the impact of a sport-for-development intervention on the physical and mental health of young adolescents in Gulu, Uganda – a post-conflict setting within a low-income country

Justin Richards, University College

Submitted for the degree of Doctor of Philosophy (Medical Sciences) – Michaelmas term, 2011

ABSTRACT

Introduction

Physical inactivity is thought to contribute to the emergence of non-communicable diseases in post-conflict settings of low-income countries. Sport-for-development (SfD) organisations in these regions claim to improve the health of programme participants. However, there is a paucity of supporting evidence. I assessed the impact of a voluntary community-based SfD intervention on the physical activity (PA), physical fitness (PF) and mental health (MH) of adolescents in Gulu, Uganda.

Methods

The Acholi Psychosocial Assessment Instrument (APAI), standing broad jump (SBJ), multi-stage fitness test (MSFT) and BMI-for-age (BFA) were adapted to the local context. I tested their feasibility and reliability with a repeat-measures design (n=70). A cross-sectional analysis of a random sample was used to assess the local needs and establish the PF and MH of the adolescents reached by the intervention (n=1464). This was also the baseline assessment for the impact evaluation. It comprised a randomised control trial (n=144) nested within a cohort study (n=1400) and triangulated by cross-sectional assessment of PA using accelerometry (n=54).

Results

The adapted PF and MH measures demonstrated good intra-tester reliability (ICC>0.75). Adolescents in Gulu predominantly had "healthy" BFA (>90%). They performed better than global norms for the SBJ ($p<0.001$), but worse for the MSFT ($p<0.05$). The girls who registered for the intervention had higher PF at baseline ($p<0.05$) and experienced no significant benefits when compared to the community. The aerobic capacity of the boys intervention group increased relative to the community ($p<0.01$), but was not significantly different to the trial control group whose PF also improved. The PA results concurred with this finding. Boys in the intervention group experienced a deterioration in MH relative to their peers ($p<0.05$).

Implications

It is feasible to apply rigorous evaluation methods to SfD interventions. Although adolescents in Gulu have poor aerobic capacity, a voluntary programme may not reach those at risk. Interpreting the impact evaluation was limited by a lack of programme development theory, but suggested that opportunities for non-competitive play may confer PF benefit without harming MH. Further investigation is warranted.

ACKNOWLEDGEMENTS

I wish to acknowledge all of those people who have made a contribution to my DPhil work and general well-being in the past three years – thank you. However, I also would like to make particular mention of a few people who have been a special part of my DPhil journey.

First of all, I thank the members of the British Heart Foundation Health Promotion Research Group in the Department of Public Health at the University of Oxford. In particular, the contribution of my supervisor – Charlie Foster – must be acknowledged. I have cherished the long leash and confidence that Charlie has afforded me to follow my interests. I have been fortunate that his support has extended beyond the role of academic supervisor and I have enjoyed his personal friendship over the past three years.

There are two other people in Oxford to whom I am greatly indebted. Uwe Ackermann was the first to offer me the "Oxford experience" and was the coordinator of my MSc in 2005-07. He continues to be a friend and mentor whose opinion I hold in the highest regard. I would also like to thank Tom Lickiss who has been a great mate and unshakeable running partner throughout my time in Oxford. His willingness to share the pain of Boars Hill in the depths of an English winter pales into insignificance when compared to the rants about work and life that I subjected him to along the way.

This DPhil would not have been possible without the tremendous support I received when completing my fieldwork in Uganda. The partner organisations – OA Projects, Youth Coalition for Peace and The Kids League – all contributed to the programme that was evaluated in Gulu. A special mention goes to Gavin Hollett whose insight and unwavering support for football in Uganda is admirable. I would also like to acknowledge all the hard

work of the staff and volunteers in the Gum Marom Kids League team in Gulu. I appreciated the leadership of Ocaya Opiyo Bangi and thank him for his patience and good humour.

The research team in Gulu were a fantastic group of young men – Martin, Kenneth, Godfrey, Patrick - whose enthusiasm and adaptability were crucial to making this study a success. They were led by Aloya Michael whose ability to make things happen in Gulu continues to defy belief. His fortitude, energy, kindness and integrity in such difficult circumstances are just part of what makes him a truly remarkable individual. I sincerely appreciate everything that Michael has done for me and I consider it a privilege to have worked with him.

There are numerous people at home in Australia that I would like to mention. Luke "Muffler" O'Mara and Monte "Jedi" MacDiarmid have been invaluable sources of support and fantastic cycling buddies with whom I could let off some steam. I am indebted to Andrew "Sac" Szydzik, Brett "Davo" Davis and their better halves for humouring and putting up with me in their homes during the write-up process. I also appreciated the support of David "Buddha" Haynes whose visit to Oxford kept me sane during the initial systematic literature review.

I consider myself incredibly fortunate to have a family whose support and love seem to have no bounds. I want to thank Mumma, Papsy, Haydes and Phoebes for their generosity at home and when visiting me in Europe. I have appreciated all the laughs, debate, regular phone calls and the continued encouragement to pursue my interests in faraway places. Thank you.

Finally, I want to thank the one person who has personally shared the DPhil journey with me on all three continents it has touched. This work could not have happened without Camille's ongoing support and advice. I am happy that I have been able to share this part of my life with someone who has been so compassionate, tenacious, loving and fun. Thanks tiger!

DECLARATION OF AUTHORSHIP

The idea to assess the impact of sport-for-development programmes on health outcomes was my own. I chose the themes and focus of the research and was the sole reviewer for the systematic review of the literature. I spent a total of nine months in Uganda on two separate fieldtrips to collect the data for this thesis. This was a challenging task and I was assisted by a team of five research assistants from Gulu who were led by Aloya Michael. John McMullen and Paul O'Callaghan conducted a parallel study at Laroo Boarding School during the initial stages of my research and contributed to collecting the mental health measurements for the pilot test. I conducted all of the statistical analyses after seeking advice from colleagues in the British Heart Foundation Health Promotion Research Group at the University of Oxford. Dr. Aiden Doherty provided technical assistance to analyse the accelerometer data using a programme that he had developed. All of the conclusions drawn from the data are my own.

This thesis has been written under the supervision of Dr. Charlie Foster who has been involved throughout the project. He contributed to initial discussions regarding the research objectives and advised me on establishing the networks required to achieve these aims. He also provided feedback on my methods, questioned my interpretation of the results and edited drafts of my written work. The final draft of this thesis was also edited by Camille Neyhouser and my parents - Ann and Roger Richards - who provided grammatical and structural advice.

Word count: 50,431 (excluding glossary, figures, tables, references and appendices)

GLOSSARY: Abbreviations and Acronyms

Δ – Change in...

1xRM – One repetition maximum

AB – Assessment briefing for teachers and pupils at selected schools

AI – Academic institution

APAI – Acholi Psychosocial Assessment Instrument

AT – APAI total

B – Boys

BFA – BMI-for-age

BIA – Bioelectrical impedance analysis

BMI – Body mass index

C – Selection of study population (11-14 years old attending school within Gulu municipality)

CC – Christ Church Primary School

CI – Confidence interval

CRC – Convention on the rights of the child

CVD – Cardiovascular disease

DOB – Date of birth

DT – Depression total

DXA – Dual-energy x-ray absorpiometry

EU – European Union

EYES – European Year of Education through Sport

EXI – Expert interviewer for the APAI

G – Girls

GMKL – Gum Marom Kids League

GPol – Gulu Police Primary School

GPrs – Gulu Prison Primary School

GPub – Gulu Public Primary School

GSHS – Global School-based Student Health Survey

GTwn – Gulu Town Primary School

HFA – Height-for-age

HIC – High-income country

HOV – Homogeneity of variance

HR – Holy Rosary Primary School

I – Intervention (i.e. the GMKL season)

ICC – Intra-class correlation coefficient

IDP – Internally displaced person

ILO – United Nations International Labour Organisation

IO – International organisation

IOC – International Olympic Committee

IW – MH assessment delivered in interview format at the time of P2

JSFD – Journal of Sport for Development

K – Kumu

KKL – Kampala Kids League

LC – Layibi Central Primary School

LL – Labour Line Primary School

LMIC – Low- and middle-income country

LRA – Lord's Resistance Army

M – Age in complete months

M1 – Initial case-study measurement

M2 – Follow-up measurement immediately after the GMKL finishes

M&E – Monitoring and evaluation

MDG – Millennium Development Goal

MET – Metabolic equivalent

MH – Mental health

MINEP – International conference of ministers and senior officials responsible for PES

ML – Ma lwor

MSFT – Multi-stage fitness test ("beep" test)

MVPA – Moderate-vigorous physical activity

NC – Adolescents register for the GMKL

NCD – Non-communicable disease

NGO – Non-governmental organisation

[O] – Eligible adolescents not registered in the GMKL

OAP – OA Projects

OxTREC – Oxford Tropical Research Ethics Committee

P – Par

P1 – Initial pilot measurement

P2 – First follow-up pilot measurement

P3 – Second follow-up pilot measurement

PA – Physical activity

PES – Physical education and sport

PF – Physical fitness

PP – Pece P7 Primary School

PrS – Private sector

PuS – Public sector

Q1 – MH assessment delivered in guided questionnaire format at the time of P1

Q2 – MH assessment delivered in guided questionnaire format at the time of P3

RA1 – Research assistant 1

RA2 – Research assistant 2

RA3 – Research assistant 3

RA4 – Research assistant 4

RC – Randomly select the GMKL participants from all adolescents who registered

RCI – Research chief investigator

RCT – Randomised control trial

RF – Risk factor

RS – Randomly select sample of schools within Gulu municipality

RTL – Research team leader

S – Selection of school and class for pilot testing

SAD – Swiss Academy for Development

SBJ – Standing broad jump

SD – Standard deviation

SDPIWG – Sport for Peace and Development International Working Group

SfD – Sport-for-development

SO – Sports organisation

SP – St Peter's Primary School

[T] – Eligible adolescents participating in the GMKL – intervention group

TKL – The Kids League

TT – Two tam

UDHS - Ugandan Demographic and Health Survey

UN – United Nations

UNESCO – United Nations Educational, Cultural and Science Organisation

UNOSDP – UN Office on Sport for Development and Peace

UPE – Universal Primary Education

V – Voluntary participation by students in selected pilot testing class

[W] – Eligible adolescents wait-listed in the GMKL – wait-list group

WHO – World Health Organisation

Y – Age in complete years

YCP – Youth Coalition for Peace

Z – Age estimation score using height and weight measurements

CHAPTER 1

Introduction

1.1) Purpose of chapter

The purpose of this chapter is to outline briefly how my previous experiences as a physiotherapist and international volunteer have shaped this DPhil. I will reflect on critical events that motivated me to commence physical activity (PA) research looking at sport-for-development (SfD) programmes in post-conflict and resource-poor communities. I will also present the evaluation framework that forms the backbone of this thesis and provide a brief overview of the ensuing chapters.

1.2) How it began – from clinical physiotherapy to sport-for-development research

"Those who think they have not time for bodily exercise will sooner or later have to find time for illness". [1]

This statement was made in 1873 by Edward Stanley during his address at Liverpool College describing "The Conduct of Life" [1]. He was the 15th Earl of Derby and a British Statesman who served as Secretary of State for Foreign Affairs. However, this comment showed considerable foresight into the role of PA in health.

I first came across this quote in 2003 when I was working as a clinical physiotherapist in Australia. I was searching for inspirational quotes related to exercise that would have some traction with my patients. Although I did not pin up the wisdom of Edward Stanley on the walls where I was working, it did start me thinking about PA and health promotion.

By the middle of 2004, I had grown frustrated with working as a physiotherapist. I felt limited by the one-on-one nature of the work and the dependence on individual compliance with home exercise programmes. I was also struggling with the curative approach of most physiotherapy and was becoming increasingly interested in preventative medicine. Having previously worked in private practice and hospital settings in a metropolitan area, I changed my workplace and began treating indigenous Australians and asylum seekers. This was my first experience of working with people from post-conflict and resource-poor settings. I found my work with adolescents during this period particularly rewarding. These experiences had a lasting impression that contributed to shaping the target population in this thesis.

The change in population reinvigorated my work, but did not appease the underlying frustrations I was experiencing with practising as a clinical physiotherapist. I took some time

off work and travelled to Papua New Guinea as a community development volunteer. This was followed by an extended period of travel in South America and Eastern Europe in 2005.

There were two common themes that emerged from this time abroad. Firstly, the easiest way for me to engage with local communities was through participating in local games and sport. This was particularly pertinent for connecting with adolescents and young men despite the language and cultural divides. Secondly, the daily lives of many people living in these countries involved levels of PA that could be assimilated with the patterns of elite sports training.

I embarked on my MSc in the Science and Medicine of Athletic Performance at the University of Oxford with the intention of applying the course content to the physically arduous lives of subsistence farmers in low- and middle-income countries (LMICs). The course coordinators – Uwe Ackermann and Tom Dawson – supported this approach. However, it was the final lecture of the MSc in 2007 that captured my attention. Charlie Foster and Steve Allender presented data about the health impact of PA in the United Kingdom and brought the words of Edward Stanley to life for me for the first time. The discussions with Charlie that ensued led to his role as the supervisor of this DPhil.

After completing the MSc I completed two internships to broaden my knowledge, experience and professional network. The first of these was at the World Health Organisation (WHO) headquarters in Geneva. I worked on developing PA recommendations for the Global Strategy on Diet, Physical Activity and Health. The focus of this work was on LMICs and I began to appreciate the global risk of non-communicable diseases (NCDs) and the emergence of the double burden of disease.

The second internship was at the headquarters of the United Nations Educational, Scientific and Cultural Organisation (UNESCO) in Paris. I contributed to the projects of the team that focused on peace and development through physical education and sport (PES). During this period I made three cursory observations:

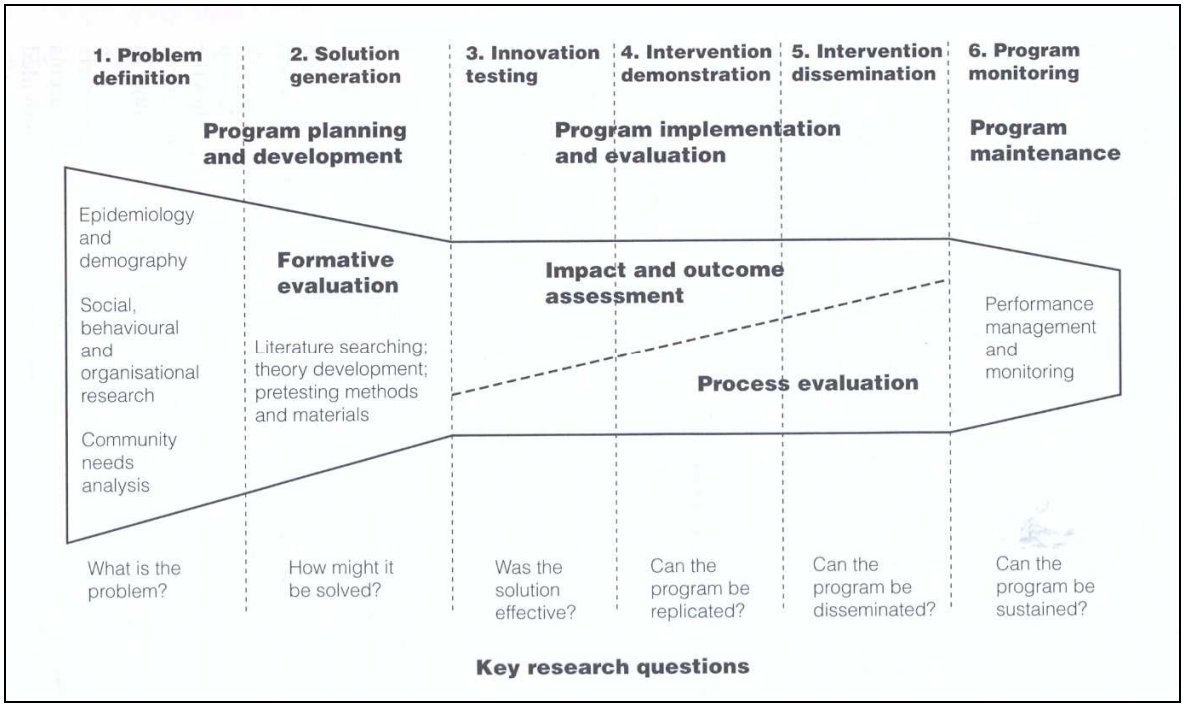
- There were a burgeoning number of organisations in the SfD sector;
- The rhetoric surrounding these organisations described SfD interventions that were essentially positive and cut across a diverse range of problems;
- There was a dearth of evidence supporting these claims.

I started my DPhil at a time when the evidence base for SfD organisations was in its infancy. I was initially interested in the novelty of peaceful play in post-conflict populations and the health implications of SfD interventions in these regions. Practically applied research also appealed to me because of its tangible outputs at the intervention level. I did not want to be confined to my desk and I was very interested in the prospect of fieldwork. Over the past three years these ideas have evolved into the programme evaluation presented in this thesis.

1.3) Approach to this DPhil – focusing on programme evaluation

Nutbeam et al (2006) described a theoretical framework for evaluating the various stages of health promotion interventions (FIGURE 1.1) [2].

FIGURE 1.1: The stages of evaluation for health promotion interventions



Source: Nutbeam et al (2006) [2]

I use this model as the health promotion theoretical framework for my research. In this thesis, I complete the first three phases for the evaluation of a SfD programme. However, the progression from "problem definition" → "solution generation" → "innovation testing" is not linear in my case-study. The components of these stages are dispersed throughout the thesis as each "key research question" is addressed.

In Part I, I describe the emerging problem of NCDs in LMICs and review the potential solutions provided by SfD interventions in post-conflict settings. I consider the role of PA and physical fitness (PF) in this context and establish a theoretical model for delivering

sustainable SfD interventions (CHAPTER 2). These facets of programme planning and development are also assessed specifically for the case-study site in Gulu, Uganda (CHAPTER 3).

In Part II, I articulate the research objectives (CHAPTER 4). These are informed by the gaps identified when defining the problem and reviewing the evidence for SfD programmes.

In Part III, I pretest selected methods and materials. They are modified according to the local context in Gulu and pilot tested for measurement quality and feasibility (CHAPTER 5).

In Part IV, I complete components of a community needs analysis using the pretested metrics in a cross-sectional sample. I also use these results to assess the potential of a SfD intervention on the outcomes of interest in Gulu according to existing programme theories (CHAPTER 6).

In Part V, I test the impact of an innovative community-based SfD intervention. I assess who is participating in the programme to describe its reach (CHAPTER 7). I also collect repeat measures after the intervention to evaluate the impact of the SfD programme on the outcomes of interest (CHAPTER 8).

In Part VI, I discuss the common themes that pervade the thesis and are associated with the original research objectives. I interpret the results in relation to the current evidence base and the existing SfD theoretical models. I also consider the strengths and weaknesses of my research and their implications on the evaluation outcomes. Several recommendations for SfD programme modification, replication and dissemination are considered and I also suggest various avenues for future research (CHAPTER 9).

1.4) Summary box

INTRODUCTION

- My interests in research have been shaped by various professional and personal experiences working in the healthcare sector, volunteering internationally, studying in Oxford and interning at UN agencies.

- The theoretical framework for evaluation of health promotion interventions, described by Nutbeam et al (2006), provides a backbone for the contents of this thesis.

PART I

PROBLEM DEFINITION AND SOLUTION GENERATION

CHAPTER 2

Background and Rationale for Research

2.1) Purpose of chapter

The purpose of this chapter is to outline the public health issues addressed in this thesis and describe the activities of stakeholders currently delivering interventions in this area. I will also review the existing evidence base and establish the rationale for my research objectives, methods and fieldwork.

2.2) Burden of non-communicable diseases and mental health disorders

WHO estimates that more than 35 million people died from non-communicable diseases (NCDs) in 2005. This represented 60% of all deaths worldwide of which 80% occurred in LMICs [3]. In several LMICs, the emergence of NCDs is part of an epidemiological transition as the prevalence and incidence of communicable diseases decreases [4,5]. However, in many parts of Africa, high levels of communicable diseases have persisted and the rise in NCDs has created a double burden of disease [6].

A silent epidemic is also occurring in mental health (MH) disorders and they represent a substantial proportion of the world's disease burden [7]. MH has been defined as "a state of well-being in which every individual realises their potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to their community" [8]. It is estimated that, in any given year, at least 20% of children and adolescents experience MH disorders that are consistent across cultures [9,10]. Although MH problems are common throughout the world, they are neglected in most African countries where they attract less than 1% of health expenditure [11,12].

It has been suggested that the rapid increase in urbanisation observed globally is contributing to the emergence of NCDs and MH disorders. Societal, environmental and behavioural factors associated with urbanisation influence the health of populations living in these settings [4,5]. The loss of traditional social support mechanisms when people relocate to an urban environment increases the risk of MH disorders [13]. Urbanisation is also associated with behavioural changes that reduce energy expenditure due to lower PA participation levels [4,5,14].

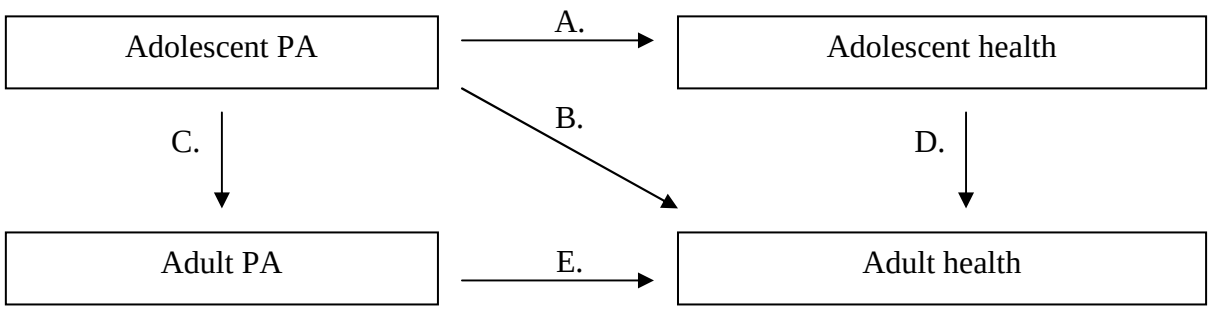
2.3) Impact of physical activity on health-related outcomes

PA has been defined by WHO as "any bodily movement produced by skeletal muscles that requires energy expenditure" [15].

Physical inactivity (i.e. lack of PA) is an independent modifiable risk factor (RF) for many common NCDs [3]. In 2002, it was estimated to be directly responsible for 1.9 million deaths and an increasing level of morbidity throughout the world [16]. More recently, physical inactivity was described as the fourth leading RF for global mortality and in 2010 was responsible for approximately 6% of all deaths around the world [17].

Participation in adequate PA reduces the risk of several NCDs in adults [16-20]. However, the relevance of this research to young adolescents is limited by the applicability of the health-related outcome measures used when investigating this relationship. The majority of the PA evidence in adults assesses mortality and disease end points not commonly experienced by young people. Research for adolescent populations instead tends to focus on RFs that are more difficult to measure and can only partially predict health in adulthood [21,22]. Blair et al (1989) depicted the potential links between adolescent PA and health outcomes at various stages in the life cycle (FIGURE 2.1) [23].

FIGURE 2.1: Links between PA and health at various stages in the life cycle



Source: Modified from Blair et al (1989) [23]

There is a paucity of longitudinal studies demonstrating a direct link between adolescent PA and adult health (link B) [21,24]. The logistical challenges of conducting these studies over a lifespan and the relative infancy of PA research have contributed to these limitations in the literature. However, there is some evidence to suggest an indirect relationship between adolescent PA and adult health outcomes (i.e. via links A → D and links C → E).

There have been several reviews that have assessed the relationship between adolescent PA and adolescent health (link A). There is evidence to suggest that appropriate PA during adolescence is associated with better outcomes for [17,21,22,24-30]:

- Adiposity / overweight / obesity
- Cardiovascular health
    - Metabolic syndrome
    - Blood pressure
    - Cholesterol and blood lipids
    - Type II diabetes
    - Cardiorespiratory fitness
- Musculoskeletal health
    - Bone mineral density
    - Muscular strength and endurance
- Mental health
    - Depression
    - Anxiety
    - Self-concept

These relationships are particularly evident for populations that demonstrate an "unhealthy" state at baseline [21,25,26]. The existing evidence indicates that the metabolic syndrome during

adolescence may be associated with negative adult health outcomes (link D) [21,31-33]. The metabolic syndrome refers to a clustering of abdominal obesity and the cardiovascular RFs identified above. This establishes an indirect link between adequate adolescent PA and adult health for these disease processes (i.e. links A → D). Although some studies suggest similar tracking of adolescent musculoskeletal and MH disorders into adulthood, the evidence is sparse and of poor quality [21].

The evidence for the tracking of PA behaviour from adolescence into adulthood (link C) suggests that it is largely dependent on the quality of the initial PA experience [21,34-37]. As previously described, the relationship between adequate adult PA with adult mortality and disease end points (link E) is well established [3,16-20]. This provides another indirect link between appropriate adolescent PA and adult health outcomes (i.e. links C → E).

More prospective longitudinal trials are indicated to further establish and track the causality for each of these relationships. However, there is currently sufficient evidence to demonstrate that engaging in quality PA is an important health behaviour for adolescents.

2.4) Health implications of physical fitness vs. physical activity

PF has been defined as "a physiological state of well-being that allows one to meet the demands of daily living (health-related PF) and/or provides the basis for sport performance (performance-related PF)" [38].

PF and PA have a reciprocating relationship. Participation in adequate PA influences PF, which in turn contributes to determining PA capacity [18,22,25]. Consequently, it is not surprising that PA and PF appear to have a similar relationship to morbidity and mortality. Although there is a genetic component to PF, it is a better predictor of health-related outcomes in adolescents than PA [18,39-42].

There are several dimensions to PF that are each associated with different health-related outcomes. This thesis focuses on body composition, muscular strength and cardiorespiratory fitness because they are the most commonly reported and have the richest evidence base.

These components of adolescent PF are markers for cardiovascular disease RFs, adiposity, metabolic syndrome and premature all-cause mortality [43-45]. There is also evidence that improvements in adolescent muscular strength may have a positive effect on musculoskeletal health and the risk of osteoporosis during adulthood [43]. Cardiorespiratory fitness may also be associated with several MH outcomes [43].

From the public health perspective, it is important that recommendations for PA address all components of PF. Focusing on adequate PA is likely to positively affect PF and avoids the complication of adjusting guidelines according to genetic differences within the population [40]. However, measurement of PA in large samples of adolescents is limited by the validity of self-report questionnaires and/or the cost and feasibility of more accurate methods [18].

2.5) Metrics for describing physical fitness and physical activity

There are numerous methods for measuring PF and PA in adolescents. Using valid and reliable metrics is integral for assessing these outcomes for comparative purposes and to inform the development of evidence-based recommendations for health [46].

2.5a) Physical fitness

The methods used to assess the body composition of adolescents include densitometry, dual-energy x-ray absorpiometry (DXA) and bioelectrical impedance analysis (BIA). Although these metrics are valid and reliable, they involve complex procedures and/or require technical equipment [47]. The most widely utilised method for estimating body composition is based on anthropometric measurements and is simpler to implement. Anthropometric measurements and age are often used to calculate the body mass index-for-age (BFA) and height-for-age (HFA) in adolescents. However, there is conjecture about the accuracy of BFA and HFA as measures of adiposity and growth. Often the waist-to-hip girth ratio and skinfold thickness are assessed to provide further information about body composition [47-50].

There is no established gold standard for measuring all the dimensions of muscular condition [51,52]. The 1x repetition maximum (1xRM) is a validated metric that is most commonly used to assess muscular strength [53]. This refers to the maximum force that a muscle group can exert in a single repetition of the action of interest and is primarily fuelled by anaerobic metabolism [54]. Methods for measuring 1xRM range from assessing a specific muscle with highly technical equipment in a controlled environment to various field tests that provide a more generalised index of strength [55].

Measurement of VO_2 max is regarded as the gold standard for describing cardiorespiratory fitness [51]. VO_2 max is an abbreviation that refers to the body's maximal aerobic capacity by

measuring the uptake, transport and use of oxygen during incremental PA [54]. It is accurately measured in a laboratory setting, but can also be approximated by various field tests [55].

2.5b) Physical activity

The "dose" of PA is described according to its duration, frequency, intensity and type [56,57].

Duration of PA refers to the length of time of a single bout of exertion [56,57]. According to Haskell et al (2007), PA in adult populations can be accumulated in blocks of 10 minutes or more throughout the day [19]. Although some recent studies have shown that the health benefits for adolescents may be improved if PA is sustained continuously for 5-10 minutes, there is no widely accepted minimum bout duration stipulated for this population [26,58,59].

Frequency refers to how often an individual engages in PA. It is usually expressed as the number of bouts over a given time period (e.g. 5x/week) [56,57].

Intensity refers to the level of exertion from PA. It varies according to activity specific PF and is typically expressed in metabolic equivalents (METs) [56,57]. METs describe the rate of energy expenditure and by convention, one unit is expressed as an oxygen uptake of 3.5 millilitres per kilogram of body weight per minute [17]. PA intensity is often categorised according to specific METs values (1 MET = quiet sitting; < 3 METs = light intensity PA; 3 - 6 METs = moderate intensity PA; > 6 METS = vigorous intensity PA) [60]:

Type of PA is typically classified according to the occupation, transport and recreation domains described by WHO [61]. The occupational domain is often not relevant for young adolescents. However, school-based PA is usually assessed in this age group and includes PES classes. Non-school based sport in the recreation domain is the focus of this thesis.

2.6) Recommendations for physical fitness and physical activity

Recommendations for PF and PA can be established by examining the dose-response relationship between these parameters and health-related outcomes.

2.6a) Physical fitness

There are several reference data sets for BFA and HFA that have been established from various national and international cohorts. The recommendations for healthy BFA and HFA in adolescents are typically defined according to percentiles and/or standard deviations from the mean of an age-equivalent reference population. Therefore, health status is classified relative to a population that is not locally specific and may be ethnically and geographically different. In this thesis, these concerns are ameliorated by using the 2007 WHO growth reference for adolescents [62]. This is the most globally representative of the established international standards and the cut-offs for each category have also been assimilated with the widely accepted adult criterion for BMI (TABLE 2.1) [62].

TABLE 2.1: Threshold values for BFA and HFA recommendations in accordance with the WHO growth reference database

BFA	HFA
Severely thin BFA<-3SD	**Severely stunted** HFA<-3SD
Thin -2SD<BFA<-3SD	**Stunted** -2SD<HFA<-3SD
"Healthy" weight -2SD<BFA<1SD	**"Normal" growth** -2SD<HFA<2SD
Overweight 1SD<BFA<2SD	**"High" growth** 2SD<HFA<3SD
Obese BFA>2SD	**"Very high" growth** HFA>3SD

Source: De Onis et al (2007) [62]

In the absence of a gold standard to assess muscular strength, no criterion-referenced recommendations for adolescent health have been established [51]. The potential of several tests that approximate 1xRM for specific muscle contractions have been described in the context of talent identification and physical performance. However, the research regarding health outcomes has focused on establishing the relationship with strength rather than describing the thresholds for benefit.

The Cooper Institute has established criterion-referenced recommendations for VO_2 max that are associated with the prevalence of metabolic syndrome (TABLE 2.2) [63]. Several studies examining the relationship between other cardiovascular RFs in various age groups have identified VO_2 max thresholds at similar levels [64-66].

TABLE 2.2: Threshold recommendations of VO_2 max for risk of metabolic syndrome

Age in years	Boys VO_2 max (ml/kg/min)			Girls VO_2 max (ml/kg/min)		
	High risk	Some risk	Healthy	High risk	Some risk	Healthy
10	≤37.3	37.4 - 40.1	≥40.2	≤37.3	37.4 - 40.1	≥40.2
11	≤37.3	37.4 - 40.1	≥40.2	≤37.3	37.4 - 40.1	≥40.2
12	≤37.6	37.7 - 40.2	≥40.3	≤37.0	37.1 - 40.0	≥40.1
13	≤38.6	38.7 - 41.0	≥41.1	≤36.6	36.7 - 39.6	≥39.7
14	≤39.6	39.7 - 42.4	≥42.5	≤36.3	36.4 - 39.3	≥39.4
15	≤40.6	40.7 - 43.5	≥43.6	≤36.0	36.1 - 39.0	≥39.1
16	≤41.0	41.1 - 44.0	≥44.1	≤35.8	35.9 - 38.8	≥38.9
17	≤41.2	41.3 - 44.1	≥44.2	≤35.7	35.8 - 38.7	≥38.8
>17	≤41.2	41.3 - 44.2	≥44.3	≤35.3	35.4 - 38.5	≥38.6

Source: The Cooper Institute (2010) [63]

There are also enough reference databases to establish global norms for several common field tests that approximate VO_2 max and 1xRM [55]. In many cases, these are sufficiently large and robust to directly compare the performance of a study sample to the mean and percentile values of age- and gender-equivalent populations [55]. However, criterion-referenced recommendations for health have not been validated for global utility for the crude results of

any field tests [67]. In this thesis, I compare the study sample to global normative values to provide an indication of fitness status. This eliminates the potential for error when the crude results of a field test are converted mathematically to the metric commonly used to describe VO_2 max or 1xRM.

2.6b) Physical activity

WHO advocates the following PA recommendations for young adolescents [17]:

1. *Children and youth aged 5–17 should accumulate at least 60 minutes of moderate to vigorous intensity PA daily.*
2. *Amounts of physical activity greater than 60 minutes provide additional health benefits.*
3. *Most of the daily physical activity should be aerobic. Vigorous-intensity activities should be incorporated, including those that strengthen muscle and bone, at least 3 times per week.*

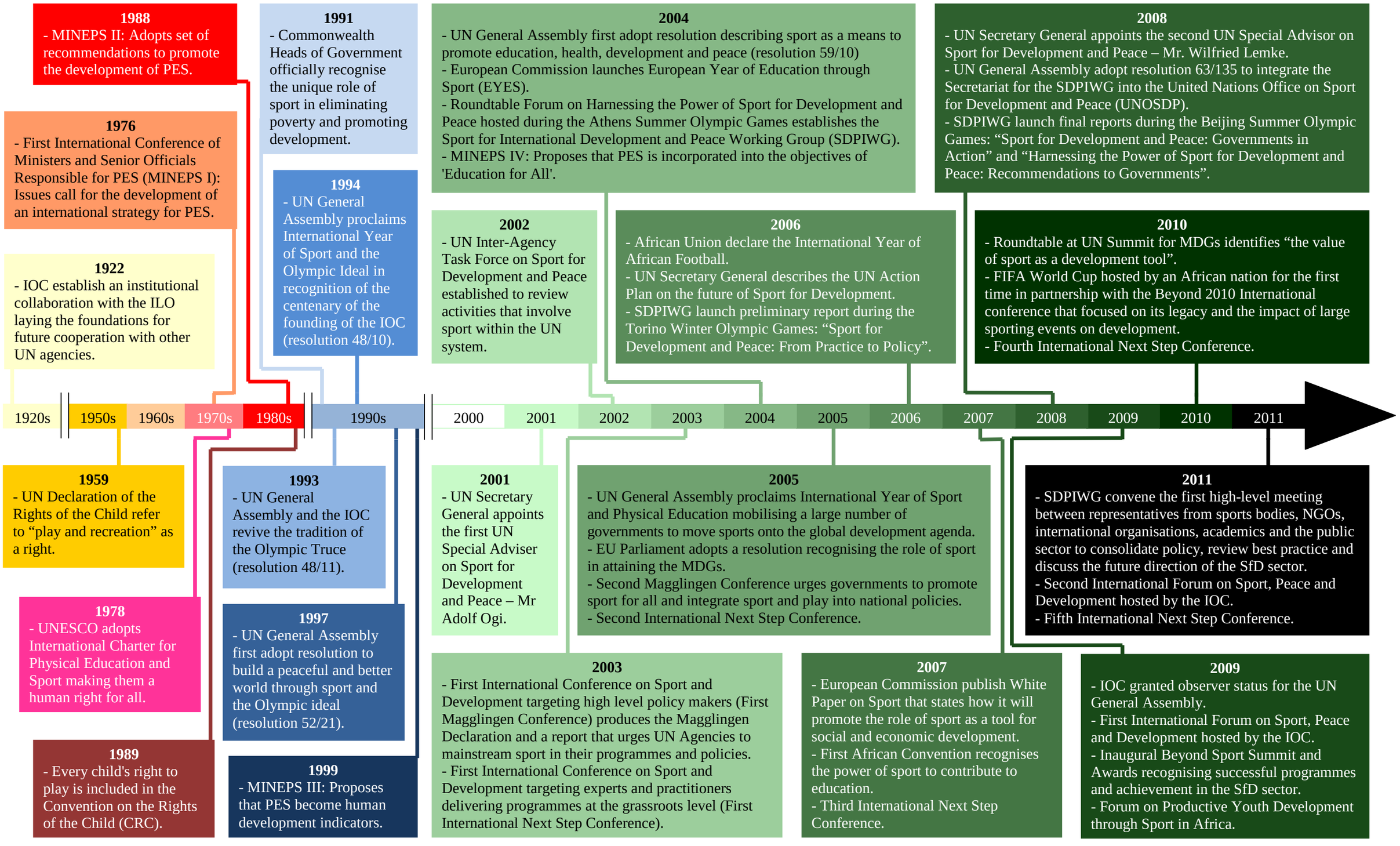
These recommendations are evidence-based and outline the PA required to improve the health outcomes described above for young adolescents (SECTION 2.3). WHO also suggests that PA for this age group should include "play, games, sports, transportation, recreation, physical education, or planned exercise, in the context of family, school and community activities" [17]. SfD programmes in LMICs address several of these recommendations and are the focus of this thesis.

2.7) Emergence and evolution of the sport-for-development sector

Organised sport is an enduring component of cultural identity and diversity that has persisted in various forms throughout history. SfD is a term that has been defined more recently as "the intentional use of sport, physical activity and play to attain specific development and peace objectives" [68].

The concept of SfD was first articulated when the International Labour Organisation (ILO) and the International Olympic Committee (IOC) signed a working agreement in 1922 [69]. Since then, there has been growing momentum to incorporate PES into development programmes in a more systematic way. The UN International Year of Sport and Physical Education in 2005 was a catalyst for rapid expansion in the SfD sector [70,71]. The subsequent growth has progressed beyond ad-hoc agreements, declarations and programmes to more widespread dissemination. I have presented the important developments in the history of the SfD sector on a timeline (FIGURE 2.2) [68,69,72-104].

FIGURE 2.2: Timeline depicting the emergence of the SfD sector



2.8) Claimed impacts of sport-for-development interventions

Stakeholders in the SfD sector currently deliver a broad array of programme models and claim various cross-cutting positive results [68]. This diverse range of interventions and outcomes is reflected in the definition for "sport" that was developed by the United Nations Inter-Agency Task Force on Sport for Development and Peace [105]:

> *"All forms of physical activity that contribute to physical fitness, mental well-being and social interaction, such as play, recreation, organised or competitive sport, and indigenous sports and games."*

The Sport for Development and Peace International Working Group (SDPIWG) claims that the impacts of SfD programmes:

- Foster child and youth development and education [106];
- Achieve health objectives [107];
- Promote gender equity [108];
- Foster inclusion, health and well-being of people with disabilities [109];
- Foster social cohesion, prevent and reduce conflict and build peace [110].

The claims of the SDPIWG literature review for the health-related impacts of SfD interventions include [107]:

- Reduced incidence of NCDs and associated RFs;
    - Cardiovascular disease
    - Type II diabetes
    - Colon and breast cancer
    - Obesity
    - Hypertension
- Reduced incidence of musculoskeletal disorders;
    - Osteoporosis and falls risk

- Reduced MH disorders;
    - Depression
    - Anxiety
    - Self-concept
- Prevention strategies for communicable diseases (e.g. HIV/AIDS, malaria, TB).

In this thesis, I focus on several health-related outcomes that are in accordance with the claims of the SDPIWG and the previously discussed evidence for adolescent PA (SECTIONS 2.3 & 2.4). The target population is young adolescents who are aged between 11-14 years and live in post-conflict settings within LMICs.

In the absence of a consistent definition for post-conflict, I described it as:

> *The period of time that occurs between the cessation of violent conflict and the "normalisation" of society.*

The notion of "normalisation" is sensitive to the variable time frames of post-conflict processes and the diverse social determinants that exist in different regions [111,112]. It can be assimilated with routine national development and occurs after the *initial stabilisation* and *transition* phases of the recovery process [111,112]. The challenges faced during the post-conflict period typically include reintegrating people into their communities, reconstructing infrastructure and reviving productive livelihoods [111,113]. The SfD sector claims social cohesion and peace-building outcomes that address all of these challenges [110]. Many organisations that implement peace-building SfD interventions in post-conflict regions also claim to achieve the previously described health impacts in this population.

2.9) Evidence for the impact of sport-for-development interventions on the health of young adolescents in post-conflict settings

Cronin (2011) completed a broad literature review to map the impact of SfD interventions and suggested that a more focused analysis of the specific claimed outcomes was needed [114]. I conducted a systematic review of the academic and grey literature to assess the evidence for the impact of SfD programmes on health-related outcomes in young adolescents in post-conflict regions. The methods and search terms reflected the diversity in SfD programmes and the broad definitions described in this thesis (APPENDIX A1). The titles of 25,282 references were identified from the initial search. This was reduced to 354 for abstract review. The full text of 58 references was read and 18 were included for final discussion.

Seven descriptive articles were identified in academic journals that referred to SfD programmes and health in a post-conflict context [115-121]. However, these papers had a sociology focus and no empirical data specific to the health-related outcomes of interest in the post-conflict context were presented. This paucity of published evidence was also reflected in five texts and book chapters that discussed SfD programmes [71,122-125].

One randomised control trial (RCT) found that creative play had no impact on the MH of adolescents aged 14-17 years in northern Uganda [126]. The Acholi Psychosocial Assessment Instrument (APAI) was used to assess local MH syndromes relevant to the local culture [127,128]. Although "sport" was part of the intervention, several other forms of play were also included and this limited the relevance of the results to the SfD sector. There was also only a minimal overlap of the ages of the subjects with the target population in this review.

Several organisations and government agencies have produced reports and documents that contribute to the rapidly expanding body of SfD grey literature. The most comprehensive of

these is a series of literature reviews commissioned by the SDPIWG assessing the evidence for SfD programmes [106-110]. Although one of these focused on health-related outcomes, it primarily cited PA intervention studies conducted in adult populations in high-income countries [107]. It is not known how generalisable these study outcomes are to LMICs and more specifically to adolescents in the post-conflict context.

Most organisations that deliver SfD programmes conduct monitoring and evaluation (M&E) action research to assess their impact. However, many of these examine other health-related outcomes and/or utilise measures of questionable reliability, validity and sensitivity. I identified no action research for SfD organisations in post-conflict populations that assessed the health outcomes of interest and/or included a control group in their design.

UK Sport recently commissioned a "before-after" study to assess the impact of 10 different SfD programmes. One of these organisations, The Kids League (TKL), specialised in delivering SfD programmes in the post-conflict region of northern Uganda. Coalter (2006) developed the M&E guidelines for this study [129]. Data were collected using an interviewer administered questionnaire and self-concept was the only health-related outcome of interest included for TKL. No objective measures for PA, PF or other MH disorders were collected. Although this study suggested a positive impact on self-concept, in the absence of a control group, it is not possible to draw any strong conclusions on causality [130].

The Swiss Academy for Development (SAD) facilitated similar evaluations for three SfD interventions for children aged 5-18 years in post-conflict areas. These programmes were in Lebanon, Sri Lanka and Nepal. The assessments included qualitative and quantitative methods. Although the focus was on psychosocial objectives, there were no empirical data for the health outcomes of interest and a large proportion of the results were anecdotal [131-133].

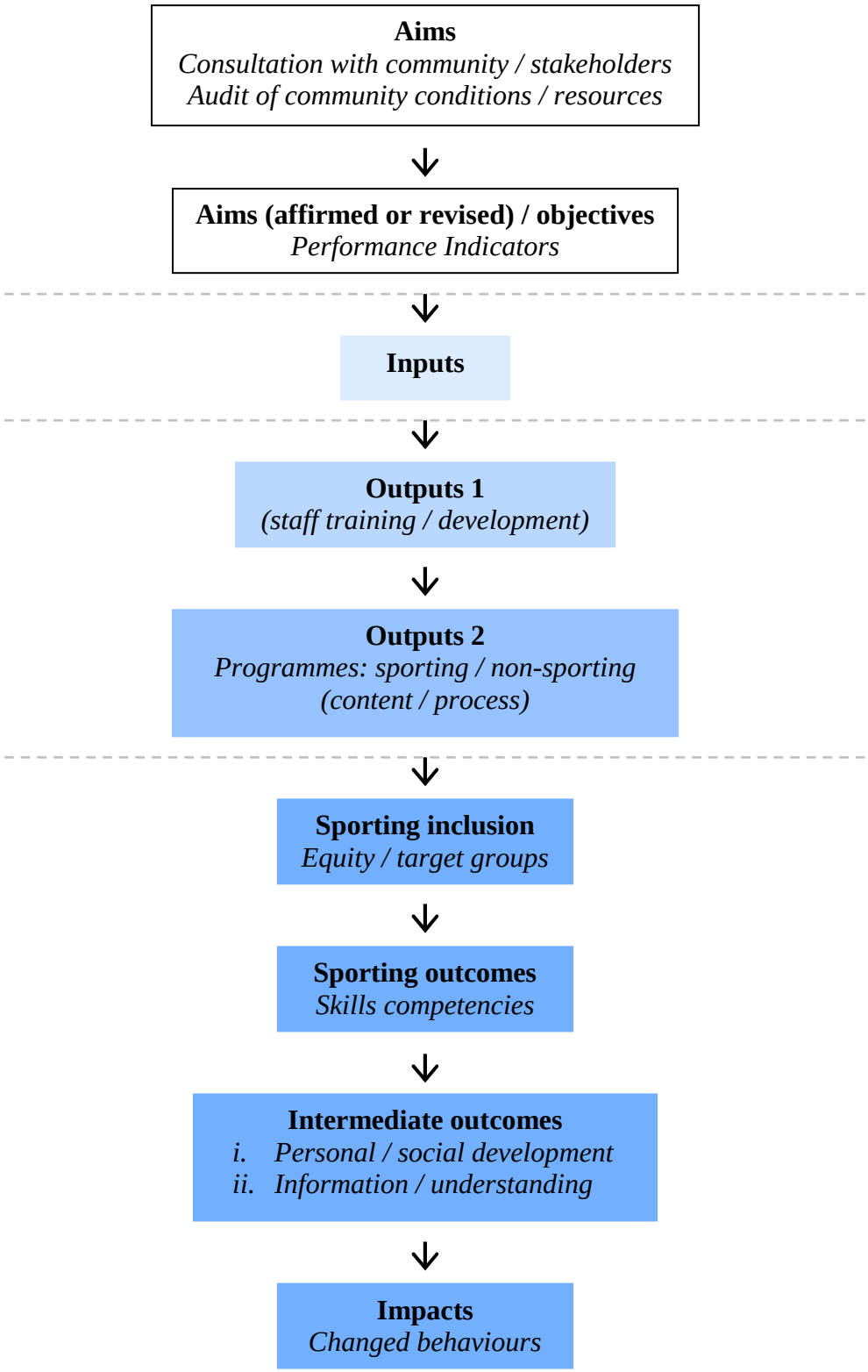
2.10) Programme theory and logic models for sport-for-development

The gap in the evidence base for the health impact of SfD interventions is accompanied by a paucity of programme theory [130]. Identifying programme objectives and building a theoretical framework that describes how these will be achieved is essential [2,134]. However, the interaction between various stakeholders in the SfD sector is poorly understood and there is rarely a concerted focus on specific objectives [115,125]. Instead programme implementers typically describe broad aims based on unsubstantiated claims about the positive effect of sport [121]. These are often adapted to the expectations of their funders rather than informed by a community needs analysis that identifies specific objectives and/or a target population [115,121,135]. The final outcomes are presumed to occur despite the lack of evaluation and/or theoretical logic models that conceptualise the processes, mechanisms and cause/effect sequences of SfD programmes [130].

The paucity of established programme theory for SfD interventions limits the potential for rigorous evaluation [2,129,130,134]. In the initial stages of the research initiative commissioned by UK sport, Coalter (2006) divided the SfD sector according to organisational objectives [129]. The "sport plus" approach focused on establishing sustainable sport programmes and using these as a vehicle to achieve broader development goals [129]. In contrast, "plus sport" organisations were less concerned about sport as an outcome, but used it as a lure for other development aims (e.g. a vaccination programme) [129]. These definitions have continued to evolve and a continuum between the two categories is now recognised [121,136].

Coalter (2006) developed the first logic model for "sport plus" programmes [129]. It was a descriptive taxonomy scheme that provided an overview of the steps taken for SfD programme implementation and M&E (FIGURE 2.3) [129].

FIGURE 2.3: SfD programmes – a logic model



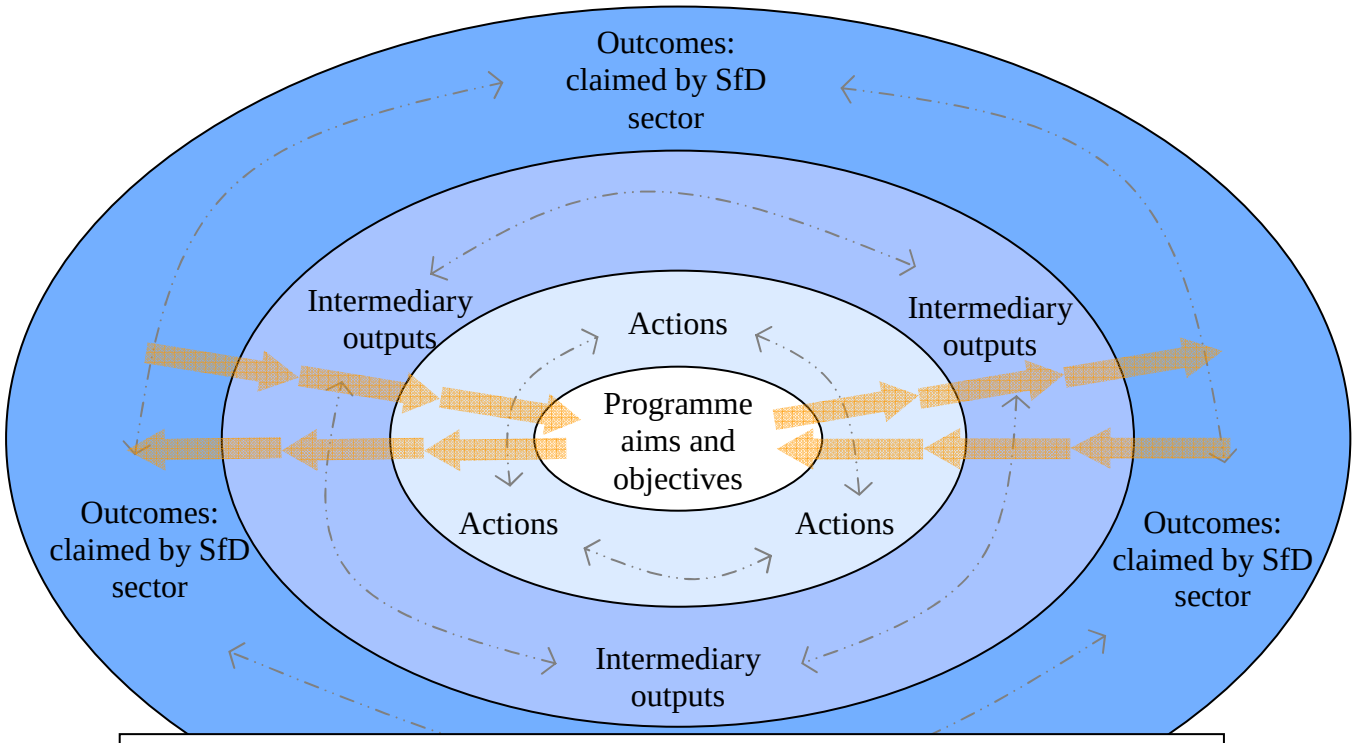
NOTE: M&E requirements are written in italics. To assist the discussion below, I have inserted line breaks (i.e. -----) and colour coding at different stages.

Source: Modified from Coalter (2006) [129]

There are several limitations to using this logic model as an evaluation framework. Although more details are outlined in the original M&E manual, it does not describe how each step is actually achieved [129]. The theoretical links from "Outputs" to "Impacts" are tenuous and indicate the challenges of applying a generic model to the diverse outcomes claimed by the SfD sector. Logic models also do not recognise the potential of ongoing feedback to influence programme development upstream. Therefore, a broader framework that articulates the mechanisms for translating SfD programme aims into impacts is indicated [121].

I have developed a programme theory for the objectives, actions, outputs and outcomes of stakeholders in the SfD sector. At the macro level, this framework provides a broad overview of the interactions that occur both within and between the described levels in the model. There is often an interdependence that develops simultaneously across each of these levels rather than a causative and chronological association (FIGURE 2.4).

FIGURE 2.4: Programme theory for SfD programmes presented at the macro level



NOTE: *The objectives, actions and outputs described at each level in the model are interrelated (e.g. intermediary outputs from one action may interact and influence other intermediary outputs). The model operates in both a feedforward and a feedback manner (e.g. actions feedforward to impact on intermediary outputs, which in turn feedback to influence the upstream actions).*

The “programme aims and objectives” have been maintained as the central starting point in this theoretical framework. These should be developed in consultation with the local community by several different stakeholders that include [68,103,104,121,125]:

- Public sector (i.e. local, district and national governments);
- Private sector (i.e. profit- and market-oriented organisations);
- International organisations (i.e. agencies of the UN and regional unions);
- Non-governmental organisations (i.e. independent non-profit organisations);
- Sport organisations (i.e. organisations that focus on sporting achievement);
- Academic institutions (i.e. education centres and institutes for learning and research).

Although the stakeholders often have interdependent “aims and objectives”, they each have their own agenda and function [68,71,125]. These are determined by the underlying purpose of each participating organisation and typically influenced by the expectations of their funding source [115,135,136]. However, it is also important that the objectives of the stakeholders are “formed in response to local needs and the intended reach of the SfD programme” [68,71,129].

I have used the term “actions” in the place of the “inputs” stage described by Coalter (2006) [129]. The “actions” of stakeholders are depicted as flexible and modifiable in response to feedback from downstream outputs and outcomes [129]. They are also tailored according to the outcomes of interest and adjustable according to variations in concurrent “actions”. Although all stakeholders contribute to the “actions” of SfD programmes, certain organisations typically take a lead role depending on the skills and expertise required.

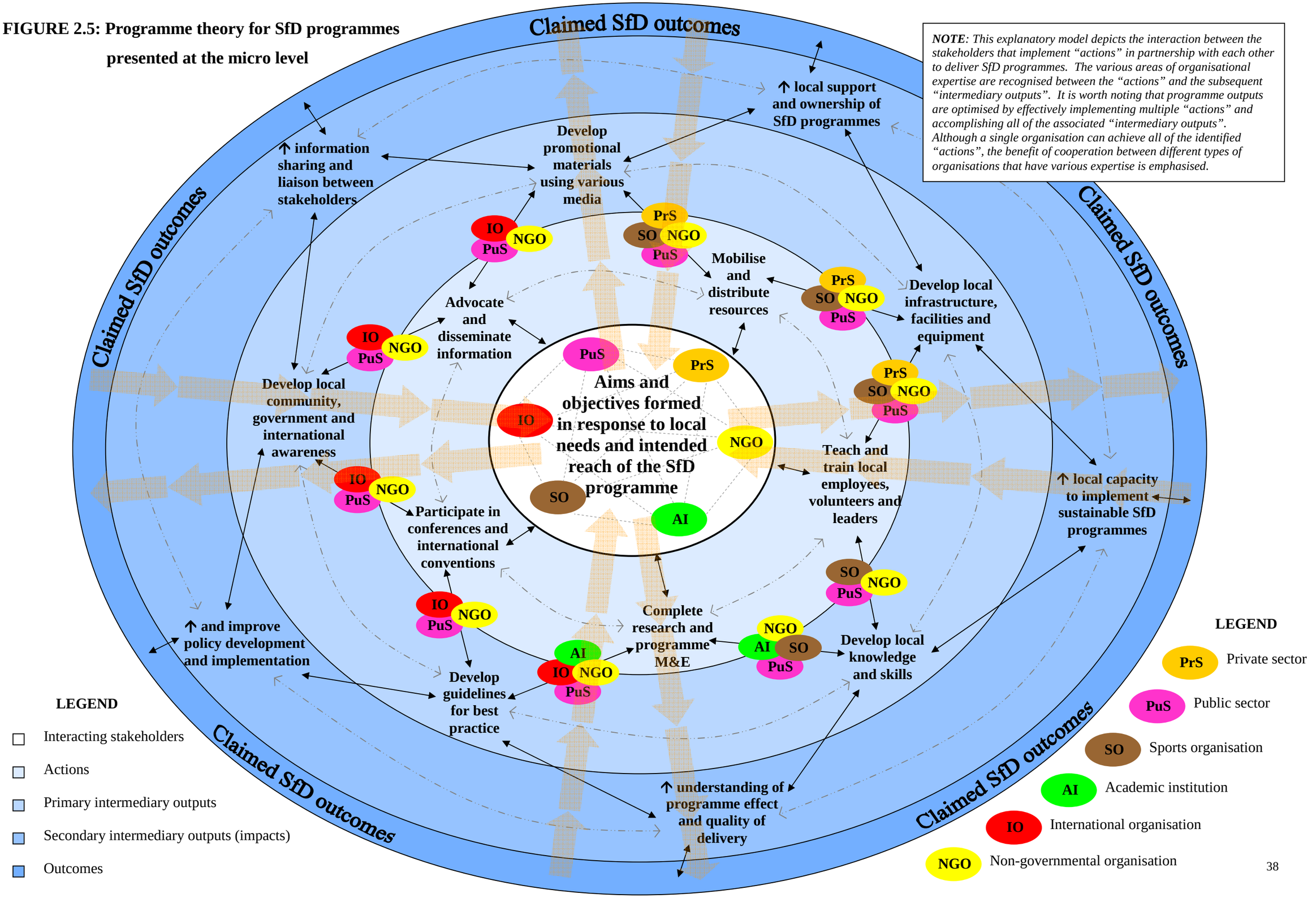
The “intermediary outputs” are accomplished as joint ventures that are led by the organisations that pilot the associated “actions”. I have used the term “intermediary outputs” to combine the terms “outputs 1” and “outputs 2” described in the Coalter (2006) model [129].

Grouping the "intermediary outputs" together illustrates that there is not necessarily a chronological relationship at this level, but rather a complex multi-directional interaction that is occurring between several outputs and short-term impacts simultaneously. These link the "actions" with the final "outcomes" of SfD interventions by developing the necessary environment, knowledge and capacity to establish a sustainable programme.

The "outcomes" in the model are those claimed by the SfD sector. This term reflects the unsubstantiated nature of the current evidence base. It combines the "sporting inclusion", "sporting outcomes", "intermediate outcomes" and "impacts" stages described by Coalter (2006) into a single term that recognises their interaction. International charters and declarations promoting sport as a human right make establishing a sustainable "sport plus" programme a development outcome in itself (FIGURE 2.2). Achieving this requires processes that reach further in the development context than indicated in the model described by Coalter (2006) [129,137]. Several "intermediary outputs" contribute to achieving multiple "claimed SfD outcomes" that develop and interact simultaneously with "sporting inclusion" and "sporting outcomes" rather than secondary to them. Therefore, the theoretical pathways at this stage vary depending on the outcome of interest and are not elaborated in this model.

I have used this theoretical framework to review the "objectives", "actions" and "intermediary outputs" of the various stakeholders that contribute to SfD programmes. Although the interactions between different organisations at each level are contextually variable, there are several notable consistencies [68]. I have incorporated these into a series of specific logic models to develop a programme theory that identifies these interactions and recognises the role of ongoing informative feedback. At the micro level, this framework explains how SfD interventions function and depicts the critical processes that contribute to establishing a sustainable programme as a basis for achieving the desired "outcomes" (FIGURE 2.5).

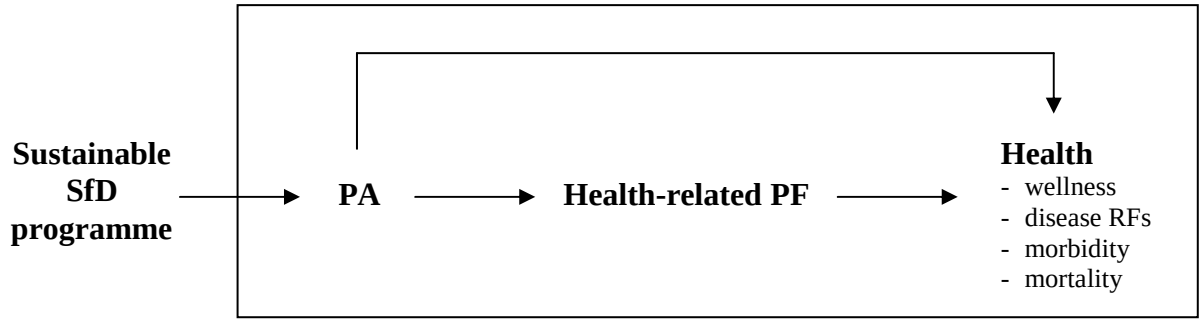
FIGURE 2.5: Programme theory for SfD programmes presented at the micro level



***NOTE**: This explanatory model depicts the interaction between the stakeholders that implement "actions" in partnership with each other to deliver SfD programmes. The various areas of organisational expertise are recognised between the "actions" and the subsequent "intermediary outputs". It is worth noting that programme outputs are optimised by effectively implementing multiple "actions" and accomplishing all of the associated "intermediary outputs". Although a single organisation can achieve all of the identified "actions", the benefit of cooperation between different types of organisations that have various expertise is emphasised.*

The final "outcomes" of interest in this thesis are the health-related outcomes previously identified (SECTION 2.8). The SDPIWG asserts that SfD interventions have a clinically significant impact on PA levels [107]. Bouchard (2001) described a direct link between PA and health as well as an indirect association mediated by PF [138]. These relationships have been subsequently developed further by various authors. However, the SDPIWG cites the original model as the theoretical explanation for the "claimed SfD outcomes" that are health-related (FIGURE 2.6) [107].

FIGURE 2.6: The theoretical framework describing the links between a sustainable SfD programme and health-related outcomes



Source: Modified from Bouchard (2001) [107,138]

2.11) Health needs in a post-conflict context

There are significant variations in the health needs of populations living in different post-conflict regions throughout the world [139,140]. Geographical location, socio-economic development, international intervention and characteristics specific to each conflict are among several influencing factors [111,141]. However, Roberts et al (2012) suggested that high levels of psychological distress and rapid urbanisation are common characteristics during the post-conflict period that increase the risks for NCDs [142].

The diversity of post-conflict settings indicates that several studies across multiple regions are necessary to establish the health-related impact of SfD programmes. The results of my systematic literature review demonstrate the relative infancy of research in this area (SECTIONS 2.9 & 2.10). Therefore, in this thesis, I will contribute to the fledgling evidence base for SfD programmes by conducting a case-study assessing PA, PF and MH outcomes in one post-conflict region.

2.12) Summary box

BACKGROUND AND RATIONALE FOR RESEARCH

- NCDs and MH disorders are significant problems throughout the world that are worsening with urbanisation and contributing to the development of the double burden of disease in many LMICs.

- Increasing PA and PF have been found to positively affect adolescent RFs for NCDs, musculoskeletal health and MH disorders that may track into adulthood.

- Measurement of PA and PF has enabled the establishment of global normative values and recommendations for health-related benefits.

- The SfD sector is rapidly expanding and claims to have similar health-related benefits that are mediated by increased PA and PF in the target population.

- There is a paucity of published evidence supporting the claimed impact of SfD programmes on health-related outcomes.

- Existing conceptual models that explain the link between SfD programmes and health-related outcomes are scarce and rudimentary. There is an indication to develop these further. I have proposed a theoretical framework describing the processes that link the needs of the community and intended reach of the programme with the actions, outputs and claimed outcomes of a SfD programme. The model described by Bouchard (2001) is most commonly cited in the SfD sector as the mechanism explaining the impact of sustainable sport programmes on the health-related outcomes assessed in this thesis.

- The diverse health needs of post-conflict communities vary throughout the world, but there are several common characteristics that may suggest an increased risk for NCDs.

CHAPTER 3

Case-Study Site – Gulu, Uganda

3.1) Purpose of chapter

The purpose of this chapter is to identify the reasons why I chose Gulu as the case-study site and provide a brief historical overview of the region. I will also describe what is known about the changing demography and epidemiological profile of the Gulu population. I focus on health-related outcomes associated with NCDs, MH and PA with a view to refining my research objectives and methods according to the local context.

3.2) Selecting the case-study site

Gulu is located in northern Uganda and is the site for the case-study presented in this thesis (APPENDIX A2). I initially considered several other locations for my fieldwork and decided on Gulu for two reasons. Firstly, it is located in the middle of a hot-spot for recent conflicts and it satisfied my definition of a post-conflict region (SECTION 2.8). Secondly, there were several organisations interested in implementing SfD interventions in northern Uganda. An exploratory fieldtrip in September 2009 indicated that the organisations located in Gulu appeared to be the most amenable to forming a partnership for evaluation (FIGURE 3.1).

FIGURE 3.1: Office sign-post for some of the SfD stakeholders active in Gulu

Photo taken by Justin Richards with consent of the subject – Gulu 2009

3.3) Local geography, history and demographics of Gulu

Uganda is situated in eastern Africa and is bordered by Sudan, Kenya, Tanzania, Rwanda and the Democratic Republic of Congo (APPENDIX A2). The capital of Uganda is Kampala and it is connected to Gulu via approximately 320km of poorly maintained paved road. During my fieldtrip in September 2009, I consulted local officials and authorities to improve my understanding of the region's history and epidemiological profile.

Although limited data are available, it is evident that the history of the area has had a profound epidemiological effect. Uganda has a complex past dominated by tribal rivalries that have caused civil unrest over the past 50 years. In 2006, northern Uganda emerged from 20 years of conflict and severe impoverishment incited by a guerrilla insurgency group known as the Lord's Resistance Army (LRA). By this time, northern Uganda had been declared the second worst "forgotten" humanitarian emergency in the world [143]. Approximately 1.84 million civilians (i.e. over 90% of the population) had been relocated to internally displaced person (IDP) camps with very poor living conditions (FIGURE 3.2) [143-145].

FIGURE 3.2: Former IDP camp located on the outskirts of Gulu municipality

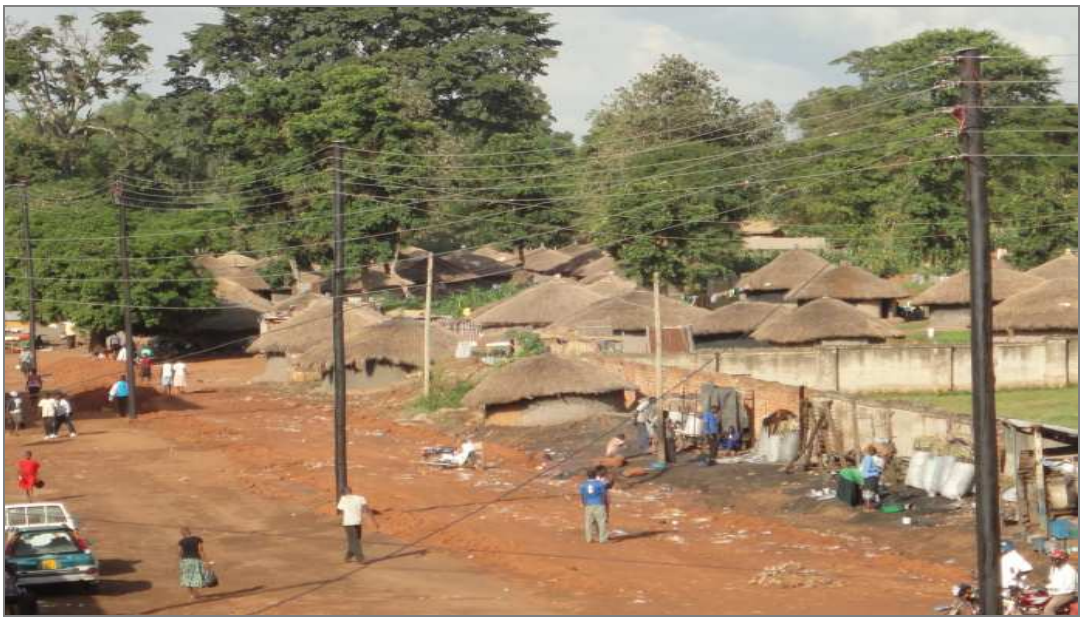


Photo taken by John McMullen (permission granted for use in this thesis) – Gulu 2010

The IDP camps were under the protection of the Ugandan national army. However, often the security was inadequate and was provided by soldiers who were not immune to accusations of severe human rights violations [146]. Abduction of children and adolescents to serve as equipment carriers, child soldiers and "wives" was a characteristic feature of the war. It is estimated that $^{1}/_{3}$ of all boys and $^{1}/_{6}$ of all girls in Gulu were abducted for at least one day during the LRA insurgency and the majority of these were held captive "in the bush" for at least two weeks [147,148].

During the post-conflict period, the majority of the population in Gulu has left overcrowded IDP camps and returned to their homes or resettled in urban environments [146]. Gulu municipality has grown and developed rapidly with the assistance of funding from international aid agencies and the support of the national government who are investing in the town as the urban and commercial hub of northern Uganda (FIGURE 3.3). This unprecedented local development and urbanisation may catalyse changes in the epidemiological profile of the region [4,5,142].

FIGURE 3.3: The infrastructure in central Gulu municipality is developing rapidly

Photo taken by John McMullen (permission granted for use in this thesis) – Gulu 2010

3.4) Emergence of non-communicable diseases in Gulu

The surveillance of health-related outcomes in Gulu is limited and the data from existing monitoring systems are of dubious quality. However, it is worth noting the changing trends in the causes of mortality reported by the Gulu District Health Office (DHO) during the post-conflict period. Although communicable diseases continue to be the primary cause of death in Gulu, cardiovascular disease (CVD) and hypertension appear to be health problems emerging in the population (TABLE 3.1) [149,150].

TABLE 3.1: Trend in cause of mortality in Gulu (>5 yrs age)

2006/2007		2007/2008		2008/2009	
Disease	**%**	**Disease**	**%**	**Disease**	**%**
HIV / AIDS	27.1	HIV / AIDS	32.7	HIV / AIDS	28.4
TB	13.7	TB	9.7	TB	16.5
Malaria	10.9	Pneumonia	5.7	Pneumonia	9.8
Pneumonia	7.8	Malaria	4.1	Anaemia	5.9
Anaemia	6.6	Septicaemia	3.2	***CVD***	***5.2***
Septicaemia	4.6	Anaemia	3.1	Malaria	4.7
Diarrhoea	4.4	Trauma	3.0	Injury	4.5
RTI	3.7	Meningitis	2.6	Diarrhoea	2.7
Meningitis	3.3	***CVD***	***2.4***	***Hypertension***	***2.6***
Trauma	2.9	Diarrhoea	1.8	Septicaemia	0.7
Others	15.2	Others	30.4	Others	16.9

Source: Gulu District Health Office (2008 & 2009) [149,150]

Despite the apparent increase in NCD incidence, it is possible that this is just an artefact of improved diagnostic procedures and awareness in Gulu health centres. It should be noted that the level of mortality attributed to these diseases in Gulu remains relatively small. However, the second highest proportion of deaths is consistently attributed to "other" causes that are not identified and it is possible that NCDs as a cause of death have been widely under-reported.

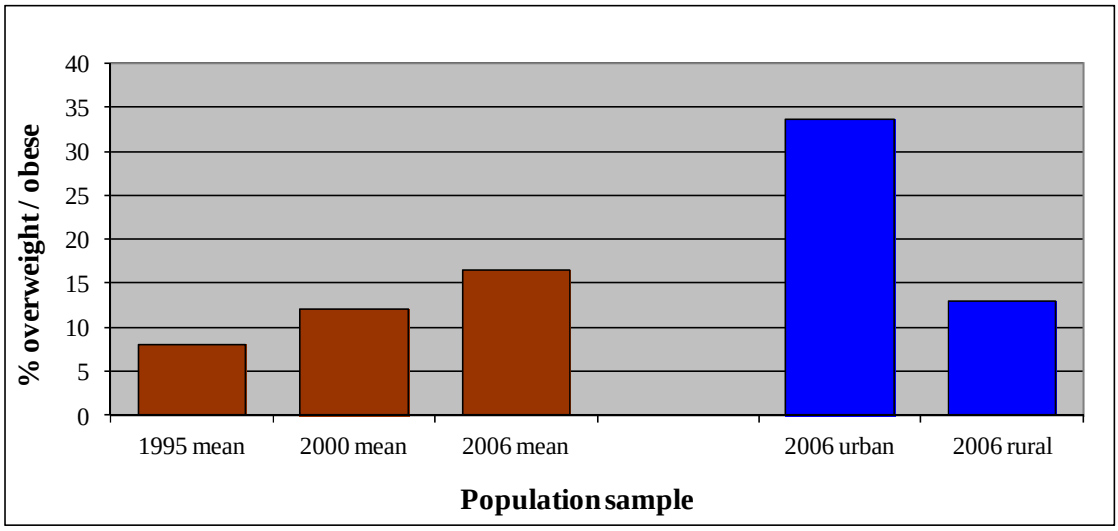
Dr Alice Lamwaka works at the Gulu Referral Hospital and supported this view. She is also the chairperson of the Gulu Diabetes Association and manages a private cardiovascular risk

centre in Gulu municipality. The existence of these services in an area of health that receives very little funding in Uganda anecdotally suggests that these diseases are emerging concerns.

The *2006 Ugandan Demographic and Health Survey* (UDHS) provides further evidence that NCDs are a growing problem in Uganda that is primarily observed in urban areas. BMI data were collected for women aged between 15-49 years. Although these are cross-sectional data from a Uganda-wide sample and are not specific to Gulu, there are two characteristics worth noting (FIGURE 3.4) [151]:

- There is a trend of an increasing percentage of overweight / obese women when comparing UDHS data from 1995 (8%), 2000 (12%) and 2006 (16.5%);
- There is a clear difference in overweight / obesity levels when comparing urban (33.7%) and rural regions (12.9%) in 2006.

FIGURE 3.4: % of overweight / obese women in Uganda (15-49 years)



Source: Uganda Bureau of Statistics (2007) [151]

The urbanisation of Gulu in recent years places it at high risk of following similar trends to other urban centres in Uganda. If the rapid change from an extended period of severe impoverishment to relative prosperity continues, it is feasible that Gulu may "fast-forward" through the epidemiological transition towards the double burden of disease [4-6,142].

3.5) Mental health in Gulu

The paucity of data on MH disorders in young adolescents may obscure their prevalence in a population that has been exposed to atrocities during an extended period of conflict. It is well established that war impacts on MH and that disorder rates tend to double after conflict and emergencies [152]. The high incidence of abduction and exposure to traumatic events during the war also appears to have contributed to mental health problems that have persisted for young adolescents during the post-conflict period in Gulu (FIGURE 3.5) [153].

FIGURE 3.5: Depiction of the experiences of a previously abducted boy from Gulu

Photo taken by Paul O'Callaghan with consent of the artist (permission granted for use in this thesis) – Gulu 2010

Annan et al (2006 & 2008) reported that although $^{2}/_{3}$ of adolescents in Gulu experienced emotional distress, they were "remarkably robust" [147,148]. Coalter (2010) also described high levels of resilience in an adolescent sample and warned against assuming a "deficit model" for interventions and evaluation [130]. However, it is important to note that in the Gulu context the benefits of resilience were secondary to high levels of traumatic exposure and MH disorders. This was demonstrated by Betancourt et al (2009) who assessed local constructs of MH in a sample of youth aged 14-17 years using the APAI. Only 12 participants in a sample of 178 were not suffering from any locally described MH syndromes (i.e. ~7%) [128].

3.6) Physical activity and sport-for-development in Gulu

There are no available data that describe PA and/or PF of young adolescents in Gulu. However, the Uganda Country Report for the *2003 Global school-based health survey* (GSHS) suggested that PES was "squeezed" out of the school curriculum during the previous 30-40 years [154]. PES was regarded as "a subject without academic significance"' and a "non-educational, non-productive use of time" [155,156]. Although PES continued to be timetabled in many Ugandan schools, "it was not seriously taught" and its "implementation, quantity and quality varied greatly between schools depending on the head teacher's attitude" [155,156].

The 2003 GSHS data supported these assertions. The results suggest that low percentages of adolescents in rural and urban settings throughout Uganda were physically active for at least 60 minutes per day (TABLE 3.2).

TABLE 3.2: Physical activity results of the 2003 Uganda GSHS

Sample		Proportion (%) and 95% confidence interval of students physically active for a total of at least 60 minutes per day on all of the last 7 days
Uganda	**Total**	15.3 (13.0 – 17.6)
	Boys	16.1 (12.3 – 19.9)
	Girls	14.4 (11.8 – 17.0)
Uganda urban	**Total**	17.7 (14.4 – 21.0)
	Boys	18.3 (12.3 – 24.3)
	Girls	16.9 (14.0 – 19.8)
Uganda rural	**Total**	12.5 (9.6 – 15.4)
	Boys	13.6 (9.7 – 17.5)
	Girls	11.4 (7.2 – 15.6)

Source: World Health Organisation – Uganda GSHS (2003) [154]

The generalisability of the GSHS data and their relevance to understanding the PA levels of young adolescents in the Gulu region is limited. The GSHS data were collected from 50 different secondary schools (25 urban and 25 rural) [154]. Since less than 18% of age-appropriate young adolescents in Uganda attended secondary school at the time when the questionnaire was completed, it is evident that a non-representative sample was assessed [157]. Although the location of the selected schools is not stipulated in the GSHS report, poor education provision and restricted access during the war in northern Uganda may have also marginalised the Gulu population in the sampling process.

However, anecdotal evidence suggests that the opportunities for adolescents to engage in adequate PA were very limited in Gulu during the war. It is likely that there was insufficient infrastructure to play sport or participate in recreational PA in the cramped IDP camps on the edge of the municipality. Landmines and the risk of abduction restricted movement and made it too dangerous for young people to assist with traditional farming practice (FIGURE 3.6).

FIGURE 3.6: Sign-post warning against landmines and unexploded ordnance in Gulu

Photo taken by Justin Richards – Gulu 2009

Although people are now safe to move freely in Gulu and most of the landmines have been cleared, several of the environmental influences on PA have persisted since the war ended.

The high attrition of human life during the conflict and the extended period of dependency on humanitarian feeding programmes have created a generational knowledge gap in farming methods. This has been accentuated by the rapid urbanisation of Gulu municipality [146].

Despite high levels of rhetorical enthusiasm for sport in Gulu, there continues to be a deficiency in practice. There is limited local expertise and the facilities required to develop sustainable programmes in northern Uganda are poorly maintained (FIGURE 3.7).

FIGURE 3.7: Homemade football and sports field for recreational PA in northern Uganda

Photos taken by Justin Richards – northern Uganda 2010

Various organisations have previously harnessed the positive perception of SfD to deliver several leagues, camps and one-day festivals in Gulu. However, these tended to be responsive to donor demands and opportunities rather than follow an established model. At the time I commenced my data collection fieldtrip in May 2010, the most recent SfD league intervention in the target population was completed 15 months prior (i.e. March 2009). Several organisations claiming a broad range of outcomes had received funding and were planning to re-establish SfD activities for young adolescents in Gulu. Although the data and anecdotal evidence presented above suggested a legitimate demand for SfD interventions, these organisations had not completed a community needs analysis or formative evaluation.

3.7) Summary box

THE CASE-STUDY SITE – GULU, UGANDA

- Gulu is a post-conflict region in northern Uganda that emerged from 20 years of war in 2006. The war was characterised by high levels of internal displacement and child abduction.

- During the post-conflict period Gulu has undergone relatively rapid socio-economic development and urbanisation.

- Although there is a paucity of evidence available on the health outcomes of interest and the available data is of dubious quality, it appears that Gulu has:
  - Consistent high levels of communicable diseases;
  - An increasing incidence of NCDs;
  - Persistent high levels of MH disorders;
  - Poor PA participation and opportunities for young adolescents.

- It is likely that Gulu will experience the double burden of disease if current patterns in health care, PA behaviour, socio-economic development and urbanisation continue.

- There is local support for the SfD sector in Gulu, but insufficient local capacity has limited the recent delivery of sustained interventions.

- A community needs analysis and formative evaluation have not been completed for the planned future programmes.

PART II

RESEARCH OBJECTIVES

CHAPTER 4

Research Hypothesis and Questions

4.1) Purpose of chapter

The purpose of this chapter is to clearly articulate the underlying hypothesis of my research and the questions that I am going to address. I will draw on the current evidence base and theoretical frameworks for the SfD sector in post-conflict settings. The gaps identified during the problem definition and solution generation stages of the SfD programme evaluation described in Part I of this thesis will also inform this process.

4.2) Research hypothesis

The underlying hypothesis of my research is that SfD programmes positively influence the PF, PA and MH of young adolescent participants in post-conflict settings. In this thesis, I test this hypothesis by identifying and addressing several research questions.

4.3) Research questions

My research questions can be broadly categorised according to the stages of evaluation described by Nutbeam et al (2006) [2]. The primary objective of this research is to contribute to the evidence base for the impact of SfD programmes on the PA, PF and MH of young adolescents living in post-conflict regions. To achieve this, I will conduct a case-study that aims to attain the highest level of evidence within the pragmatic constraints of the Gulu context. I have articulated specific research questions within each stage of the evaluation that are related to the gaps and assumptions identified in Part I of this thesis.

4.3a) What is the problem and how might it be solved?

Prior to assessing the impact of a SfD intervention it is important to establish appropriate measurement protocols. In Chapter 5, I identify the metrics that I will use to assess the outcomes of interest in Gulu. The objective is to pretest the measurement methods to ensure they can be practically applied within the constraints of locally available resources and capacity without significantly compromising their quality. Therefore, the primary research questions of Chapter 5 are:

- How do previously validated measurement protocols for PF, PA and MH need to be adapted to be delivered in the Gulu context to evaluate a SfD intervention?
- Are the adapted measurement methods reliable when applied to adolescents in Gulu?

The outcomes of Chapter 5 will inform the data collection methods for a cross-sectional analysis of a sample of young adolescents in Gulu. The objective of Chapter 6 is to assess the baseline levels of PF and MH in the target population. This is part of a community needs analysis that compares the results to global norms and considers if subjects previously abducted during the war score differently to their peers. It also provides a formative evaluation of whether a SfD intervention that focuses on improving PF is likely to impact on MH. The specific research questions addressed in Chapter 6 are:

- How does the performance of young adolescents in Gulu compare to global norms?
- Is a history of abduction associated with current deficits in adolescent PF or MH in Gulu?
- How does adolescent PF relate to MH prior to commencing the intervention in Gulu?

4.3b) Was the intervention effective?

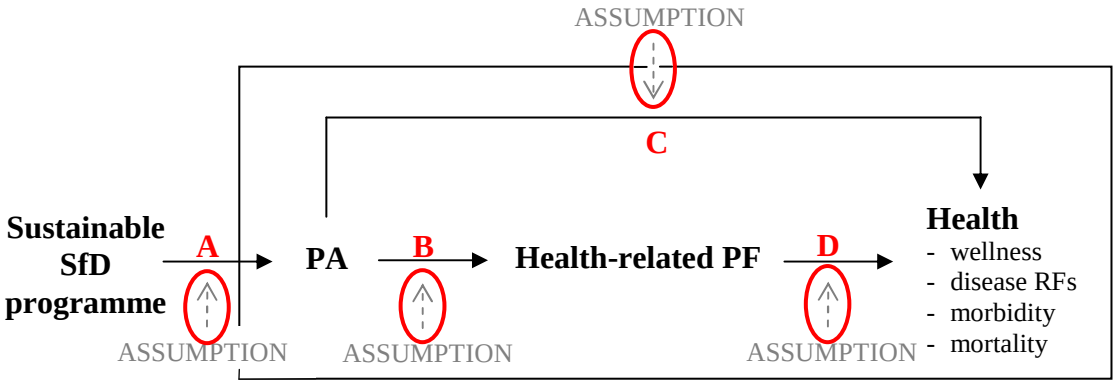
The results from Chapters 5 and 6 should inform the implementation and evaluation of a subsequent voluntary community-based SfD programme in Gulu. The objective is to use the measurement methods and baseline measures to assess the impact of the intervention. The research question addressed in Chapter 7 is:

- Does the SfD intervention reach the adolescents most in need or is there a self-selection bias at the time of registration for those who are already the fittest / healthiest in Gulu?

In Chapter 8, I will focus on the assumptions in the theoretical framework commonly cited in the SfD literature linking SfD programmes with health outcomes (FIGURE 4.1) [107,138]. The objective is to assess the impact of the intervention. The specific research questions are:

- Is a "dose" of the SfD intervention in Gulu discernible when measured by accelerometry and does the programme affect the overall levels of adolescent PA (ASSUMPTION A)?
- How does the SfD intervention affect the PF and MH of the adolescent participants in Gulu (ASSUMPTIONS B, C & D)?

FIGURE 4.1: The assumptions in the theoretical framework describing the links between a sustainable SfD programme and health-related outcomes



Source: Modified from Bouchard (2001) [107,138]

4.3c) Can the intervention be replicated, disseminated and sustained?

In Chapter 9, I will interpret the results of the Gulu case-study in the context of the strengths and weaknesses of my evaluation and the observed processes of programme delivery. The objective is to contribute to the fledgling evidence base for SfD programmes in the academic literature. I will also endeavour to make a practical contribution to informing best-practice for various stakeholders that fund, deliver, evaluate and benefit from similar programmes. The main research questions addressed are:

- How can the results of the evaluation in Gulu be explained in the context of the current evidence base and established theoretical frameworks?
- What are the critical components and necessary conditions for a SfD programme to successfully achieve and sustain PA, PF and MH outcomes in Gulu and other settings?

The practical implications of this research have been an important motivating factor for me throughout my DPhil. I believe that they are an integral part of improving the quality of delivery and accountability in a sector that has only recently begun to realise the value of effective and meaningful programme evaluation.

4.4) Research roadmap

I have developed a timeline that provides an overview of my DPhil research (FIGURE 4.2).

FIGURE 4.2: Summary timeline of my DPhil research activities

Year	Month	Research Activity	Location
2009	January ⋮ August	Background reading and problem definition Identify potential partner organisations	Oxford
	September	Exploratory fieldtrip	Gulu
	October ⋮ December	Systematic literature review and theory development	Oxford
2010	January ⋮ April	Develop evaluation design University of Oxford ethical approval	Oxford
	May	Gulu University ethical approval	Gulu
	June	Pilot testing of measurement protocols	Gulu
	July August	Local stakeholder meetings, community needs analysis and baseline assessment	Gulu
	September ⋮ November	Intervention delivery	Gulu
	December	Follow-up assessment	Gulu
2012	January	Data entry	Oxford
	February ⋮ June	Data cleaning and analysis	Oxford
	July ⋮ December	Update systematic literature review and thesis write-up	Oxford

Further details and timeframes for each component of my fieldwork are provided in the ensuing chapters. However, at this point it is important to note that my intention of independently evaluating the SfD intervention in Gulu was not realised. When I arrived in Uganda in April 2010, I was informed by the local partner that I was required to coordinate the delivery of the intervention and to help secure funding for its implementation. Although my intimate involvement with the programme management provided a valuable insight, I was unable to formally develop and assess process indicators for intervention delivery. Consequently, overseeing the implementation of a SfD programme in Gulu became a major objective of my DPhil that compromised some components of the evaluation.

4.5) Summary box

RESEARCH HYPOTHESIS AND QUESTIONS

- The underlying hypothesis of my research is that SfD programmes positively influence the PF, PA and MH of young adolescent participants in post-conflict settings.

- To test this hypothesis, I have established a series of research questions for the evaluation of a SfD programme in Gulu that will:
  - Develop locally adapted measurement methods that can feasibly and reliably assess the PA, PF and MH of young adolescents;
  - Analyse the local needs and establish whether the intervention is likely to have an effect on the PA, PF and MH of young adolescents;
  - Establish if the intervention reaches young adolescents most in need;
  - Assess the effect of the intervention on young adolescent PA, PF and MH within the context of the current evidence base and established theoretical frameworks;
  - Identify the essential processes in intervention implementation for successfully achieving and sustaining PA, PF and MH outcomes;
  - Contribute to the academic evidence base and guide future programme delivery.

- Several components of my proposed evaluation were compromised when coordinating the delivery of the intervention unexpectedly became a major objective of my DPhil.

PART III

PRESTESTING METHODS AND MATERIALS

CHAPTER 5

Feasibility and Reliability of Adapted Protocols to Measure the Physical Fitness, Mental Health and Physical Activity of Young Adolescents in Gulu, Uganda

5.1) Purpose of chapter

The purpose of this chapter is to identify and pretest appropriate measurement techniques for a SfD evaluation in Gulu. I will adapt standardised measurement protocols for PF, MH and PA to ensure they are compatible with local capacity and the resources available. The social, environmental and political norms in Gulu are also considered. I will then assess the feasibility and reliability of the modified measurement methods for subsequent use in a needs assessment and impact evaluation of a SfD programme.

5.2) Introduction

Applying established measurement methods for PF, MH and PA to a large population of adolescents living in a post-conflict setting of a LMIC is challenging. There are several pragmatic demands that are either unique to this setting or more pronounced than in the developed countries where most testing protocols are designed. There are also ethical considerations when working with this population that impact on measurement methods.

My initial fieldtrip to northern Uganda in September 2009 was exploratory and informed several of the decisions described in this chapter. Gulu has a variable and rapidly changing environment that requires an adaptable approach to programme evaluation. Although flexibility in study methods contrasts with the principles of scientific rigor, compromises are often necessary in this setting. The physical, social and political landscapes in Gulu pose overlapping logistical challenges that influence research implementation and application.

The population in Gulu is transient as people continue to relocate after the war [146]. However, poor infrastructure for public transport in northern Uganda limits mobility. If the objective is to assess a large population from a wide area, the researcher must be independently mobile and travel to the participants for testing. Motorcycles are the primary means of motorised transport and provide the most convenient access in the community. Therefore, measurement tools must be suitably portable to be transported safely by motorcycle. They must also be capable of rapid set-up and disassembly at various assessment sites.

Primary schools are the most reliable and convenient locations to access the target population in Gulu. Since Universal Primary Education (UPE) was introduced in 1997, enrolment has reached very high levels throughout the community [158]. Although working through schools is the most efficient way to measure large numbers of subjects, attendance can be haphazard

and timetable changes are often unpredictable. The facilities and support of the staff are also variable at different schools. Class sizes range between 50-130 students and school lessons are largely focused on the core subjects that are examined at the end of each term. PES is not one of these subjects and despite its inclusion in the weekly timetable most schools do not implement these classes at the P6 level [154,155,159]. Therefore, it is important that the measurement methods can engage a large number of subjects simultaneously with nominal assistance from school staff. They should cause minimal disruption to the school lessons and must be readily applicable in the diverse settings typically found in Gulu primary schools.

Optimising utility of the assessment protocol in Gulu primary schools requires minimal dependence on modern technology. Although electricity is available in several locations in Gulu, its supply is temperamental and is not widespread in schools. Generators are often used during power outages or for special events, but their use is limited by regular depletion of fuel provisions throughout the municipality. This also affects the use of computers and communication with mobile phones that require regular charging. Therefore, measurement methods must rely minimally on electronic devices and/or ongoing remote communication.

Access to primary schools requires a clear understanding of the local bureaucratic processes. It is necessary to obtain approval from district and municipal government officials as well as the school principals and class teachers. Although there is a clear hierarchy of authority that is strictly enforced, responsibility is often deferred. Similar challenges are faced when "negotiating" ethical clearance for research from local authorities. Obtaining approval from all of the relevant officials is an extended process that highlights cultural differences in time management, planning and accountability.

Several other cultural differences need to be considered when adapting the measurement protocols. Physical contact with the opposite gender in public is taboo and typically causes severe embarrassment in Gulu. Therefore, measurement methods must minimise the risk of violating this cultural norm.

"Agreeability" is also a common trait in Acholi culture that can often confuse attempts to obtain honest and realistic answers to direct questions. This is particularly the case when discussing sensitive topics (e.g. mental health) and may be compounded by language barriers. Although English is the official language in Uganda, Luo is the dialect primarily used for verbal communication in Gulu and the first learnt by most Acholi children. English comprehension skills vary significantly within the population. Consequently, measurement methods should minimise the use of questionnaire-based tools that have not been established and validated locally in Luo.

Acholi cultural norms also influence how children are named. Most people in Gulu have two names. The first is chosen by their parents. The second is typically the same as their father's first name and is adapted according to the child's gender. There is little diversity and this complicates tracking procedures when conducting research in northern Uganda. Therefore, clear subject identification is required.

Security issues must also be considered in the local context. The open display of novel and/or valuable equipment may increase the risk of theft and compromise the personal safety of the user. The introduction of unfamiliar apparatus may also instigate fears about its "spiritual" impact on health and well-being in the community. It is important to be sensitive to these issues when designing the measurement methods.

The feasibility of implementing many of the existing standardised protocols is also limited by the local capacity and resources available in Gulu. Dr. Paul Onek from the Gulu District Health Office described in disparaging terms the way international researchers often overcome this problem by "carrying" in the necessary equipment and staff to conduct their studies. Alternatively, sometimes local assistants (e.g. programme evaluators / tertiary students / researchers from within the study population) are trained to utilise the tools "carried" in by the researcher. The impact of this training is limited if the tool is then "carried" out when the study is finished or if there is insufficient local knowledge and motivation for it to be sustained.

I believe that researchers have an ethical obligation to utilise methods that are sustainable within the study population. The researcher should endeavour to use affordable equipment sourced from within the community and engage established institutions that work through reputable systems. The focus should be on strengthening existing legitimate channels rather than creating a parallel structure that may undermine local programme and research development. As a result, the process of research can also contribute to capacity building within the study population [114]. This takes time and is contrary to the "fly-in, fly-out" approach that is commonly utilised in challenging settings such as Gulu [160,161].

These concerns contribute to the ethical limitations of conducting research in Gulu. However, in settings like northern Uganda there is often insufficient infrastructure, materials, personnel or systems to fully support this approach. If this is the case and the research is ethically warranted, a practical compromise may be necessary.

One of the underlying objectives when adapting measurement methods to the logistical and ethical challenges faced in Gulu is to make the change as subtle as possible. Although the

selected protocols have been assessed for reliability and validity in their original form, it is important to consider the data quality for the modified techniques [44,47,51,128,162-166]. It is feasible to test the reliability of these adapted methods, but the capacity to apply a gold standard comparison to assess the validity in this setting is limited. The same factors that force the modification of the measurement protocols also pose problems for rigorous validity testing.

In this chapter, I will assess the feasibility of PF, MH and PA measures in Gulu in accordance with the logistical and ethical limitations that I have described above. I apply these parameters to the initial selection of measurement methods and to inform their subsequent adaptation. I also assess the quality of the data obtained with a focus on the reliability of the modified protocols. The aim of these procedures is to pretest the proposed methods for a subsequent community needs analysis (CHAPTER 6) and impact evaluation of a SfD programme in Gulu (CHAPTERS 7 & 8).

5.3) Methods

As a feasibility study, the initial stages were characterised by fluid modification of the measurement methodology. This was in response to pragmatic and ethical limitations as they were identified. The details of this process are explained in the feasibility section of the results. The final measurement methods that evolved from this process and underwent reliability testing are described below.

5.3a) Study design

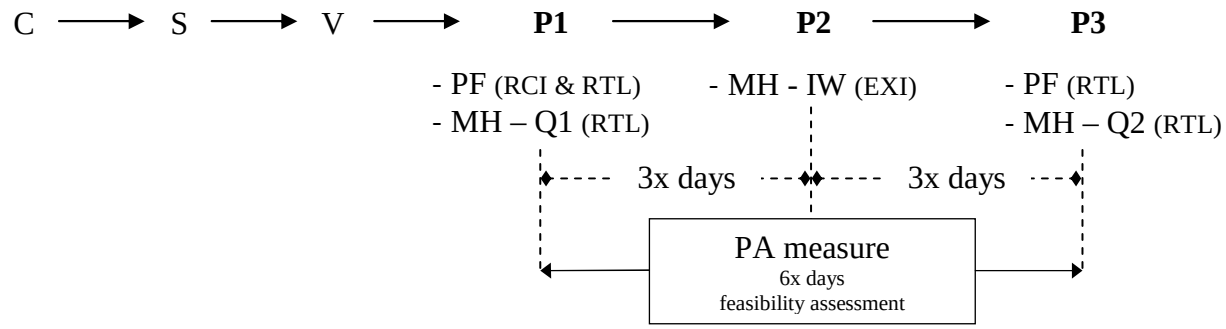
The study comprised a series of repeat measures over a period of six days. The target population (C) comprised adolescents who were eligible to participate in the intervention of a subsequent SfD evaluation. This required subjects to be aged 11-14 years. The P6 class at one school was selected (S) to complete the study and participation was voluntary (V).

The research chief investigator (RCI) and research team leader (RTL) both completed initial PF measures on the recruited subjects (P1). The RTL also completed MH measures at this time in the form of a questionnaire (Q1). Prior to P1, the RTL had received training from an expert assessor (EXI) who had contributed to the development of the MH assessment tool during its initial pilot testing in interview form.

Three days later, repeat measures for MH were taken in interview form (IW) by the EXI (P2). After a further three days, repeat measures for PF were completed by the RTL (P3). The MH measures were also repeated by the RTL in a guided questionnaire format (Q2).

PA measurement commenced one day prior to P1 and was completed at the time of P3. This was a period of one week and enabled six full days of measurement throughout the pilot testing period (FIGURE 5.1).

FIGURE 5.1: Diagram representing subject recruitment and study design



LEGEND:

Subject recruitment

C	Selection of study population (11-14 years old attending school within Gulu municipality)
S	Selection of school and class for pilot testing
V	Voluntary participation by students in selected pilot testing class

Study design and data collection

P1	Initial pilot measurement (PF, MH)
P2	First follow-up pilot measurement (MH)
P3	Second follow-up pilot measurement (PF, MH)
Q1	MH assessment delivered in guided questionnaire format at the time of P1
IW	MH assessment delivered in interview format at the time of P2
Q2	MH assessment delivered in guided questionnaire format at the time of P3
RCI	Research chief investigator
RTL	Research team leader
EXI	Expert interviewer for the MH assessment tool

5.3b) Sample size calculation

A sample size calculation was not completed due to the paucity of adequate data to estimate a relevant difference or effect size. Therefore, convenience and pragmatic restraints dictated the number of subjects in the current study. The sample size in a previous study assessing the validity and reliability of the APAI in Gulu also guided the MH assessment [128].

5.3c) Subject recruitment

Recruiting and testing adolescents through primary schools provided the only feasible and reliable option in Gulu. Transporting and retaining subjects for multiple assessments at an independent testing centre was not logistically feasible. Providing incentives for participation compromised ethical principles of conducting research in a resource-poor setting.

Laroo Boarding School was selected for convenience of subject tracking and repeat measures. Selecting a boarding school for this pilot study facilitated access to young adolescents on multiple occasions during a short period of time with minimal disruption to school activities. It was also deemed the most secure environment to trial PA measurement for the first time in the community. In a setting where there was a diverse range of ages at each educational level, the P6 class was selected because it most closely matched the target population.

There were 35 boys and 49 girls enrolled in the selected class who were informed about the study and invited to take part. Consent was obtained from the participants as well as the adult responsible for their welfare within the boarding school system. All of the subjects were included in the PF assessment. However, the PA assessment was restricted to six subjects (6 boys) because of the limited availability of testing equipment. Recruitment for the MH assessment was restricted to 20 subjects (11 boys, 9 girls) because the EXI deemed that this was the maximum number that could be interviewed per day by one assessor.

5.3d) Subject identification and tracking procedures

The following information was collected at P1, P2 and P3 to facilitate subject identification and tracking during the study:

- Subject's name
- Gender
- Mother's / guardian's name
- Father's / guardian's name
- School
- Class
- Division of residence

All of these details were kept in strict confidence by the assessors. The electronic files were password locked and the hard copies of the assessment sheets were securely stored.

5.3e) Outcome variables and metrics

i) Physical fitness

The outcome variables for the PF measures comprised:

- Body composition
- Muscular power/strength
- Cardiorespiratory fitness

Body composition was assessed using anthropometric measures to calculate BFA and HFA. BFA provided an indication of thinness / overweight / obesity and HFA was used as a gauge of stunting as a result of chronic malnutrition [166]. Calculating BFA and HFA required the measurement of height, weight and age. These were the metrics assessed in this study. Initial plans to assess waist-to-hip ratio and skinfold measurements were abandoned because they were not feasible.

Muscular strength was tested using a 1xRM of the standing broad jump (SBJ). The SBJ has been advocated as a practical, efficient and resourceful index of muscular fitness in youth [165].

Cardiorespiratory fitness was measured with the multi-stage fitness test (MSFT). The standardised MSFT has been found to be a valid and reliable test for estimating adolescent VO_2 max [51,162-164]. Performance on the MSFT during adolescence has also been associated with cardiovascular health profile later in life [44].

ii) Mental health

Initial plans to assess self-concept were abandoned because they were not feasible. The APAI was used to assess other facets of MH associated with PA.

The APAI was identified during the systematic literature review (SECTION 2.9). I chose to use the APAI because of its intrinsic link with local language, culture, behaviours and attitudes. In contrast to other tools validated for global utility and commonly used in similar settings, the APAI provided an emic account of MH and was not born from a "Western" conceptual model. It was developed and validated in Luo specifically for local communities by a team that comprised international researchers working with local psychosocial workers. Initial qualitative studies identified and defined several local MH syndromes evident in youths who were 10-17 years old [127]. A 60-item interview measurement tool for these disorders was subsequently developed and quantitative studies established its validity and reliability [128].

In this study, all of the APAI items that described the local MH syndromes of interest were retained to be trialled in a questionnaire format (TABLE 5.1). These 44 items described the local constructs of Two Tam (TT), Kumu (K), Par (P) and Ma Lwor (ML).

The translation of the APAI into English for analysis and international dissemination enabled several parallels to be drawn between the local syndromes and MH disorders commonly diagnosed in developed countries. TT, K and P have been described as "depression-like" syndromes [127]. The combined score for the APAI items that contribute to each of these syndromes can be used to describe a "depression-like" total (DT). Similarly, ML has several "anxiety-like" characteristics [127]. These local syndromes were selected because of the current evidence associating PA and sport with depression and anxiety [25,107]. An APAI total (AT) score for all 44 items was also assessed in this study (TABLE 5.1).

TABLE 5.1:
Items of the APAI that contribute to each of the mental health outcome variables - two tam, kumu, par, ma lwor, depression-like total and APAI total [127,128]
(NOTE: * Indicates item is included for the outcome variable described)

		Two tam	Kumu	Par	Ma lwor	Depres-like total	APAI total
1	**Atamo madwong** *I have a lot of thoughts*	*		*		*	*
2	**Atiko par** *I have constant worries*	*				*	*
3	**Koma weng rem arema** *I have pain all over my body*	*				*	*
4	**Wiya pe tiyo maber** *My brain is not functioning well*	*				*	*
5	**Atamo ni konya pe** *I think I am of no use*	*				*	*
6	**Atamo ni a dene** *I think about suicide*	*		*		*	*
7	**Atito peko ma atye kwede** *I talk constantly about my problems*	*				*	*
8	**Abedo kena-kena** *I sit alone*	*	*			*	*
9	**Ape ki miti me kwan** *I lose interest in school*	*				*	*
10	**Wiya bara** *I get headaches*	*	*			*	*
11	**Doga pe mito cam** *I lose my appetite*		*	*	*	*	*
12	**Awinyo ma lit I cwinya** *I feel a lot of pain in my heart*		*			*	*
13	**Abedo ma nongo ateno tika** *I sit with my cheek in my palm*		*			*	*
14	**Akok ka atye kena** *I cry when I'm alone*		*			*	*
15	**Pe anino I dye wor** *I do not sleep at night*		*		*	*	*
16	**Ading** *I am disobedient*		*	*		*	*
17	**Awinyo koyo** *I feel cold*		*			*	*
18	**Atiko butu (I dye ceng)** *I lie down all the time during the day*		*			*	*
19	**Aparo par madwong** *I have lots of worries*		*	*		*	*
20	**Amito bedo kena** *I want to be alone*			*	*	*	*
21	**Akejo oyot** *I am easily annoyed*			*		*	*
22	**Amako wiya** *I hold my head*			*		*	*
23	**Tama bedo kama bor I ot kwan** *I lose concentration in class*			*		*	*
24	**Amato kongo** *I drink alcohol*			*	*	*	*
25	**Ayeto luwota** *I insult friends*			*		*	*
26	**Pe amoto dano** *I don't greet people*			*	*	*	*
27	**Pe atamo atir** *I don't think straight*			*		*	*
28	**Ajwat kekena** *I mutter to myself*			*		*	*
29	**Pe ageno dano** *I don't trust*			*		*	*
30	**Awinyo ni ape kigin mo wek akonye kede** *I feel I can do nothing to help myself*			*		*	*
31	**Amako ludito matek** *I cling to elders*				*		*
32	**Atamo ni anyima peke** *I think I do not have a future*				*		*
33	**Atiko ngwec angweca** *I am constantly running around*				*		*
34	**Pe amito woo** *I don't like noise*				*		*
35	**Atamo ni dano tye ka ryema** *I think people are chasing me*				*		*
36	**Cwinya gone oyot** *I have a fast heart rate*				*		*
37	**Alworo bedo kena** *I fear being alone*				*		*
38	**Cwinya cwer** *I feel sad*	*				*	*
39	**Atamo jami maraco** *I think of bad things*	*				*	*
40	**Agoro** *I am weak*	*	*	*		*	*
41	**Pe amito lok** *I don't feel like talking*	*	*			*	*
42	**Wiya maro wil ngo ma atamo** *I like to forget*	*				*	*
43	**Atiko kok akoka** *I cry continuously*	*		*		*	*
44	**Pe aparo kedi a kwo nyo a too** *I do not care whether I live or die*	*				*	*

iii) Physical activity

PA was measured with accelerometers. Only boys were assessed due to the limited number of accelerometers available and their intended use in the subsequent studies. At the time of this study, accelerometry was seen as a valid and reliable method for measuring PA duration, frequency and intensity under free-living conditions [167]. The accelerometers were biaxial and the epoch length was set to 10 seconds. Results were expressed in terms of activity "counts", which could be converted to indicate different levels of PA intensity.

iv) History of abduction

The following question was included at P1 to indicate history of abduction:

Gi maki i kare me lweny? / *Were you abducted during the war?*

Kuu ☐ *No* **Eyo** ☐ *Yes*

5.3f) Data collection procedures

Data collection was initially planned to take place during the PES classes. However, flexibility was required when it became evident that teachers did not follow this schedule and PES was not actually implemented at the P6 level. Therefore, all pilot testing was shifted to after the final class at the end of the school day.

A hand written diary was kept by each of the assessors to record the feasibility of applying the adapted methods in the field. Notes were taken at the time of the test and during debriefing sessions after P1 and P3. All deviations from the planned testing procedure and their cause were recorded. The information from the field diaries and debriefing sessions was collated in one notebook and summarised for each outcome measure. This included reference to all modifications that were made to the measurement protocols and the reasons for these changes.

For the quantitative assessment, each subject was issued with a recording sheet prior to commencing measurement. The information for subject identification, tracking and history of abduction was completed by each subject on their recording sheet with the guidance of the assessors. The class and PES teachers also provided help for the subjects if required. The testing protocols that were adapted in response to the feasibility assessment were then applied in the order of height, weight, SBJ, MSFT and APAI (APPENDIX A3 – A6).

The APAI interview was conducted according to the protocol described by Betancourt et al (2009) [128]. The answers were recorded by the expert assessor on a separate recording sheet for each subject.

All of the quantitative data for the PF and MH measurements were entered into PASW Statistics 18 for analysis.

The PA assessment followed the protocol that was adapted in response to the feasibility assessment (APPENDIX A7). The data were stored by each accelerometer until they were downloaded using the ActiLife Data Analysis Software v5.5.5 at the end of the testing period for validation and analysis.

5.3g) Data analysis and presentation

The feasibility results for PF, MH and PA were presented in narrative form. All of the quantitative data were cleaned and checked for outliers prior to commencing the analysis. The underlying assumptions for each statistical test were also verified before proceeding. The reporting of descriptive statistics for each of the outcome variables included the crude mean, 95% confidence interval (CI) and sample size (n).

i) Physical fitness

The data were stratified according to gender and descriptive statistics were calculated for height, weight, SBJ and MSFT. Although the method of reporting the MSFT results categorised the data, based on previous convention, they were treated as a continuous variable for statistical analysis [168-170]. The results for all of the PF outcome variables were presented in a summary table.

Inter-tester (i.e. RCI:P1 vs. RTL:P1) and intra-tester (i.e. RTL:P1 vs. RTL:P3) reliability for height, weight, SBJ and MSFT were assessed using an intra-class correlation coefficient (ICC). A two-way random analysis for absolute agreement of a single measure (significance value $p<0.05$) was used for each of the comparisons of the crude data. The results of the reliability testing using an ICC were presented in a table stratified by gender.

The reliability of the self-reported age variables was assessed by calculating a percentage of concordance when comparing P1 vs. P3. The 95% CI was also established and the results were tabulated.

ii) Mental health

Descriptive statistics for each of the MH outcome variables were calculated and a summary was tabulated. Several protocols were then utilised to assess the intra-tester and inter-method reliability of the MH outcome variables.

To assess the intra-tester reliability (i.e. RTL:P1 vs. RTL:P3), Betancourt et al (2009) set a precedent of using a correlation coefficient [128]. These results indicated the strength of the association between the repeat measures. To enable direct comparison with the results of

Betancourt et al (2009), a Pearson's r correlation coefficient was used to assess intra-tester association of the crude results for the APAI in questionnaire format.

The same method was used to assess inter-method association between interview and questionnaire format (i.e. IW vs. Q1). To facilitate cross-study comparison, the correlation coefficients found by Betancourt et al (2009) were included in the summary results table.

However, the assessment of reliability requires testing for repeated measure agreement. This was assessed for the intra-tester reliability with an ICC of the crude results for each of the MH outcome measures. A two-way random analysis for absolute agreement of a single measure (significance value $p<0.05$) was utilised.

Establishing inter-method agreement is more complex and a battery of tests was used. Lee et al (1989) described a protocol to statistically evaluate agreement between two quantitative methods that required [171]:

1) No statistically significant difference between the means;
2) The lower limit of the 95% CI of the ICC to be >0.75;
3) No marked additive or non-additive systematic bias.

A paired t-test was used to assess the difference between the crude means of IW vs. Q1. The ICC was a two-way mixed analysis for absolute agreement of a single measure (significance value $p<0.05$). Bias was assessed visually by plotting the difference between the methods against their mean for each subject (i.e. Bland-Altman plot) [171,172].

However, the merits of this approach have been questioned [173]. Bland et al (1986) did not advocate the comparison of means and use of an ICC to assess inter-method reliability [172].

Although they agreed with the use of a Bland-Altman plot to assess bias, they believed it had greater utility in assessing inter-method agreement. Bland et al (1986) outlined a practical approach for comparing the results of two methods that measure the same variable [172]:

1) Complete a Bland-Altman plot for the crude results of each variable;
2) Calculate the limits of agreement as two standard deviations above and below the mean difference between the groups;
3) Calculate the precision of the limits of agreement as the 95% CI around the upper and lower limits;
4) Test if the difference between the methods was normally distributed and/or if the normality was improved by applying a Ln transformation to the original data;
5) If indicated, repeat steps 1-3 with the Ln transformed data and back-transform (antilog) the results;
6) Interpret the limits of agreement for the crude and Ln transformed data to assess the interchangeability of the methods in a clinical context.

The results of all the procedures used to assess inter-method agreement of the MH outcome variables were summarised in a series of tables and plots presented in chronological order.

The internal consistency for the APAI questionnaire total and each of the sub-scales was assessed by calculating a cronbach's α statistic for the Q1 scores. The results were tabulated.

iii) Physical activity

In addition to field notes, the feasibility of using accelerometry was assessed by applying the following criteria for obtaining valid measures for analysis [174]:

- A valid week required at least four days of wear (including at least one weekend day and three week days);

- A valid day required at least 10 hours of wear time per day;
- Non-wear time was defined as 20 consecutive minutes of "0" readings.

These criteria were recommended as an appropriate balance between feasibility and validity [174]. Descriptive statistics for the accelerometry measures were tabulated according to the validity criteria.

5.3h) Ethical approval

The University of Oxford Tropical Research Ethics Committee (OxTREC) provided ethical approval for the proposed study.

Research clearance and a recommendation to the Uganda National Council for Science and Technology were obtained from the Ethics Review Committee of Gulu University. Official documentation of this approval was only granted in exchange for an "administrative fee" of $US200. Ironically, this "negotiation" was paid in cash and there was no official receipt issued by the chairman receiving the funds.

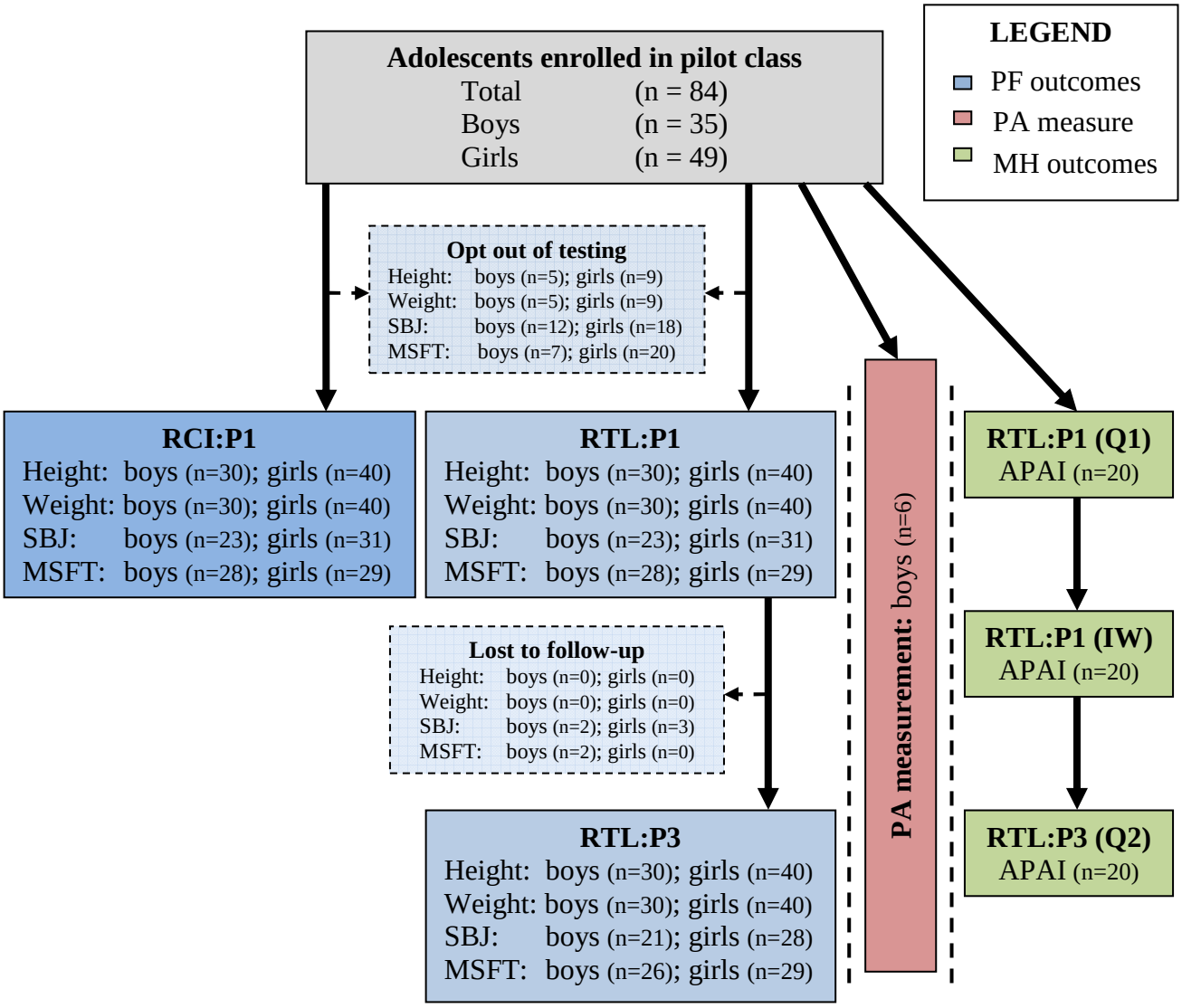
Approval to access schools and conduct testing was granted by the Gulu District Education Officer, Municipality Education Officer, District Sports Officer and the Head Teachers of the target schools. The process of organising meetings was rarely scheduled and appointments were seldom kept. Instead, the only reliable way to meet with these officials was to wait outside their office until they were available. This process took anywhere between 5 minutes to several days and often became more efficient once a relationship had been established based on mutual benefit.

5.4) Results

5.4a) Sample distribution

All of the enrolled students were present at school throughout the pilot testing period. A variable number of participants volunteered to participate in each of the PF tests. The six volunteers for PA measurement were randomly selected and the MH assessment was completed by the first 20 participants (11 boys, 9 girls). The number of subjects lost to follow-up at P2 and P3 varied according to the outcome being measured (FIGURE 5.2).

FIGURE 5.2: Flow diagram of the sample for the pilot study in Gulu



5.4b) Physical fitness

i) Feasibility assessment

The selected field tests provided an indication of body composition, musculoskeletal condition and cardiorespiratory fitness. They required minimal modification from their validated form to be ethically and pragmatically feasible in the Gulu context.

Standardised protocols use calibrated stadiometers and digital scales for measurements to the closest 0.1cm for height and 0.1kg for weight. These devices were not available in Gulu and would be difficult to transport using a motorcycle. Most digital scales also depend on a reliable power supply and are a valuable item susceptible to theft. Instead of this equipment a locally acquired tailor's measuring tape that could be attached to any wall and mechanical scales were used. The measurement resolution was also reduced to the closest 1cm for height and 1kg for weight (APPENDIX A3).

Age to the nearest month is easily calculated for adolescents who know their DOB. However, birthdays are not significant occasions in Acholi culture and the calculation of age in months was adapted to allow for poor DOB recall (APPENDIX A3).

The assessment of waist-to-hip ratio and skinfold were abandoned because they required culturally inappropriate physical contact between the assessors and the subjects. The nature of the contact for these tests was particularly provocative in Acholi culture regardless of gender and caused excessive embarrassment. The equipment required to measure skinfold accurately was also not locally available.

The standardised protocol for assessing the SBJ typically uses either a jumping board next to a level sand pit or an assessment mat clearly marked with a start line and distance indicators.

These facilities and equipment were not available in Gulu. Although it would have been possible to construct a measurement mat using local materials, transporting it using a motorcycle would have been difficult. A locally sourced tailor's measuring tape secured by two large nails and a starting line marked with ash were used instead. These were set up in an outdoor area that was flat, dry and covered by a thin layer of loose dirt to show the foot markings after each jump (APPENDIX A4).

The MSFT is typically measured using a standardised audio recording in a 20-metre testing area demarcated using a measuring tape and marking cones. In Gulu the audio recording was provided to the assessors on a cassette, CD, DVD and USB stick. After assessing the options, a small battery powered cassette player and portable speaker that was audible outdoors by 20-30 participants were purchased in Gulu. The testing area was measured using a 20-metre length of rope that was sourced locally. Although the marking cones could be purchased in Gulu, the cost was prohibitively expensive to be sustainable. Their value also made them a target for theft. Therefore, the marking cones were replaced by plastic bowls purchased at the Gulu market and placed upside-down to demarcate the testing area. However, the utility of the bowls also made them a target for theft. This problem was averted by cutting holes in the bottom of the bowls before using them for the MSFT assessment (APPENDIX A5).

Initial plans to conduct all of the PF tests in a specific order were abandoned at P1 due to logistical and/or environmental interruptions (e.g. weather, demands of school staff, local time management). Consequently, flexibility in the order of testing was necessary despite its potential effect on the results. However, assessing the SBJ immediately after the MSFT was avoided to minimise the effect of fatigue on the results. Improved coordination with the school and fortuitous weather conditions meant that testing at P3 proceeded in the previously planned order (i.e. height → weight → SBJ → MSFT).

ii) Descriptive statistics

The mean and 95% CI for height, weight, SBJ and MSFT were calculated for all of the subjects who participated in the pilot testing (TABLE 5.2).

TABLE 5.2: Crude mean and 95% CI of the PF outcomes in the pilot study in Gulu - stratified by gender, assessor and test

Outcome variable	Gender	Assessor and test		
		RTL:P1	**RCI:P1**	**RTL:P3**
Height (cm)	Boys	**163.20** (159.29↔167.11) n=30	**163.10** (159.08↔167.12) n=30	**163.10** (159.22↔166.98) n=30
	Girls	**159.58** (157.66↔161.50) n=40	**159.75** (157.77↔161.73) n=40	**159.70** (157.76↔161.64) n=40
Weight (kg)	Boys	**49.57** (45.87↔53.27) n=30	**49.73** (46.06↔53.40) n=30	**49.77** (46.02↔53.52) n=30
	Girls	**50.40** (48.05↔52.75) n=40	**50.53** (48.15↔52.91) n=40	**50.43** (48.07↔52.79) n=40
SBJ (cm)	Boys	**195.70** (185.38↔206.02) n=23	**195.78** (185.44↔206.12) n=23	**197.43** (187.00↔207.86) n=21
	Girls	**163.97** (157.08↔170.86) n=31	**164.03** (157.07↔170.99) n=31	**166.50** (162.18↔174.74) n=28
MSFT (level complete)	Boys	**5.86** (4.96↔6.76) n=28	**5.89** (4.98↔6.80) n=28	**6.77** (6.05↔7.49) n=26
	Girls	**3.31** (2.69↔3.93) n=29	**3.24** (2.64↔3.84) n=29	**4.34** (3.72↔4.96) n=29

NOTE: RTL = research team leader, RCI= research chief investigator, P1 = pilot measurement 1, P3 = pilot measurement 3.

iii) Inter-tester reliability

I found that the inter-tester reliability was satisfactory for all of the PF outcome variables assessed in this study. The underlying assumptions for the inter-tester reliability analyses were satisfied (APPENDIX A8). Excellent inter-tester reliability was demonstrated in both genders for the crude results of height (ICC: boys = 0.999, girls = 0.993), weight (ICC: boys = 0.997, girls = 0.996), SBJ (ICC: boys = 1.000, girls = 1.000) and MSFT (ICC: boys = 0.997, girls = 0.988). These were highly significant [$p<0.001$] for all of the variables (TABLE 5.3).

iv) Intra-tester reliability

I found that the intra-tester reliability was satisfactory for all of the PF outcome variables assessed in this study. The underlying assumptions for the intra-tester reliability analyses were satisfied (APPENDIX A8). Excellent intra-tester reliability was also demonstrated in both genders for the crude results for height (ICC: boys = 0.999, girls = 0.996) and weight (ICC: boys = 0.998, girls = 0.985). The ICC was lower for both genders when measuring SBJ (ICC: boys = 0.976, girls = 0.895) and MSFT (ICC: boys = 0.821, girls = 0.787). However, all of these results remained highly significant [$p<0.001$] for both genders (TABLE 5.3).

TABLE 5.3: Inter-tester (RTL:P1 vs. RCI:P1) and intra-tester (RTL:P1 vs. RTL:P3) reliability of the crude PF outcomes in the pilot study in Gulu - stratified by gender

Outcome variable	Gender	Inter-tester reliability		Intra-tester reliability	
		n	ICC (single measure)	n	ICC (single measure)
Height (cm)	Boys	30	0.999	30	0.999
	Girls	40	0.993	40	0.996
Weight (kg)	Boys	30	0.997	30	0.998
	Girls	40	0.996	40	0.985
SBJ (cm)	Boys	23	1.000	21	0.976
	Girls	31	1.000	28	0.895
MSFT (level complete)	Boys	28	0.997	26	0.821
	Girls	29	0.988	29	0.787

NOTE: ICC was a two-way random analysis for absolute agreement.
NOTE: All of the results were statistically significant at a level of $p<0.001$

v) Reliability of self-reported age

I found that DOB recall was poor in the study sample, but the subjects were able to report their age in years more reliably. Of the 70 adolescents who were asked at P1, three did not report to know their age in years and 57 did not know their DOB. For those who did report to know how old they were, the concordance with self-reported age at P3 was 88.06% for age in years compared to 61.54% for age in months (DOB). Therefore, 84.29% of all subjects asked were able to report their age reliably in years and 11.43% were consistent when identifying their DOB (TABLE 5.4).

TABLE 5.4: Concordance of self-reported age in the pilot study in Gulu (P1 vs. P3)

Outcome variable	n		Concordant responses at P1 and P3		
	Number of subjects who were asked	Number of subjects who responded	Total number	% of subjects who were asked (95% CI)	% of subjects who responded (95% CI)
Age in years	70	67	59	84.29 (73.84↔91.17)	88.06 (77.91↔94.08)
Age in months (DOB)	70	13	8	11.43 (5.66↔21.21)	61.54 (35.41↔82.40)

NOTE: The 95% CIs were calculated using the modified Wald method [175].

5.4c) Mental health

i) Feasibility assessment

The initial plans to measure self-concept were abandoned due to the absence of an appropriately validated assessment instrument. Coalter (2006) developed an interview-based tool to assess the "Western" construct of self-concept in young adolescents and had it translated into Luo [129]. Although it had been previously used in a SfD evaluation in Gulu, pilot testing of this instrument to assess measurement quality in the local context was very limited. Concerns about its validity stemmed from feedback received in Gulu regarding its translation and the Acholi cultural trait of "agreeability". No other assessment tools for self-concept had been previously translated into Luo and adequately validated in northern Uganda for inclusion in this study.

The APAI had previously only been delivered in Gulu as an interview. Using a questionnaire format made it possible to collect data from a relatively large number of subjects in a short period of time. As a locally developed tool, its delivery was both practically and ethically feasible.

However, there were several challenges that were addressed during the pilot phase. Several subjects had trouble comprehending the written questionnaire correctly and/or copied the answers of their friends sitting next to them. This was addressed by rearranging the participants in the classroom randomly before commencing the questionnaire. The RTL then guided the students through each item and encouraged them to complete the questions independent of their peers. The class and PES teachers were asked to assist any participants who continued to have difficulty and to help maintain discipline in the classroom. These actions also reduced the number of items that were accidentally missed or ambiguously answered by the participants (APPENDIX A6).

Initial plans to conduct the MH testing before assessing PF were abandoned at P1 due to logistical and/or environmental interruptions (e.g. weather, demands of school staff, local time management). However, improved coordination with the school and fortuitous weather conditions meant that testing at P3 proceeded in the previously planned order (i.e. MH → PF).

ii) Descriptive statistics

The crude mean and 95% CI for all of the MH outcome variables were calculated for the 20 subjects (11 boys, 9 girls) who were randomly selected to complete the APAI (TABLE 5.5).

TABLE 5.5: Crude mean and 95% CI of the MH outcomes in the pilot study in Gulu – stratified by assessor and test

Outcome variable	Assessor and test		
	RTL:Q1	**EXI:IW**	**RTL:Q2**
TT (score)	**15.60** (11.11↔20.09) n=20	**13.80** (11.05↔16.55) n=20	**17.30** (13.29↔21.31) n=20
K (score)	**11.90** (8.46↔15.34) n=20	**10.00** (7.70↔12.30) n=20	**11.75** (8.39↔15.11) n=20
P (score)	**13.75** (9.05↔18.45) n=20	**12.35** (9.26↔15.44) n=20	**15.95** (10.72↔21.18) n=20
ML (score)	**9.15** (6.06↔12.24) n=20	**7.20** (5.22↔9.18) n=20	**9.50** (6.19↔12.81) n=20
DT (score)	**31.95** (22.47↔41.43) n=20	**27.90** (21.93↔33.87) n=20	**34.75** (25.27↔44.23) n=20
AT (score)	**38.50** (27.28↔49.72) n=20	**33.00** (25.85↔40.15) n=20	**41.50** (29.89↔53.11) n=20

NOTE: RTL = research team leader, EXI = expert APAI interviewer, Q1 = APAI questionnaire 1, Q2 = APAI questionnaire 2, IW = APAI interview.

iii) Intra-tester reliability

I found that the intra-tester reliability was satisfactory for all of the MH outcome variables assessed in this study. The underlying assumptions for the intra-tester reliability analyses were satisfied (APPENDIX A9).

Pearson's r correlation coefficient demonstrated good intra-tester association for the crude results of all the MH outcome variables (TT = 0.805, K = 0.777, P = 0.823, ML = 0.773, DT = 0.839, AT = 0.835). These were highly statistically significant [$p<0.001$]. The questionnaire intra-tester association was comparable to the Pearson's r correlation coefficient described for the APAI in interview format (TABLE 5.6) [128].

The ICC of the crude results for all of the MH outcome variables demonstrated satisfactory intra-tester agreement (TT = 0.796, K = 0.785, P = 0.811, ML = 0.779, DT = 0.839, AT = 0.836). These were also highly statistically significant [$p < 0.001$] and suggested that there was adequate intra-tester reliability to deliver the APAI in questionnaire form (TABLE 5.6).

TABLE 5.6: Intra-tester reliability (Q1 vs. Q2) and inter-method reliability (Q1 vs. IW) of the crude MH outcomes in the pilot study in Gulu contrasted with previous reliability assessment of the APAI in interview format

Outcome variable	n	Type of reliability assessed (test)	Research study	Statistical test	Test statistic	p
TT (score)	19	Inter-tester (IW)	Betancourt	Pearson r	0.860	-
	30	Intra-tester (IW)	Betancourt	Pearson r	0.790	-
	20	Intra-tester (Q1 vs. Q2)	Richards	Pearson r	0.805	<0.001
	20	Intra-tester (Q1 vs. Q2)	Richards	ICC	0.796	<0.001
	20	Inter-method (Q1 vs. IW)	Richards	Pearson r	0.836	<0.001
K (score)	19	Inter-tester (IW)	Betancourt	Pearson r	0.920	-
	30	Intra-tester (IW)	Betancourt	Pearson r	0.830	-
	20	Intra-tester (Q1 vs. Q2)	Richards	Pearson r	0.777	<0.001
	20	Intra-tester (Q1 vs. Q2)	Richards	ICC	0.785	<0.001
	20	Inter-method (Q1 vs. IW)	Richards	Pearson r	0.806	<0.001
P (score)	19	Inter-tester (IW)	Betancourt	Pearson r	0.780	-
	30	Intra-tester (IW)	Betancourt	Pearson r	0.790	-
	20	Intra-tester (Q1 vs. Q2)	Richards	Pearson r	0.823	<0.001
	20	Intra-tester (Q1 vs. Q2)	Richards	ICC	0.811	<0.001
	20	Inter-method (Q1 vs. IW)	Richards	Pearson r	0.839	<0.001
ML (score)	19	Inter-tester (IW)	Betancourt	Pearson r	0.620	-
	30	Intra-tester (IW)	Betancourt	Pearson r	0.680	-
	20	Intra-tester (Q1 vs. Q2)	Richards	Pearson r	0.773	<0.001
	20	Intra-tester (Q1 vs. Q2)	Richards	ICC	0.779	<0.001
	20	Inter-method (Q1 vs. IW)	Richards	Pearson r	0.763	<0.001
DT (score)	19	Inter-tester (IW)	Betancourt	Pearson r	0.810	-
	30	Intra-tester (IW)	Betancourt	Pearson r	0.810	-
	20	Intra-tester (Q1 vs. Q2)	Richards	Pearson r	0.839	<0.001
	20	Intra-tester (Q1 vs. Q2)	Richards	ICC	0.839	<0.001
	20	Inter-method (Q1 vs. IW)	Richards	Pearson r	0.879	<0.001
AT (score)	19	Inter-tester (IW)	Betancourt	Pearson r	0.740	-
	30	Intra-tester (IW)	Betancourt	Pearson r	0.840	-
	20	Intra-tester (Q1 vs. Q2)	Richards	Pearson r	0.835	<0.001
	20	Intra-tester (Q1 vs. Q2)	Richards	ICC	0.836	<0.001
	20	Inter-method (Q1 vs. IW)	Richards	Pearson r	0.886	<0.001

NOTE: ICC was a two-way mixed analysis for absolute agreement.

NOTE: IW = APAI interview, Q1 = APAI questionnaire 1, Q2 = APAI questionnaire 2.

NOTE: Betancourt et al (2009) did not list p values for the Pearson's r correlation coefficients calculated in their study [128].

iv) Inter-method reliability

I found satisfactory inter-method association, but poor inter-method agreement for all of the MH outcomes assessed in this study. The underlying assumptions for the inter-method reliability analyses were satisfied (APPENDIX A9).

Pearson's r correlation coefficient demonstrated good inter-method association for the crude results of all the MH outcome variables (TT = 0.836, K = 0.806, P = 0.839, ML = 0.763, DT = 0.879, AT = 0.886). These were highly statistically significant [$p < 0.001$]. The inter-method Pearson's r correlation coefficient for the APAI questionnaire in this pilot study was comparable to the inter-tester and intra-tester association described by Betancourt et al (2009) for the APAI in interview format (TABLE 5.6).

Inter-method agreement was first assessed by the protocol described by Lee et al (1989) [171]. When comparing Q1 vs. IW, there was no statistically significant difference between the crude means for TT [p=0.201], K [p=0.091], P [p=0.322], ML [p=0.075], DT [p=0.136] or AT [p=0.084]. All of the MH outcome variables passed this stage of the protocol (TABLE 5.7).

However, the lower limit of the 95% CI for the ICC tests was <0.75 for the crude results of all the MH outcome variables (TT = 0.456, K = 0.427, P = 0.512, ML = 0.335, DT = 0.530, AT = 0.529). All of the APAI sub-scales failed this stage of the Lee et al (1989) inter-method reliability testing protocol (TABLE 5.7).

TABLE 5.7: Comparison of repeat measures using the crude mean and the 95% CI of the ICC to assess the inter-method (Q1 vs. IW) reliability of the MH outcomes in the pilot study in Gulu

Outcome variable	n	Paired t-test		ICC
		Test statistic	p	(95% CI)
TT (score)	20	1.325	0.201	0.737 (0.456↔0.886)
K (score)	20	1.782	0.091	0.725 (0.427↔0.881)
P (score)	20	1.018	0.322	0.770 (0.512↔0.901)
ML (score)	20	1.884	0.075	0.668 (0.335↔0.854)
DT (score)	20	1.556	0.136	0.781 (0.530↔0.907)
AT (score)	20	1.825	0.084	0.785 (0.529↔0.909)

NOTE: ICC was a two-way mixed analysis for absolute agreement.
NOTE: IW = APAI interview, Q1 = APAI questionnaire 1.

The Bland-Altman plots indicated a bias when comparing the crude data of the two methods for all of the outcome variables. As scores increased, there appeared to be a greater relative increase in Q1 when compared to IW. The limits of agreement for the crude data were also wide for all of the outcome variables. This suggested poor agreement between Q1 and IW (FIGURES 5.3 – 5.8).

FIGURE 5.3: Bland-Altman plot for the crude TT scores in the pilot study in Gulu (Q1 vs. IW)

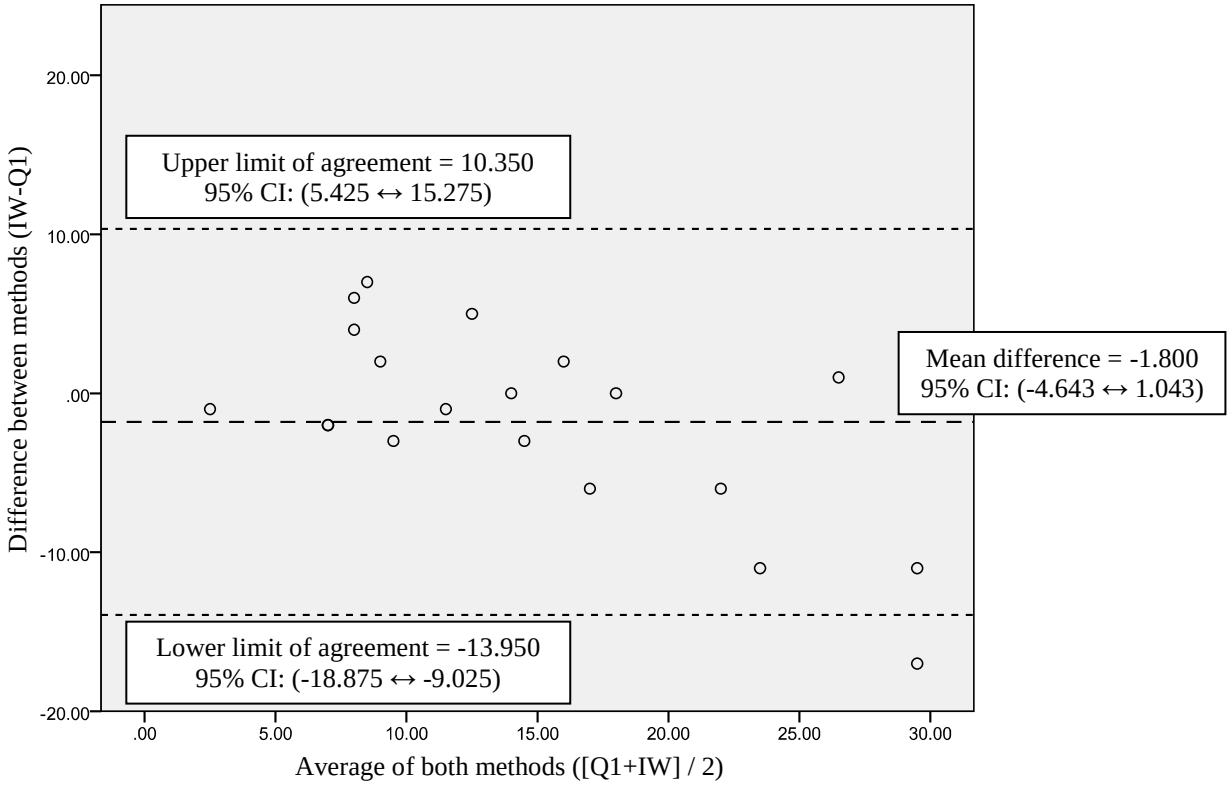


FIGURE 5.4: Bland-Altman plot for the crude K scores in the pilot study in Gulu (Q1 vs. IW)

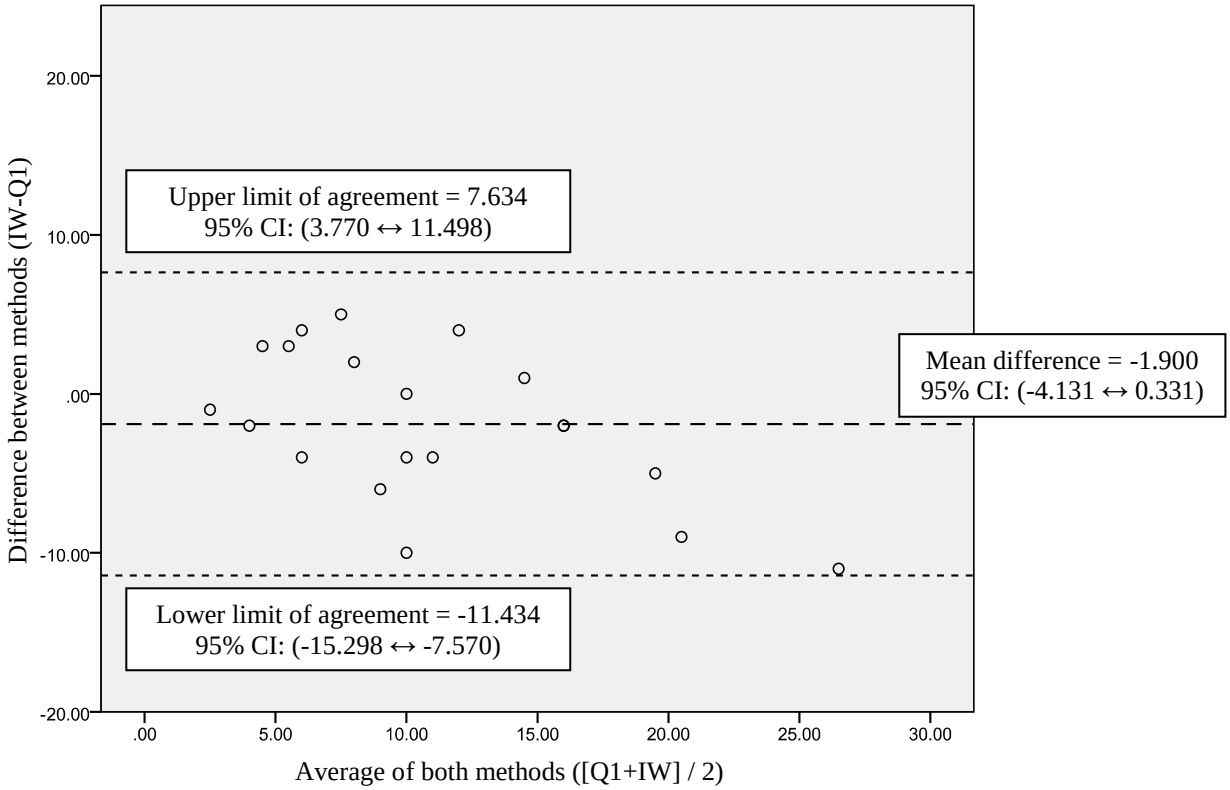


FIGURE 5.5: Bland-Altman plot for the crude P scores in the pilot study in Gulu (Q1 vs. IW)

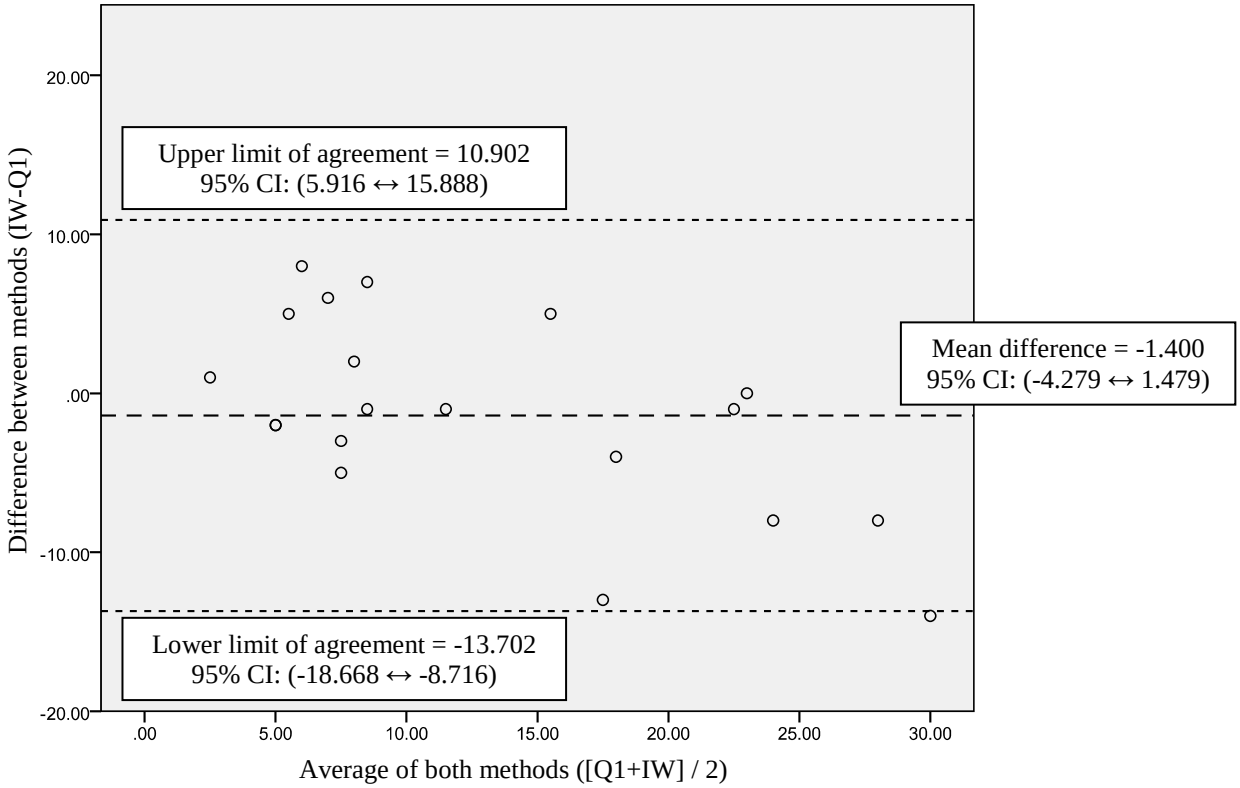


FIGURE 5.6: Bland-Altman plot for the crude ML scores in the pilot study in Gulu (Q1 vs. IW)

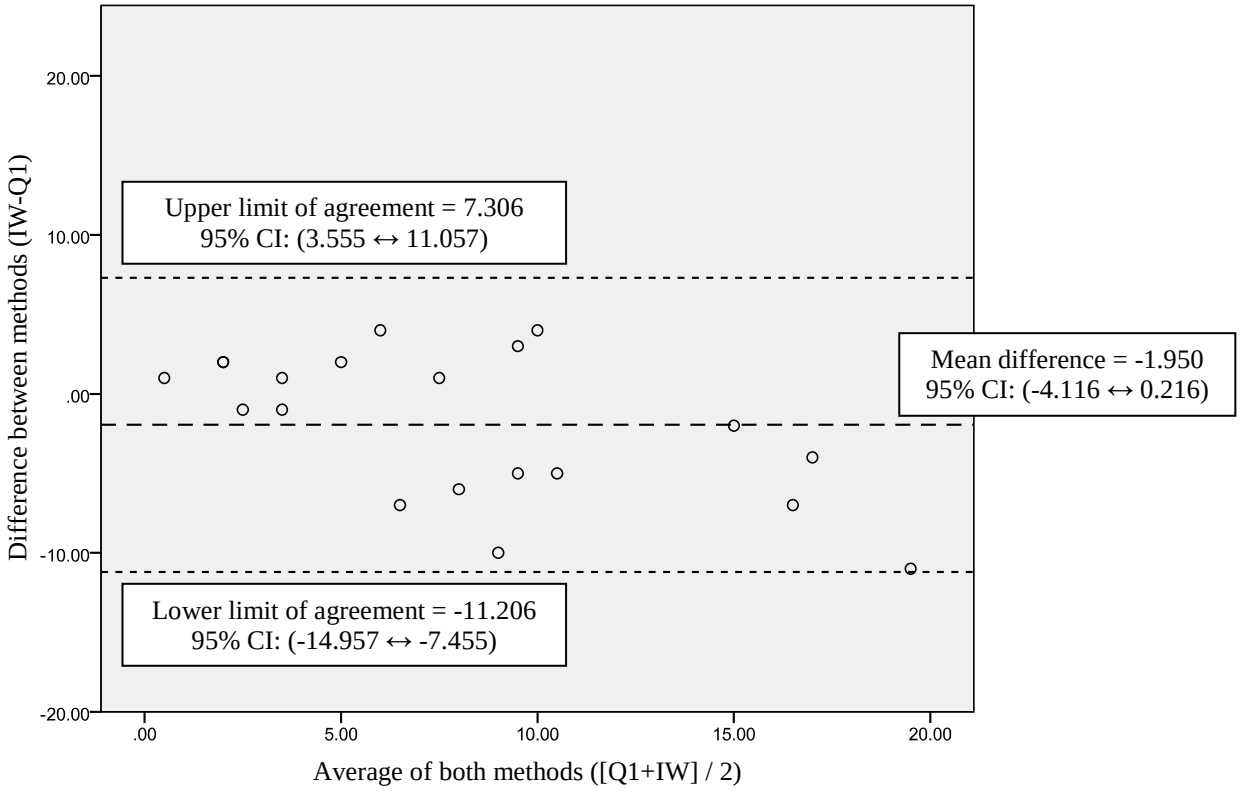


FIGURE 5.7: Bland-Altman plot for the crude DT scores in the pilot study in Gulu (Q1 vs. IW)

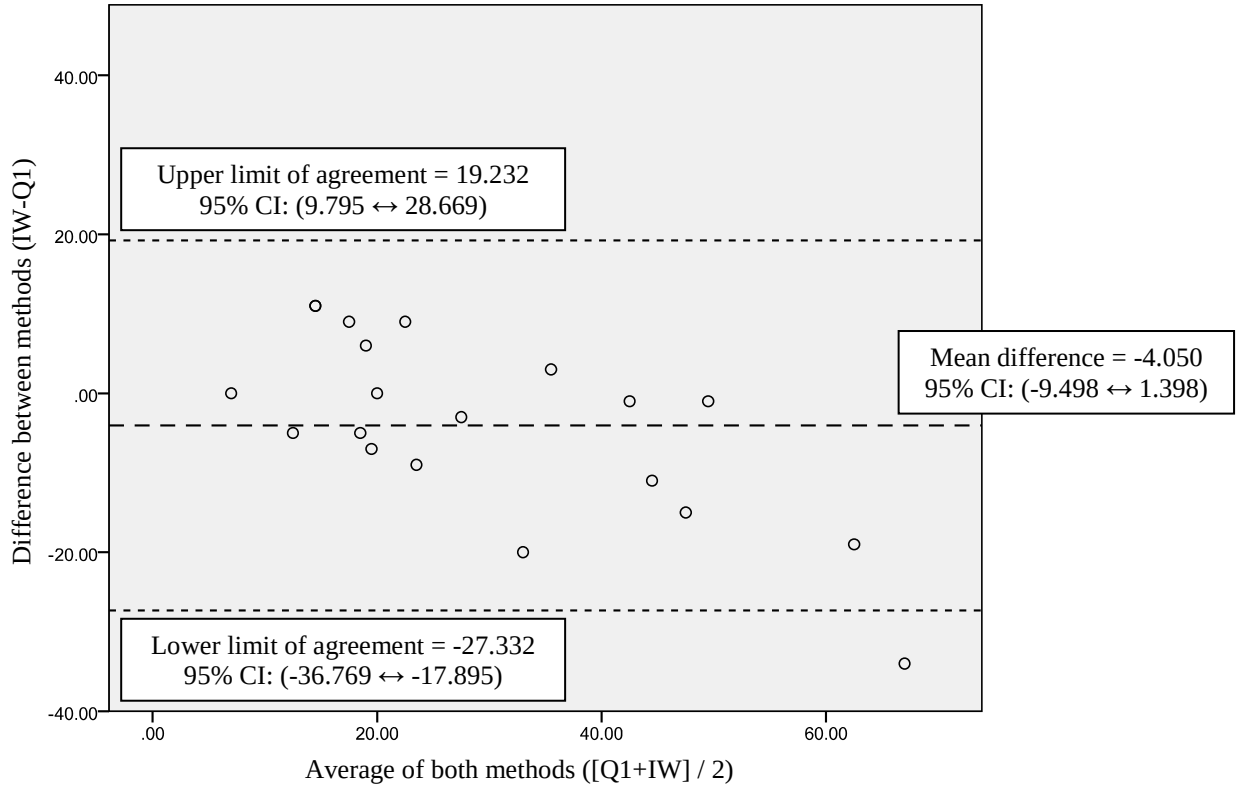
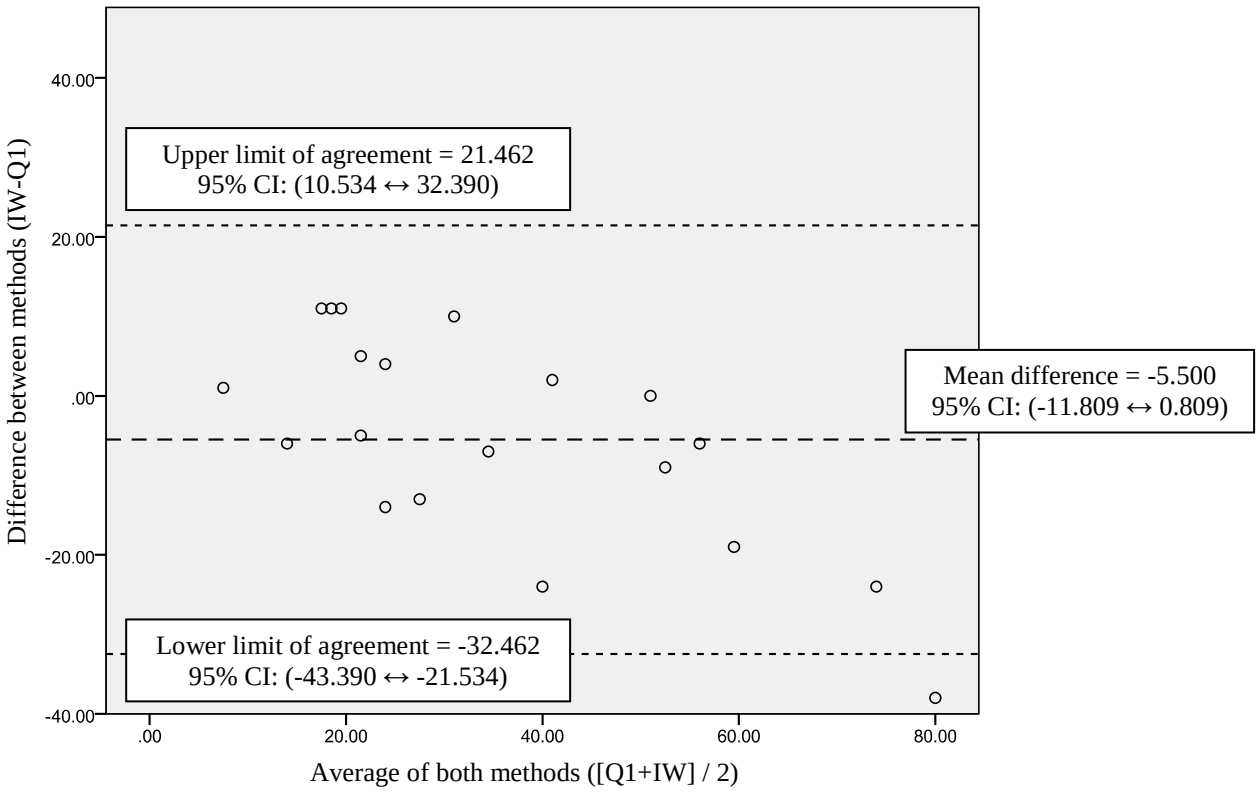


FIGURE 5.8: Bland-Altman plot for the crude AT scores in the pilot study in Gulu (Q1 vs. IW)



Assessment of the normality of the difference between methods for the crude data returned mixed results. Statistical testing for normality suggested a Gaussian distribution, but visual examination of the histograms was less convincing (APPENDIX A10).

Applying a Ln transformation to the data had a variable effect on the normality of the difference between methods. This was evident in both the statistical analyses and the histograms (APPENDIX A10). The limited sample size may have contributed to this ambiguity. However, the marginal changes that occurred as a result of the Ln transformation warranted further investigation as described by Bland et al (1986).

The assessment of the limits of agreement and their precision for the Ln transformed data returned similar results to the crude analyses (TABLE 5.8).

TABLE 5.8: Inter-method reliability (Q1 vs. IW) testing using the "limits of agreement" for the Ln transformed data in the pilot study in Gulu

Outcome variable	n	95% of questionnaire results will fall between: L x interview score ↔ U x interview score (precision of limits: 95% CI)	
TT (Ln score)	20	L = 0.42 (0.14-0.59)	U = 2.21 (1.58-3.09)
K (Ln score)	20	L = 0.32 (0.21-0.48)	U = 2.50 (1.65-3.80)
P (Ln score)	20	L = 0.30 (0.18-0.50)	U = 3.72 (2.23-6.18)
ML (Ln score)	19	L = 0.29 (0.16-0.50)	U = 4.28 (2.43-7.53)
DT (Ln score)	20	L = 0.42 (0.30-0.59)	U = 2.21 (1.58-3.09)
AT (Ln score)	20	L = 0.42 (0.31-0.59)	U = 2.09 (1.51-2.88)

NOTE: L and U indicate the range that 95% of questionnaire results will fall between as a percentage of the interview data. For example, 95% of the TT questionnaire results will be between 0.42 - 2.21 x the TT interview results. This means that the questionnaire may differ from the interview by 58% below to 121% above. Considering the range for the precision of the limits of agreement, in the best-case scenario, 95% of the questionnaire results for TT will be between 0.59 - 1.58 x the interview result (i.e. between 41% below and 58% above) [172].

Therefore, poor inter-method agreement was indicated regardless of whether the difference between the methods was normally distributed. Even in the best-case scenario, the range of scores when comparing Q1 vs. IW was too large to consider the methods clinically interchangeable [172].

v) Internal consistency

Cronbach's α for TT (α = 0.889), K (α = 0.845), P (α = 0.912), ML (α = 0.850), DT (α = 0.949) and AT (α = 0.956) demonstrated internal consistency levels appropriate for this research (TABLE 5.9) [176,177].

TABLE 5.9: Internal consistency of the APAI questionnaire scores at Q1 in the pilot study in Gulu

Outcome variable	n	Cronbach's α	Internal consistency
TT	20	0.889	moderate-high
K	20	0.845	moderate-high
P	20	0.912	high
ML	20	0.850	moderate-high
DT	20	0.949	high
AT	20	0.956	high

NOTE: Internal consistency – moderate-high ($0.8 \leq \alpha < 0.9$); high ($\alpha \geq 0.9$) [176,177].

5.4d) Physical activity

i) Feasibility assessment

PA was not assessed with locally sourced equipment and facilities. Although it would have been possible to deliver a self-report PA questionnaire standardised for global utility, no standardised instrument had previously been translated into Luo and validated in the local context. It is not known how the age of the subjects, translation process and cultural norms

may have affected the quality of the results [178]. Therefore, accelerometers were used to provide a more objective assessment of PA in the Gulu context.

The accelerometers generated a lot of interest in the local community and the measurement protocol was adapted to optimise compliance and minimise risk. The attention drawn to the accelerometers was reduced by covering them in strapping tape and hiding them underneath the clothing of the subjects. This was made easier by only including boys in this component of the study. It was important to emphasise that the accelerometers were safe to wear after traditional doctors raised significant concern in the community about the dangers of the "magic" box. Wear time also improved when the subjects regularly interacted with the assessors and were encouraged to take personal responsibility to increase compliance (APPENDIX A7).

ii) Measurement compliance

Of the six accelerometers that were distributed, five of the subjects satisfied the wear time criteria for the week (TABLE 5.10). The subject who failed to meet these requirements registered activity counts on five days (four week days and one weekend day) during the week, but had not reached the minimum of 10 hours of wear time on any of these days.

TABLE 5.10: Accelerometer compliance according to valid days of wear in the pilot study in Gulu

Subject number	Valid days	Valid week days	Valid weekend days	Valid measure
1	5	4	1	Yes
2	6	4	2	Yes
3	7	5	2	Yes
4	0	0	0	No
5	6	4	2	Yes
6	5	4	1	Yes

5.5) Preliminary discussion

5.5a) Principal findings

I developed a method to assess body composition, muscle strength and cardiorespiratory fitness that was pragmatic and ethical in the Gulu context. However, this required field testing of previously validated metrics and the subtle adaptation of these measurement methods according to locally available resources and capacity. The modified protocols developed in this study for height, weight, SBJ and MSFT demonstrated good inter- and intra-tester reliability. Most participants did not know their DOB, but could report their age in years consistently.

I was also able to assess MH pragmatically and ethically in a P6 classroom in Gulu using a guided questionnaire. This was adapted from an interview assessment tool previously validated in the target population. The measurement protocol developed in this study for TT, K, P, ML, DT and AT demonstrated satisfactory intra-tester reliability. Although there was a clear association between the interview and questionnaire results, the inter-method agreement appeared to be poor. The internal consistency of the APAI in questionnaire form was good.

Satisfactory levels of compliance were noted for the use of accelerometers by young adolescents in Gulu. This improved if the measurement protocol was delivered using methods that were sensitive to the local context and the subjects were monitored regularly by the assessors.

5.5b) Limitations and strengths

There were several challenges to conducting research in Gulu that influenced the methods and results of this study. Although many of these difficulties introduced potential bias and

compromised the validity of the results, the successful adaptation of the research methods to the local context was perhaps the greatest strength of this study.

The validity and reliability of the original measurement methods for the selected PF indicators had been established prior to commencing this research [44,47,51,162-166]. Although the modification to these metrics for application in Gulu was minimal, it may have affected the measurement quality. The inability to validate these by comparing them to gold standards was a significant limitation of the current study [179].

Similar difficulties were faced for the MH assessment. Betancourt et al (2009) used a novel approach to validate the APAI delivered in interview format in Gulu [128]. The poor inter-method agreement in this study may indicate that delivering the APAI as a guided questionnaire significantly compromised the validity of the tool. However, it is also possible that response bias and measurement error were reduced by using a questionnaire that enabled the subjects to answer independently of any interviewer influence [2]. In a parallel study on a larger sample, strong associations were found between the APAI questionnaire scores and post-traumatic stress symptoms measured using the Impact of Events Scale [153]. Both studies suggest that internal consistency for the APAI questionnaire is better than interview [128].

In the current study, the inter-method reliability assessment for the APAI was also an inter-tester analysis because different assessors conducted the interviews and guided questionnaires. Although Betancourt et al (2009) described good inter-tester reliability for the APAI in interview format, the statistical analyses only assessed association [128]. I did not analyse the inter-tester reliability independently from method because one assessor guided all of the APAI questionnaires simultaneously. Therefore, it is not possible to make any definitive statements about inter-method or inter-tester reliability for the APAI.

There was also no assessment of the sensitivity of the adapted measurement protocols. Therefore, the high levels of intra- and inter-tester reliability in this study may have been an artefact of relatively insensitive measurement methods [179]. This is of particular concern in the modified measurements for BFA and HFA because the resolution of the underlying metrics (i.e. height, weight, age) was reduced when compared to the established protocols.

The quality of the results for all of the outcomes may have also been compromised by the flexibility in the testing order at P1 [162,164]. Completing the PF tests first may have impacted on the MH assessment and/or vice versa [29,30,180]. However, variation in the testing order between P1 and P3 did not prevent satisfactory intra-tester reliability for all of the outcomes.

All of these limitations resulted from the compromises that were necessary to ensure the methods implemented were pragmatically and ethically feasible in the Gulu context. The challenges for selecting and delivering the metrics in this study were described in the results (SECTION 5.4). To summarise, the measurement methods used in this study aimed to:

- Apply protocols that had been appropriately validated previously and required minimal adaptation to the local context and environment;
- Utilise equipment that was transportable by motorcycle and could be rapidly assembled / disassembled;
- Enable efficient measurement of large cohorts within variable school environments;
- Build local capacity using resources and equipment that matched the skills and facilities available in the community;
- Avoid compromising the security of the subjects, assessors and equipment;
- Ensure all actions were culturally appropriate according to local law, customs, language, behavioural norms and spiritual beliefs.

Research that assesses similar outcomes in resource-poor settings in LMICs is sparse [55,168-170,181,182]. Most of the published studies either do not report the details of the metrics utilised and/or fail to adequately assess their quality [130,183-189]. They have also typically applied a "fly in, fly out" approach to measurement [160,161]. Therefore, consideration of the local context and sustainability when assessing the reliability of the PF and MH indicators was a notable strength of this study.

However, this approach was not appropriate for the assessment of PA of young adolescents in Gulu. The intention to use locally available and sustainable resources was violated when the researcher "carried" in the accelerometers. Although the subjects appeared to be adequately compliant, the selection of a boarding school setting for this study may have affected the generalisability of these results to other schools in Gulu.

The boarding school setting may have also caused systematic bias in the results observed for the PF and MH outcomes [179,190]. Laroo Boarding School had a disproportionately high population of "war-affected youth" who had been targeted by several previous researchers. Teachers and pupils living on site facilitated the 100% class attendance rate observed in this study. It was also the only government school in the region that had staff trained in psychosocial counselling and established after-school programmes. Although the boarding school setting enabled the testing procedures to be shifted to after school, this did affect the recruitment of subjects. Students opted in to PF and MH assessment after school rather than opted out of testing embedded in PES classes that theoretically they all attended. The influence of these factors on the distribution of the crude results is unknown for each of the outcome variables.

Concerns about the distribution of the data were evident when assessing the sample normality (APPENDICES A8 & A9) [179]. The limited sample size, mixed gender results and subject recruitment procedures may have also contributed to the distribution of the APAI results. However, in the absence of sufficient data to conduct a sample size calculation, these parameters and procedures were informed by measurement feasibility and the precedents set by Betancourt et al (2009) [128].

The normality of the MSFT data for the girls sample was affected by the high number of subjects who dropped out before completing level 2 of the test. There was a 1.0km/hr increment between levels 1 – 2 of the MSFT (i.e. double the 0.5km/hr change between all other levels). Therefore, the normality improved when the results were expressed according to ***completed level*** rather than the conventional method of reporting ***speed at the highest completed level.*** This has not typically been a problem in previous MSFT studies because fewer participants scored so poorly [168-170].

Several other broad methodological challenges were faced that required adaptation to cultural and behavioural differences. Consideration of local capacity in planning and time management was a key factor in designing a study that minimised loss to follow-up. This also depended on effective subject tracking, which was achieved despite the existence of several common names in the sample. There was also a clear disconnect in the local concept of accountability that had to be overcome to obtain ethical approval. Maintaining high levels of scientific rigour while negotiating these difficulties highlighted the value of enlisting local assistance and establishing strong relationships with the community. The promotion of local capacity and ownership for subsequent research was an important success of this study [114].

5.5c) Mechanisms and implications

This study demonstrated that it was feasible to select and adapt assessment protocols for the outcomes of interest according to the pragmatic and ethical challenges faced in Gulu. However, not all measurement methods could be modified appropriately for delivery in this setting. Assessing the quality of the metrics that were adapted and/or "carried" in by the researcher provided an indication of their utility in subsequent studies.

The questionable validity of the modified PF and MH metrics limits comparisons to other populations and their use for diagnostic purposes [179,190]. The interview and questionnaire methods of delivering the APAI are also not interchangeable. Therefore, when applying the methods developed in this study, the results should be interpreted with caution when comparing them to established normative values or other studies that used the original measurement protocols.

The intra- and inter-tester reliability for the PF outcomes indicates that they can be used to assess change within a longitudinal study. Their utility in this context is limited by the sensitivity of the measures [179]. Satisfactory intra-tester reliability for the APAI enables a similar application provided the questionnaire is guided by the same person on each occasion. There should be one facilitator for all MH assessments in the subsequent studies because the inter-tester reliability of the APAI is not known. This also reduces the risk of systematic error when several assessors deliver the APAI.

Applying the modified metrics in multiple non-boarding schools for the subsequent studies may also require further adaptations. Embedding the assessment of PF and MH within school PES classes will facilitate an opt-out approach, which potentially captures a more representative sample. This should optimise attendance and minimise the potential loss to

follow-up of subjects that do not live on the school premises. These factors ought to be considered when calculating the sample size for future research using the results of the current study. It is also recommended that a consistent testing order and location within each school are used to reduce the effect of any systematic error between sites [162,164]. These adjustments to the assessment protocol require locally facilitated communication with target schools, relevant officials and the wider community prior to data collection. Garnering local support for the research is also critical for the security and compliance of students more regularly exposed to the broader community whilst wearing the accelerometers.

Several concerns also persist regarding the data distribution and analysis. The normality of the APAI data in future studies is likely to be improved by recruiting a larger sample that is not selected on a "first in, first served" basis. Stratifying these data according to gender may also improve the distribution and would enable separate analyses for boys and girls [190].

Future studies assessing the MSFT in the target population should report the results according to ***level complete*** to improve the sample distribution. However, comparisons to normative data and other studies will require these results to be transformed to ***speed at the highest completed level*** prior to analysis [55,168,170]. The disproportionate representation of girls at level one in this pilot sample suggested that the modified MSFT was insensitive to differences in performance at this level. The cause of the relatively high prevalence of low scores may be due to poor cardiorespiratory fitness and/or apathy towards the test [55,168-170]. It is important that these factors are considered when interpreting the results if similar patterns are observed in the subsequent community needs analysis and impact evaluation of a SfD programme.

5.6) Summary box

PRETESTING METHODS - FEASIBILITY AND RELIABILITY

- I developed pragmatic and ethical methods to assess the PF (height, weight, SBJ, MSFT) and MH (APAI) of young adolescents in Gulu using measurement protocols adapted to the local context. There was also good compliance with PA measurement using accelerometers.

- All of the adapted measurement methods for the PF parameters had good intra- and inter-tester reliability.

- Poor recall of DOB may affect future studies in Gulu that assess BFA and HFA of young adolescents.

- The APAI questionnaire had good intra-tester reliability and internal consistency, but the results did not appear to be interchangeable with data collected by interview.

- The results of this study can inform sample size calculations that ensure future research assessing PF and/or MH in Gulu is adequately powered.

- The adapted PF and MH tests should be implemented in a consistent order for the subsequent community needs analysis and evaluation of a SfD intervention in Gulu. An opt-out approach embedded in the PES classes of various schools may improve subject recruitment and minimise selection bias. The results should be stratified according to gender and interpreted with caution, particularly in cross-sectional analysis that compares the results to other studies.

- Enlisting local support in Gulu is critical for the effective application of the modified measurement methods that were developed in this study.

PART IV

COMMUNITY NEEDS ANALYSIS

CHAPTER 6

Baseline Measurements for Physical Fitness and Mental Health of Young Adolescents in Gulu, Uganda

6.1) Purpose of chapter

The purpose of this chapter is to describe the current PF and MH status of young adolescents living in Gulu. I will use measurement protocols developed in Chapter 5 to conduct a cross-sectional assessment of a sample within the municipality. The results are contrasted with global normative values to indicate comparative health status. The outcomes will also be used to identify if a history of abduction during the war is associated with current MH and PF profiles. The baseline measures of this case-study also provide an insight into the relationship between PF and MH in this context. The results describe some of the relevant needs in Gulu and inform a subsequent SfD intervention.

6.2) Introduction

Africa is often depicted by development agencies as a uniformly desperate community in need of continuous humanitarian support. Images of malnourished children suffering from communicable diseases saturate the portrayal of the health challenges that are faced throughout the continent. Although Africa continues to bear a heavy burden from acute humanitarian crises, the health problems that emerge during the epidemiological transition are of increasing concern [4-6].

It is evident that the widespread perception of humanitarian suffering and the accompanying epidemiological profile is not the reality throughout Africa. Although large parts of the continent remain geopolitically and environmentally fragile making them sensitive to natural and man-made perturbations, other regions are rapidly developing with relative stability. Often this diversity is seen within the same community [4-6]. This highlights the importance of consulting key stakeholders in local communities to identify health concerns and priorities [2].

For an extended period the health priorities in Gulu have been influenced by international donors focussed on communicable disease management. Gulu is the largest urban centre in northern Uganda and the *transition* from the humanitarian to the development phase in the post-conflict period has been rapid [111,112]. It has been fuelled by a large NGO community with the assistance of UN agencies and the rhetorical support of the public sector. The relative stability in Gulu since the end of the war in 2006 has brought significant reductions in child and adult mortality [149,150,191,192]. However, little is known about other health-related outcomes in the region and there is increasing international concern about the risk of NCDs in post-conflict settings [142]. Although there has been a growing interest in the emergence of these disease processes within the community, there has been no coordinated action [193].

The risk factors for "lifestyle" diseases in Gulu have been particularly neglected in young people. An anecdotal image of non-obese youths who walk large distances everyday and are consequently physically fit has endured. However, there are no data to substantiate this claim in comparison to global norms. Testing these assumptions is a critical step towards establishing a baseline understanding of NCD risk factors in the target population [194]. Identifying the PF and MH needs in Gulu can also inform the implementation of future interventions relevant to these health outcomes.

Additional consideration should also be afforded to at-risk groups in Gulu. Adolescents who were abducted during the war in northern Uganda are regularly described as an isolated and marginalised population. They often receive particular attention from NGOs active in the region. However, the process of targeting these adolescents for intervention may actually accentuate stigmatisation by the wider community rather than promote reintegration [147,148]. Little is known about how previous experience of being "in the bush" is associated with current health status. This is particularly pertinent for experiences affecting MH outcomes that remain a taboo topic within the Acholi culture [147,148]. Confidential identification of adolescents abducted during the war provides an opportunity for cross-sectional comparison to their non-abducted peers. Describing current PF and MH discrepancies associated with a history of abduction should indicate if there is a specific need for targeted SfD programmes.

However, the potential impact of a SfD intervention on the PF and MH of young adolescents in Gulu is unknown. As previously discussed, PF has a reciprocating relationship with PA that may explain its association with MH outcomes (SECTION 2.4). Review level evidence suggests that physically active adolescents are fitter and leaner as well as less depressed and anxious [17,21,22,25,26,29]. However, these assertions are based almost exclusively on research conducted in environments and populations very different to northern Uganda.

Although the mechanism linking PA to improved PF is well established, it is more tenuous when considering MH outcomes [29,30,180]. I believe that the assumption that MH is associated with PA and PF in the same way in Gulu as it is in developed and peaceful communities requires further scrutiny. Describing this relationship in the target population may provide an initial indication of whether increasing PA and PF will potentially benefit MH in this setting.

In this chapter, I will conduct part of a community needs analysis for young adolescents in Gulu. Typically a broader approach is used that includes participatory planning through focus group discussions and workshops with key local stakeholders [2]. Qualitative approaches are also commonly applied to gain an insight into the capacity for change in population health determinants and define intervention target groups [2]. However, the health outcomes that I assess were previously established by the intervention funders and implementers with minimal consultation of the local community. Time and resource limitations also shape my approach by restricting engagement of programme personnel and beneficiaries in the problem definition phase of my evaluation. Consequently, I use quantitative methods to establish the current PF and MH of adolescents in Gulu whose lives and experiences vary considerably.

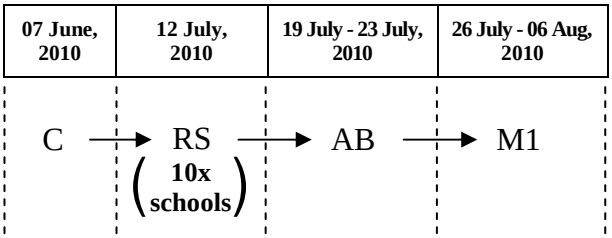
I endeavour to account for several potential confounding factors. For example, the exposure to the conflict is influenced by age of the subjects [147]. The education quality, facilities and child support systems are vastly different among schools. There are also clear socio-economic and infrastructure divides among the divisions of Gulu municipality. Therefore, I adjust for the potential discrepancies in PF and MH of the target population caused by these factors when assessing the local needs and establishing baseline levels for a subsequent SfD intervention evaluation (CHAPTERS 7 & 8).

6.3) Methods

6.3a) Study design

The current study had a cross-sectional design. The target population (C) was young adolescents who were eligible to participate in the Gum Marom Kids League (GMKL). This was a SfD intervention that was delivered within Gulu municipality for adolescents aged between 11-14 years. 10 primary schools in Gulu municipality were randomly selected to participate in the study (RS). The P6 class teachers, PES instructors and pupils at each of the schools were briefed about the assessment procedures in the week prior to commencing testing (AB). The measurements were then collected over a two-week period (M1). Subjects recruited for the study undertook measurement for PF and MH (FIGURE 6.1).

FIGURE 6.1: Diagram representing subject recruitment and study design



LEGEND:

Subject recruitment

C	Selection of study population (11-14 years old attending school within Gulu municipality)
RS	Randomly select sample of schools within Gulu municipality
AB	Assessment briefing for teachers and pupils at selected schools

Study design and data collection

M1	Initial case-study measurement (PF, MH)

The data collected at M1 were compared to global norms and used to assess if there were any relationships between the dependent variables. Stratifying for history of abduction and comparing the two groups made it possible to identify if there were any persistent differences.

6.3b) Sample size calculation

In the absence of more appropriate data for adolescents in Gulu, the sample size calculations were based on the MSFT results collected by the RTL at P1 during pilot testing (SECTION 5.4b). The MSFT scores were used because it was the primary outcome variable and aerobic fitness is less sensitive to change in adolescents than measures of muscle strength [28]. Therefore, a larger sample size was required to detect a relevant difference in the MSFT and it was a more appropriate reference point [179,195]. BFA was not used because it was not the primary outcome variable and anecdotally there appeared to be minimal variation in Gulu with low levels of obesity. There were also no previous data that could be used to indicate the relevant difference or effect size for the APAI scores.

When comparing the means of two groups, the sample size calculation used a standard formula based on the relevant difference (δ), standard deviation (σ) of the variable and the acceptable probability of a type I (α) or type II error (β) [179,195,196]:

$$n = \frac{2 \times [z_{(1-\alpha/2)} + z_{(1-\beta)}]^2}{[\delta / \sigma]^2}$$

z represented a standard normal distribution.

α represented the probability of a type I error and was taken as 0.05.

β represented the probability of a type II error and was taken as 0.15.

σ represented the within-group standard deviation and was taken as 1.215 for boys and 1.026 for girls based on the results given by the RTL at P1.

δ represented the expected effect size of the intervention and was taken as 5% of the mean MSFT score given by the RTL at P1 for each gender (i.e. boys = 10.929; girls = 9.534) [28].

Therefore, for the boys sample:

$$n = \frac{2 \times [1.96 + 1.0364]^2}{[0.05 \times 10.929 / 1.215]^2}$$

n = 88.77 (89) subjects in each group

Therefore, for the girls sample:

$$n = \frac{2 \times [1.96 + 1.0364]^2}{[0.05 \times 9.534 / 1.026]^2}$$

n = 83.18 (84) subjects in each group

These calculations were used to decide how many schools were assessed at M1 to ensure a sufficient number of subjects were measured for all of the relevant analyses in Chapters 6, 7 and 8. Therefore, it was important to take into account average school attendance. The Gulu Municipality Education Officer indicated that the average daily school absenteeism for those enrolled in P6 was approximately 10%. The RTL estimated that a further 5% of the attending students would either opt out of testing or be ineligible for the GMKL. The sample calculations for the number of subjects in each group were adjusted proportionally (boys: n = 103.82 (104), girls: n = 97.29 (98)) [195].

Several of the statistical analyses in Chapters 6, 7 and 8 used this sample size calculation. However, it assumed that the groups being compared had an equal number of subjects. A standard formula using the n and the anticipated ratio (k) between the two uneven groups was used to calculate the adjusted total number of participants (ń) when necessary [195]:

$$ń = \frac{2n [1 + k]^2}{4k}$$

The sample size for each group was then assessed as:

Group 1: $n = ń / [1 + k]$ Group 2: $n = k \times ń / [1 + k]$

Although all of the statistical analyses were considered prior to collecting data at M1, the details for each are presented separately in the relevant chapters.

In the current study, the comparison to global norms involved contrasting a single sample to a known population value. Therefore, the required sample size was half of that calculated above because only one mean was estimated (TABLE 6.1) [196].

The comparison between abducted vs. NOT abducted was adjusted for an expected uneven sample size. Annan et al (2006 & 2008) estimated that $^1/_3$ of all boys and $^1/_6$ of all girls in Gulu were abducted during the war [147,148]. These ratios were used to adjust the required sample sizes for this comparison (TABLE 6.1) [195].

Assessing the relationships between the dependent variables required the application of an alternative formula based on the relevant difference of a correlation coefficient (r) and the acceptable α and β values [197]:

$$n = \left[\frac{[z_{(1-\alpha/2)} + z_{(1-\beta)}]}{| ź_{r0} - ź_{r1} |} \right]^2 + 3$$

z represented a standard normal distribution.

α represented the probability of a type I error and was taken as 0.05.

β represented the probability of a type II error and was taken as 0.15.

r0 represented the correlation coefficient under the null hypothesis and was taken as 0.

r1 represented the correlation coefficient under the alternative hypothesis and was taken as 0.4.

ź represented the Fisher z transformation to create z score for a correlation coefficient and can be sourced from a standardised reference table for the r value of interest [197,198].

Therefore, for assessing the relationship between the dependent variables:

$$n = \left[\frac{[1.96 + 1.0364]}{|0 - 0.4236|}\right]^2 + 3$$

n = 53.04 (54) subjects with measures for both variables

This calculation was used to describe the minimum sample size required for assessing the relationship between the dependent variables (TABLE 6.1).

TABLE 6.1: Calculated minimum sample sizes for the statistical analyses in the community needs assessment in Gulu - stratified by gender

Statistical analysis	Gender	Required sample size (n)		
		Group 1	Group 2	Total
Comparison to global normative values	Boys	Gulu sample: 51.91 (52)	NA	**52**
	Girls	Gulu sample: 48.65 (49)	NA	**49**
Comparison of abducted vs. NOT abducted	Boys	Abducted: 69.21 (70)	NOT abd.: 207.64 (208)	**278**
	Girls	Abducted: 56.75 (57)	NOT abd.: 340.52 (341)	**398**
Relationship between dependent variables	Boys	Data for both variables being compared: 53.04 (54)		**54**
	Girls	Data for both variables being compared: 53.04 (54)		**54**

The results of these calculations were considered along with the number of pupils enrolled in P6 when deciding how many schools were assessed at M1 (SECTION 6.3c).

6.3c) Subject recruitment

The existence of UPE in Uganda made primary schools the most reliable place to access and track adolescents aged 11-14 years. The District Education Officer provided a comprehensive list of primary schools in Gulu municipality. A total of 10 schools were selected based on the number of P6 students enrolled at each and the required sample size for the relevant analyses (SECTIONS 6.3b, 7.3b & 8.3b). These schools were selected from the original list using random number allocation in Microsoft Excel. The head teachers at each of these schools were invited to volunteer their P6 classes for measurement at M1. All of the selected schools accepted the invitation and scheduled a testing time within school hours. The 10 target schools selected for M1 measurement were:

- Christ Church Primary School (CC)
- Gulu Police Primary School (GPol)
- Gulu Prison Primary School (GPrs)
- Gulu Public Primary School (GPub)
- Gulu Town Primary School (GTwn)
- Holy Rosary Primary School (HR)
- Labour Line Primary School (LL)
- Layibi Central Primary School (LC)
- Pece P7 Primary School (PP)
- St Peter's Primary School (SP)

These were all government schools within Gulu municipality (APPENDIX A2). There were a total of 873 boys and 1058 girls enrolled in P6 at the 10 target schools. At M1, all members of the P6 classes in the selected schools were invited to participate. Opt-out consent was obtained from the subjects prior to commencing the assessment of the PF and MH outcomes.

6.3d) Subject identification procedures

The procedures for subject identification comprised collecting the same information at M1 as described for pilot testing in Chapter 5 (SECTION 5.3d).

6.3e) Outcome variables and metrics

The outcome variables and metrics comprised the same tools and techniques as described for PF and MH in the pilot testing in Chapter 5 (SECTION 5.3e).

6.3f) Data collection procedures

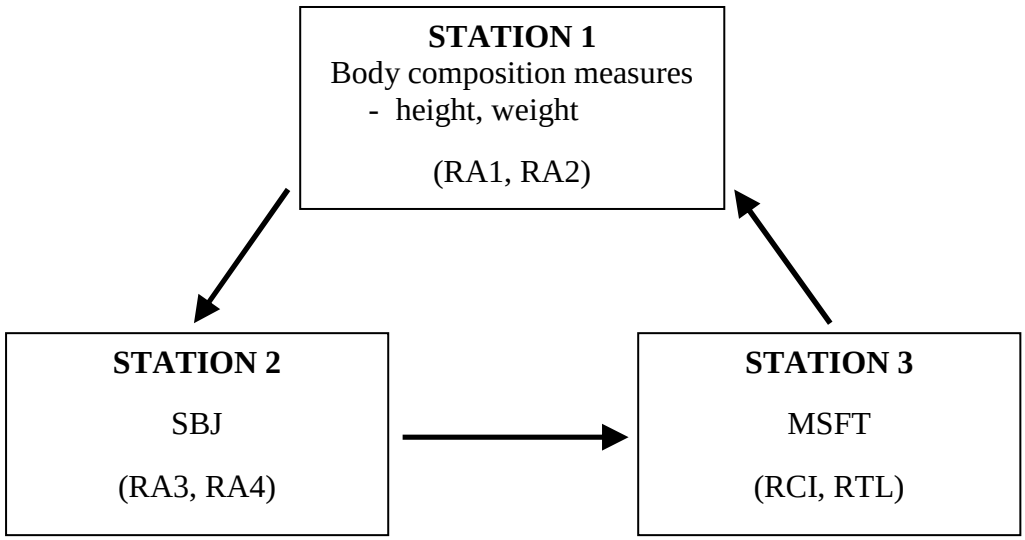
The quantitative data collection followed the same procedures as described for pilot testing in Chapter 5 (SECTION 5.3f).

However, the testing procedure was embedded into the PES timetables for P6 at each of the schools. It was directed by the RCI who worked with the RTL and four additional research assistants (RA1, RA2, RA3, RA4). P6 and PES teachers from each school contributed to maintaining student discipline throughout the measurement procedures.

Measurements were taken from one P6 class at a time. Whenever practically feasible the APAI assessment took place within the classroom before completing PF testing outside in the school grounds.

The pupils from each class were divided into three groups and assigned to an initial testing station for PF measurement (FIGURE 6.2). Each group performed the test to which they were initially assigned and then rotated to the next station. This was repeated once more so that every group had completed all three measurements. Each PF test was assessed by the same members of the research team to reduce risk of inter-rater variability [164].

FIGURE 6.2: Diagram of the stations for collecting the PF measurements in Gulu



On several occasions testing was not completed due to logistical and/or environmental interruptions (e.g. weather, variable class timetable, demands of school staff, local time management). When this occurred, the assessment sheets were collected by the assessors and test completion was rescheduled at a convenient time within the two-week measurement window.

6.3g) Data analysis and presentation

All of the quantitative data were cleaned and checked for outliers prior to commencing the analysis. The underlying assumptions for each statistical test were also verified before proceeding. Unless otherwise stated, the reporting of descriptive statistics for each of the variables included the crude mean, 95% CI and sample size (n).

i) Sample distribution

The sample was described in terms of gender, age, division of residence, school attended and history of abduction during the war. Disparities in the distribution of these independent variables within the sample were identified and presented in a series of tables and graphs.

The sample was also stratified according to these independent variables and descriptive statistics were calculated for each level. A one-way ANOVA test for the crude mean of each outcome variable identified if there was a significant inter-level difference for age, division, school or history of abduction at M1. The results were presented graphically and summarised in a table.

ii) Comparison to normative data

The data were stratified according to gender and descriptive statistics were calculated for all of the PF outcome variables. The descriptives for BFA and HFA were presented as proportions in each of the body composition categories defined by WHO. The SBJ and MSFT were also stratified by age and described using the sample mean. However, to enable a direct comparison to normative values the results for the MSFT were first converted to speed at the highest completed level (km/hr). Summaries of the descriptive statistics were tabulated.

The statistical analyses for BFA and HFA comprised a comparison between the Gulu and the WHO reference populations. Both samples were presented graphically to demonstrate their proportional distribution in each of the BFA and HFA categories defined by WHO. The 95% CI for the proportions in the Gulu sample was used to identify significant deviations from the normative values. 99% CI was also calculated to indicate greater statistical certainty of variation from the global norms.

For the SBJ and MSFT, an independent t-test was used to contrast the crude mean of the Gulu case-study results to the normative values. Normative data for the SBJ were described by Tomkinson et al (2007) [170]. Olds et al (2006) established the normative data for the MSFT that were used in this study [168]. The results were presented graphically.

In the absence of appropriate normative data for the MH outcomes, no descriptive or comparative statistics were completed for these outcome variables.

iii) History of abduction

Descriptive statistics for all outcome variables stratified according to history of abduction and gender were calculated. The results were presented in a summary table.

An independent t-test was used to compare the crude means of each outcome variable at M1 for abducted vs. NOT abducted subjects. The results were presented graphically.

A univariate ANOVA was used to repeat this analysis adjusted for age, division and school. Significant findings were tabulated.

iv) Relationship between outcome variables

The data were stratified according to gender and descriptive statistics that included all eligible subjects were calculated for each of the PF and MH outcome variables. The results were presented in a summary table.

The crude results for each of the PF variables were plotted against the MH outcomes. Linear and non-linear lines of best fit were drawn on each of the scatterplots to assess visually if a relationship existed.

Pearson's r correlation coefficients were calculated to assess the strength of the linear relationship between the crude results for each of the PF variables with the MH outcomes.

Partial correlation coefficients were used to repeat this analysis adjusted for age, division, school and history of abduction. The results were tabulated.

The scatterplots, Pearson's r and partial correlation coefficients were then repeated to assess the relationship between each of the body composition outcome variables (BFA, HFA) vs. the indicators for muscular strength (SBJ) and cardiorespiratory fitness (MSFT).

6.3h) Ethical approval

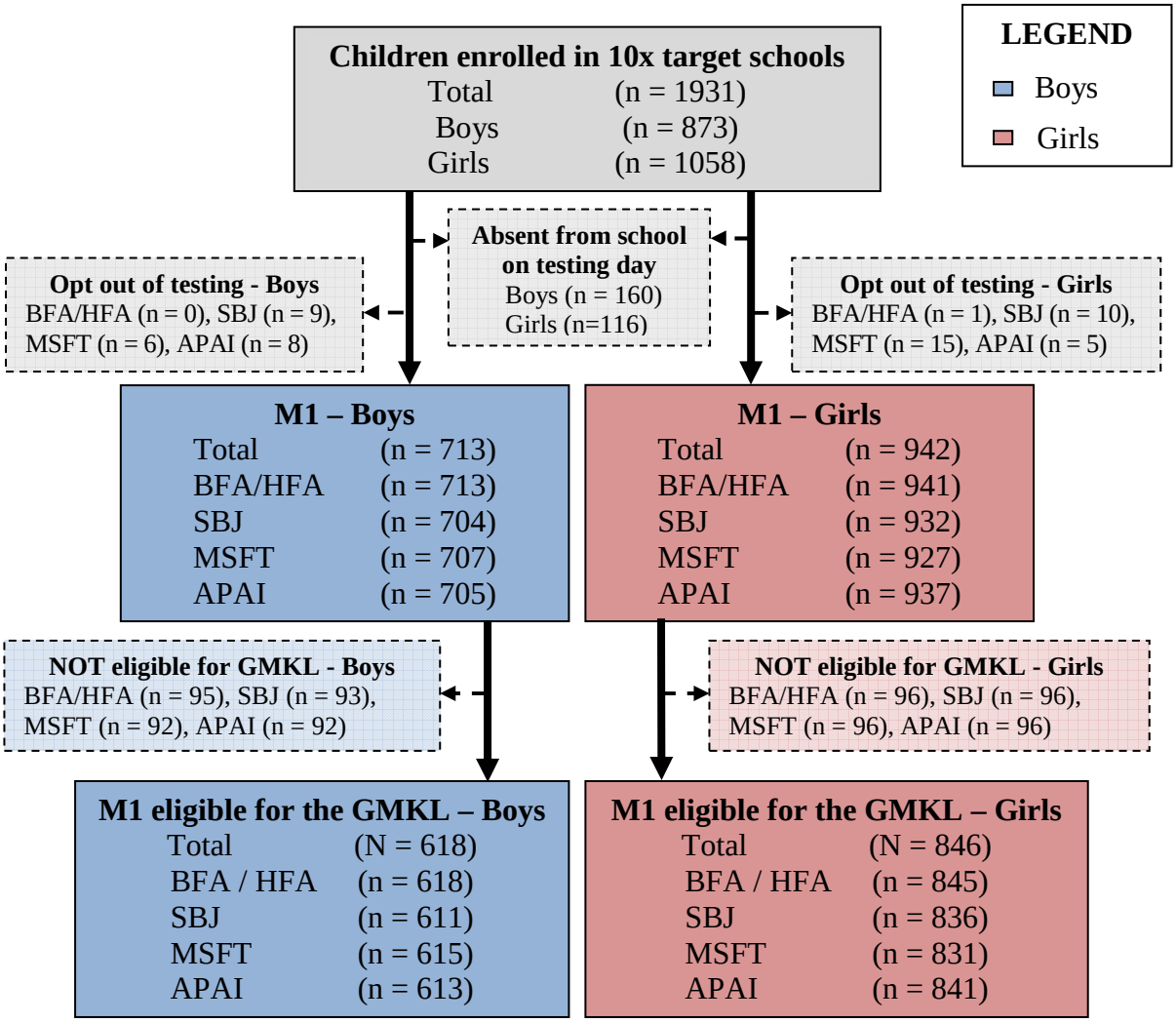
Ethical approval was obtained from the same entities as described for pilot testing in Chapter 5 (SECTION 5.3h).

6.4) Results

6.4a) Sample distribution

A total of 713 boys and 942 girls in P6 were present and measured at M1. This indicated higher school absenteeism than expected (boys ~18.3%, girls ~11.0%). 95 boys and 96 girls who were measured at M1 were ineligible for the GMKL because of their age. This was also a higher proportion than expected (boys ~13.3%, girls ~10.2%). Adolescents ineligible for the GMKL were excluded from the subsequent analysis. The total number of subjects included in the final analysis for each gender is represented by "N". A variable number of subjects opted out of testing for each of the outcomes at M1 (FIGURE 6.3).

FIGURE 6.3: Flow diagram of the sample for the community needs analysis at M1 in Gulu



The self-reported age of the subjects measured at M1 was not distributed evenly within the eligible range for both genders (TABLE 6.2). 13-year-old adolescents (n=527, 36.00%) comprised the greatest proportion of the sample and 11-year-old youths (n=136, 9.29%) were the least represented.

TABLE 6.2: Gulu sample distribution at M1 - stratified by gender and age

Age in years	Boys (B)		Girls (G)		Total (B+G)	
	n	% of all B	n	% of all G	n	% of all B+G
11	44	7.12	92	10.87	136	9.29
12	141	22.82	177	20.92	318	21.72
13	234	37.86	293	34.63	527	36.00
14	199	32.20	284	33.57	483	32.99
Total	**618**	**100.00**	**846**	**100.00**	**1464**	**100.00**

The subjects measured at M1 were not evenly distributed according to the division in which they lived, history of abduction and their mean age (TABLE 6.3).

Pece division (boys: n=283, 45.79%; girls: n=436, 51.54%) comprised the greatest proportion of the sample and Bardege division (boys: n=48, 7.77%; girls: n=57, 6.74%) was the least represented for both genders.

The proportion of adolescents reporting a history of abduction varied according to division (TABLE 6.3). The incidence of abduction for boys was highest in Pece (boys: n=49, 17.31%) and lowest in Laroo (boys: n=11, 8.94%). For girls, the highest incidence of abduction was reported in Layibi (girls: n=41, 19.62%) and the lowest was for Bardege (girls: n=6, 10.53%). There was no gender difference in the incidence of abduction in Pece (girls: n=76, 17.43%)

and Bardege (boys: n=5, 10.42%). Girls reported a higher incidence of abduction than boys in Laroo (girls: n=24, 16.67%) and Layibi (boys: n=22, 13.41%).

The mean age in the sample from Pece (boys: 12.78yrs; girls: 12.72yrs) was younger than in the other divisions. There was no overall difference in the mean age of subjects who were abducted vs. NOT abducted for either gender (boys: abducted = 13.01yrs, NOT abducted = 12.94yrs; girls: abducted = 12.93yrs, NOT abducted = 12.91yrs). However, abducted girls from Layibi (girls: 13.44yrs) and boys from Laroo (boys: 13.55yrs) tended to be older than those NOT abducted in these divisions (TABLE 6.3).

TABLE 6.3: Gulu sample distribution and mean age at M1 - stratified by gender, division and history of abduction

Gender	Division	% of N - same sex	Abducted		Mean age in years (95% CI)		
			n	% of division	Abducted	NOT abducted	All - same sex
Boys (N=618)	Bardege (n=48)	7.77	5	10.42	13.20 (12.47↔13.93)	13.02 (12.80↔13.24)	13.04 (12.83↔13.25)
	Laroo (n=123)	19.90	11	8.94	13.55 (13.14↔13.96)	13.04 (12.88↔13.20)	13.09 (12.94↔13.24)
	Layibi (n=164)	26.54	22	13.41	12.95 (12.57↔13.33)	13.14 (12.99↔13.29)	13.12 (12.98↔13.26)
	Pece (n=283)	45.79	49	17.31	12.90 (12.61↔13.19)	12.76 (12.64↔12.88)	12.78 (12.67↔12.89)
	Total (N=618)	**100.00**	**87**	**14.08**	**13.01 (12.81↔13.21)**	**12.94 (12.86↔13.02)**	**12.95 (12.88↔13.02)**
Girls (N=846)	Bardege (n=57)	6.74	6	10.53	13.00 (12.37↔13.63)	13.16 (12.94↔13.38)	13.14 (12.94↔13.34)
	Laroo (n=144)	17.02	24	16.67	12.79 (12.33↔13.25)	12.97 (12.80↔13.14)	12.94 (12.78↔13.10)
	Layibi (n=209)	24.70	41	19.62	13.44 (13.21↔13.67)	13.16 (13.03↔13.29)	13.22 (13.10↔13.34)
	Pece (n=436)	51.54	76	17.43	12.68 (12.45↔12.91)	12.73 (12.62↔12.84)	12.72 (12.62↔12.82)
	Total (N=846)	**100.00**	**147**	**17.38**	**12.93 (12.77-13.09**	**12.91 (12.84↔12.98)**	**12.91 (12.84↔12.98)**

The number of young adolescents measured was different at each of the 10 target schools (TABLES 6.4 & 6.5). Although GPol had the most subjects in both genders at M1 (boys: n=110; girls: n=154), the proportion that each school contributed to the total sample also varied according to gender. Despite this variation, the number of girls outnumbered the boys at all schools except SP.

There appeared to be differences between schools in the incidence of history of abduction (TABLES 6.4 & 6.5). Although PP had the highest number of subjects reporting a history of abduction for both genders (boys: n=18, 28.13%; girls: n=21, 25.61%), the incidence also appeared to vary according to gender.

It appeared that the mean age of previously abducted boys attending GPrs (boys: 13.73yrs) and girls attending LC (girls: 13.88yrs) was significantly older than their classmates. There were also significant differences in the mean age of the subjects according to school. The subjects from GPol had the lowest mean age (boys: 12.19yrs; girls: 12.02yrs) and LC had the highest (boys: 13.32yrs; girls: 13.47yrs). The age differences between schools also appeared to vary according to gender (TABLES 6.4 & 6.5).

TABLE 6.4: Gulu boys sample distribution and mean age at M1 - stratified by school and history of abduction

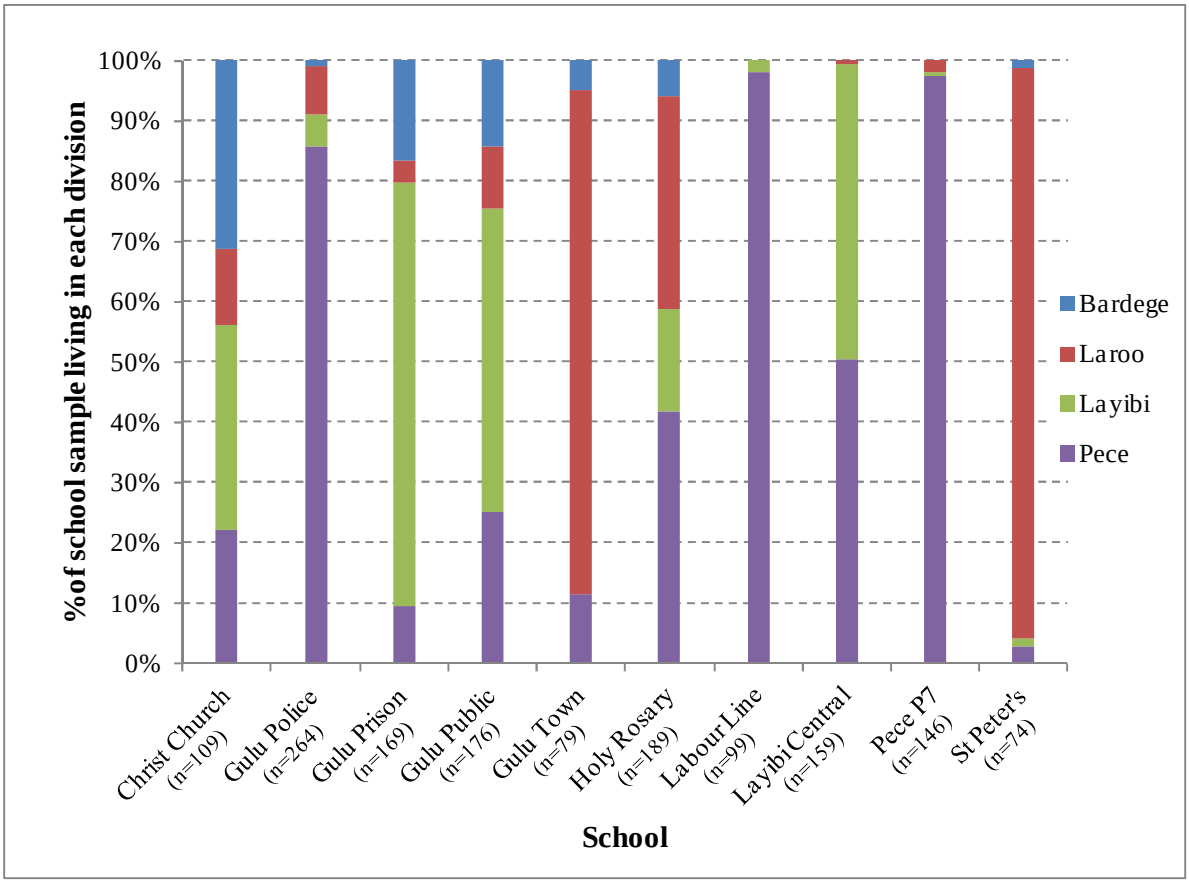
School	% of N - boys	Abducted		Mean age in years (95% CI)		
		n	% of school	Abducted	NOT abducted	All - same sex
CC (n=49)	7.93	4	8.16	13.25 (12.31↔14.19)	12.91 (12.65↔13.17)	12.94 (12.69↔13.19)
GPol (n=110)	17.80	22	20.00	12.23 (11.82↔12.64)	12.18 (11.99↔12.37)	12.19 (12.02↔12.36)
GPrs (n=71)	11.49	11	15.49	13.73 (13.35↔14.11)	13.22 (13.01↔13.43)	13.30 (13.11↔13.49)
GPub (n=72)	11.65	5	6.94	13.00 (12.38↔13.62)	13.21 (13.05↔13.37)	13.19 (13.03↔13.35)
GTwn (n=35)	5.66	3	8.57	14.00 -	13.25 (12.96↔13.54)	13.31 (13.03↔13.59)
HR (n=79)	12.78	10	12.66	13.10 (12.64↔13.56)	12.68 (12.46↔12.90)	12.73 (12.52↔12.94)
LL (n=33)	5.34	6	18.18	12.83 (12.04↔13.62)	13.11 (12.76↔13.46)	13.06 (12.74↔13.38)
LC (n=66)	10.68	7	10.61	13.29 (12.73↔13.85)	13.32 (13.13↔13.51)	13.32 (13.14↔13.50)
PP (n=64)	10.36	18	28.13	13.17 (12.74↔13.60)	13.02 (12.83↔13.21)	13.06 (12.88↔13.24)
SP (n=39)	6.31	1	2.56	14.00 -	13.24 (13.01↔13.47)	13.26 (13.04↔13.48)
Total (N=618)	**100.00**	**87**	**14.08**	**13.01 (12.81↔13.21)**	**12.94 (12.86↔13.02)**	**12.95 (12.88↔13.02)**

TABLE 6.5: Gulu girls sample distribution and mean age at M1 - stratified by school and history of abduction

School	% of N - girls	Abducted		Mean age in years (95% CI)		
		n	% of school	Abducted	NOT abducted	All - same sex
CC (n=60)	7.09	1	1.67	14.00 -	12.80 (12.57↔13.03)	12.82 (12.59↔13.05)
GPol (n=154)	18.20	38	24.68	12.03 (11.75↔12.31)	12.02 (11.86↔12.18)	12.02 (11.88↔12.16)
GPrs (n=98)	11.58	18	18.37	13.44 (13.11↔13.77)	13.38 (13.20↔13.56)	13.39 (13.23↔13.55)
GPub (n=104)	12.29	22	21.15	13.23 (12.94↔13.52)	12.95 (12.76↔13.14)	13.01 (12.85↔13.17)
GTwn (n=44)	5.20	5	11.36	13.20 (12.47↔13.93)	12.97 (12.69↔13.25)	13.00 (12.74↔13.26)
HR (n=110)	13.00	18	16.36	12.78 (12.29↔13.27)	12.95 (12.73↔13.17)	12.92 (12.72↔13.12)
LL (n=66)	7.80	3	4.55	12.67 (11.36↔13.98)	12.87 (12.65↔13.09)	12.86 (12.65↔13.07)
LC (n=93)	10.99	16	17.20	13.88 (13.71↔14.05)	13.39 (13.21↔13.57)	13.47 (13.32↔13.62)
PP (n=82)	9.69	21	25.61	13.00 (12.64↔13.36)	13.15 (12.94↔13.36)	13.11 (12.93↔13.29)
SP (n=35)	4.14	5	14.29	13.40 (12.62↔14.18)	13.30 (13.03↔13.57)	13.31 (13.06↔13.56)
Total (N=846)	**100.00**	**147**	**17.38**	**12.93 (12.77↔13.09)**	**12.91 (12.84↔12.98)**	**12.91 (12.84↔12.98)**

The proportion of subjects measured at M1 who live in each division varied among the target schools (FIGURE 6.4). It is evident that most participants attended schools within the division in which they lived. However, schools that were located close to a division boundary and/or more centrally also attracted adolescents from the neighbouring divisions (APPENDIX A2).

FIGURE 6.4: Gulu sample distribution at M1 – stratified by school attended and division of residence



The difference in sample distribution and the potential confounding of these independent variables were considered in the subsequent statistical analyses. There were also variable differences between groups at M1 for several of the outcomes when comparing stratified groups according to age, division and school (APPENDIX A11). To control for these factors in the subsequent statistical analyses, the data were stratified according to gender and a model that adjusted for age, division, school and history of abduction was applied when appropriate.

6.4b) Comparison to normative data

i) Descriptive statistics

The proportion and 95% CI of the Gulu case-study data were calculated in each of the BFA and HFA categories defined by WHO. The proportions for the reference population used to establish the WHO normative cut-offs in each of these categories were also reported (TABLES 6.6 & 6.7) [62].

TABLE 6.6: Proportion and 95% CI of the Gulu sample at M1 in each BFA category according to the WHO global normative values - stratified by gender

BFA category	Proportion of Gulu sample (%)		Global norm proportion (%)
	Boys (N=618)	Girls (N=846)	
Severely thin	**1.29** (0.61↔2.58) n=8	**0.47** (0.14↔1.26) n=4	**0.14**
Thin	**4.85** (3.40↔6.87) n=30	**2.84** (1.89↔4.21) n=24	**2.14**
"Healthy" weight	**90.29** (87.69↔92.40) n=558	**88.77** (86.46↔90.73) n=751	**81.86**
Overweight	**3.56** (2.34↔5.36) n=22	**7.80** (6.17↔9.82) n=66	**13.59**
Obese	**0.00** (0.00↔0.75) n=0	**0.12** (0.00↔0.74) n=1	**2.28**

NOTE: BFA normative values sourced from the WHO guidelines (2007) [62].

NOTE: The 95% CIs were calculated using the modified Wald method [175].

TABLE 6.7: Proportion and 95% CI of the Gulu sample at M1 in each HFA category according to the WHO global normative values - stratified by gender

HFA Category	Proportion of Gulu sample (%)		Global norm proportion (%)
	Boys (N=618)	Girls (N=846)	
Severely stunted	**1.13** (0.50↔2.37) n=7	**0.35** (0.07↔1.09) n=3	**0.14**
Stunted	**5.02** (3.53↔7.05) n=31	**1.06** (0.53↔2.04) n=9	**2.14**
"Normal" growth	**93.04** (90.74↔94.81) n=575	**97.52** (96.21↔98.39) n=825	**95.45**
2SD<HFA<3SD	**0.65** (0.19↔1.72) n=4	**1.06** (0.53↔2.04) n=9	**2.14**
HFA>3SD	**0.16** (0.00↔1.01) n=1	**0.00** (0.00↔0.55) n=0	**0.14**

NOTE: HFA normative values sourced from the WHO guidelines (2007) [62].

NOTE: The 95% CIs were calculated using the modified Wald method [175].

The mean and 95% CI for the SBJ and MSFT were calculated for the Gulu case-study and global normative data (TABLE 6.8) [168,170].

TABLE 6.8: Crude mean and 95% CI of the SBJ and MSFT for the Gulu sample at M1 and global normative values - stratified by gender and age

Sample		SBJ (cm)		MSFT (km/hr)	
		Gulu sample	Global norms	Gulu sample	Global norms
Boys (N=618)	11	**165.48** (160.84↔170.12) n = 44	**154** (153.55↔154.45) n = 10045	**10.82** (10.51↔11.13) n = 44	**10.716** (10.697↔10.735) n = 15480
	12	**171.59** (169.12↔174.06) n = 141	**165** (164.64↔165.36) n = 15313	**10.69** (10.53↔10.85) n = 141	**10.951** (10.936↔10.966) n = 24544
	13	**182.47** (179.97↔184.97) n = 231	**176** (175.57↔176.43) n = 12052	**10.95** (10.82↔11.07) n = 232	**11.166** (11.151↔11.181) n = 27535
	14	**195.11** (192.46↔197.76) n = 195	**187** (186.56↔187.44) n = 12415	**11.32** (11.18↔11.46) n = 198	**11.517** (11.501↔11.533) n = 27106
Girls (N=846)	11	**158.13** (155.04↔161.22) n = 91	**145** (144.57↔145.43) n = 9933	**9.65** (9.47↔9.83) n = 90	**10.135** (10.119↔10.151) n = 15446
	12	**160.05** (157.73↔162.37) n = 175	**152** (151.65↔152.35) n = 16198	**9.69** (9.56↔9.82) n = 174	**10.241** (10.228↔10.254) n = 24255
	13	**165.02** (162.93↔167.11) n = 292	**157** (156.59↔157.41) n = 12143	**9.87** (9.76↔9.98) n = 291	**10.220** (10.207↔10.233) n = 27110
	14	**166.50** (164.38↔168.62) n = 278	**160** (159.59↔160.41) n = 12139	**9.65** (9.51↔9.78) n = 276	**10.311** (10.298↔10.324) n = 24924

NOTE: SBJ normative values sourced from Tomkinson et al (2007) [170].

NOTE: MSFT normative values sourced from Olds et al (2006) [168].

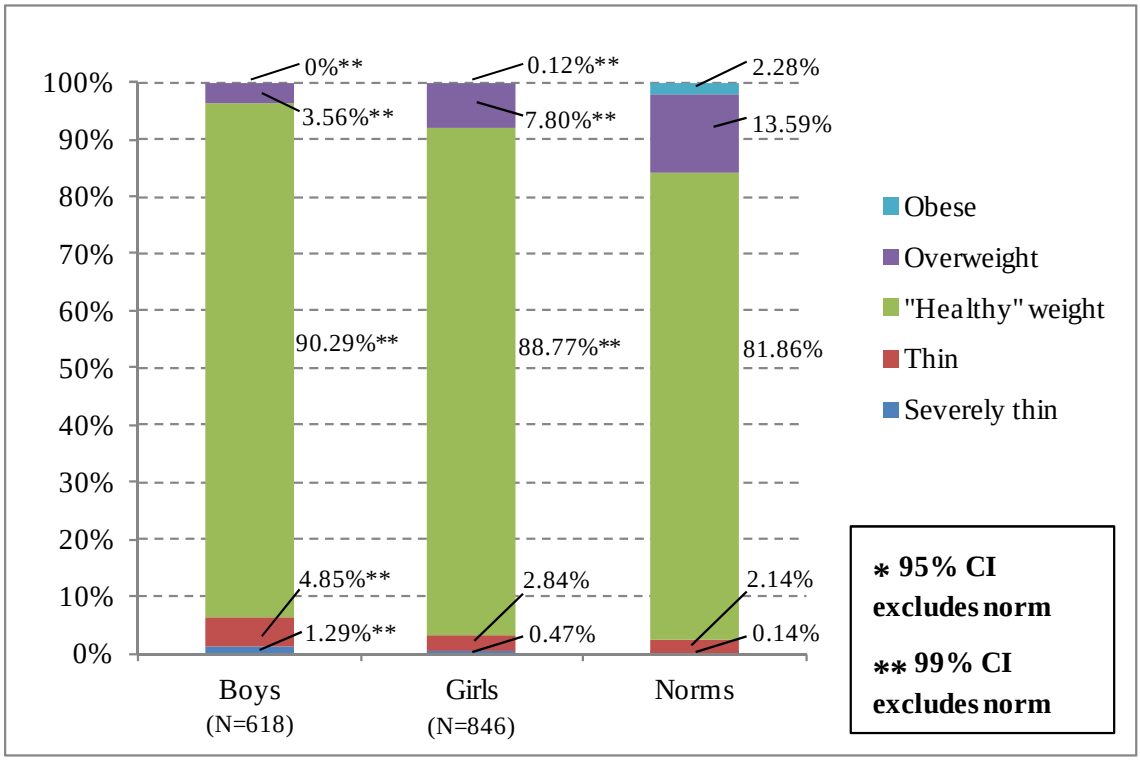
ii) Statistical findings

I found that young adolescents in the Gulu sample predominantly had a "healthy" body composition, performed well in the SBJ and achieved poor results for the MSFT.

There were no underlying assumptions for the methods used to compare BFA and HFA to normative values (APPENDIX A12). When examining BFA, there was a high prevalence of subjects in the "healthy" weight range in the Gulu sample (boys = 90.29%, girls = 88.77%).

The proportion of Gulu girls overweight (7.80%) or obese (0.12%) was lower than global norms, but exceeded those who were thin (2.84%) or severely thin (0.47%). No obese boys were identified in the Gulu sample and the proportion who were overweight (3.56%) was also very low. There was a higher prevalence of Gulu boys who were thin (4.85%) and severely thin (1.29%) than observed in the Gulu girls sample or the boys global normative population (FIGURE 6.5).

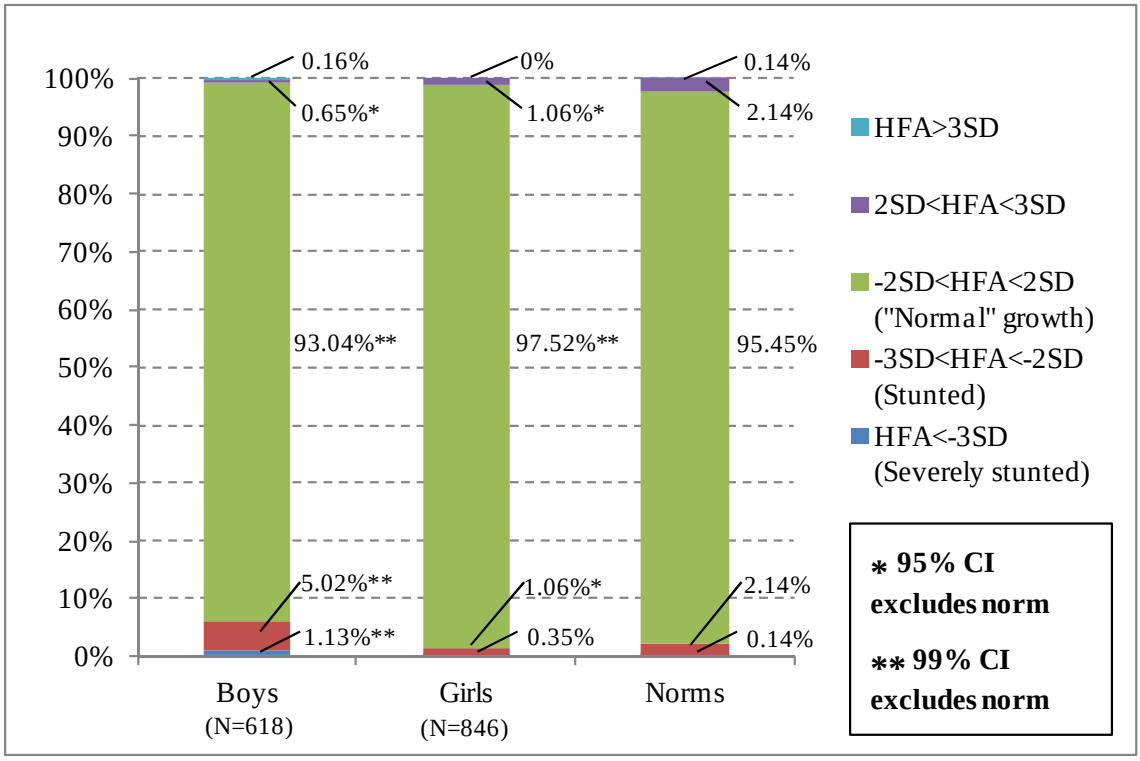
FIGURE 6.5: Graph comparing BFA proportions of Gulu sample at M1 to the WHO normative data - stratified by gender



Similar patterns were seen for HFA at M1. A high proportion of the Gulu adolescents demonstrated “normal” growth (boys = 93.04%, girls = 97.52%). However, this was lower than global norms for the boys. The girls sample had very low levels of stunting (1.06%) and severe stunting (0.35%). The prevalence of stunted (5.02%) and severely stunted (1.13%) growth in the Gulu boys was higher than observed in the Gulu girls sample and the global normative population. The proportion of subjects with a HFA greater than 2 standard

deviations above the mean was very low in both the girls (1.06%) and the boys (0.81%) Gulu samples (FIGURE 6.6).

FIGURE 6.6: Graph comparing HFA proportions of Gulu sample at M1 to the WHO normative data - stratified by gender



The assumption of homogeneity of variance (HOV) was violated for all age groups and both genders for the SBJ (APPENDIX A12). These violations resulted in a "conservative" statistical test, which reduced the risk of a type I error [199,200].

The mean score for both the boys and the girls in the Gulu sample was significantly better than global normative values in all age groups [$p<0.001$] (FIGURES 6.7 & 6.8). The statistical certainty of these results was strengthened by the "conservative" nature of the tests [199,200].

FIGURE 6.7: Graph comparing the crude mean and 95% CI of the boys SBJ in the Gulu sample at M1 to global normative data - stratified by age

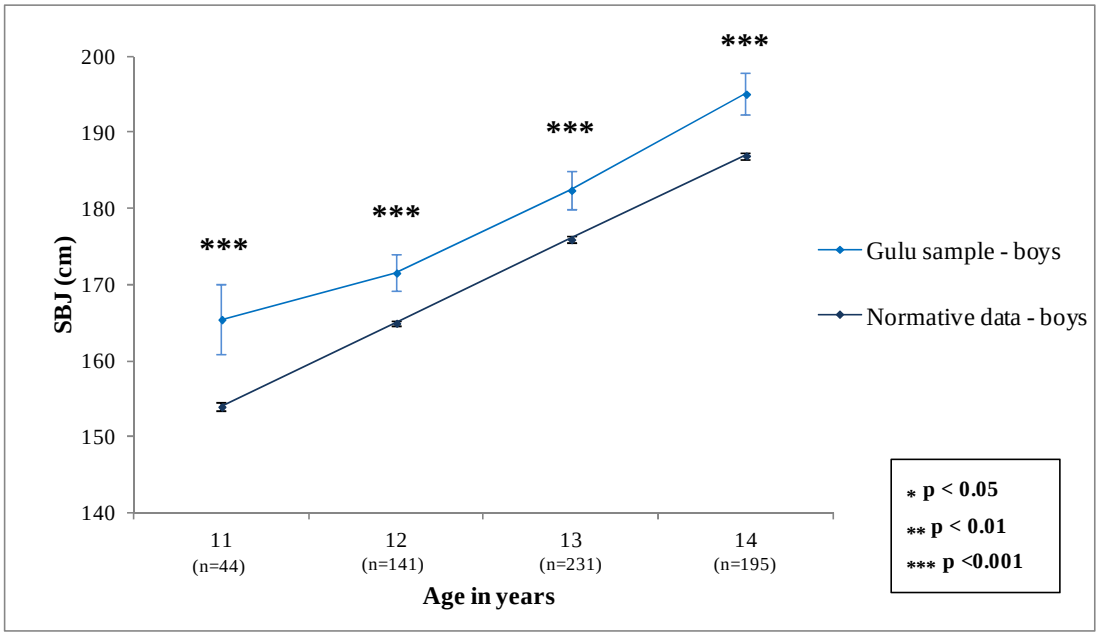
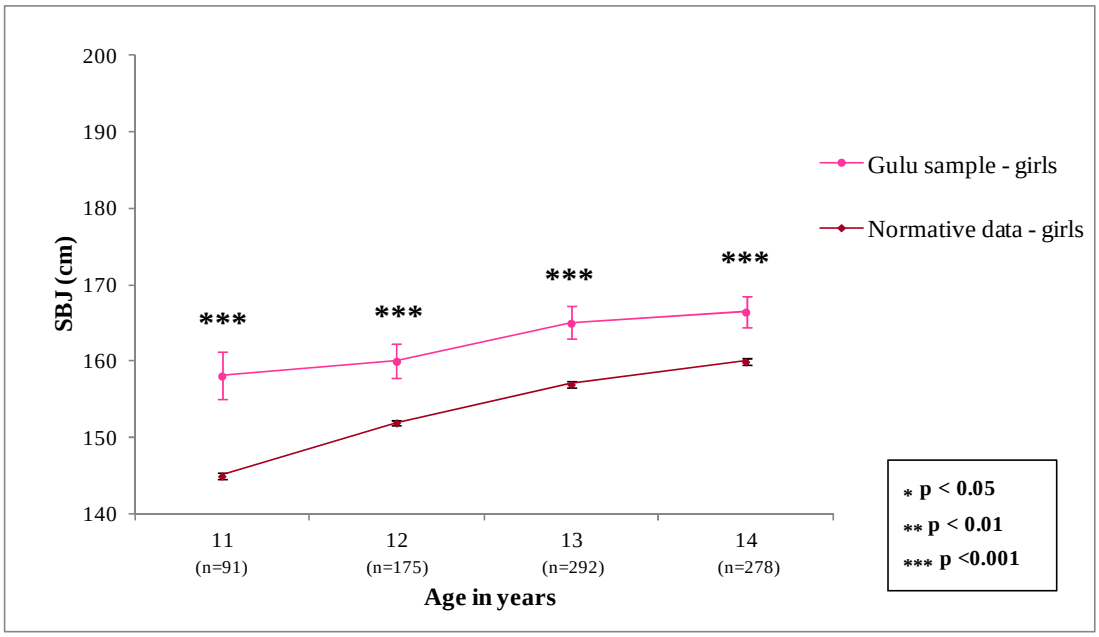


FIGURE 6.8: Graph comparing the crude mean and 95% CI of the girls SBJ in the Gulu sample at M1 to global normative data - stratified by age



With the exception of the 11-year-old boys and 14-year-old girls, all other age groups violated the assumption of HOV for the MSFT (APPENDIX A12). These violations resulted in a "conservative" statistical test, which reduced the risk of a type I error [199,200].

The mean score for the 11-year-old boys in the Gulu sample was not significantly different to the global normative values [p=0.574]. The boys aged 12 [p=0.012], 13 [p=0.011] and 14 [p=0.035] years performed significantly worse than the global normative values (FIGURE 6.9). The mean score for all age groups in the girls Gulu sample was significantly worse than the normative data [p<0.001] (FIGURE 6.10). The statistical certainty of the significant results was strengthened in the samples where a "conservative" test was applied [199,200].

FIGURE 6.9: Graph comparing the crude mean and 95% CI of the boys MSFT in the Gulu sample at M1 to global normative data - stratified by age

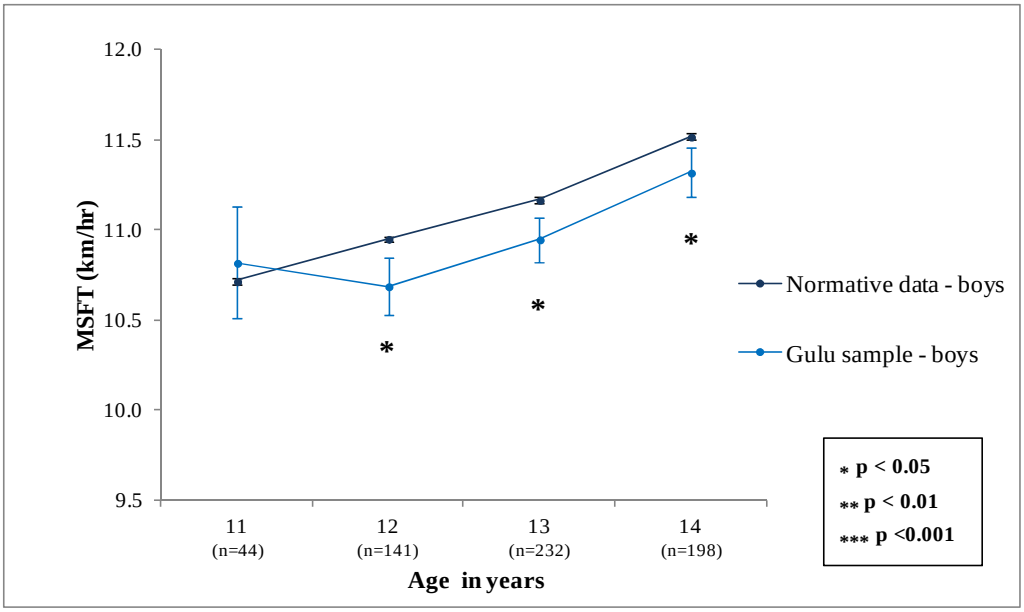
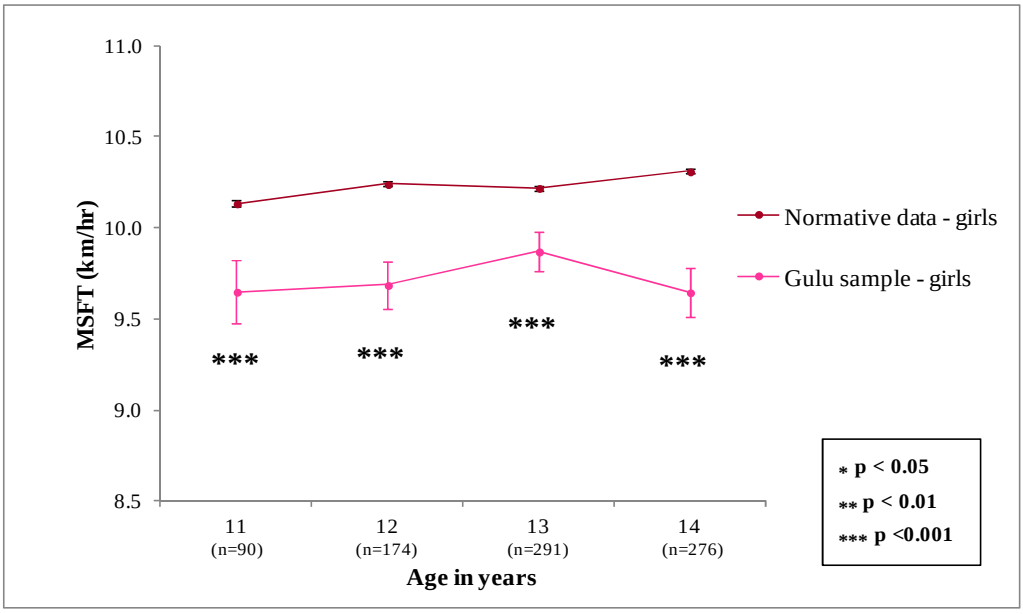


FIGURE 6.10: Graph comparing the crude mean and 95% CI of the girls MSFT in the Gulu sample at M1 to global normative data - stratified by age



6.4c) History of abduction

i) Descriptive statistics

The mean and 95% CI for all outcome variables were calculated for subjects who were abducted vs. NOT abducted (TABLES 6.9 & 6.10).

TABLE 6.9: Crude mean and 95% CI of the PF outcomes for the Gulu sample at M1 - stratified by gender and history of abduction

History of abduction		BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)
Boys (N=618)	**Yes** (n=87)	**-0.659** (-0.863↔-0.455) n=87	**-0.462** (-0.666↔-0.258) n=87	**183.64** (179.19↔188.09) n=84	**5.85** (5.43↔6.27) n=86
	No (n=531)	**-0.642** (-0.717↔-0.567) n=531	**-0.454** (-0.541↔-0.367) n=531	**182.63** (180.88↔184.38) n=527	**6.04** (5.87↔6.21) n=529
Girls (N=846)	**Yes** (n=147)	**-0.177** (-0.318↔-0.037) n=147	**-0.015** (-0.162↔0.133) n=147	**165.43** (162.83↔168.03) n=144	**3.62** (3.34↔3.90) n=143
	No (n=699)	**-0.221** (-0.288↔-0.154) n=698	**-0.071** (-0.137↔-0.004) n=698	**163.37** (162.04↔164.70) n=692	**3.57** (3.44↔3.70) n=688

TABLE 6.10: Crude mean and 95% CI of the MH outcomes for the Gulu sample at M1 - stratified by gender and history of abduction

History of abduction		TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Boys (N=618)	**Yes** (n=87)	**15.19** (13.72↔16.66) n=85	**11.51** (10.24↔12.78) n=85	**14.58** (13.22↔15.94) n=85	**10.88** (9.76↔12.00) n=85	**32.36** (29.40↔35.32) n=85	**40.00** (36.52↔43.48) n=85
	No (n=531)	**10.81** (10.31↔11.31) n=528	**8.19** (7.78↔8.60) n=528	**11.11** (10.61↔11.61) n=528	**8.37** (8.00↔8.74) n=528	**23.18** (22.19↔24.17) n=528	**29.36** (28.20↔30.52) n=528
Girls (N=846)	**Yes** (n=147)	**19.77** (18.62↔20.92) n=145	**15.39** (14.44↔16.34) n=145	**17.84** (16.74↔18.94) n=145	**12.28** (11.52↔13.04) n=145	**40.43** (38.26↔42.60) n=145	**49.15** (46.62↔51.68) n=145
	No (n=699)	**14.26** (13.76↔14.76) n=696	**10.97** (10.56↔11.38) n=696	**14.03** (13.54↔14.52) n=696	**9.70** (9.36↔10.04) n=696	**30.01** (29.01↔31.01) n=696	**37.24** (36.08↔38.40) n=696

ii) Statistical findings

I found that the PF of young adolescents with a history of abduction was not significantly different to their peers, but their MH was worse than those who were NOT abducted.

The underlying assumptions for the analysis of the PF outcome variables of abducted vs. NOT abducted subjects were satisfied (APPENDIX A13).

When analysing the crude means, there was no significant difference at M1 between the scores of the boys who were abducted vs. NOT abducted for BFA [p= 0.874], HFA [p=0.943], SBJ [p=0.675] and MSFT [p=0.395]. The girls sample demonstrated similar results for BFA [p= 0.593], HFA [p=0.490], SBJ [p=0.200] and MSFT [p=0.779] (FIGURES 6.11 – 6.14).

FIGURE 6.11: Comparison of the crude mean and 95% CI of BFA for adolescents in the Gulu sample at M1 who were abducted and NOT abducted during the war - stratified by gender

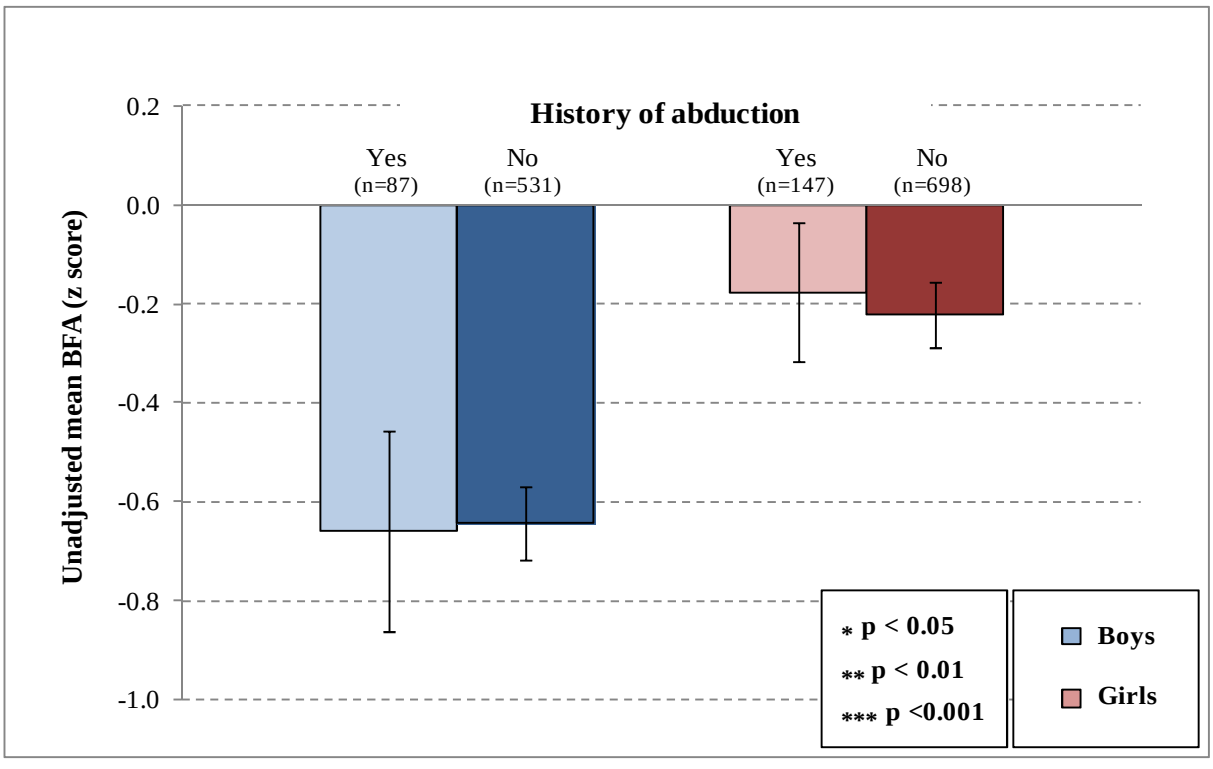


FIGURE 6.12: Comparison of the crude mean and 95% CI of HFA for adolescents in the Gulu sample at M1 who were abducted and NOT abducted during the war - stratified by gender

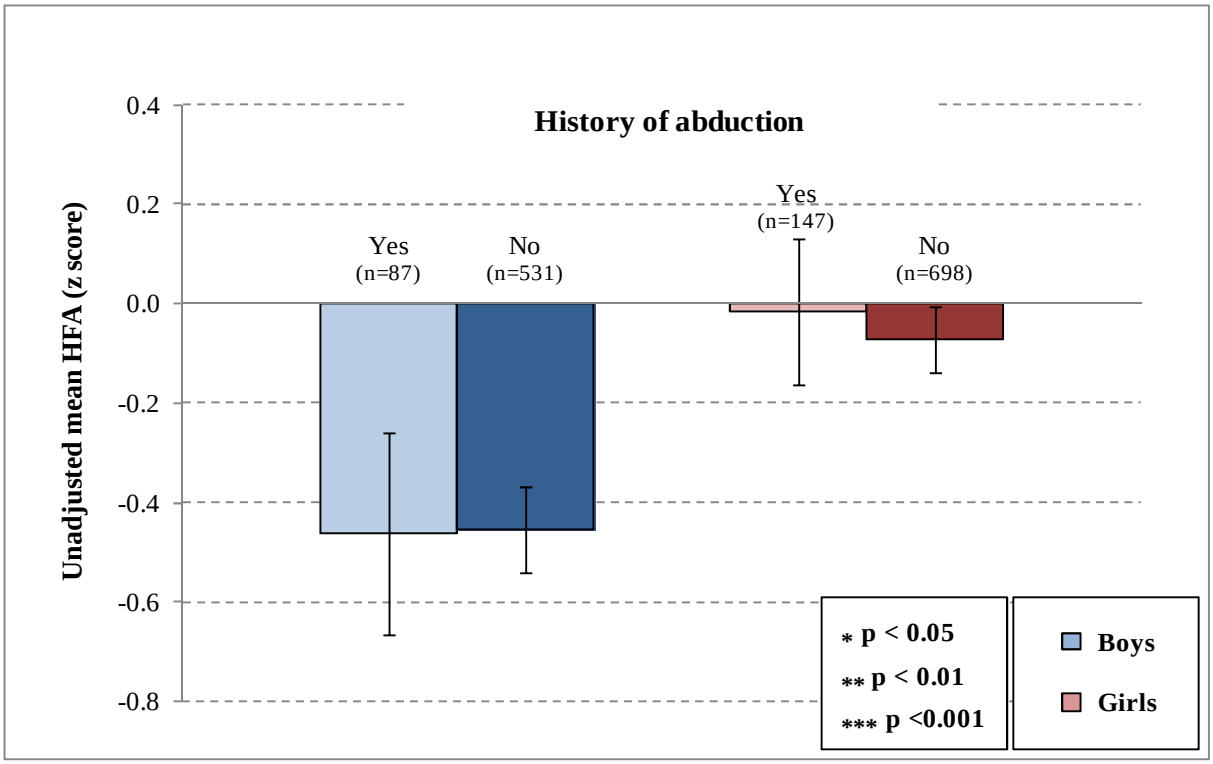


FIGURE 6.13: Comparison of the crude mean and 95% CI of the SBJ for adolescents in the Gulu sample at M1 who were abducted and NOT abducted during the war - stratified by gender

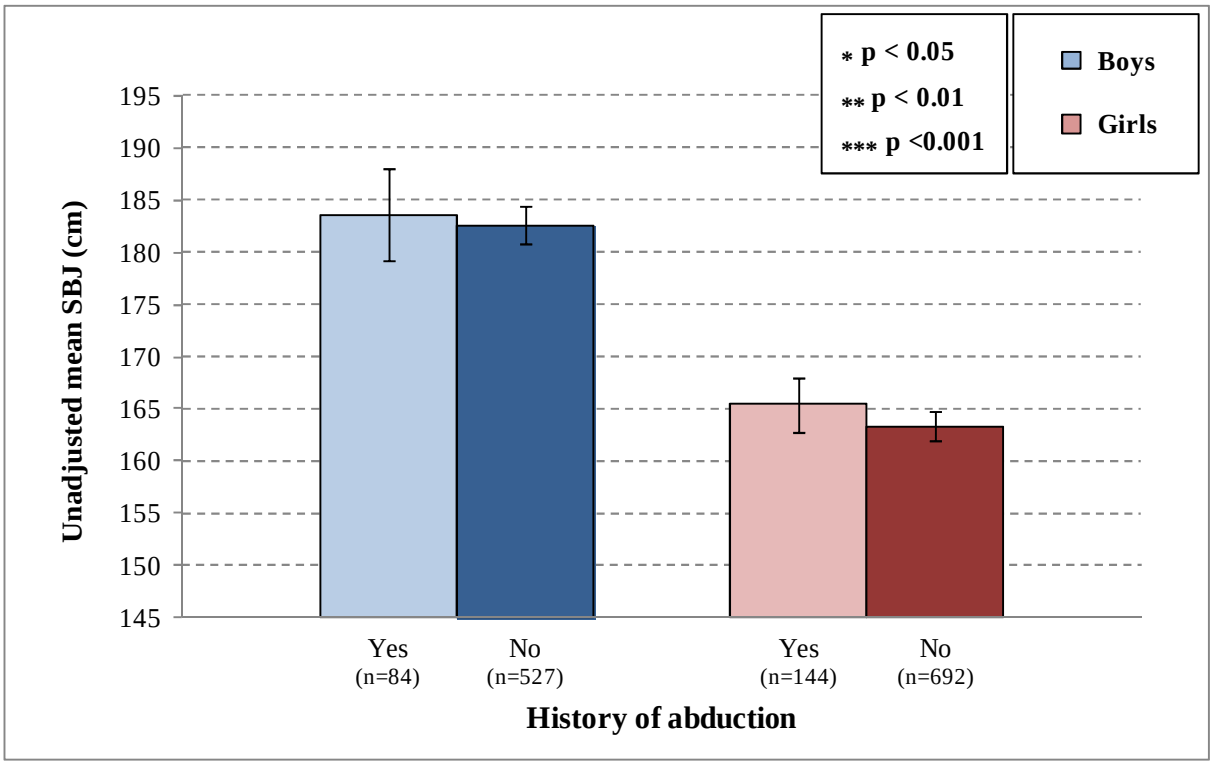
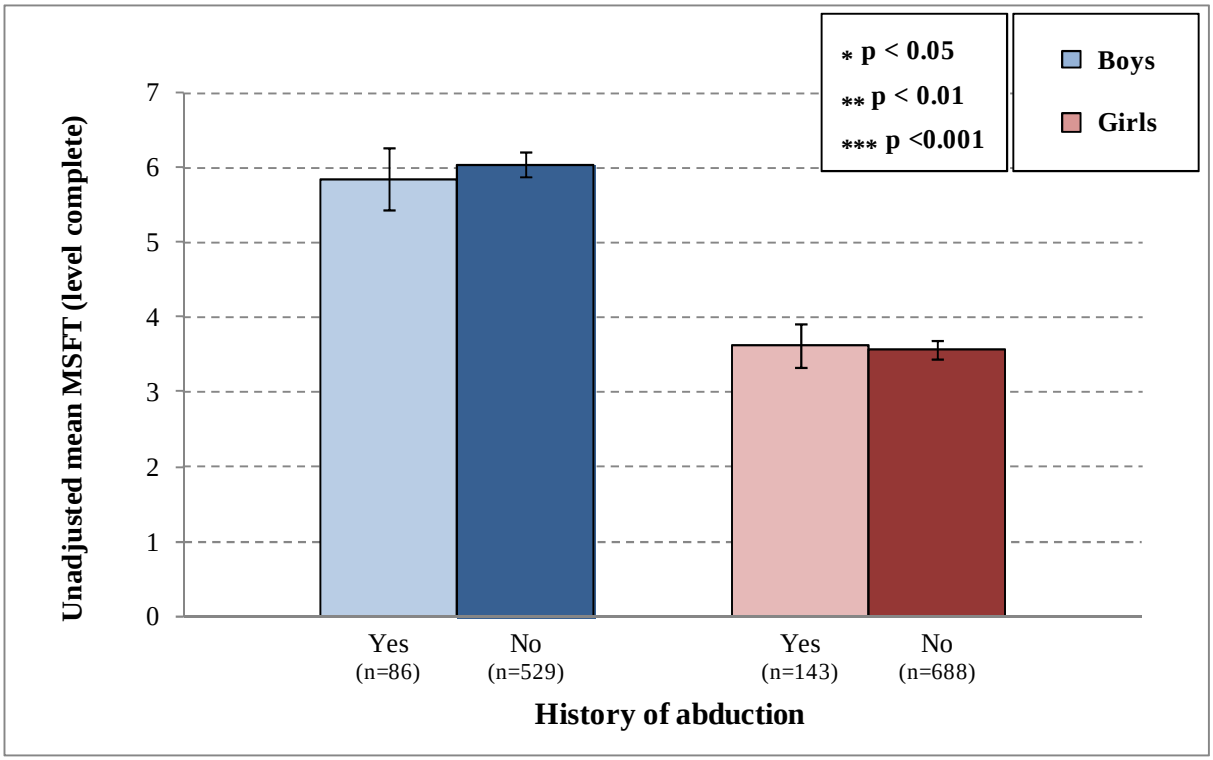


FIGURE 6.14: Comparison of the crude mean and 95% CI of the MSFT for adolescents in the Gulu sample at M1 who were abducted and NOT abducted during the war - stratified by gender



With the exception of ML in the boys sample, the underlying assumptions for the analysis of the MH outcome variables of abducted vs. NOT abducted subjects were satisfied (APPENDIX A13). The violation of the assumption of HOV for boys ML resulted in a "liberal" statistical test and an increased risk of a type I error [199,200]. Statistical testing proceeded as planned for all other variables.

The mean score of the boys who were abducted was significantly worse than those NOT abducted for all of the MH outcomes when analysing the crude data [p<0.001]. However, the "liberal" nature of the statistical test for ML in the boys sample suggested that this result should be interpreted with caution [199,200]. The girls who reported a history of abduction also had crude means that were significantly worse than their non-abducted peers for all of the MH outcomes [p<0.001] (FIGURES 6.15 & 6.16).

FIGURE 6.15: Comparison of the crude mean and 95% CI of the boys MH outcomes for adolescents in the Gulu sample at M1 who were abducted and NOT abducted during the war

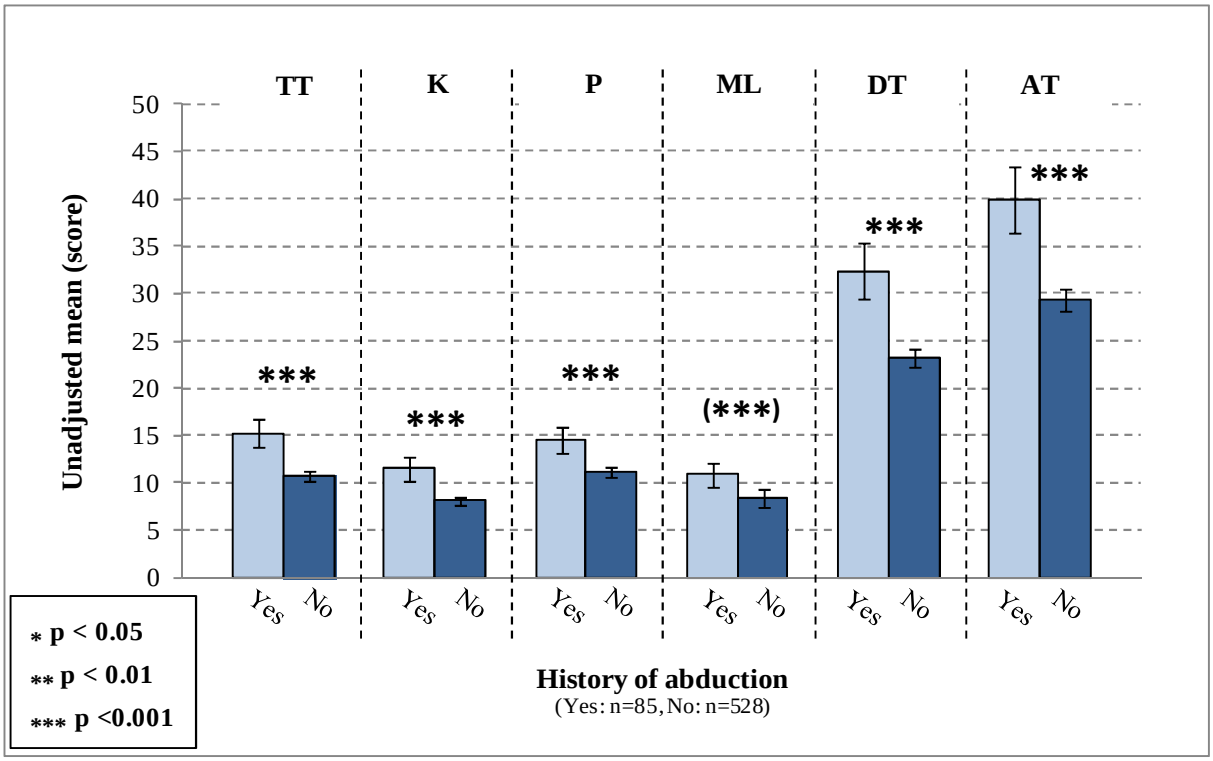
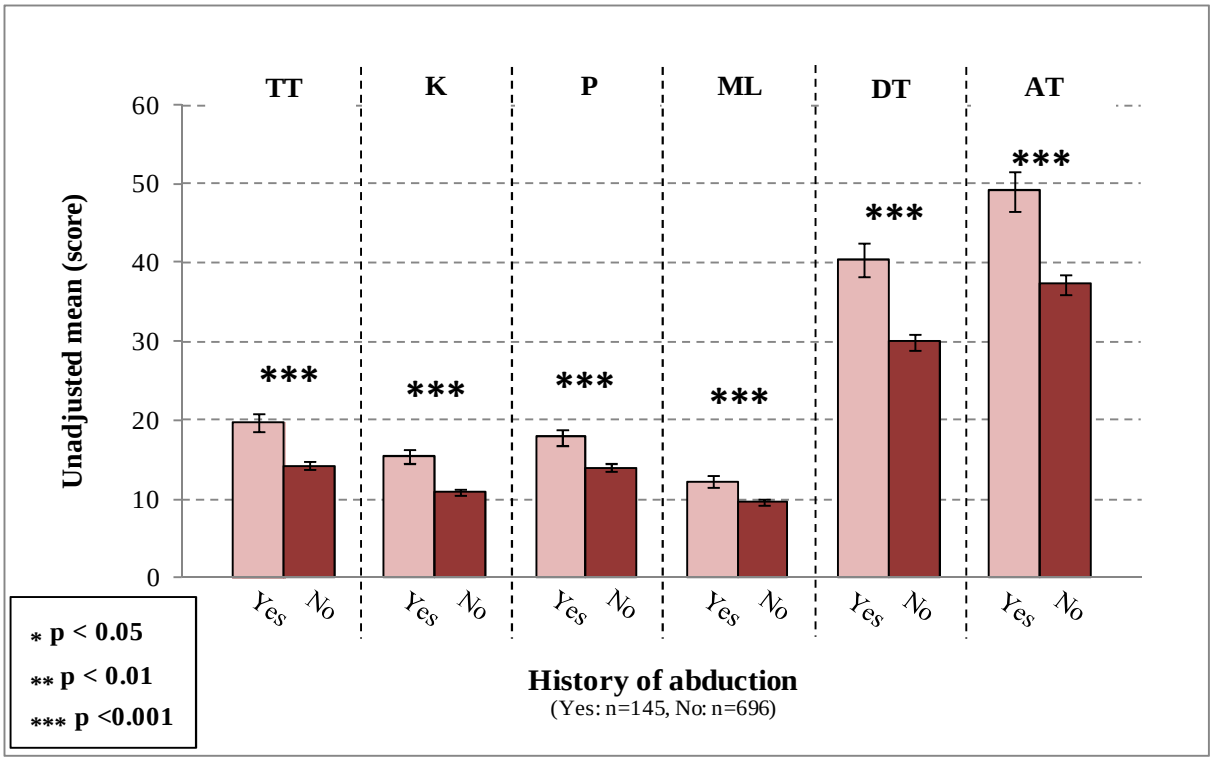


FIGURE 6.16: Comparison of the crude mean and 95% CI of the girls MH outcomes for adolescents in the Gulu sample at M1 who were abducted and NOT abducted during the war



The statistical analyses were repeated after adjusting the data for age, division and school. The differences in the mean scores for subjects abducted vs. NOT abducted remained insignificant for all of the PF outcomes for both the boys sample (BFA [p=0.169], HFA [p=0.712], SBJ [p=0.495], MSFT [p=0.846]) and the girls sample (BFA [p=0.365], HFA [p=0.967], SBJ [p=0.309], MSFT [p=0.341]).

In the boys sample, the difference between the group means was no longer statistically significant for K [p=0.094] and P [p=0.081]. The statistically significant difference was weakened but maintained for TT [p=0.009], ML [p=0.048], DT [p=0.012] and AT [p=0.015] (TABLE 6.11). The result for ML in the boys sample must be interpreted with caution due to the "liberal" nature of the F-test [199,200].

The mean scores of girls who reported a history of abduction remained worse than those NOT abducted at a highly significantly level for K, ML, DT and AT [p<0.001]. The statistically significant difference between the groups was weakened but maintained for TT [p=0.001] and P [p=0.004] (TABLE 6.11).

TABLE 6.11: Univariate ANOVA comparing the mean of the MH outcomes for abducted and NOT abducted adolescents in the Gulu sample at M1 - stratified by gender and adjusted for age, division and school

Gender	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Boys (N=618)	F=6.877 **	F=2.811	F=3.067	F=3.926 (*)	F=6.412 *	F=5.970 *
Girls (N=846)	F=10.602 **	F=14.361 ***	F=8.506 **	F=17.477 ***	F=12.099 ***	F=14.341 ***

NOTE: Means compared using a model III sum of squares analysis (unbalanced groups with no empty cells for the selected fixed factors).

NOTE: * Denotes a significant difference between abducted vs. NOT abducted adolescents (* p<0.05, ** p<0.01, *** p<0.001).

6.4d) Relationship between outcome variables

i) Descriptive statistics

The mean and 95% CI for all outcome variables were calculated (TABLES 6.12 & 6.13).

TABLE 6.12: Crude mean and 95% CI of the PF outcomes in the Gulu sample at M1 - stratified by gender

Gender	BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)
Boys (N=618)	**-0.645** (-0.715↔-0.574) n=618	**-0.455** (-0.535↔-0.375) n=618	**182.77** (181.14↔184.40) n=611	**6.02** (5.86↔6.18) n=615
Girls (N=846)	**-0.213** (-0.274↔-0.153) n=845	**-0.061** (-0.122↔0.000) n=845	**163.72** (162.53↔164.91) n=836	**3.58** (3.46↔3.70) n=831

TABLE 6.13: Crude mean and 95% CI of the MH outcomes in the Gulu sample at M1 - stratified by gender

Gender	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Boys (N=618)	**11.41** (10.92↔11.90) n=613	**8.65** (8.24↔9.06) n=613	**11.59** (11.11↔12.07) n=613	**8.72** (8.36↔9.08) n=613	**24.45** (23.47↔25.43) n=613	**30.84** (29.70↔31.98) n=613
Girls (N=846)	**15.21** (14.73↔15.69) n=841	**11.73** (11.33↔12.13) n=841	**14.69** (14.23↔15.15) n=841	**10.15** (9.83↔10.47) n=841	**31.80** (30.85↔32.75) n=841	**39.29** (38.19↔40.39) n=841

ii) Statistical findings

I found that the there was no relationship between the PF vs. MH outcomes variables. However, there were weak positive associations between body composition (BFA, HFA) vs. SBJ for the boys and a weak inverse relationship between BFA vs. MSFT for the girls.

Scatterplots of the M1 results demonstrated no relationship between the crude results of the PF vs. MH outcome variables (APPENDIX A14). The use of parametric procedures was

supported by the assumptions tests for statistical assessment of these relationships (APPENDIX A15). Pearson's r correlation coefficients did not suggest a linear relationship between the crude results of the PF vs. MH outcome variables (TABLE 6.14).

TABLE 6.14: Pearson's r correlation coefficient between the crude PF and MH outcomes in the Gulu sample at M1 - stratified by gender

Compared outcome variables		Boys (N=618)		Girls (N=846)	
		n	Pearson r	n	Pearson r
BFA (z score)	TT (score)	613	0.000	841	0.067
	K (score)	613	-0.042	841	0.023
	P (score)	613	-0.068	841	0.040
	ML (score)	613	-0.058	841	0.007
	DT (score)	613	-0.028	841	0.058
	AT (score)	613	-0.035	841	0.048
HFA (z score)	TT (score)	613	0.030	841	-0.005
	K (score)	613	0.012	841	0.011
	P (score)	613	-0.038	841	0.019
	ML (score)	613	-0.036	841	0.000
	DT (score)	613	0.000	841	0.010
	AT (score)	613	-0.015	841	0.009
SBJ (cm)	TT (score)	609	0.030	836	0.079
	K (score)	609	0.043	836	0.041
	P (score)	609	-0.001	836	0.050
	ML (score)	609	-0.042	836	0.075
	DT (score)	609	0.034	836	0.070
	AT (score)	609	0.011	836	0.075
MSFT (level complete)	TT (score)	611	0.036	829	0.017
	K (score)	611	0.043	829	-0.023
	P (score)	611	0.014	829	-0.002
	ML (score)	611	0.026	829	0.027
	DT (score)	611	0.041	829	0.000
	AT (score)	611	0.040	829	0.011

Partial correlation coefficients between the PF vs. MH outcome variables that were adjusted for age, division, school and history of abduction produced similar results (TABLE 6.15).

TABLE 6.15: Partial correlation coefficient between the PF and MH outcomes in the Gulu sample at M1 - stratified by gender and adjusted for age, division, school and history of abduction

Compared outcome variables		Boys (N=618)		Girls (N=846)	
		n	Partial correlation	n	Partial correlation
BFA (z score)	TT (score)	613	-0.015	841	0.023
	K (score)	613	-0.056	841	-0.009
	P (score)	613	-0.074	841	0.008
	ML (score)	613	-0.048	841	-0.004
	DT (score)	613	-0.041	841	0.017
	AT (score)	613	-0.043	841	0.010
HFA (z score)	TT (score)	613	0.033	841	-0.023
	K (score)	613	0.013	841	-0.002
	P (score)	613	-0.038	841	0.008
	ML (score)	613	-0.037	841	-0.008
	DT (score)	613	0.002	841	-0.006
	AT (score)	613	-0.014	841	-0.006
SBJ (cm)	TT (score)	609	0.012	836	0.037
	K (score)	609	0.021	836	0.009
	P (score)	609	-0.006	836	0.019
	ML (score)	609	-0.029	836	0.062
	DT (score)	609	0.019	836	0.031
	AT (score)	609	0.006	836	0.039
MSFT (level complete)	TT (score)	611	0.049	829	0.008
	K (score)	611	0.050	829	-0.033
	P (score)	611	0.024	829	-0.006
	ML (score)	611	0.040	829	0.027
	DT (score)	611	0.053	829	-0.008
	AT (score)	611	0.056	829	0.005

Visual assessment of scatterplots comparing the crude results for body composition (BFA, HFA) vs. muscular strength (SBJ) and cardiorespiratory fitness (MSFT) suggested associations that were marginal at best (APPENDIX A14). The assumptions for statistical assessment of these relationships supported the use of parametric procedures (APPENDIX A15). In concordance with visual assessment of the scatterplots, Pearson's r correlation coefficients of the crude results suggested some weak linear relationships. The results for the

boys SBJ had a weak positive correlation with BFA [Pearson's r = 0.289] and HFA [Pearson's r = 0.301]. The correlation between the girls MSFT and BFA was negative and weak [Pearson's r = -0.207]. The Pearson's r correlation coefficients for the results of the other outcome variables were negligible (TABLE 6.16).

TABLE 6.16: Pearson's r correlation coefficient between the crude results for body composition (BFA, HFA) and indicators of muscle strength (SBJ) and cardiorespiratory fitness (MSFT) in the Gulu sample at M1 - stratified by gender

Compared outcome variables		Boys (N=618)		Girls (N=846)	
		n	Pearson r	n	Pearson r
BFA (z score)	SBJ (cm)	611	0.289	836	-0.034
	MSFT (level complete)	615	0.097	831	-0.207
HFA (z score)	SBJ (cm)	611	0.301	836	0.166
	MSFT (level complete)	615	0.133	831	0.036

Partial correlation coefficients that compared these outcome variables and were adjusted for age, division, school and history of abduction produced similar results (TABLE 6.17).

TABLE 6.17: Partial correlation coefficient between the results for body composition (BFA, HFA) and indicators of muscle strength (SBJ) and cardiorespiratory fitness (MSFT) in the Gulu sample at M1 - stratified by gender and adjusted for age, division, school and history of abduction

Compared outcome variables		Boys (N=618)		Girls (N=846)	
		n	Partial correlation	n	Partial correlation
BFA (z score)	SBJ (cm)	611	0.289	836	-0.079
	MSFT (level complete)	615	0.093	831	-0.213
HFA (z score)	SBJ (cm)	611	0.330	836	0.159
	MSFT (level complete)	615	0.129	831	0.033

6.5) Preliminary discussion

6.5a) Principal findings

There were gender-, age-, division- and school-level differences in the distribution of the sample that had a variable effect on each of the measured outcomes. This was also the case for subjects reporting a history of abduction. To account for these discrepancies, I stratified the data by gender and adjusted for age, division, school and history of abduction in the statistical analyses.

The results for BFA and HFA varied in comparison to global norms. The girls sample had a relatively high proportion in the "healthy" range with low levels of overweight and stunted growth [62]. Although the boys also had a high percentage in the "healthy" weight range, there were some signs of acute and chronic malnourishment when compared to global norms [62].

The results for the other PF outcomes were also variable. The Gulu sample appeared to perform better in the SBJ and worse for the MSFT than established global norms [168,170]. This suggested that the sample was anaerobically adapted and had poor aerobic capacity.

A history of abduction appeared to have no lasting effect on any of the PF outcome variables. However, all of the MH indicators measured in this study demonstrated that the girls who were abducted during the war were significantly worse than their peers. There also appeared to be lasting negative effects on TT, ML, DT and AT for the boys abducted during the war.

There was no relationship between MH and PF in the current study. The positive associations between body composition (BFA, HFA) and muscular strength (SBJ) for boys were weak. There was also a weak negative correlation between BFA and the MSFT for the girls sample.

6.5b) Limitations and strengths

The biggest weakness in this study was the use of measurement methods that were modified from their validated form. This compromised the certainty of the cross-sectional analysis comparing the current study to other research that used the original assessment protocols [179]. Therefore, the strength of the conclusions from the comparison to global norms was limited. It also had implications when assessing the associations between the dependent variables.

There were several challenges to completing this study that further compromised measurement validity and sample recruitment. Although all 10 randomly selected schools agreed to conduct testing during scheduled PES classes, varying levels of coercion were required when this became a concrete task. The assessors prioritised morning assessments because often students did not return to school after lunch. This was of particular concern during the local wet season, which also affected the capacity to conduct the PF tests outdoors.

There was also uncertainty about the testing time frames due to several unforeseen events. For example, the Ugandan government provided only 3 weeks' notice prior to changing the dates for the end-of-school term due to an upcoming by-election. This shortened the data collection window by one week because the dates for P6 examinations and the Gulu Primary School Music Festival were brought forward.

These complications affected subject recruitment and data collection. Contrary to initial planning, it became necessary to visit multiple schools each day during the M1 assessment period. Return visits were required at most sites depending on student attendance, weather and staff compliance at the scheduled times. Although flexibility made the sample larger and more representative, the testing order became dictated by opportunity and convenience. However, the impact of fatigue on the PF results was minimised by never conducting the

MSFT immediately prior to the SBJ assessment. The haphazard testing procedures fitted with local planning and time management practices, but compromised academic rigour.

The subject recruitment procedures affected the distribution of the sample according to age, school, division and history of abduction. Assessing only P6 students determined the age distribution. The number of adolescents recruited from each division was influenced by the uneven size and dispersion of the randomly selected target schools within Gulu municipality (APPENDIX A2). The previous activities of the LRA and resettlement patterns also meant that the risk of abduction during the war varied according to the division in which the subjects lived [146]. Although the disproportionate distribution of the sample was a weakness in this study, statistical methods to adjust for these differences were utilised [179,190].

Despite higher levels of school absenteeism and the measurement of more ineligible subjects than expected, the sample size appeared to be adequate for the analyses in the current study. The compromises in statistical power caused by adjusting the data for potential confounders was offset by oversampling [179,201].

There were also some concerns about the comparative data used in this study. No age-appropriate Gulu or Uganda specific norms were identified for any of the outcome variables at the time of the current study. It was possible to use the data from other African or LMIC samples as age-appropriate normative comparators for several of the PF outcomes. However, there were few studies and this approach introduced assumptions about the determinants of these variables for which there were limited data [55,168-170]. The development of global recommendations also suggested that using regionally and/or socio-economic specific norms may have been an inherently discriminatory process (SECTION 2.6).

Normative values for the MH outcomes of the APAI questionnaire had not been established due to a paucity of data. Poor inter-method agreement meant that comparison to previously collected APAI interview data was not appropriate (SECTION 5.4civ) [126,128]. The biggest strength of the APAI was its relevance to the Gulu context and the local construct of MH. However, this emic characteristic limited its generalisability and comparability to other settings and MH measures. Although the APAI outcomes were translated to "depression-like" and "anxiety-like" syndromes, they were not the same as the constructs of depression and anxiety [127]. Therefore, comparing the outcomes of the current study to results from different populations using etic scales for depression and anxiety was of limited utility [153].

Comparing abducted vs. NOT abducted subjects was also limited by the paucity of data for the MH outcomes. In the absence of criterion-referenced parameters for the APAI in questionnaire format, it was not possible to interpret the clinical relevance of the results [179].

The use of correlations to assess the relationship between the dependent variables preserved the richness of the continuous data. However, the linearity and non-specific interpretation of correlation analysis was a potential weakness in the statistical methods used in this study. Although the accompanying scatterplots addressed these concerns, alternative statistical approaches that required data categorisation were considered. Various factors limited the application of these methods for each of the outcome variables.

The logical cut-offs for BFA and HFA corresponded to the health-related categories identified by WHO [62]. However, high proportions within the "healthy" range in the Gulu sample limited the power of categorical statistical analysis in the "unhealthy" groups. Alternatively, classifying BFA and HFA according to percentiles would have inevitably included individuals from the "healthy" range in the "unhealthy" groups at both ends of the spectrum.

Criterion-referenced recommendations for adolescent health have not been established for the SBJ and MSFT field tests (SECTION 2.6a). The health-related thresholds described for VO_2 max were not used in this study because of the unknown conversion error for the modified MSFT field test results [63-66]. Tomkinson et al (2008) did describe a series of age-related normative results for the SBJ and MSFT according to percentiles [55]. However, for the purpose of this study, there was limited value in introducing percentile categories (e.g. quintiles) to continuous data that were not based on clinically relevant parameters.

There were similar concerns for the MH data. Although Bolton et al (2007) identified a diagnostic cut-off score for each local syndrome, these were based on the results of APAI interview data [126]. Poor inter-method agreement between interview and questionnaire results for the APAI precluded using the same threshold scores in this study (SECTION 5.4civ).

6.5c) Mechanisms and implications

The concerns regarding the measurement validity weaken any conclusions about the local needs that can be drawn from the results of the current study. However, it is important to note that the metric modifications were minimised and were integral to implementing the desired study design in the local context (CHAPTER 5). When comparing the Gulu sample to global norms the magnitude of the statistical difference was sufficient to suggest significance despite the minor modifications to the measurement methods.

The most globally representative normative values derived from large databases were utilised for the statistical analyses in this study. Although several organisations have established norms for BFA and HFA, the WHO standards were the most internationally representative [62]. The global norms for the MSFT were based on the results from 109 studies in 37

countries and represented all continents [168]. The SBJ normative values were derived from the most diverse database available and included the results from 23 European countries [170].

There are numerous potential explanations for the PF results of the Gulu sample compared to global norms. It was expected that the high proportion of adolescents with a "healthy" body weight would be associated with performances that were better than the global norms for the SBJ and MSFT [55,168-170,185,202,203]. This was the case for the SBJ, but the under-performance of the Gulu sample in the MSFT indicated poor aerobic capacity and did not fit this theory.

Malnourishment and stunted growth is associated with poor PF outcomes [185]. However, in the current study poor performance was isolated to the MSFT. This suggested that there was no systemic limiting factor for PF. Only the boys sample had significant levels of stunting and their relative performance in the MSFT was better than the girls. The scatterplots and correlation analyses between these variables also suggested that malnourishment did not contribute to the poor MSFT results in the Gulu sample (APPENDIX A14).

There may have been a genetic predisposition for adolescents in Gulu to perform well in the SBJ and poorly on the MSFT. The differences in the diversity of the normative populations for the SBJ and MSFT (i.e. European vs. Global) complicated this issue [168,170]. In the current study it was not possible to determine the genetic contribution to the results. However, it was unlikely that heredity fully explained the divergence of the SBJ and MSFT results in Gulu relative to the selected norms [40,204].

Measures of intrinsic and extrinsic motivation were not considered when formulating the normative values for the SBJ and MSFT [168,170]. They were also not directly measured in the current study. However, the absence of a relationship between the PF and MH variables

suggested that the affect of intrinsic motivation on the SBJ and MSFT performance was not a major concern for the Gulu sample. Despite questionable measurement validity and biased sampling, Coalter (2010) also described high levels of self-efficacy in Gulu youth [130].

Several extrinsic environmental factors cited in the literature may have affected subject motivation and the PF results in this study. These include the cultural perception of sport, engagement in PA and local climate [55,168-170]. It was important to consider these factors to ensure the needs assessment remained contextually relevant.

A broad lack of familiarity with sustained exertion may have contributed to the poor results in the MSFT [55]. Although it appeared that adolescents in Gulu engaged in high volumes of moderate PA through active transport (i.e. walking), the opportunities to engage in vigorous PA (VPA) were limited. This premise was supported anecdotally by the haphazard implementation of PES classes, limited farming practices, poor infrastructure for recreational PA and the absence of community sport programmes (SECTION 3.6).

The results of the Ugandan GSHS also suggested that 15-year-old students engaged in low volumes of PA (SECTION 3.6) [154]. However, this was based on questionnaire data from a biased sample that was not age-appropriate [154]. Further investigation is indicated to measure and differentiate adolescent moderate-vigorous PA (MVPA) in Gulu.

The hot and humid climate in Gulu may have contributed to an aversion to physical exertion. The effect would be more pronounced in the MSFT than the SBJ because it required a sustained effort. This fits with the results of the current study. However, previous research suggests that local climate is actually a confounder for other socio-cultural issues and/or geographical anomalies [168,170].

Poor test familiarity may have also affected the MSFT results in the Gulu sample [55,169,170]. Although the intra-tester reliability for the MSFT was satisfactory, there was a trend suggesting improved performance at the second assessment (SECTION 5.4biv). It is not known how many of the subjects in the global norms meta-analysis were also being exposed to the MSFT for the first time [168]. However, the volume of data that contributed to formulating the normative values was intended to attenuate these sampling inconsistencies and minimise the impact [168-170]. The SBJ also required a higher level of motor skill than the MSFT to optimise performance and the effect of test familiarity is typically more marked as the task becomes more technical [55]. Therefore, it is unlikely that test familiarity contributed to the comparative differences of the SBJ and MSFT results in the Gulu sample.

Cultural norms and gender-based attitudes also discouraged girls from participating in sport when opportunities were available in Gulu [130]. Being overweight has traditionally been seen as attractive in Uganda and its prevalence is increasing [130,151]. There were more overweight girls in the Gulu sample than boys. This could explain the relatively poor performance of girls in the MSFT when compared to the boys in the current study [55]. There was either a true difference in the cardiorespiratory fitness of boys vs. girls relative to the global norms and/or a disparity in their exertion during the actual MSFT.

It was not possible to compare the MH results for the entire Gulu sample to normative values due to the paucity of appropriate data. However, the subjects who had been abducted during the war scored significantly worse for most MH outcomes than their peers in the current study. The statistical difference was sufficient to suggest significance despite concerns about the measurement validity. It is likely that a history of abduction affected subsequent MH and

that these results indicate good face validity for the APAI in questionnaire form. However, it is not possible to make causative statements due to the cross-sectional nature of this study.

These results were in contrast to the psychosocial findings of previous researchers in Gulu. Coalter (2006 & 2010) and Annan et al (2006 & 2008) measured different MH outcomes that were based on a "Western" construct for MH [129,130,147,148]. They described high levels of robustness in Gulu youth and warned against the assumed psychosocial deficiency of the previously abducted sub-population [130,147,148]. However, these results are not directly comparable to the current study and it is likely that the high levels of resilience observed were secondary to traumatic exposures that have caused underlying MH problems.

The contrasting results may also be due to the questionable measurement validity in the current study. However, previous research that established the validity of the APAI in interview format identified a high prevalence of MH syndromes in Gulu [128]. Although the APAI metric in the current study was modified, it retained its internal consistency and relevance to the local construct of MH [127]. This may have contributed to the different psychosocial conclusions that were drawn when using other MH assessment tools in Gulu. Therefore, using metrics appropriate to the local context to report on psychosocial outcomes is critical in future studies.

It is not known whether the statistically significant differences between abducted vs. NOT abducted subjects in this study were of clinical importance. However, a previous study arbitrarily described a variation in APAI interview score of 10 as clinically relevant [126]. Although this was not based on health-related criteria and may not be transferable to the questionnaire data in the current study, the crude difference between abducted vs. NOT abducted for AT was more than 10 for both genders.

Considering the diverging relationships that PF and MH had with history of abduction, it is not surprising that there was no association between any of the MH and PF outcomes. Although the cross-sectional nature of the study precludes statements about causation, this result suggested that it was unlikely that a SfD intervention would impact on MH outcomes in Gulu. This premise assumes that the fittest adolescents in Gulu are the most active. It is based on the mechanism described by Bouchard (2001) commonly cited in the SfD literature, which suggests improvements in MH are mediated by change in PA and PF [107,138]. Therefore, the need for SfD programmes to address MH problems in Gulu appears minimal. However, Coalter (2010) describes a broader role for sport in SfD programmes that may influence MH via other avenues [130]. Further investigation is warranted to assess the impact of a SfD intervention on PA, PF and MH outcomes.

In future studies, the distribution of the sample may be improved by using proportional randomisation to select a more balanced number of schools and subjects from within each division. Assessing multiple class levels at each school should ensure each age group of young adolescents is also appropriately represented relative to the Gulu population.

The results from the current study can be used as baseline measures in subsequent research, provided the data are adjusted for the differences in the baseline measures (i.e. at M1). The disproportionate sample distribution according to age, division, school and history of abduction should also be controlled for in the final statistical analysis. The associations between body composition (BFA, HFA) and the SBJ and MSFT were too weak to be considered robust to the questionable validity of the metrics. Therefore, it is not necessary to adjust for BFA or HFA as linear covariates in future studies that assess the reach of a SfD programme (CHAPTER 7) and its impact on the outcomes of interest in Gulu (CHAPTER 8).

6.6) Summary box

COMMUNITY NEEDS ANALYSIS – GULU, UGANDA

- Subject recruitment procedures caused disparities in the sample distribution according to gender, age, school, division and history of abduction. These factors were adjusted for statistically.

- It is not known how pragmatic changes to the measurement methods and study design impacted on the validity of the results.

- The magnitude of the statistical difference when comparing PF to global norms was sufficient to suggest significance despite the concerns about measurement validity. The same can be said for the relationship between MH and history of abduction. However, the weak associations identified between several of the outcome variables were discounted due to the concerns about measurement validity.

- The comparison of the Gulu PF results to global norms and the hypothesised explanatory factors can be summarised as:
  - Large volumes of moderate-intensity PA contributed to a high prevalence of "healthy" BFA in both genders;
  - Very good performance in the SBJ by both genders was due to the high prevalence of "healthy" BFA and because the test did not require a sustained effort of VPA;
  - Poor performance in the MSFT by the boys was because the benefit of "healthy" BFA was outweighed by limited opportunities to engage in bouts of VPA that were sufficient to improve cardiorespiratory fitness and/or accustom the subjects to sustained exertion during the test;
  - Very poor performance in the MSFT by the girls was caused by the same reasons and exacerbated by cultural norms that discouraged participation and/or exertion in sport.

- A history of abduction was strongly associated with persistent MH disorders in young adolescents in Gulu. This relationship was particularly pronounced in girls.

- The negligible association between PF and MH suggested that a SfD intervention depending on PA and/or PF mediators to improve MH is unlikely to be successful.

- The results of this study describe the PF and MH needs in Gulu using metrics that are appropriate to the local context. This data can be used to inform the objectives and implementation of a SfD intervention.

- Further investigation is warranted to assess if a SfD intervention does actually impact on PA, PF and MH. The results from the current study can be used as baseline measures for this evaluation provided the disproportionate sample distribution is adjusted for in the final statistical analysis.

PART V

INNOVATION TESTING

CHAPTER 7

Reach of a Sport-for-Development Intervention for Young Adolescents in Gulu, Uganda

7.1) Purpose of chapter

The purpose of this chapter is to evaluate the reach of a SfD programme in Gulu. I present a cross-sectional analysis of the PF and MH data collected in Chapter 6 to compare young adolescents who register for a voluntary SfD programme to their peers. The results will be used to assess who the intervention is reaching and if it is addressing the community needs.

7.2) Introduction

The UN International Year of Sport and Physical Education in 2005 catalysed an unprecedented increase in interest and funding for the SfD sector [81]. Sport is now promoted as an inherently positive intervention that reaches people in diverse societies and provides an economy of remedies for various problems [135,137]. The number of new organisations specialising in delivering SfD programmes continues to grow rapidly [70,71]. Many other established organisations are also diversifying and introducing sports activities to facilitate access to this expanding and potentially lucrative field [125,129].

The emergence of SfD as a prominent player on the international agenda has prompted discussions about the different approaches used by various organisations. Coalter (2006) described the terms "sport plus" and "plus sport" to classify the increasing number of stakeholders delivering interventions (SECTION 2.10) [122,129]. The focus of "sport plus" organisations is to develop sport and use it as a vehicle for achieving their aims [122,129,136]. The "plus sport" approach is less concerned about the sustainability of sport as an intervention, but uses it as a hook to engage people in other programmes that have development objectives [122,129,136]. However, it is widely accepted that there is a continuum between these categories and the attraction of sport is the common denominator that influences the reach of SfD interventions at both ends of the spectrum [121,122,129,136,137].

Monitoring of implementation is the most common way that programme reach is described in the SfD sector [136]. This often involves reporting the number of beneficiaries and/or the amount of equipment distributed. However, the focus should be on ***who*** is reached rather than just the number of children living in the communities where the intervention is implemented [129,205]. This depends on the allure of sport in the target population and the recruitment procedures [121].

The GMKL intervention evaluated in this thesis is a community-based "sport plus" programme in Gulu municipality. The participants voluntarily register and the target population is influenced by the objectives of the contributing stakeholders. These focus on peace-building and improving adolescent health.

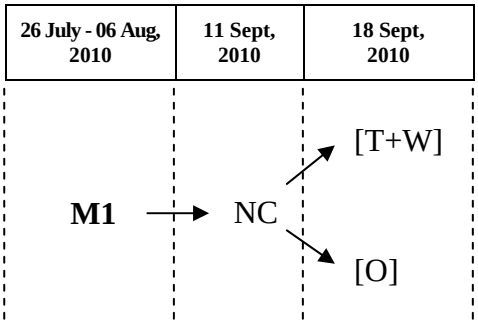
In this chapter, I endeavour to explore who is registering for the GMKL in Gulu according to differences in age, division and school. I will also consider if a history of abduction during the war and the associated stigma are barriers for accessing the programme in the community. It is also important to assess if the participants include adolescents who have low PF and/or poor MH. The aim is to establish if the GMKL programme is actually reaching those who have the greatest need when considering the outcomes of interest. The results will contribute to interpreting the outcomes of the subsequent impact evaluation of the intervention (CHAPTER 8).

7.3) Methods

7.3a) Study design

The study was cross-sectional and used the data that were collected at M1 (CHAPTER 6). The GMKL was the intervention evaluated in the study (APPENDIX A16). The voluntary subscription of young adolescents for the GMKL was a non-random allocation process (NC). A total of 678 young adolescents (511 boys, 167 girls) registered for the programme. The subjects measured at M1 who registered for the GMKL formed group [T+W]. Those who did not register for the programme but were measured at M1 were in group [O].

FIGURE 7.1: Diagram representing subject recruitment and study design



LEGEND:
Subject recruitment

NC	Adolescents register for the GMKL (NOTE: non-random)
[T+W]	Eligible adolescents registered in the GMKL
[O]	Eligible adolescents NOT registered in the GMKL

Study design and data collection

M1	Initial case-study measurement (PF, MH)

7.3b) Sample size calculation

The sample size calculation for comparing two groups that was initially described in Chapter 6 was relevant to the current study (SECTION 6.3b) [179,195,196]. However, the comparison between registered vs. NOT registered was adjusted for an expected uneven sample size. Based on previous experience, the GMKL coordinators estimated that approximately $^{1}/_{4}$ of the

boys and $^{1}/_{10}$ of the girls in the community would register for the league. These ratios were used to adjust the required sample sizes for this comparison (TABLE 7.1) [195].

TABLE 7.1: Calculated minimum sample sizes for the statistical analysis in the programme reach assessment in Gulu - stratified by gender

Statistical analysis	Gender	Required sample size (n)		
		Group 1	Group 2	Total
Comparison of registered vs. NOT registered	Boys	Registered: 64.89 (65)	NOT reg.: 259.55 (260)	**325**
	Girls	Registered: 53.51 (54)	NOT reg.: 535.10 (536)	**590**

It is important to note that the sample size calculations were completed before M1 (SECTION 6.3b). The results were considered along with the number of pupils enrolled in P6 when deciding how many schools were assessed at M1 (SECTION 6.3c).

7.3c) Subject recruitment

The same subjects as recruited in Chapter 6 were included in this study (SECTION 6.3c).

7.3d) Subject identification and tracking procedures

The procedures for subject identification and tracking comprised collecting the same information at M1 and NC as described for pilot testing in Chapter 5 (SECTION 5.3d).

7.3e) Outcome variables and metrics

The same outcome variables and metrics as assessed in Chapter 6 were used in this study (SECTION 6.3e).

7.3f) Data collection procedures

The data collected at M1 in Chapter 6 were used in this study (SECTION 6.3f).

7.3g) Data analysis and presentation

All of the quantitative data were cleaned and checked for outliers prior to commencing the analysis. The underlying assumptions for each statistical test were also verified before proceeding. The reporting of descriptive statistics for each of the variables included the crude mean, 95% CI and sample size (n).

i) Sample distribution

The sample of eligible adolescents who completed M1 was described according to registration for the GMKL. The results were stratified by gender, age, division, school and history of abduction. The absolute number of subjects and the proportional representation at each level of the stratified data were tabulated to identify disparities in the distribution of the sample at the time of registration.

ii) Reach of Gum Marom Kids League (self-selection bias)

Descriptive statistics for all outcome variables stratified according to registration for the GMKL and gender were calculated. The results were presented in a summary table.

An independent t-test was used to compare the crude means of each outcome variable at M1 for registered vs. NOT registered subjects. The results were presented graphically.

A univariate ANOVA was used to repeat this analysis adjusted for age, division, school and history of abduction. Significant findings were tabulated.

7.3h) Ethical approval

Ethical approval was obtained from the same entities as described for pilot testing in Chapter 5 (SECTION 5.3h).

7.4) Results

7.4a) Sample distribution

A total of 618 boys and 846 girls eligible for the GMKL were measured at M1. Of these young adolescents, there were 146 boys and 81 girls who registered for the GMKL [T+W]. This proportion was slightly lower than expected (boys ~23.6%, girls ~9.6%). Therefore, there were 472 boys and 765 girls who were measured at M1 and did not register for the GMKL [O]. A variable number of these subjects opted out of testing for each of the outcomes measured at M1 (FIGURES 7.2 & 7.3).

FIGURE 7.2: Flow diagram of the sample for the boys programme reach at M1 in Gulu

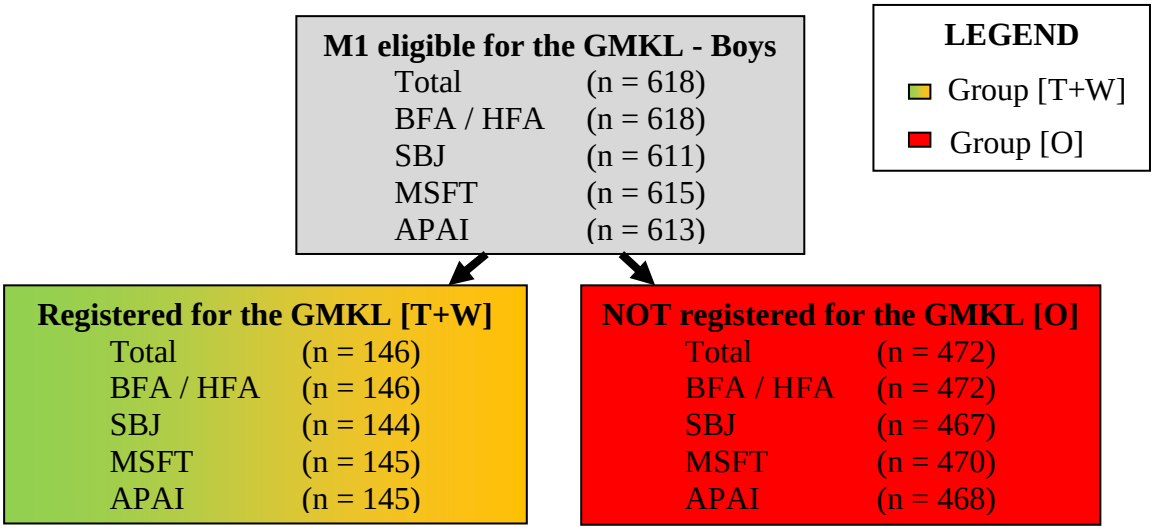
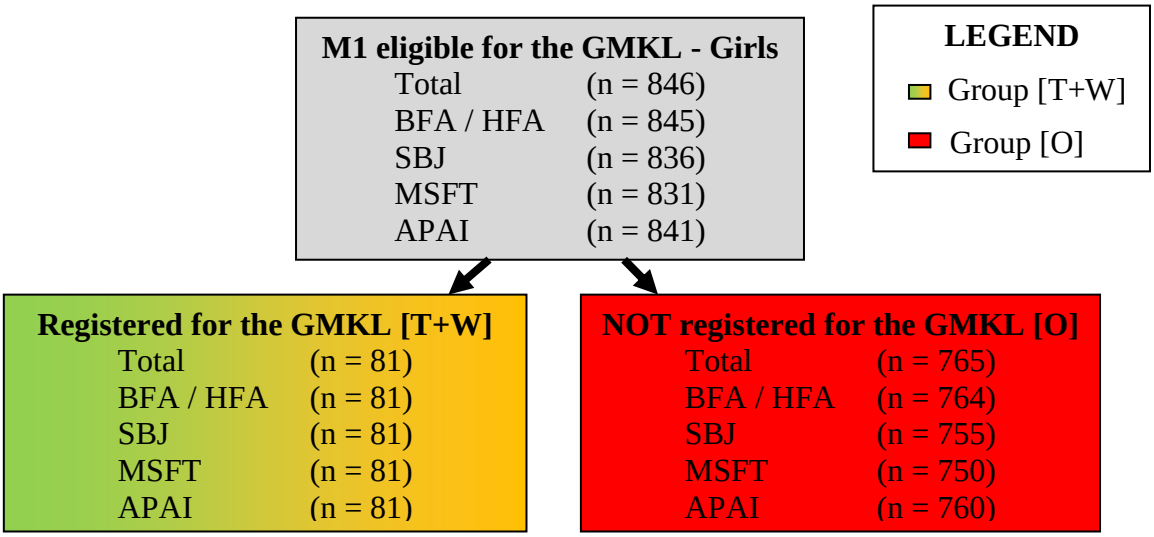


FIGURE 7.3: Flow diagram of the sample for the girls programme reach at M1 in Gulu



When considering the adolescents who were measured at M1, the number and proportion who registered for the GMKL were higher for boys than for girls (boys: n=146, 23.62%; girls: n=81, 9.57%). There was also variation in the number of subjects who registered for the GMKL according to age, division and school that followed different patterns between the genders (TABLE 7.2).

Although 11-year-old boys were the least represented age group at M1, of the adolescents who were measured, they were the most likely to register for the GMKL (boys: n=14, 31.82%). In contrast, 11-year-old girls measured at M1 had the lowest proportion who registered for the GMKL (girls: n=6, 6.53%). 13-year-old adolescents captured at M1 had the highest number of registrants for the GMKL in both genders (boys: n=63, 26.92%, girls: n=35, 11.95%). This was also the highest proportion in the girls sample (TABLE 7.2).

There were clear differences in the registration rates between the divisions when comparing the adolescents measured at M1 (TABLE 7.2). Laroo division had the lowest proportion of subjects measured at M1 who registered for the GMKL in both genders (boys: n=22, 17.89%; girls: n=5, 3.47%). Boys measured at M1 from Layibi had the highest proportion who registered for the GMKL (boys: n=49, 29.88%). Bardege was the least represented division for girls at M1, but had the highest proportion registered for the GMKL (girls: n=9, 15.79%).

A school-level variation in registration for the GMKL was also evident (TABLE 7.2). Although there were gender-based variations in the registration rates from different schools, GPub (boys: n=27, 37.50%; girls: n=29, 27.88%) and LL (boys: n=16, 48.48%; girls: n=12, 18.18%) had the highest proportion of participants from M1 in both genders. GPol was the largest school contributing to M1, but had the lowest overall registration rate for the GMKL (boys: n=12, 10.91%; girls: n=4, 2.60%).

A history of abduction did not appear to be a barrier to adolescents registering for the GMKL in either gender (TABLE 7.2). Of the subjects measured at M1, there was a higher proportion register for the GMKL who reported abduction (boys: n=25, 28.74%; girls: n=16, 10.88%) than those who were NOT abducted (boys: n=121, 22.79%; girls: n=65, 9.30%).

TABLE 7.2: Sample distribution for subjects measured at M1 who registered for the GMKL in Gulu - stratified by age, division, school and history of abduction

Independent variable		Total no. of boys (n_{Tb})	Boys registered for the GMKL		Total no. of girls (n_{Tg})	Girls registered for the GMKL	
			n	% of n_{Tb}		n	% of n_{Tg}
Age in years	11	44	14	31.82	92	6	6.52
	12	141	28	19.86	177	17	9.60
	13	234	63	26.92	293	35	11.95
	14	199	41	20.60	284	23	8.10
Division	Bardege	48	12	25.00	57	9	15.79
	Laroo	123	22	17.89	144	5	3.47
	Layibi	164	49	29.88	209	30	14.35
	Pece	283	63	22.26	436	37	8.49
School	CC	49	13	26.53	60	2	3.33
	GPol	110	12	10.91	154	4	2.60
	GPrs	71	25	35.21	98	5	5.10
	GPub	72	27	37.50	104	29	27.88
	GTwn	35	4	11.43	44	3	6.82
	HR	79	17	21.52	110	15	13.64
	LL	33	16	48.48	66	12	18.18
	LC	66	16	24.24	93	8	8.60
	PP	64	8	12.50	82	3	3.66
	SP	39	8	20.51	35	0	0.00
History of abduction	Yes	87	25	28.74	147	16	10.88
	No	531	121	22.79	699	65	9.30
Total		**618**	**146**	**23.62**	**846**	**81**	**9.57**

NOTE: n_{Tb} is the total number of boys who completed M1 for each of the sub-populations identified within the sample.

NOTE: n_{Tg} is the total number of girls who completed M1 for each of the sub-populations identified within the sample.

The difference in sample distribution for gender, age, division, school and history of abduction supported their inclusion as adjustment factors in the subsequent data analyses.

7.4b) Reach of Gum Marom Kids League (self-selection bias)

i) Descriptive statistics

The mean and 95% CI of all outcome variables at M1 were calculated for the subjects who were registered vs. NOT registered for the GMKL (TABLES 7.3 & 7.4).

TABLE 7.3: Crude mean and 95% CI of the PF outcomes for the Gulu sample at M1 - stratified by gender and registration for the GMKL

Registered for the GMKL		BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)
Boys (N=618)	Yes (n=146)	**-0.631** (-0.762↔-0.499) n=146	**-0.463** (-0.621↔-0.304) n=146	**185.90** (182.59↔189.21) n=144	**6.25** (5.94↔6.56) n=145
	No (n=472)	**-0.649** (-0.732↔-0.566) n=472	**-0.452** (-0.545↔-0.360) n=472	**181.80** (179.93↔183.67) n=467	**5.94** (5.76↔6.12) n=470
Girls (N=846)	Yes (n=81)	**-0.261** (-0.449↔-0.073) n=81	**0.172** (-0.015↔0.359) n=81	**166.53** (162.67↔170.39) n=81	**4.23** (3.83↔4.63) n=81
	No (n=765)	**-0.208** (-0.273↔-0.144) n=764	**-0.086** (-0.150↔-0.022) n=764	**163.42** (162.17↔164.67) n=755	**3.51** (3.39↔3.63) n=750

TABLE 7.4: Crude mean and 95% CI of the MH outcomes for the Gulu sample at M1 - stratified by gender and registration for the GMKL

Registered for the GMKL		TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Boys (N=618)	Yes (n=146)	**10.65** (9.65↔11.65) n=145	**8.09** (7.27↔8.91) n=145	**10.90** (9.94↔11.86) n=145	**8.57** (7.78↔9.36) n=145	**22.96** (20.93↔24.99) n=145	**29.20** (26.81↔31.59) n=145
	No (n=472)	**11.65** (11.09↔12.21) n=468	**8.82** (8.35↔9.29) n=468	**11.80** (11.25↔12.35) n=468	**8.77** (8.36↔9.18) n=468	**24.91** (23.80↔26.02) n=468	**31.35** (30.05↔32.65) n=468
Girls (N=846)	Yes (n=81)	**15.37** (13.88↔16.86) n=81	**11.86** (10.50↔13.22) n=81	**15.25** (13.63↔16.87) n=81	**10.35** (9.32↔11.38) n=81	**32.37** (29.18↔35.56) n=81	**39.95** (36.28↔43.62) n=81
	No (n=765)	**15.19** (14.68↔15.70) n=760	**11.72** (11.31↔12.13) n=760	**14.63** (14.15↔15.11) n=760	**10.13** (9.80↔10.46) n=760	**31.74** (30.75↔32.73) n=760	**39.22** (38.07↔40.37) n=760

ii) Statistical findings

I found that voluntary registration for the GMKL was affected by a self-selection bias for subjects who performed better in the SBJ and MSFT. The underlying assumptions for the analysis of all outcome variables for adolescents registered vs. NOT registered for the GMKL were satisfied (APPENDIX A17).

When analysing the crude means for the PF outcome variables, there was no significant difference at M1 between registered vs. NOT registered for boys BFA [p= 0.831], HFA [p=0.916] and MSFT [p=0.104]. Boys who registered for the GMKL jumped significantly further at M1 in the SBJ [p=0.036]. The girls who registered for the GMKL performed significantly better at M1 on the MSFT [p<0.001] and had a higher mean HFA [p=0.014]. There was no significant difference between the groups for girls BFA [p=0.619] and SBJ [p=0.130] (FIGURES 7.4 – 7.7).

FIGURE 7.4: Comparison of the crude mean and 95% CI of BFA for adolescents in the Gulu sample at M1 who were registered and NOT registered for the GMKL - stratified by gender

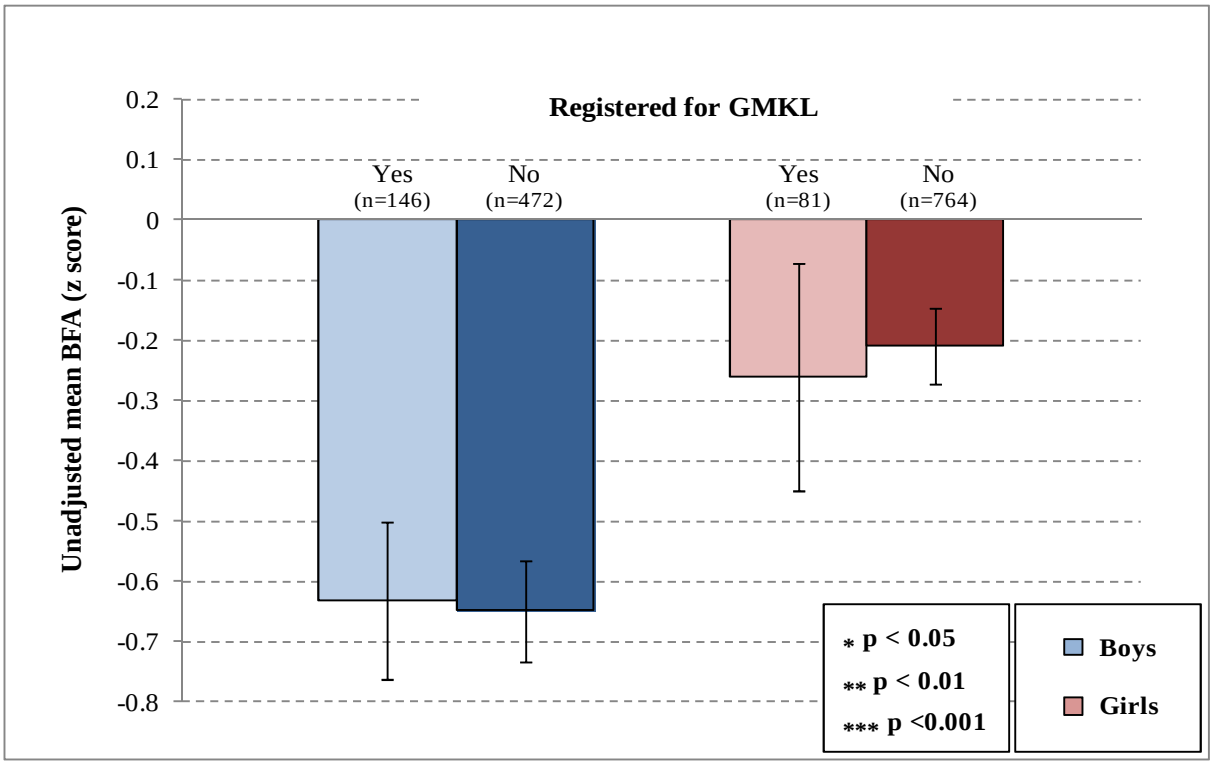


FIGURE 7.5: Comparison of the crude mean and 95% CI of HFA for adolescents in the Gulu sample at M1 who were registered and NOT registered for the GMKL - stratified by gender

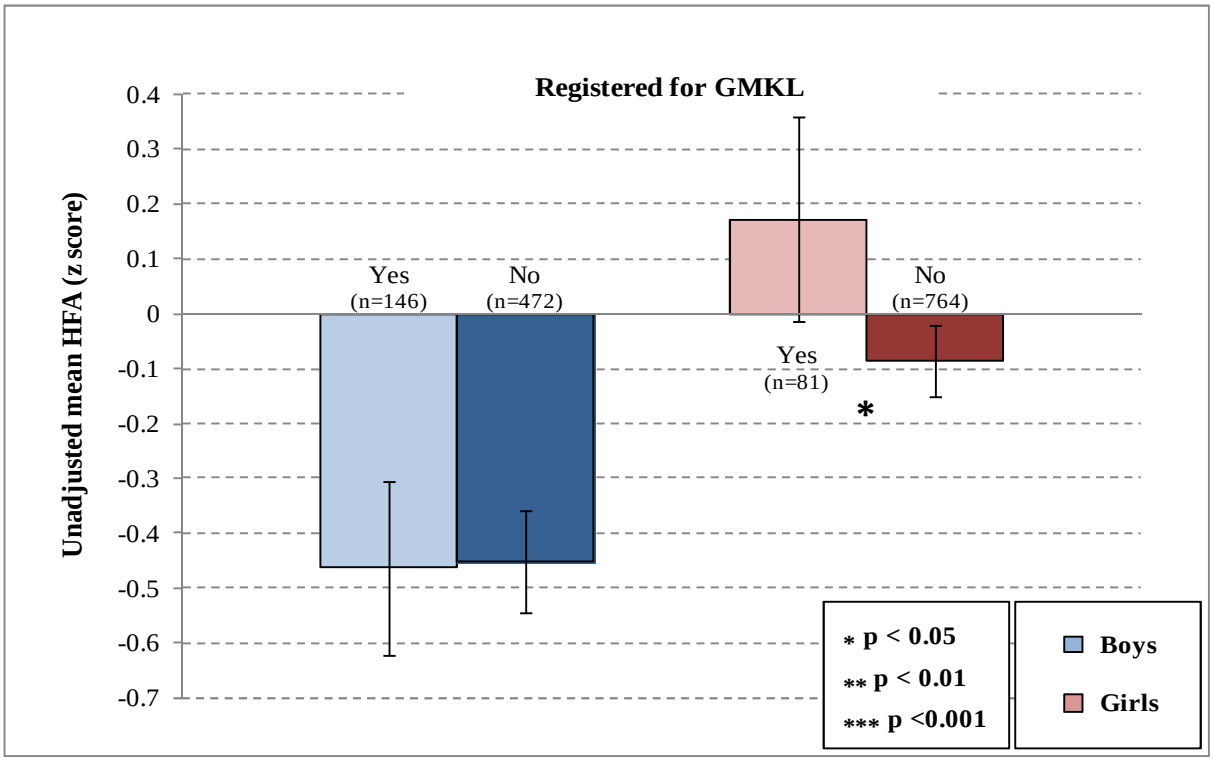


FIGURE 7.6: Comparison of the crude mean and 95% CI of the SBJ for adolescents in the Gulu sample at M1 who were registered and NOT registered for the GMKL - stratified by gender

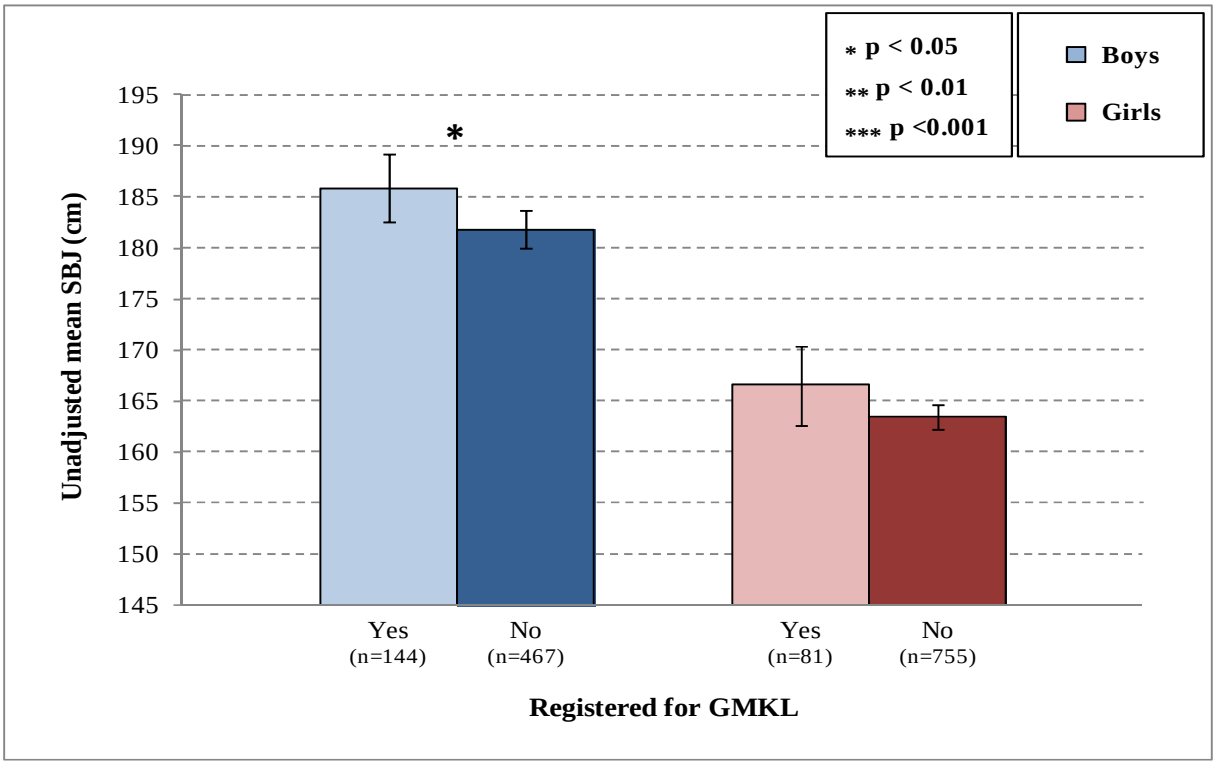
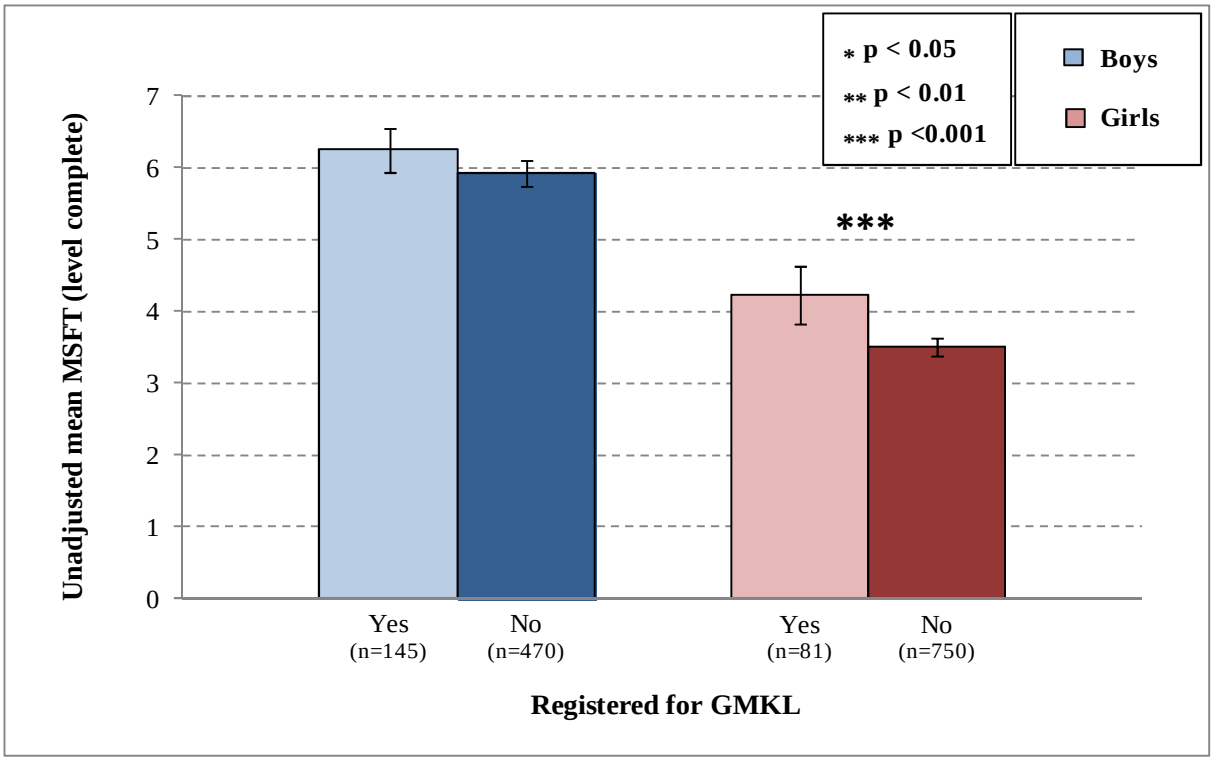


FIGURE 7.7: Comparison of the crude mean and 95% CI of the MSFT for adolescents in the Gulu sample at M1 who were registered and NOT registered for the GMKL - stratified by gender



There was no significant difference at M1 between the crude means of the boys who were registered vs. NOT registered for TT [p=0.090], K [p=0.134], P [p=0.119], ML [p=0.643], DT [p=0.095] and AT [p=0.118]. The girls sample demonstrated similar insignificant MH differences for TT [p=0.832], K [p=0.833], P [p=0.438], ML [p=0.686], DT [p=0.702] and AT [p=0.700] (FIGURES 7.8 & 7.9).

FIGURE 7.8: Comparison of the crude mean and 95% CI of the boys MH outcomes for adolescents in the Gulu sample at M1 who were registered and NOT registered for the GMKL

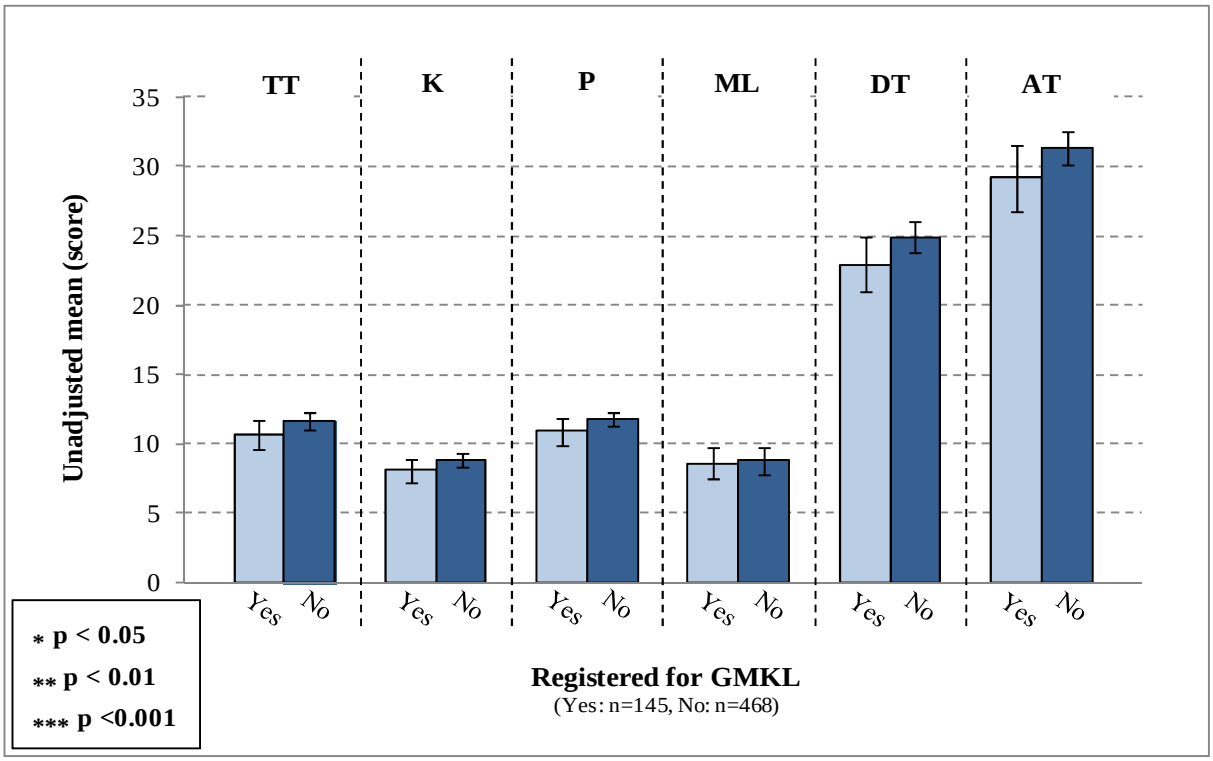
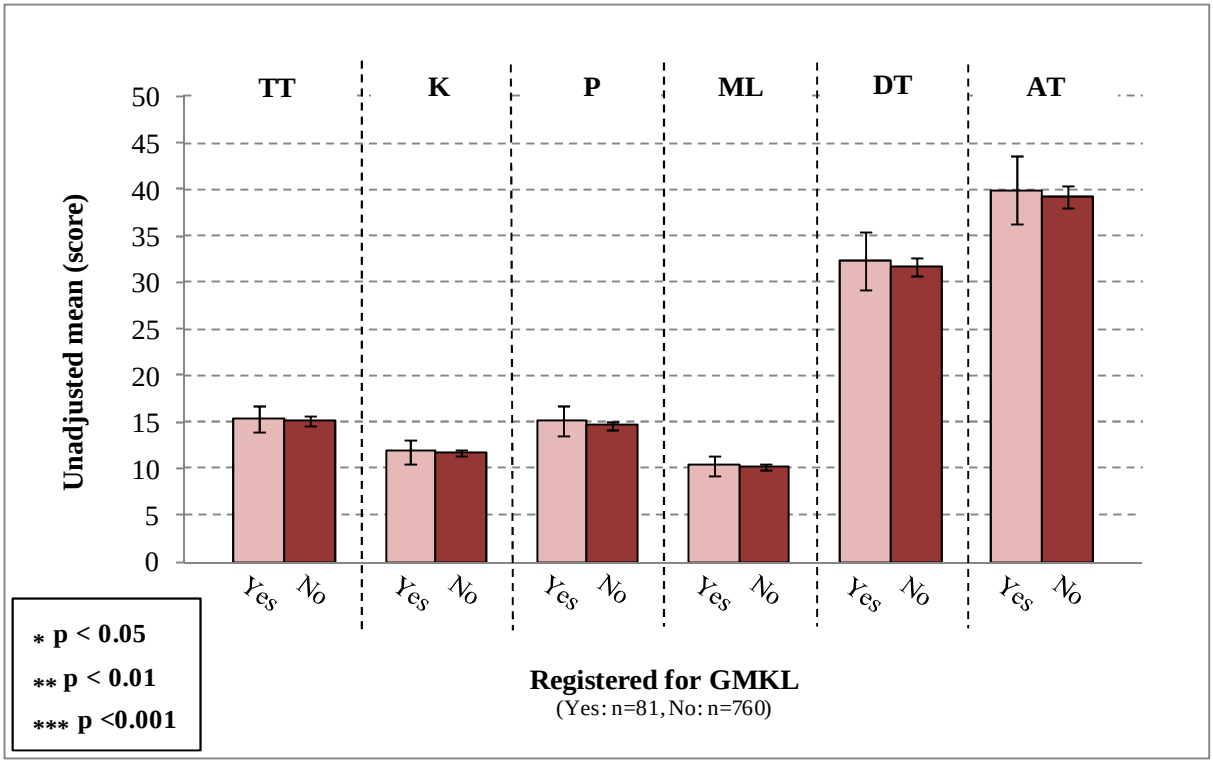


FIGURE 7.9: Comparison of the crude mean and 95% CI of the girls MH outcomes for adolescents in the Gulu sample at M1 who were registered and NOT registered for the GMKL



The statistical analyses were repeated after adjusting the data for age, division, school and history of abduction. The differences between the means for subjects registered vs. NOT registered remained insignificant for all of the MH outcomes for both the boys sample (TT [p=0.417], K [p=0.754], P [p=0.743], ML [p=0.424], DT [p=0.912], AT [p=0.856]) and the girls sample (TT [p=0.629], K [p=0.937], P [p=0.827], ML [p=0.587], DT [p=0.804], AT [p=0.803]).

When considering the PF outcomes for the adjusted data in the boys sample, the difference between the group means was no longer statistically significant for the SBJ [p=0.115]. It also remained insignificant for boys BFA [p=0.929], HFA [p=0.299] and MSFT [p=0.262]. The adjusted data for the girls sample demonstrated similar results to the crude analyses for BFA [p=0.753]. The statistically significant difference between the group means was strengthened for girls HFA [p=0.007] and was maintained at a highly significant level for the MSFT [p<0.001]. The adjusted data also indicated that girls who registered for the GMKL performed significantly better at M1 in the SBJ [p=0.042] (TABLE 7.5).

TABLE 7.5: Univariate ANOVA comparing the mean of the PF outcomes for registered and NOT registered adolescents in the Gulu sample at M1 - stratified by gender and adjusted for age, division, school and history of abduction

Gender	BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)
Boys (N=618)	F=0.008	F=1.080	F=2.499	F=1.264
Girls (N=846)	F=0.099	F=7.274 ******	F=4.158 *****	F=11.137 *******

NOTE: Means compared using a model III sum of squares analysis (unbalanced groups with no empty cells for the selected fixed factors).

NOTE: ***** Denotes a significant difference between registered vs. NOT registered adolescents (* p<0.05, ** p<0.01, *** p<0.001)

7.5) Preliminary discussion

7.5a) Principal findings

There were differences in the distribution of the sample in each registration group that varied according to gender, age, division and school. A history of abduction did not appear to be a barrier to participating in the GMKL and there were higher registration rates for the boys who had been in "the bush" when compared to their peers.

After adjusting for these differences, the girls reached by the GMKL were significantly taller and performed better in the SBJ and MSFT. There was no significant self-selection bias for boys registering for the GMKL. However, there was a trend suggesting that the boys who performed better in the SBJ and MSFT were more likely to register.

7.5b) Limitations and strengths

The current study used the same data and cross-sectional methods as described in Chapter 6. Consequently, it shared several of the same limitations and strengths (SECTION 6.5b). This included concerns about the measurement validity of the modified metrics and the unknown clinically relevant effect size for each of the outcomes [179]. However, there were some additional limitations associated with the registration process. These contributed to anomalies in the sample distribution between the registered vs. NOT registered groups and were considered when applying the statistical methods [179,199].

The high levels of school absenteeism and measurement of ineligible subjects reported in Chapter 6 also affected the sample size in the current study (SECTION 6.4a). This was compounded by lower levels of registration than expected. However, the sample size appeared to be adequate for the analyses in this study. The compromises in statistical power

caused by adjusting the data for potential confounders continued to be offset by oversampling [179,201].

The greatest strengths of the current study were related to its design and the use of locally adapted assessment methods. Embedding the measurements in the school system was the most effective way to collect data from a representative cross-section of the community to assess the reach of the GMKL. It also provided a reliable means for identifying and tracking subjects measured at M1 who registered for the programme.

7.5c) Mechanisms and implications

Articulating ***who*** is actually reached by SfD interventions is a common oversight in programme evaluations [135]. Implementing organisations often make inflated and broad claims based on factually correct data that are poorly represented [135,205]. This makes it difficult to compare the reach of the current study to previous evaluations of SfD programmes. Therefore, the results were interpreted with reference to my fieldwork observations and the existing evidence base for sport participation.

The current season of the GMKL reached 400 young adolescents (240 boys, 160 girls), 32 volunteer coaches, six volunteer referees and seven paid employees in Gulu municipality. Although it was only statistically significant for the girls sample, it appeared that those who were physically fitter were more likely to register for the programme. This self-selection bias was probably caused by the voluntary registration process [179,190]. Review level evidence suggests that competence is a key determinant of sport participation in young people [206,207]. Therefore, it is likely that adolescents voluntarily registering for a SfD intervention would already engage in more PA and have higher PF than their peers.

The disparities in subject recruitment at M1 described in Chapter 6 and the disproportionate distribution in the registration groups according to age, school and division also affected the reach of the programme. It is likely that the location and distance to the registration sites were influential on these factors (APPENDIX A2) [208]. This may have contributed to the high registration rates for students at GPub and LL as well as the poor recruitment levels from SP and GTwn. The exceptions to this observation could be explained by fieldwork observations of differences at the school level. GPol had an academic focus and conducted some Saturday morning classes that coincided with the registration period. PP hosted separate peace-building activities that were organised by another NGO in Gulu. This attracted many of the PP students and took place at the same time as the GMKL registration. Since most adolescents attended a school close to where they lived (SECTION 6.4a), these factors also explained the disparities in registration between divisions.

The stigma associated with a history of abduction did not appear to affect the proportion that voluntarily registered for the GMKL. Previous psychosocial interventions in Gulu specifically for abducted adolescents compromised their reintegration in society, but may have contributed to making them more willing to register for the GMKL [147,148]. Annan et al (2006 & 2008) warned against programmes that isolated previously abducted adolescents and these results indicated that targeted recruitment for this population may not be necessary for SfD interventions in Gulu [147,148]. However, it is not known whether there was a health difference between the previously abducted adolescents who were registered vs. NOT registered for the GMKL. There was an insufficient sample size to investigate this further.

The results of the current study indicated several ways to improve the reach of future GMKL interventions. Firstly, the implementing organisations should consider whether the current model is appropriate for their aims and the community needs. The “sport plus” approach is

the most suitable for PF and PA objectives and improvement in these outcomes may also mediate MH benefits [107,120,122,138]. The community needs analysis indicated a general deficit in aerobic capacity and persistent MH problems in adolescents who were abducted during the war (CHAPTER 6). Therefore, the "sport plus" approach was appropriate for these outcomes in the current study. However, the self-selection bias inherent to the voluntary community-based registration process meant that the GMKL was unlikely to engage those most in need.

It was also evident from the needs assessment that a SfD intervention aimed at PA and PF is unlikely to have an impact on MH in the target population. Consequently, the MH objectives may be better served by a "plus sport" approach [122,137]. This may also be the case for the peace-building objectives of the implementing partners. A community needs analysis and baseline measures for peace-building were not included in this thesis due to limited resources, expertise and explanatory theoretical models. Therefore, it is not possible to comment on whether the allure of sport effectively engaged those most in need of a peace-building intervention.

Despite this, there is sufficient evidence to suggest that a review of the registration process may be beneficial for the GMKL. Voluntary participation did not specifically target the population most at risk according to the programme objectives and/or the results of the community needs analysis. Therefore, the reach of the intervention was completely dependent on the attraction of playing sport and who was available to participate in the programme. Active promotion of the GMKL in response to the community needs analysis prior to registration may improve the representation of at-risk groups. This requires a clear understanding of the barriers to participation [206-208].

The distribution of the registrants suggested that future programmes should introduce more registration sites to ensure better accessibility for young adolescents in all divisions of the municipality. Gender equity also needs to be addressed and may require concerted efforts to build community support for girls playing sport. Schools should be included in the planning process to improve engagement and minimise schedule clashes. However, these are likely to be difficult tasks considering the infrastructure available in Gulu, local gender norms and challenges associated with planning and time management (CHAPTER 3).

The implications on future research of the limitations and strengths in the current study are similar to those described in Chapter 6 (SECTION 6.6c). This includes adjusting for age, school, division and history of abduction in the statistical analyses when using these data. Considering ***who*** was actually reached in the context of the local needs and programme objectives is also important when interpreting the results of the subsequent GMKL impact evaluation (CHAPTER 8).

7.6) Summary box

REACH OF THE GUM MAROM KIDS LEAGUE IN GULU

- The concerns about measurement validity and sample distribution identified in Chapters 5 and 6 persisted in the current study. Age, division, school and history of abduction were adjusted for in the statistical analysis.

- Voluntary registration for the GMKL resulted in a self-selection bias for subjects who performed better in the SBJ and MSFT. This was more pronounced in the girls sample.

- Division- and school-level differences in the registration proportions were caused by the location of the registration sites and specific activities at GPol and PP.

- A history of abduction did not appear to be a barrier to registering for the GMKL.

- Voluntary community-based registration for SfD interventions may not effectively reach those who have the greatest need if improving PF and PA is the objective.

- Future GMKL programmes in Gulu should modify recruitment methods to reach a more representative sample and address the needs of the community by:
  - Introducing more registration sites in all divisions of the municipality;
  - Engaging schools in the intervention planning process;
  - Building local support for the participation of girls in sport;
  - Promoting the programme to at-risk groups prior to registration.

- The subsequent impact evaluation for the GMKL should:
  - Consider ***who*** was actually reached by the intervention in the context of the local needs and the programme objectives;
  - Use statistical methods that adjust for disparities in the sample distribution.

CHAPTER 8

Impact of a Sport-for-Development Intervention on the Physical Fitness, Mental Health and Physical Activity of Young Adolescents in Gulu, Uganda

8.1) Purpose of chapter

The purpose of this chapter is to evaluate the impact of a SfD programme in Gulu. I will use the measurement protocols developed in Chapter 5 to conduct a longitudinal assessment of PF and MH in young adolescents. The data collected in Chapter 6 form the baseline measures and the initial group allocation for the study was described in Chapter 7. PA will also be measured in a sub-sample of the subjects to triangulate the results of the longitudinal analysis and assess the "dose" of the intervention. The results will be used to indicate the programme impact on the PF, MH and PA of the GMKL participants in comparison to their peers.

8.2) Introduction

The cross-cutting nature of sport means that it is sold as a "cure-all" in the development sector [106-110]. It is portrayed by many stakeholders as an intervention that has an essentially positive impact and is as simple as organising sports activities, playing educational games or providing equipment covered in messages that promote the development agenda [121,136,137]. However, often little consideration is given to the sustainability of these processes and the actual effect of the programme on the beneficiaries [70,71,121].

The paucity of evidence and theoretical frameworks for SfD interventions has contributed to a growing tension between implementing and funding partners (SECTIONS 2.9 & 2.10) [71,115,125]. Increasing competition for funds and general scepticism towards the NGO sector has led to a demand for greater accountability in programme delivery [115,125,205]. In response, the role of M&E for SfD interventions has been forced to evolve from the “something is better than nothing” approach that has previously pervaded humanitarian and development organisations [137,209]. It is now widely accepted that the evaluation of programme impact in the SfD sector is in need of significant overhaul [120,135-137].

However, M&E in most of the SfD sector remains rudimentary. There appears to be limited scope for independent evaluation that assesses programme effectiveness [135,136]. Existing reports from SfD organisations tend to reinforce the positive rhetoric that is evident throughout the sector and are often aimed at securing further funding [70,120,125]. There is typically a focus on overly ambitious objectives and outcomes to satisfy the demands of donors rather than addressing the self-determined aims of the community and implementing organisations [115,121,125,135,136]. This is often portrayed through narrative stories of success accompanied by photo opportunities, which capture smiling and active children for promotional reasons [137].

Academic practitioners conduct research that has a practical focus and are ideally placed to contribute to introducing scientific rigour to M&E design and outcome measurement [205]. Researchers can contribute to programme development by establishing true causal relationships using methods that adequately assess the quality of intervention delivery and the subsequent level of impact [205]. However, it is important to recognise the distinctions between the perspectives of the scientists and practitioners [2].

The impact evaluation of the GMKL intervention in this thesis uses scientific methods to assess the effectiveness of the programme, but does not inform the implementers about the process of delivery [2]. Achieving the optimal level of evidence is balanced with the pragmatic and ethical constraints of assessing the programme outcomes [205]. This is informed by my fieldtrip experiences in Gulu and extensive planning with the organisations delivering the intervention. It also takes into account the results of the measurement method pretesting (CHAPTER 5) and the community needs analysis (CHAPTER 6).

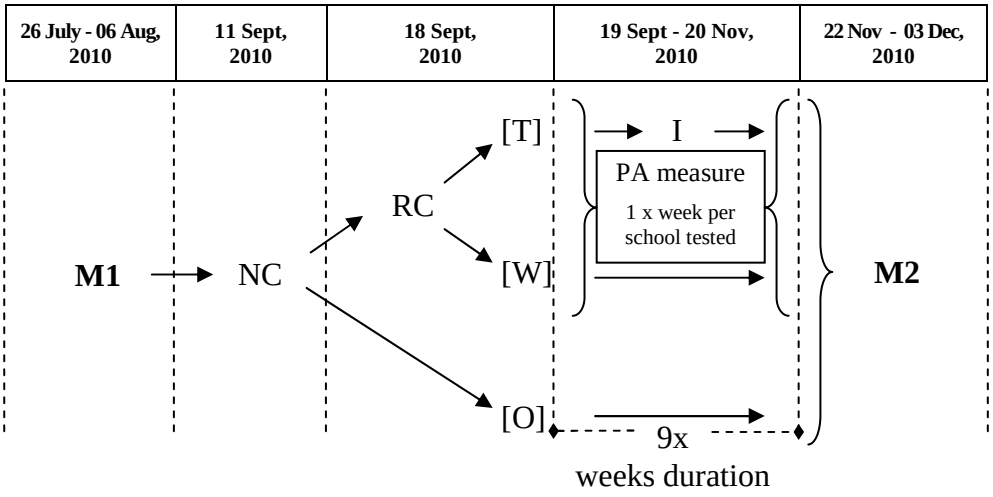
In this chapter, I work with the organisations delivering the GMKL in Gulu to design an impact evaluation that is feasible and causes minimal disruption to programme delivery. I use metrics that are adapted to the resources and capacity available in Gulu. Stringent academic and ethical standards are observed with respect for Acholi cultural norms and ongoing local development. In doing this, I endeavour to make a meaningful contribution to the evidence base for the impact of SfD programmes implemented in post-conflict environments within developing countries.

8.3) Methods

8.3a) Study design

The study was designed as an impact evaluation of the GMKL (I). It built on the data collected at M1 (CHAPTER 6) and the subsequent intervention registration process (CHAPTER 7). It comprised an RCT nested within a cohort study and triangulated with a cross-sectional study. These three components were methodologically interdependent (FIGURE 8.1).

FIGURE 8.1: Diagram representing subject recruitment and study design



LEGEND:
Subject recruitment

NC	Adolescents register for the GMKL (NOTE: non-random)
RC	Randomly select the GMKL participants from all adolescents who registered
[T]	Eligible adolescents participating in the GMKL – intervention group
[W]	Eligible adolescents wait-listed in the GMKL – wait-list group (NOTE: Boys only)
[O]	Eligible adolescents NOT registered in the GMKL

Study design and data collection

M1	Initial case-study measurement (PF, MH)
M2	Follow-up measurement immediately after the GMKL finishes (PF, MH)
I	Intervention (i.e. the GMKL season)

The GMKL was an 11-week SfD programme coordinated by the RCI and delivered as a joint venture between local and international NGOs. The GMKL model was a fusion of the programmes previously delivered by the funding and implementing partners in Gulu. Several facets of the intervention evolved during its development and implementation that were in response to unexpected logistical and/or environmental interruptions (e.g. weather, demands of programme staff and volunteers, local time management and planning, funding provision). The GMKL model described was the final programme that was actually delivered and evaluated (APPENDIX A16).

Prior to the intervention evaluated in this case-study, there had been no community-based SfD league for adolescents aged 11-14 years delivered in Gulu for a period of 15 months. Therefore, the assessment of PF and MH at M1 was taken as a true baseline for the RCT and cohort study (FIGURE 8.1).

The over-subscription of boys for the GMKL programme presented an experimental opportunity. The boys who registered for the GMKL were randomly allocated (RC) into an intervention [T] or wait-list [W] group. The randomisation procedures in this study were computer-generated using Microsoft Excel number allocation. All of the girls who registered were included in the intervention since their number was approximately equal to the places available in the GMKL programme. These girls formed group [T]. The subjects measured at M1 who did NOT register for the GMKL formed group [O].

After implementing the intervention, repeat measures (M2) were taken in each of the 10 target schools previously identified (SECTION 6.3c). Tracking the subjects from M1 to M2 in each of the groups enabled a within-group comparison of the outcome variables over time.

The cohort study contrasted groups [T] vs. [W+O] in the boys sample and groups [T] vs. [O] in the girls sample. The RCT compared the boys intervention [T] vs. wait-list [W] groups. The "disappointment effect" of being wait-listed could also be assessed by analysing the difference between the boys in groups [W] vs. [O].

The purpose of collecting PA data was to provide cross-sectional validation for the impact of the GMKL intervention and to assess its "dose". It was conducted in a sub-sample of the boys in groups [T] and [W]. A group of seven subjects were assessed each week during the intervention. PA measurement commenced on Friday afternoon and was completed one week later on the following Friday morning. This enabled six full days of measurement for the boys from each of the schools included in the PA assessment (FIGURE 8.1).

8.3b) Sample size calculation

The sample size (n) for the PA cross-sectional study was dictated by the duration of the intervention and the capacity to collect seven accelerometer measurements per week.

The sample size calculation for comparing two groups that was initially described in Chapter 6 was relevant to the current study (SECTION 6.3b) [179,195,196]. However, the RCI and RTL estimated that there would be an additional 5% loss to follow-up due to school absenteeism and/or opt out of measurement at M2. The sample calculations for the number of subjects in each group were adjusted proportionally (boys: n = 109.28 (110), girls: n = 102.41 (103)) [195].

In the current study, the within-group comparisons required half the sample size of that when comparing independent groups [196,197]. Therefore, the required sample size for these analyses was adjusted accordingly (TABLE 8.1).

The between-group analyses were adjusted for an expected uneven sample size. Based on previous experience, the GMKL coordinators estimated that twice as many boys as places available in the league would register. Therefore, it was expected that the RCT would divide the registered boys approximately evenly between groups [T] vs. [W] and they would each comprise $^{1}/_{8}$ of the boys in the community. The GMKL coordinators also expected that the $^{1}/_{10}$ registration rate for girls in the community would be insufficient to establish a control group. These ratios were used to adjust the required sample sizes for each of the groups in the RCT and cohort study comparisons (TABLE 8.1) [195].

TABLE 8.1: Calculated minimum sample sizes for the statistical analyses in the impact evaluation of the GMKL intervention in Gulu – stratified by gender

Statistical analysis	Gender	Required sample size (n)		
		Group 1	Group 2	Total
Within-group comparisons	Boys	Measures at M1 and M2: 54.64 (55)		**55**
	Girls	Measures at M1 and M2: 51.21 (52)		**52**
Between-group comparison - cohort study	Boys	[T]: 61.47 (62)	[W+O]: 491.76 (492)	**554**
	Girls	[T]: 56.33 (57)	[W+O]: 563.23 (564)	**621**
Between-group comparison - GMKL impact (RCT)	Boys	[T]: 109.28 (110)	[W]: 109.28 (110)	**220**
	Girls	NA	NA	NA
Between-group comparison - "disappointment effect"	Boys	[W]: 61.47 (62)	[O]: 491.76 (492)	**616**
	Girls	NA	NA	NA

NOTE: When calculating the required total n for the "disappointment effect" the number of subjects who would be allocated to group [T] was also taken into account

It is important to note that the sample size calculations were completed before M1 and GMKL registration (SECTION 6.3b). The results were considered along with the number of pupils enrolled in P6 when deciding how many schools were assessed at M1 (SECTION 6.3c).

8.3c) Subject recruitment

All subjects eligible for the GMKL and tested at M1 in Chapter 6 were recruited into the longitudinal studies assessing PF and MH (SECTION 6.3c).

Boys who registered for the GMKL and were measured at M1 were eligible for PA assessment. Target schools were allocated one week of PA measurement for every 10 eligible boys in attendance (FIGURE 8.1). This was to ensure the capacity to measure seven subjects per week was fulfilled in schools where attendance rates were as low as 70% at M1.

The order in which schools undertook PA measurement was generated randomly. For convenience, consecutive weeks were used in schools that underwent more than one week of PA assessment (TABLE 8.2).

Separate lists for groups [T] and [W] were generated at each school. The boys on these lists were ordered randomly. Every week PA measurements were conducted for the first three subjects on each of the group lists at the selected school. The seventh measure each week was randomly allocated to the next boy on the list in either group [T] or [W]. Boys who were absent from school at the time when PA measurement commenced were replaced by the next subject on the list who was in attendance (TABLE 8.2).

TABLE 8.2: Randomly selected order for PA measurement in the eligible target schools in Gulu and the number of boys tested per group

School	No. registered for the GMKL	Week for PA measurement	No. of PA measures	
			Group [T]	Group [W]
CC	13	3	3	4
GPol	12	7	4	3
GPrs	25	8 & 9	7	7
GPub	27	4 & 5	7	7
GTwn	4	-	0	0
HR	17	6	3	4
LL	16	1	4	3
LC	16	2	4	3
PP	8	-	0	0
SP	8	-	0	0
Total	**146**		**32**	**31**

8.3d) Subject identification and tracking procedures

The procedures for subject identification and tracking comprised collecting the same information at M1, NC and M2 as described for pilot testing in Chapter 5 (SECTION 5.3d). This information was also used to minimise sample contamination of the subjects in groups [T], [W] and [O] throughout the intervention period.

At M2 the subjects were also asked if they registered for the GMKL and if so, whether they participated in the programme or were wait-listed. These responses were cross-checked with the details described above. In cases of ambiguity, the class teachers assisted the assessors and the subject was approached within the school setting for clarification.

8.3e) Outcome variables and metrics

The outcome variables and metrics comprised the same tools and techniques described for PF, PA and MH in the pilot testing in Chapter 5 (SECTION 5.3e). However, HFA was not included in the current study because it is a measure of chronic malnourishment and it was not expected to change as a result of the SfD intervention. When calculating the age of the subjects for BFA at M2, it was adjusted to be 3 months older than at M1 (i.e. M = [Yx12]+9).

8.3f) Data collection procedures

The data collection at M2 followed the same procedures as described for the assessment at M1 in Chapter 6 (SECTION 6.3f). For the PA measurements, the assessors collected the accelerometers from the subjects each Friday and downloaded the data stored from the preceding week using the ActiLife Software v5.5. The accelerometers were then charged and calibrated before they were redistributed to a new group of boys in the afternoon.

8.3g) Data analysis and presentation

All of the quantitative data were cleaned and checked for outliers prior to commencing analysis. The underlying assumptions for each statistical test were also verified before proceeding. The reporting of descriptive statistics for each of the variables included the crude mean, 95% CI and sample size (n). Only subjects who completed both M1 and M2 for the outcomes of interest were included in the analysis and the level of programme uptake was not accounted for in the statistical tests.

i) Sample distribution

The sample of eligible adolescents who completed M1 and M2 was described according to group allocation for the GMKL. The results were stratified according to gender, age, division, school and history of abduction. The absolute number of subjects and the proportional representation at each level of the stratified data were tabulated to identify disparities in the distribution of the sample in each experimental group.

ii) Impact of the Gum Marom Kids League (cohort study)

Descriptive statistics at M1 and M2 were calculated for all outcome variables stratified according to gender and cohort study group (i.e. boys: [T] vs. [W+O]; girls: [T] vs. [O]). The results were summarised and tabulated.

A paired t-test was used to assess the within-group changes between M1 and M2 for the crude means of all the outcome variables (i.e. boys: M1 vs. M2 for [T] and [W+O]; girls: M1 vs. M2 for [T] and [O]). The results were presented graphically.

A repeated measures ANOVA was used to compare the change (Δ) from M1 to M2 in the group means for those in the intervention vs. control (i.e. boys: Δ[T] vs. Δ[W+O]; girls: Δ[T] vs. Δ[O]). The analysis was initially completed using the crude data for each outcome. It

was then repeated adjusting for baseline values (Model 1). The analysis was then repeated again adjusting for baseline, age, division, school and history of abduction (Model 2). The results for all of the between-group analyses that assessed the intervention effect in the cohort study were summarised and tabulated.

iii) Impact of the Gum Marom Kids League (randomised control trial)

The data analysis and presentation for the RCT followed the same procedures as the cohort study (SECTION 8.3gii). However, the sample was only boys and the descriptive statistics were stratified according to intervention group (i.e. boys: [T] vs. [W] vs. [O]). All of the within-group changes in crude mean were examined (i.e. boys: M1 vs. M2 for [T], [W], [O]). The assessment of the between-group differences focused on the impact of the intervention compared to being wait-listed (i.e. boys: Δ[T] vs. Δ[W]) and the "disappointment effect" (i.e. boys: Δ[W] vs. Δ[O]).

iv) Impact of the Gum Marom Kids League (accelerometry)

The validation of accelerometer wear time followed the same procedure as described in Chapter 5 (SECTION 5.3giii). However, subjects were only included in the analysis if they had valid accelerometer wear time on the GMKL activity day on the weekend.

The accelerometry data were interpreted using a .NET programme developed by Dr. Aiden Doherty at the University of Oxford. This translated and summarised the accelerometer counts according to the PA intensity values described by Freedson (1998) [210]. The number of minutes doing MVPA and VPA were reported for each of the participants who had valid measures (SECTION 5.3giii). Three time frames for MVPA and VPA were considered:

- Daily average on valid days;
- Daily average during the GMKL training times (4-7pm);
- Total on the GMKL activity day (i.e. game day).

The MVPA and VPA were also presented graphically for each subject using "heat maps" depicting the entire week of wear [211]. They were used to attempt to identify "doses" of the GMKL intervention. The "heat maps" captured bouts of MVPA and VPA that were at least of 10 minutes duration. Higher intensity PA was presented in a darker shade. In the absence of bout length guidelines for adolescents, a 10-minute duration was chosen because recent studies indicated that this may influence the benefit for adiposity and bone health [26,58,59]. It also matched the evidence base and the WHO recommendations for adult accumulation of PA [17]. Up to 2 minutes of "down time" below the relevant threshold was accepted within each 10-minute bout to allow for the sporadic nature that PA is often accrued by adolescents [26,212].

Descriptive statistics for each time frame stratified according to group [T] vs. [W] were calculated. The results were presented in a summary table.

The daily MVPA and frequency of VPA bouts were assessed for each subject to compare the activity levels of both groups against the WHO recommendations (SECTION 2.6b) [17].

The patterns of MVPA and VPA bouts were also observed and matched with the GMKL schedule to assess if a "dose" of the intervention was discernible in group [T].

Independent t-tests were used to compare the crude means for MVPA and VPA of groups [T] vs. [W] for each time frame. The results were presented graphically.

8.3h) Ethical approval

Ethical approval was obtained from the same entities as described for pilot testing in Chapter 5 (SECTION 5.3h).

8.4) Results

8.4a) Sample distribution

Of the boys measured at M1 and registered for the GMKL, 74 were randomly allocated to group [T] and 72 were in the wait-list [W] group. The girls intervention [T] group originally comprised all 81 subjects who registered for the programme. There were an insufficient number of girls to create a wait-list [W]. Group [O] had 472 boys and 765 girls who were measured at M1. A variable number of these subjects opted out of testing for each of the outcomes assessed (FIGURES 8.2 & 8.3).

After the intervention, 757 boys and 969 girls were measured at M2 in the target schools. 73 boys and 79 girls in group [T] were measured at both M1 and M2. Repeat measures were also collected for 71 wait-listed boys. Group [O] had 450 boys and 727 girls measured at both M1 and M2. The overall proportional loss to follow-up was lower than expected (boys ~3.9%, girls ~4.7%). However, it was higher for the PF outcomes than for the MH assessment. The total number of subjects included from each group in the final analysis for each gender is represented by "N" (FIGURES 8.2 & 8.3). A variable number of subjects were lost to follow-up for each outcome variable because they either opted out of assessment or were absent from school on the day of M2 testing (FIGURES 8.2 & 8.3).

A total of 32 boys in group [T] and 31 boys in group [W] were randomly selected for PA assessment. Of these, 30 valid measures were accumulated from group [T] and 24 were collected from group [W]. Six of the invalid measures were caused by an accelerometer that became faulty after week 3 of the intervention (5x [W], 1x [T]). The remaining three invalid measures were due to insufficient wear time (FIGURE 8.2).

FIGURE 8.2: Flow diagram of the sample for the boys longitudinal and accelerometer studies in Gulu

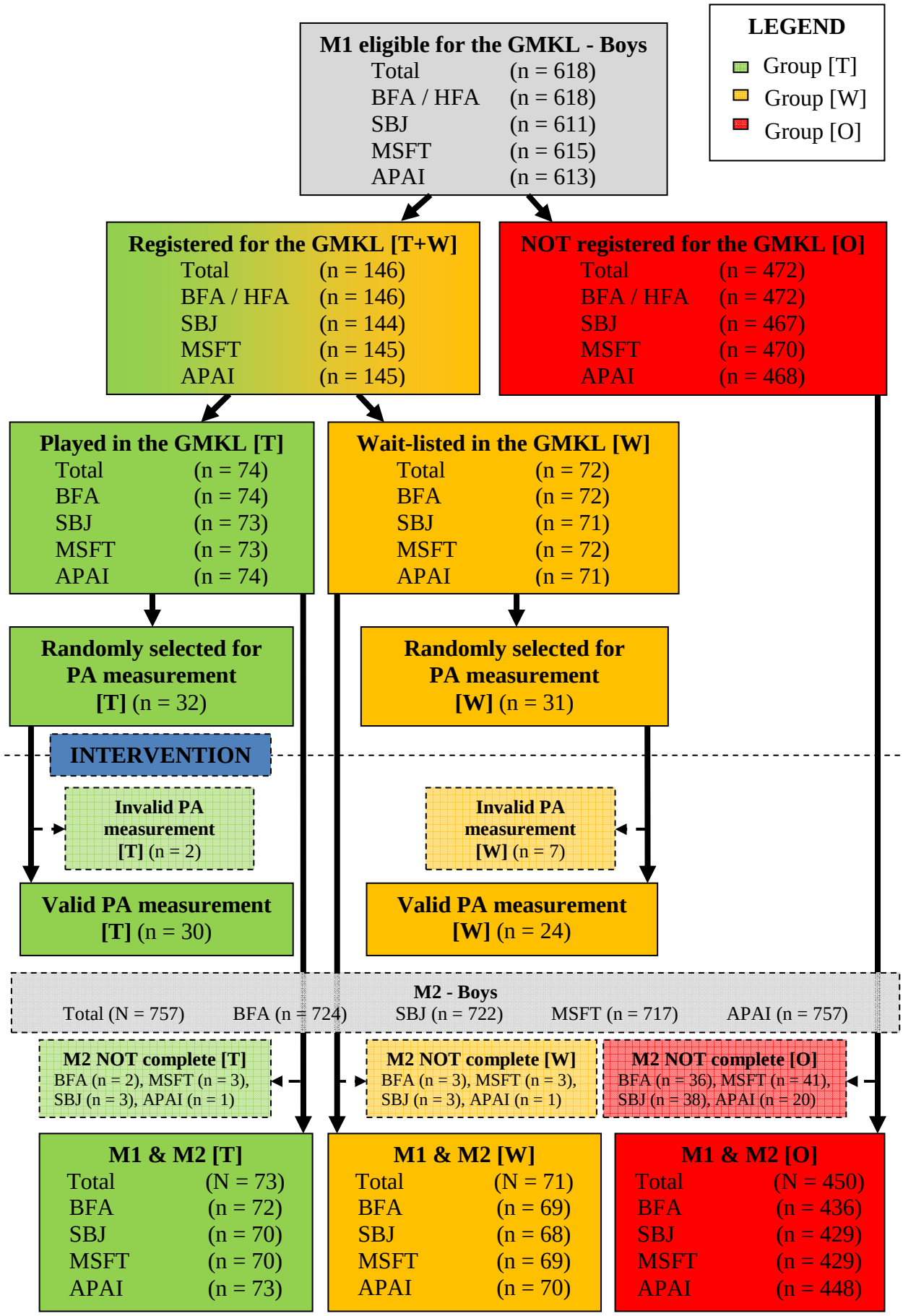
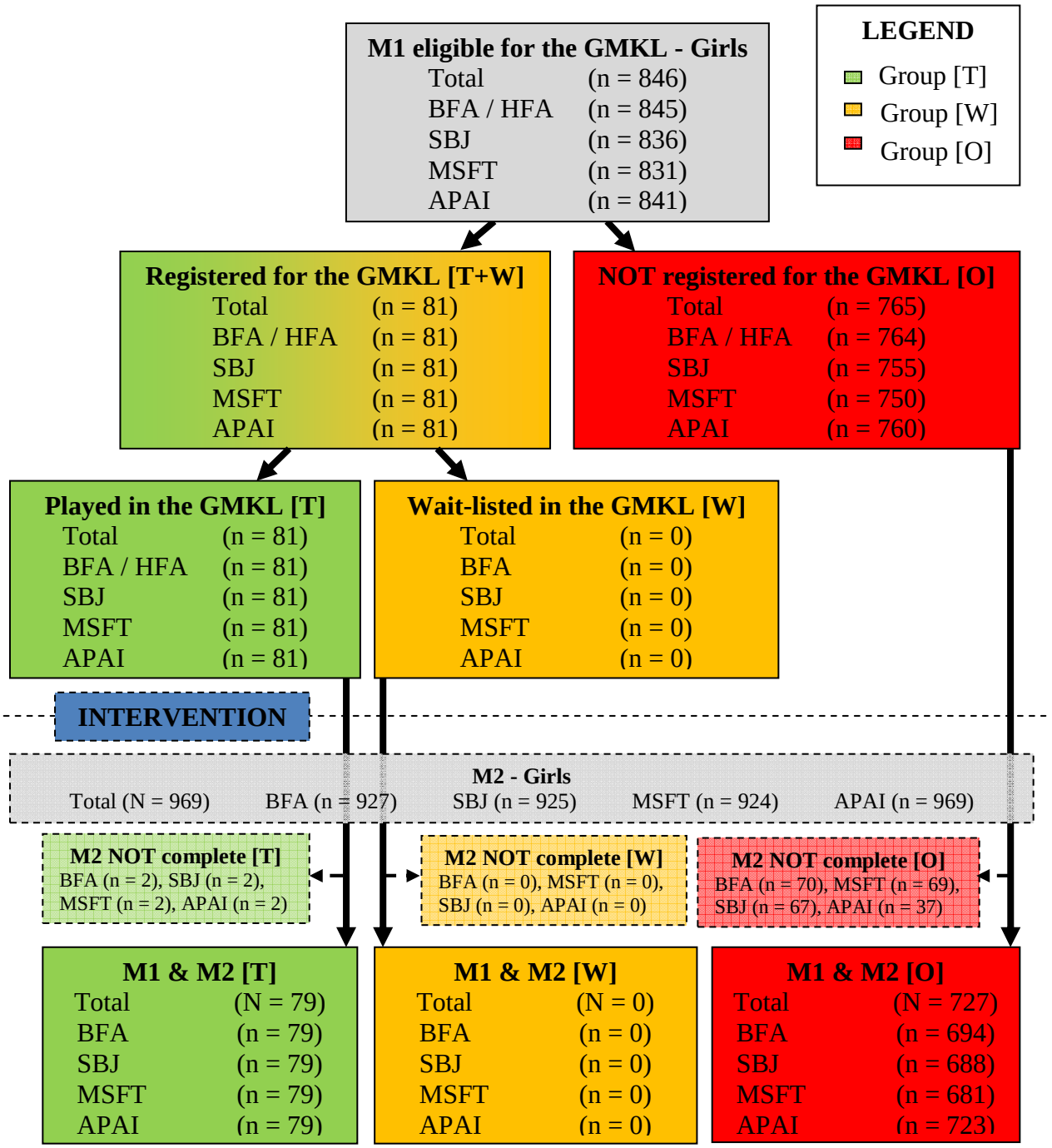


FIGURE 8.3: Flow diagram of the sample for the girls longitudinal study in Gulu



The process of random allocation was successful in separating the boys registered for the GMKL relatively evenly between groups [T] vs. [W] according to age, school and history of abduction. However, there was some disparity between the intervention and wait-list groups for boys from different divisions (TABLE 8.3). Of particular note was the discrepancy in the number of boys in groups [T] vs. [W] from Laroo ([T]: n=14, 11.67%; [W]: n=8, 6.67%).

The low number of boys lost to follow-up at M2 and the successful random allocation process meant that the proportion of boys measured in groups [T] and [W] broadly followed the same trends as the GMKL registrants described previously (SECTION 7.4a). This was most evident in the persistent school-level variation of the proportion of boys in the intervention and wait-list groups (TABLE 8.3). However, the differences in the proportion of boys from each age group and division were attenuated in group [T] and became relatively balanced. Similar changes were noted for the difference between the proportions of boys abducted vs. NOT abducted in the intervention group. Conversely, the proportional differences between levels for the same independent variables in group [W] were extenuated and became more unbalanced (TABLE 8.3).

Group [W+O] was the control group in the cohort study for boys. Therefore, its sample distribution reciprocated the patterns described for group [T]. Proportional differences were evident between schools. The age, division and abduction history of the boys in group [W+O] were represented by relatively consistent proportions across all levels of these independent variables (TABLE 8.3). Since there were only two groups in the boys cohort study, sub-populations that were over-represented in group [T] were reciprocally under-represented in group [W+O].

Due to the limited loss to follow-up, the sample distribution of boys in group [O] was similar to that of the boys who did NOT register for the GMKL. This is represented by the reciprocal of the proportions described for the male registrants for the GMKL (SECTION 7.4a). There were age-, division- and school-level differences in the proportion of boys in group [O] (TABLE 8.3).

TABLE 8.3: Sample distribution according to intervention group for the boys in Gulu measured at M1 and M2 – stratified by age, division, school and history of abduction

Independent variable		Total no. of boys (n_T)	Group [T]		Group [W]		Group [O]		Group [W+O]	
			n	% of n_T	n	% of n_T	n	% of n_T	n	% of n_T
Age in years	11	44	6	13.64	8	18.18	30	68.18	38	86.36
	12	137	15	10.95	13	9.49	109	79.56	122	89.05
	13	229	33	14.41	29	12.66	167	72.93	196	85.59
	14	184	19	10.33	21	11.41	144	78.26	165	89.67
Division	Bardege	46	7	15.22	5	10.87	34	73.91	39	84.78
	Laroo	120	14	11.67	8	6.67	98	81.67	106	88.33
	Layibi	152	21	13.82	26	17.11	105	69.08	131	86.18
	Pece	276	31	11.23	32	11.59	213	77.17	245	88.77
School	CC	48	6	12.50	7	14.58	35	72.92	42	87.50
	GPol	108	6	5.56	6	5.56	96	88.89	102	94.44
	GPrs	67	13	19.40	11	16.42	43	64.18	54	80.60
	GPub	69	14	20.29	13	18.84	42	60.87	55	79.71
	GTwn	33	2	6.06	2	6.06	29	87.88	31	93.94
	HR	72	8	11.11	9	12.50	55	76.39	64	88.89
	LL	33	9	27.27	7	21.21	17	51.52	24	72.73
	LC	63	7	11.11	8	12.70	48	76.19	56	88.89
	PP	62	3	4.84	5	8.06	54	87.10	59	95.16
	SP	39	5	12.82	3	7.69	31	79.49	34	87.18
History of abduction	Yes	81	11	13.58	12	14.81	58	71.60	70	86.42
	No	513	62	12.09	59	11.50	392	76.41	451	87.91
Total		**594**	**73**	**12.29**	**71**	**11.95**	**450**	**75.76**	**521**	**87.71**

NOTE: n_T is the total number of boys who completed M1 and M2 for each of the sub-populations identified within the sample.

The girls study did not have a wait-list for the intervention. Therefore, the distribution in the girls sample for group [T] was similar to that observed in the female registrants for the GMKL (SECTION 7.4a). There were small deviations from the proportions described above as a result of loss to follow-up at M2. However, there was no difference in the overall trends observed for distribution of the sample at the level of age, division, school and history of abduction for group [T] in the girls sample (TABLE 8.4). Since there were only two groups,

sub-populations that were over-represented in group [T] were reciprocally under-represented in the control group [O].

TABLE 8.4: Sample distribution according to intervention group for the girls in Gulu measured at M1 and M2 – stratified by age, division, school and history of abduction

Independent variable		Total no. of girls (n_T)	Group [T]		Group [O]	
			n	% of n_T	n	% of n_T
Age in years	11	90	6	6.67	84	93.33
	12	173	16	9.25	157	90.75
	13	282	35	12.41	247	87.59
	14	261	22	8.43	239	91.57
Division	Bardege	54	9	16.67	45	83.33
	Laroo	139	5	3.60	134	96.40
	Layibi	198	29	14.65	169	85.35
	Pece	415	36	8.67	379	91.33
School	CC	60	2	3.33	58	96.67
	GPol	152	4	2.63	148	97.37
	GPrs	87	4	4.60	83	95.40
	GPub	101	29	28.71	72	71.29
	GTwn	42	3	7.14	39	92.86
	HR	106	15	14.15	91	85.85
	LL	62	11	17.74	51	82.26
	LC	87	8	9.20	79	90.80
	PP	76	3	3.95	73	96.05
	SP	33	0	0.00	33	100.00
History of abduction	Yes	141	15	10.64	126	89.36
	No	665	64	9.62	601	90.38
Total		**806**	**79**	**9.80**	**727**	**90.20**

NOTE: n_T is the total number of girls who completed M1 and M2 for each of the sub-populations identified within the sample.

The difference in sample distribution for gender, age, division, school and history of abduction supported their inclusion as adjustment factors in the subsequent data analysis.

8.4b) Impact of the Gum Marom Kids League (cohort study)

i) Descriptive statistics

The mean and 95% CI for all outcome variables were calculated for the boys in groups [T] and [W+O] (TABLES 8.5 & 8.6).

TABLE 8.5: Crude mean and 95% CI of the PF outcomes for the boys in the Gulu sample at M1 and M2 - stratified by time and cohort study group

Group	Time	BFA (z score)	SBJ (cm)	MSFT (level complete)
[T] (n=73)	M1	**-0.663** (-0.849↔-0.477)	**187.89** (182.93↔192.85)	**6.34** (5.89↔6.79)
	M2	**-0.759** (-0.954↔-0.565)	**181.67** (176.63↔186.71)	**7.00** (6.56↔7.44)
		n=72	n=70	n=70
[W+O] (n=521)	M1	**-0.674** (-0.751↔-0.597)	**181.41** (179.63↔183.19)	**5.95** (5.78↔6.12)
	M2	**-0.687** (-0.765↔-0.610)	**181.63** (179.72↔183.54)	**6.70** (6.52↔6.88)
		n=505	n=497	n=498

TABLE 8.6: Crude mean and 95% CI of the MH outcomes for the boys in the Gulu sample at M1 and M2 - stratified by time and cohort study group

Group	Time	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
[T] (n=73)	M1	**9.95** (8.61↔11.29)	**7.49** (6.35↔8.63)	**10.03** (8.74↔11.32)	**8.19** (7.16↔9.22)	**21.21** (18.53↔23.89)	**27.33** (24.16↔30.50)
	M2	**11.58** (9.98↔13.18)	**8.79** (7.40↔10.18)	**11.34** (9.84↔12.84)	**8.79** (7.66↔9.92)	**24.40** (21.19↔27.61)	**31.30** (27.46↔35.14)
		n=73	n=73	n=73	n=73	n=73	n=73
[W+O] (n=521)	M1	**11.56** (11.02↔12.10)	**8.73** (8.29↔9.17)	**11.76** (11.23↔12.29)	**8.78** (8.39↔9.17)	**24.76** (23.69↔25.83)	**31.19** (29.94↔32.44)
	M2	**10.70** (10.18↔11.22)	**7.68** (7.26↔8.10)	**10.36** (9.83↔10.89)	**7.74** (7.38↔8.10)	**22.15** (21.10↔23.20)	**28.01** (26.80↔29.22)
		n=518	n=518	n=518	n=518	n=518	n=518

The mean and 95% CI for all outcome variables were calculated for girls in groups [T] and [O] (TABLES 8.7 & 8.8).

TABLE 8.7: Crude mean and 95% CI of the PF outcomes for the girls in the Gulu sample at M1 and M2 - stratified by time and cohort study group

Group	Time	BFA (z score)	SBJ (cm)	MSFT (level complete)
[T] (n=79)	M1	**-0.280** (-0.470↔-0.091)	**165.82** (162.03↔169.61)	**4.22** (3.81↔4.63)
	M2	**-0.326** (-0.490↔-0.161)	**166.84** (163.12↔170.56)	**4.87** (4.46↔5.28)
		n=79	n=79	n=79
[O] (n=727)	M1	**-0.238** (-0.305↔-0.170)	**163.02** (161.72↔164.32)	**3.50** (3.37↔3.63)
	M2	**-0.284** (-0.350↔-0.219)	**162.87** (161.42↔164.32)	**4.53** (4.38↔4.68)
		n=694	n=688	n=681

TABLE 8.8: Crude mean and 95% CI of the MH outcomes for the girls in the Gulu sample at M1 and M2 - stratified by time and cohort study group

Group	Time	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
[T] (n=79)	M1	**15.06** (13.59↔16.53)	**11.63** (10.30↔12.96)	**14.89** (13.31↔16.47)	**10.11** (9.11↔11.11)	**31.62** (28.53↔34.71)	**39.13** (35.55↔42.71)
	M2	**13.90** (12.48↔15.32)	**11.03** (9.82↔12.24)	**13.51** (12.19↔14.83)	**9.35** (8.48↔10.22)	**29.58** (26.75↔32.41)	**37.10** (33.89↔40.31)
		n=79	n=79	n=79	n=79	n=79	n=79
[O] (n=727)	M1	**15.08** (14.56↔15.60)	**11.62** (11.20↔12.04)	**14.55** (14.06↔15.04)	**10.09** (9.75↔10.43)	**31.54** (30.53↔32.55)	**39.01** (37.85↔40.17)
	M2	**14.35** (13.83↔14.87)	**11.14** (10.72↔11.56)	**13.75** (13.25↔14.25)	**9.23** (8.89↔9.57)	**29.86** (28.84↔30.88)	**36.87** (35.69↔38.05)
		n=723	n=723	n=723	n=723	n=723	n=723

ii) Statistical findings

I found that the GMKL had a positive effect on the performance of the boys in the MSFT when compared to the control group, but a negative impact on their crude MH scores. The changes in PF and MH for the girls in group [T] were not significantly different to their peers.

With the exception of the MSFT for the girls in group [O], the underlying assumptions for the statistical analyses of within-group changes in each of the cohort study treatment groups were satisfied (APPENDIX A18). The violation of the assumption of HOV for the MSFT for girls in group [O] resulted in a "liberal" statistical test and an increased risk of a type I error [199,200]. Statistical testing proceeded as planned for all other variables.

When examining the crude means, the boys in group [T] improved significantly between M1 and M2 for the MSFT [p=0.035]. There was also a highly significant improvement in performance over time by the boys in group [W+O] for the MSFT [p<0.001]. No significant change was noted for BFA in either the intervention group [p=0.169] or control [p=0.535]. There was a significant deterioration in performance for the boys in group [T] for the SBJ [p=0.011], but no change in the control group [p=0.796] (FIGURES 8.4 – 8.6).

FIGURE 8.4: Crude change from M1 to M2 in the mean and 95% CI of BFA for the boys in groups [T] and [W+O] of the Gulu sample

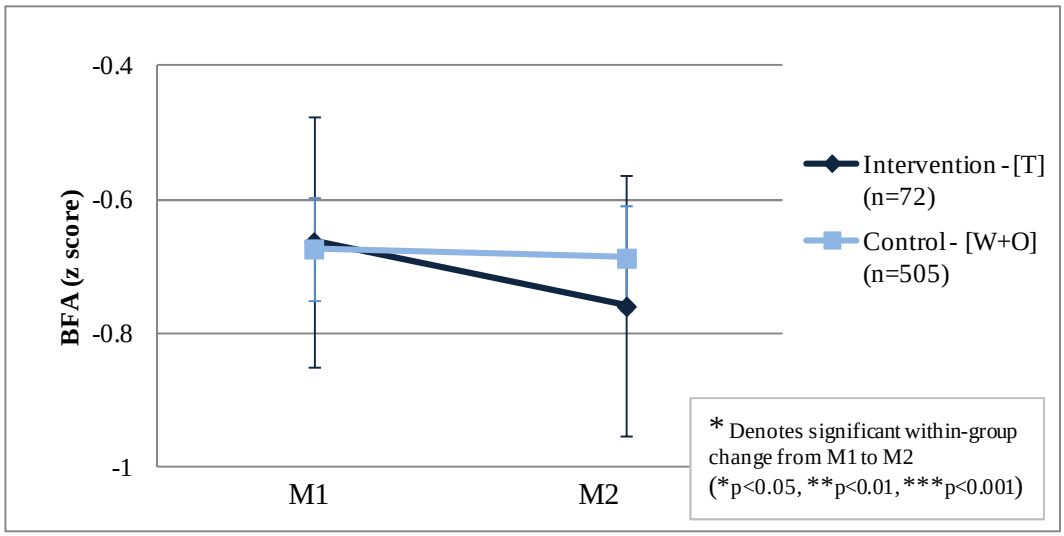


FIGURE 8.5: Crude change from M1 to M2 in the mean and 95% CI of the SBJ for the boys in groups [T] and [W+O] of the Gulu sample

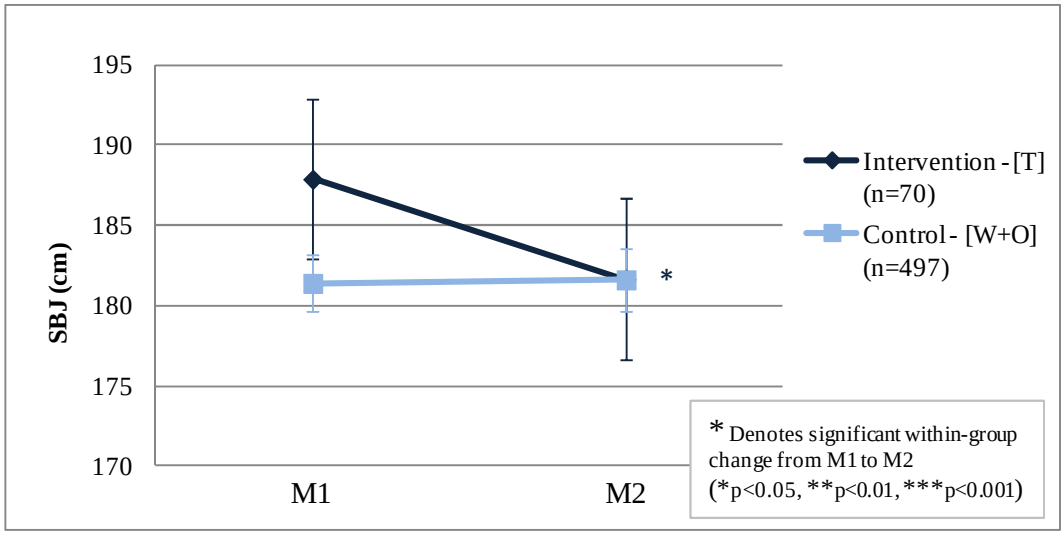
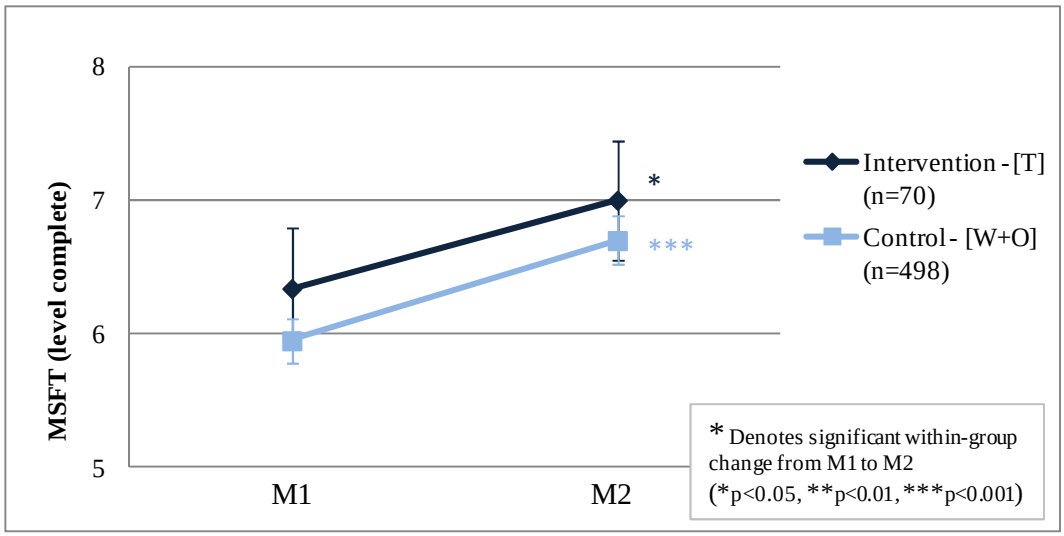


FIGURE 8.6: Crude change from M1 to M2 in the mean and 95% CI of the MSFT for the boys in groups [T] and [W+O] of the Gulu sample



The crude means for the girls in group [T] demonstrated no significant change in BFA [p=0.217]. There was a decrease in BFA for group [O] that was statistically significant [p=0.001]. No significant change was noted for the SBJ in either the intervention group [p=0.499] or control [p=0.836]. There was a highly significant improvement in performance by the girls in groups [T] and [O] for the MSFT [p<0.001] (FIGURES 8.7 – 8.9). However, the "liberal" nature of the statistical test for the MSFT for the girls in group [O] weakens the statistical certainty for this variable and the result should be interpreted with caution [199,200].

FIGURE 8.7: Crude change from M1 to M2 in the mean and 95% CI of BFA for the girls in groups [T] and [O] of the Gulu sample

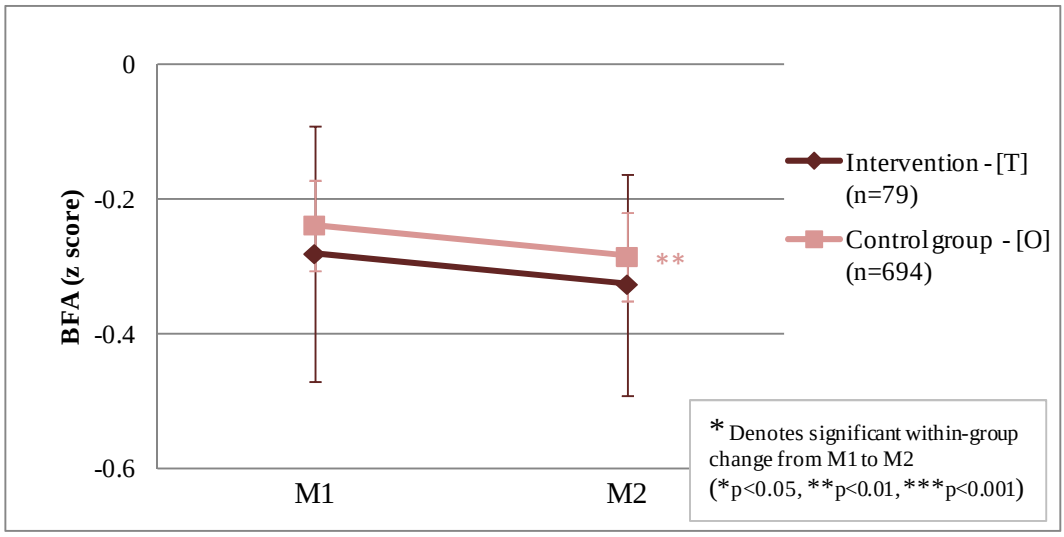


FIGURE 8.8: Crude change from M1 to M2 in the mean and 95% CI of the SBJ for the girls in groups [T] and [O] of the Gulu sample

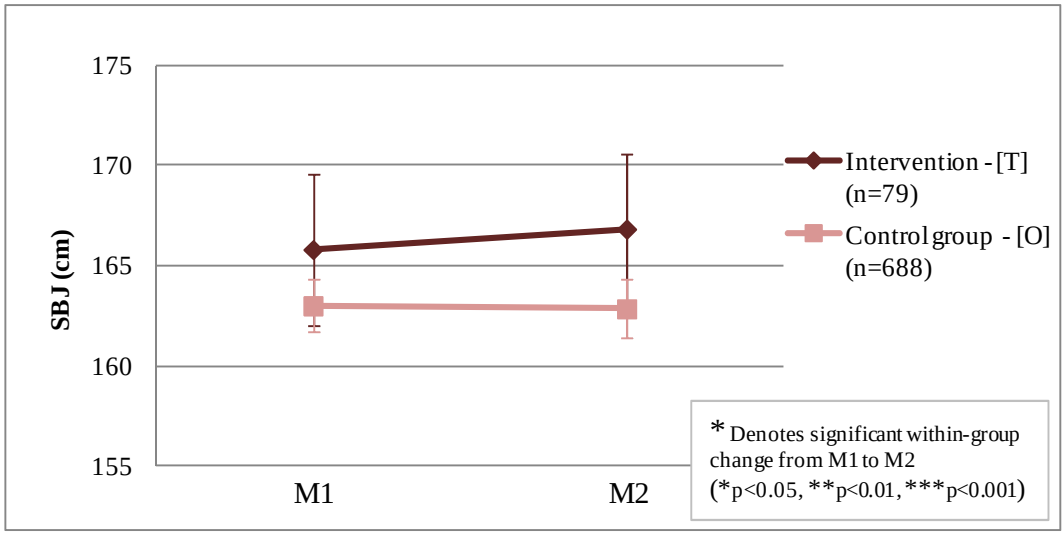
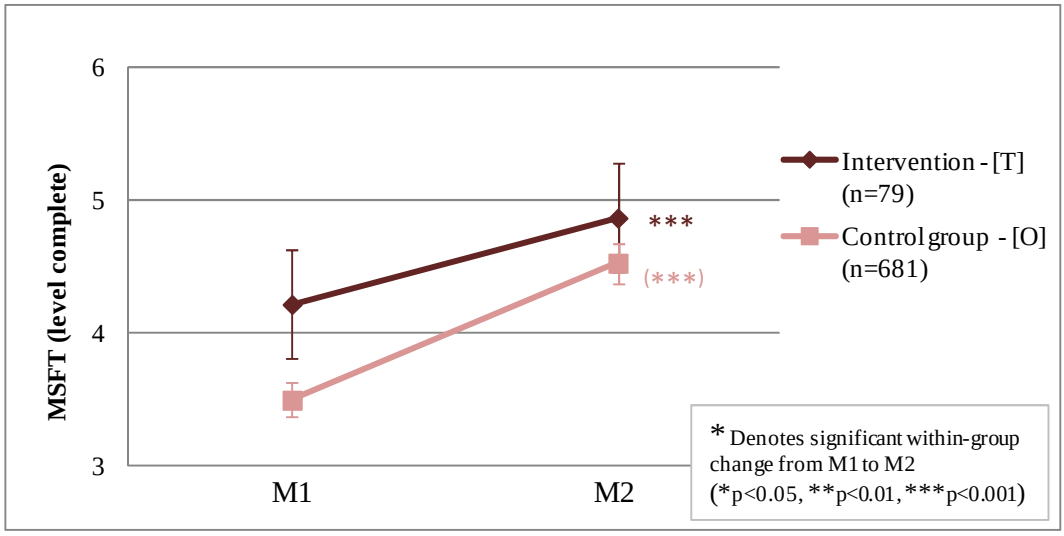


FIGURE 8.9: Crude change from M1 to M2 in the mean and 95% CI of the MSFT for the girls in groups [T] and [O] of the Gulu sample



There was no significant change in the crude means for the boys in group [T] for TT [p=0.085], K [p=0.078], P [p=0.109], ML [p=0.344], DT [p=0.076] and AT [p=0.067]. The boys in group [W+O] experienced a highly significant improvement for K, P, ML, DT and AT [p<0.001]. The improvement for boys in group [W+O] for TT was also significant [p=0.005] (FIGURES 8.10 – 8.15).

FIGURE 8.10: Crude change from M1 to M2 in the mean and 95% CI of TT for the boys in groups [T] and [W+O] of the Gulu sample

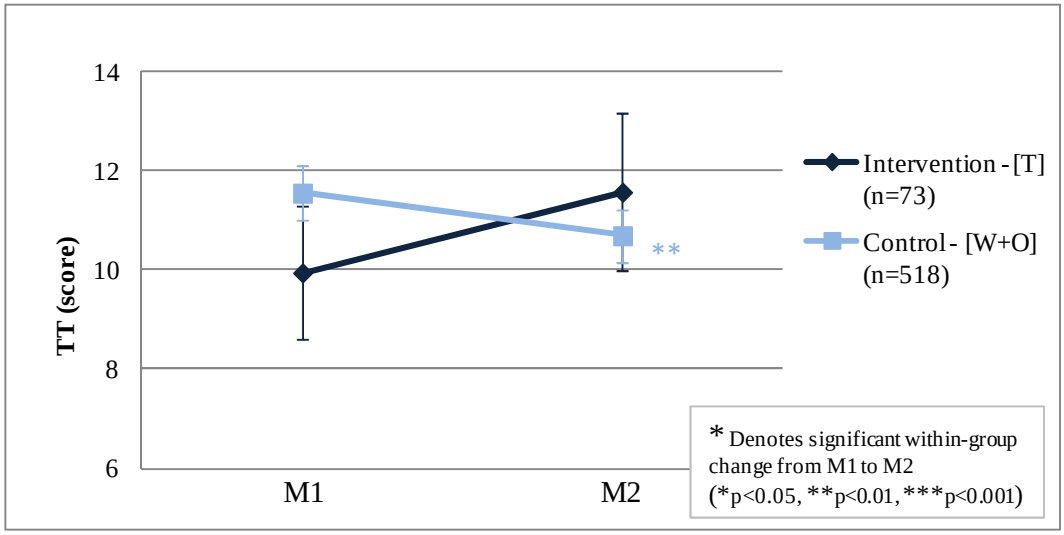


FIGURE 8.11: Crude change from M1 to M2 in the mean and 95% CI of K for the boys in groups [T] and [W+O] of the Gulu sample

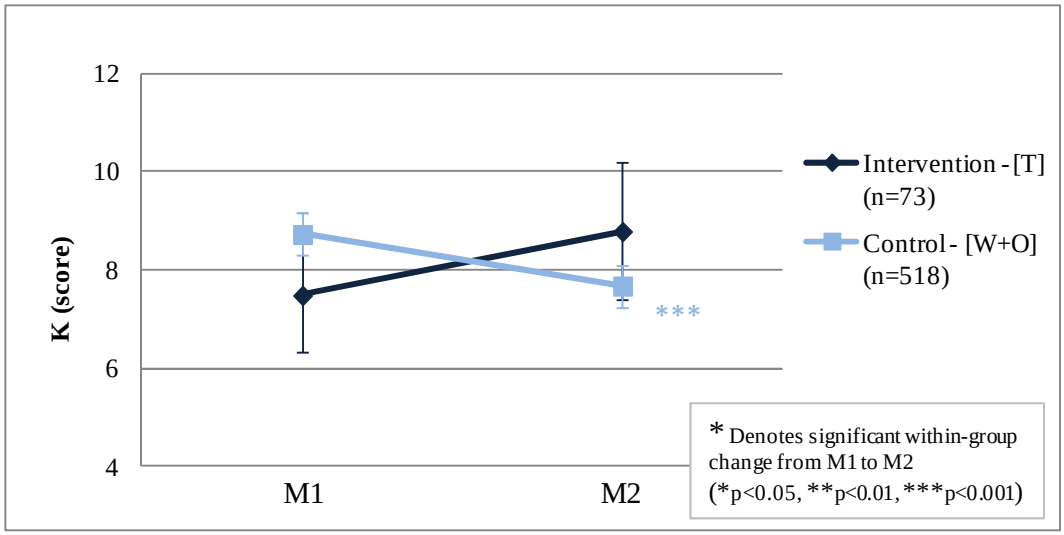


FIGURE 8.12: Crude change from M1 to M2 in the mean and 95% CI of P for the boys in groups [T] and [W+O] of the Gulu sample

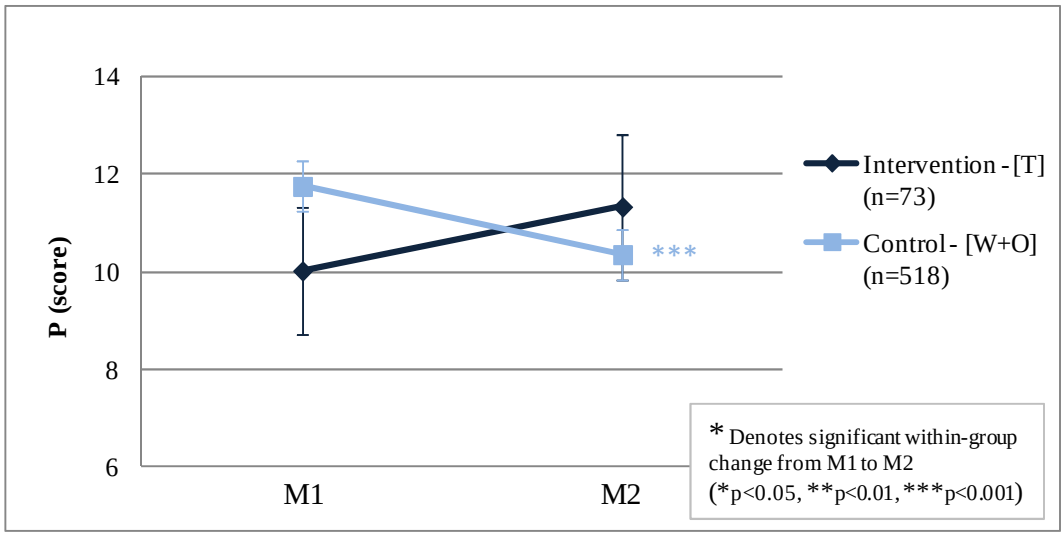


FIGURE 8.13: Crude change from M1 to M2 in the mean and 95% CI of ML for the boys in groups [T] and [W+O] of the Gulu sample

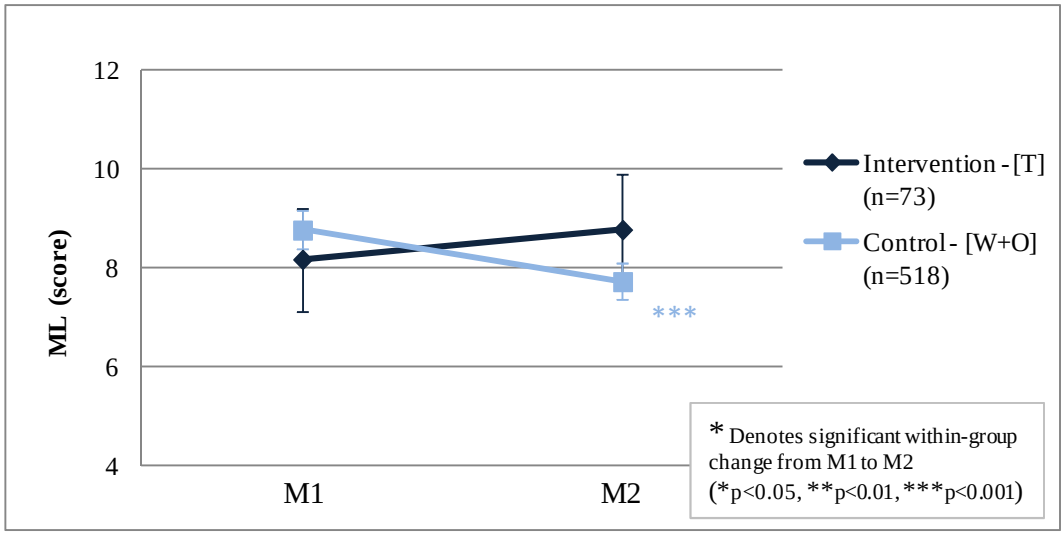


FIGURE 8.14: Crude change from M1 to M2 in the mean and 95% CI of DT for the boys in groups [T] and [W+O] of the Gulu sample

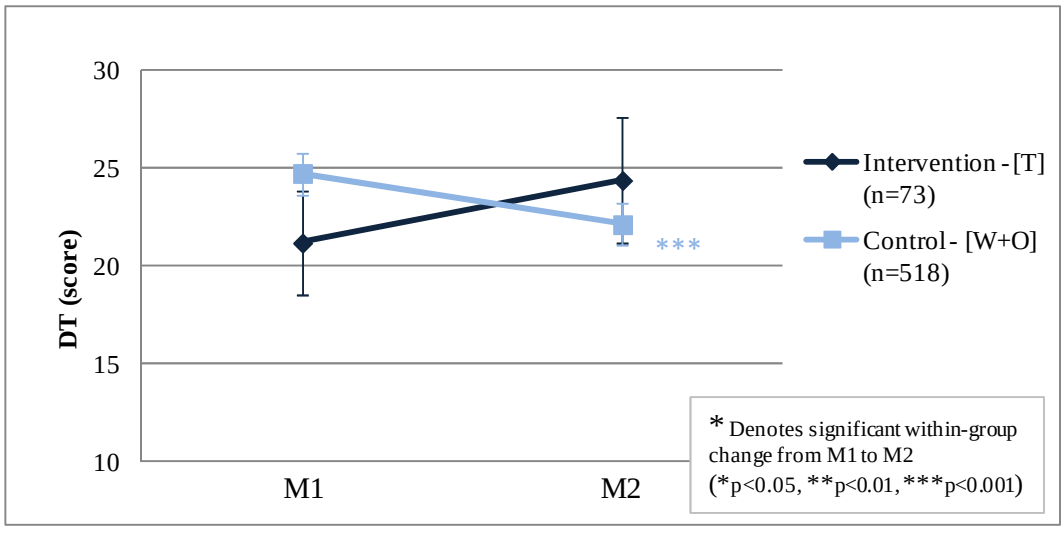
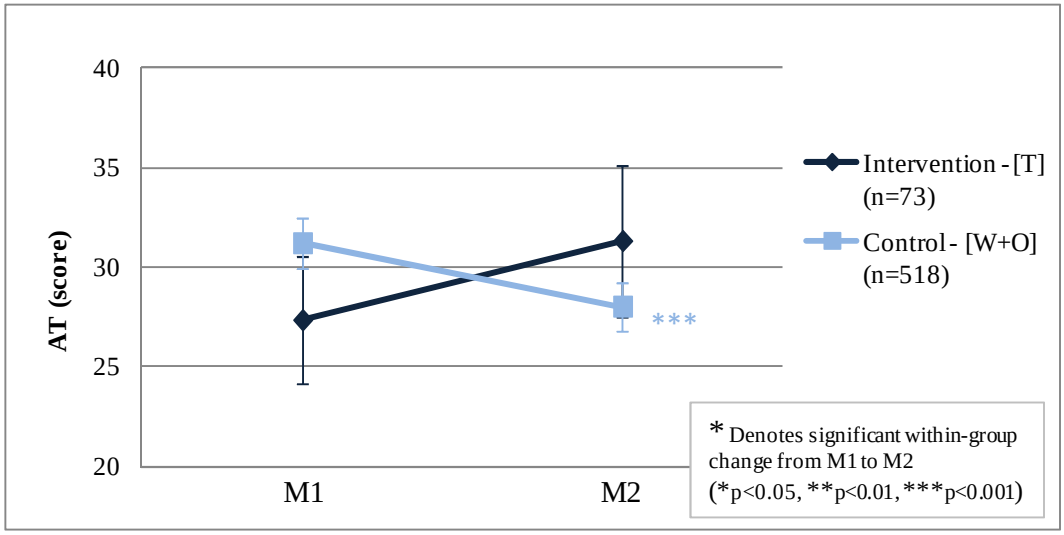


FIGURE 8.15: Crude change from M1 to M2 in the mean and 95% CI of AT for the boys in groups [T] and [W+O] of the Gulu sample



There was no significant change in the crude means for the girls in group [T] for TT [p=0.134], K [p=0.334], P [p=0.055], ML [p=0.223], DT [p=0.155] and AT [p=0.228]. The girls in group [O] experienced significant improvements for T [p=0.014], K [p=0.043], P [p=0.004], DT [p=0.003] and AT [p=0.001]. The improvement for girls in group [O] for ML was highly significant [p<0.001] (FIGURES 8.16 – 8.21).

FIGURE 8.16: Crude change from M1 to M2 in the mean and 95% CI of TT for the girls in groups [T] and [O] of the Gulu sample

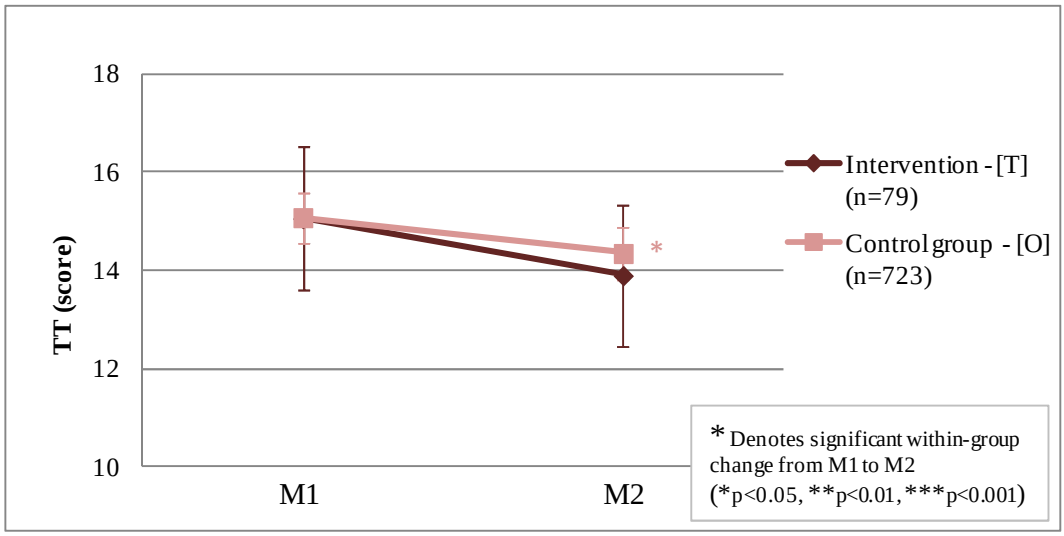


FIGURE 8.17: Crude change from M1 to M2 in the mean and 95% CI of K for the girls in groups [T] and [O] of the Gulu sample

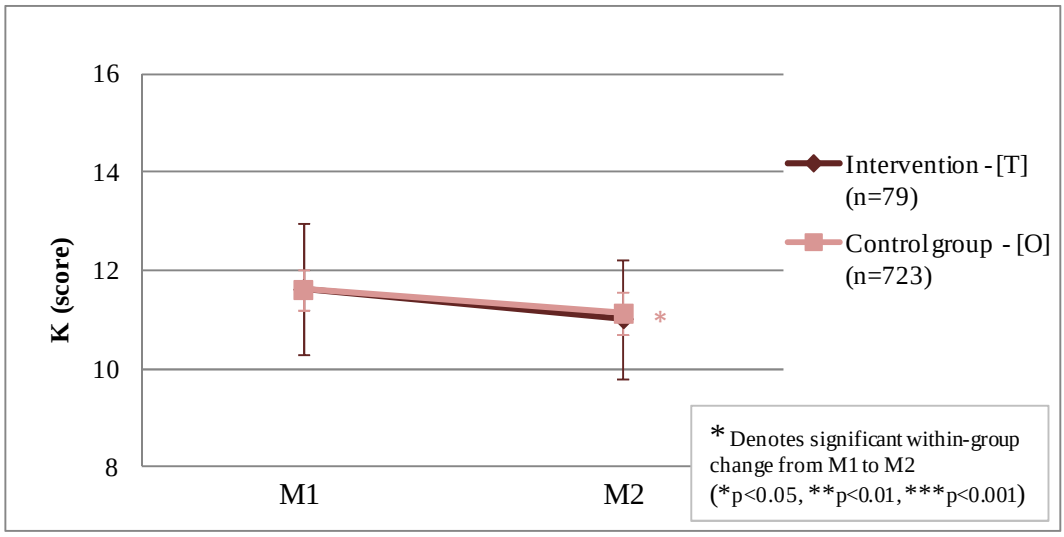


FIGURE 8.18: Crude change from M1 to M2 in the mean and 95% CI of P for the girls in groups [T] and [O] of the Gulu sample

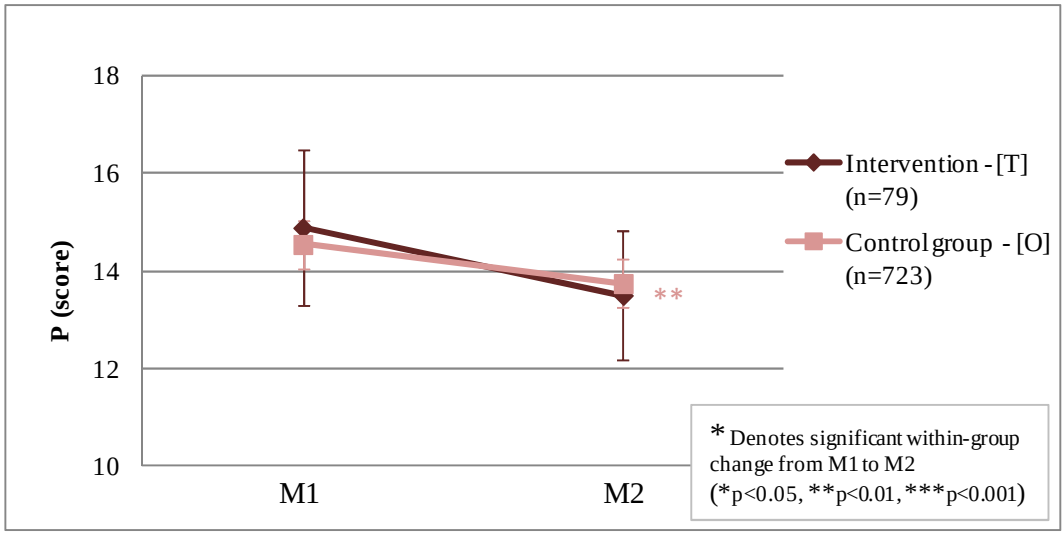


FIGURE 8.19: Crude change from M1 to M2 in the mean and 95% CI of ML for the girls in groups [T] and [O] of the Gulu sample

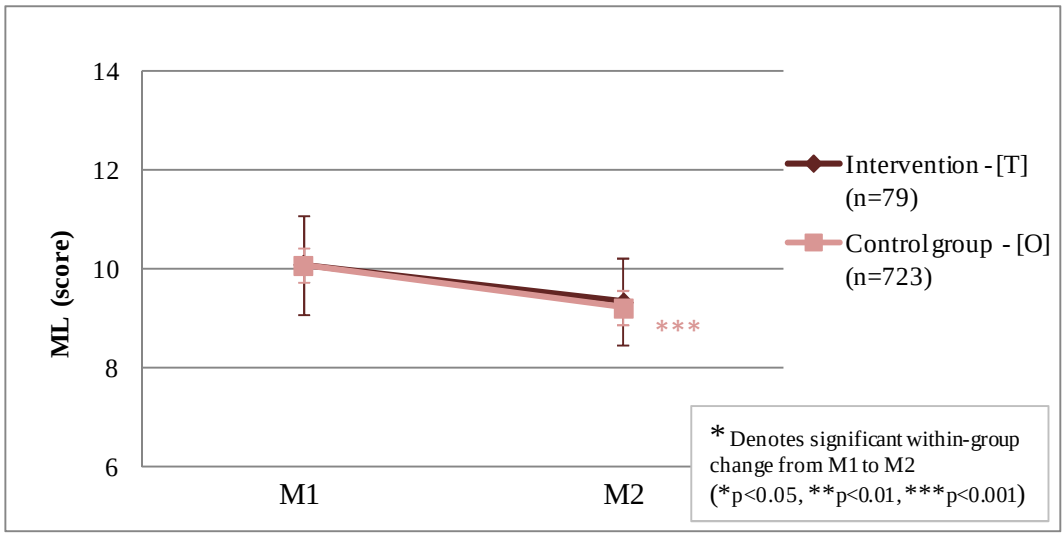


FIGURE 8.20: Crude change from M1 to M2 in the mean and 95% CI of DT for the girls in groups [T] and [O] of the Gulu sample

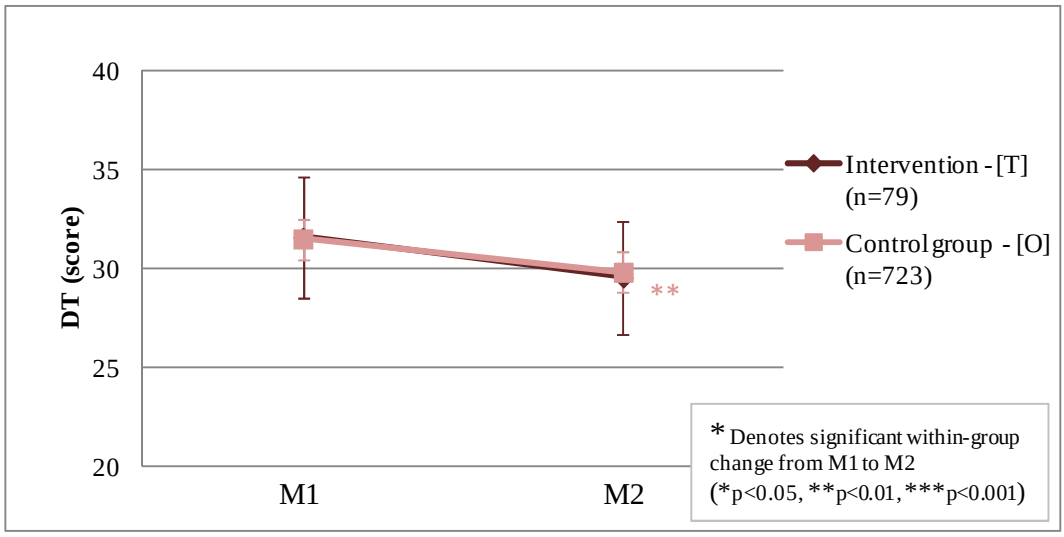
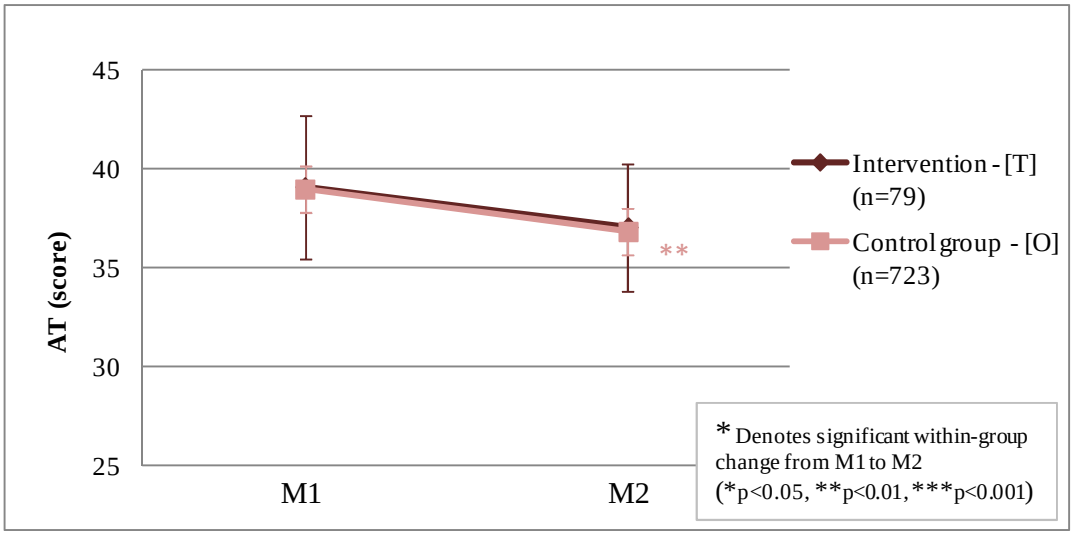


FIGURE 8.21: Crude change from M1 to M2 in the mean and 95% CI of AT for the girls in groups [T] and [O] of the Gulu sample



The underlying assumptions for the between-group statistical analyses were satisfied (APPENDIX A19).

When considering the boys crude means, the improvements in group [W+O] were significantly better than the changes in group [T] for the SBJ [p=0.008], TT [p=0.005], K [p=0.001], P [p=0.002] and ML [p=0.013]. There was also a highly significant between-group difference for DT and AT [p<0.001]. The changes that occurred in the control and intervention groups were not significantly different for BFA [p=0.194] or MSFT [p=0.766].

Adjusting for baseline values using the Model 1 analysis resulted in no significant between-group differences for any of the PF outcomes in the boys sample (BFA [p=0.185], SBJ [p=0.072], MSFT [p=0.520]). The statistical certainty for the MH outcomes was weakened in Model 1, but continued to indicate greater improvement for the boys in the control group for all variables (TT [p=0.048], K [p=0.008], P [p=0.031], ML [p=0.019], DT [p=0.011] and AT [p=0.005]).

When the Model 2 analysis was applied, the improvement of the boys in group [T] for the MSFT was significantly greater than for the control group [p=0.003]. There was no between-group difference for BFA [p=0.471] or SBJ [p=0.770]. The between-group differences for all of the MH outcomes also became statistically insignificant (TT [p=0.366], K [p=0.108], P [p=0.396], ML [p=0.345], DT [p=0.232], AT [p=0.201]).

There were no significant between-group differences for any of the outcome variables when comparing the girls groups crude means (BFA [p=0.977], SBJ [p=0.594], MSFT [p=0.141], TT [p=0.641], K [p=0.867], P [p=0.509], ML [p=0.872], DT [p=0.841], AT [p=0.954]).

Similar results were observed when the crude values were adjusted for baseline in Model 1 for the girls (BFA [p=0.926], SBJ [p=0.219], MSFT [p=0.812], TT [p=0.562], K [p=0.849], P [p=0.604], ML [p=0.819], DT [p=0.836], AT [p=0.915]).

All of the between-group differences remained insignificant in the girls sample when the Model 2 analysis was applied (BFA [p=0.390], SBJ [p=0.834], MSFT [p=0.065], TT [p=0.331], K [p=0.620], P [p=0.440], ML [p=0.881], DT [p=0.532], AT [p=0.721]).

A summary of these results was tabulated (TABLES 8.9 & 8.10).

TABLE 8.9: Repeated measures ANOVA comparing the change in the mean from M1 to M2 between boys in groups [T] and [W+O] of the Gulu sample for each of the outcomes

Data analysed	BFA (z score)	SBJ (cm)	MSFT (level complete)	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Crude data	F=1.691	F=7.029 ##	F=0.088	F=7.926 ##	F=10.877 ##	F=9.486 ##	F=6.159 #	F=11.551 ###	F=12.724 ###
Model 1	F=1.759	F=3.260	F=0.415	F=3.919 #	F=7.053 ##	F=4.668 #	F=5.540 #	F=6.450 #	F=8.109 ##
Model 2	F=0.521	F=0.086	F=9.095 ##	F=0.820	F=2.588	F=0.720	F=0.895	F=1.433	F=1.641

NOTE: # Denotes a significant difference between "change in [T]" vs. "change in [W+O]" (# p<0.05, ## p<0.01, ### p<0.001).

TABLE 8.10: Repeated measures ANOVA comparing the change in the mean from M1 to M2 between girls in groups [T] and [O] of the Gulu sample for each of the outcomes

Data analysed	BFA (z score)	SBJ (cm)	MSFT (level complete)	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Crude data	F=0.001	F=0.284	F=2.176	F=0.218	F=0.028	F=0.437	F=0.026	F=0.040	F=0.003
Model 1	F=0.009	F=1.511	F=0.056	F=0.336	F=0.036	F=0.269	F=0.053	F=0.043	F=0.011
Model 2	F=0.740	F=0.044	F=3.428	F=0.948	F=0.246	F=0.596	F=0.022	F=0.391	F=0.127

NOTE: # Denotes a significant difference between "change in [T]" vs. "change in [O]" (# p<0.05, ## p<0.01, ### p<0.001)

NOTE: Comparisons completed with a split-plot design using a model III sum of squares analysis (unbalanced groups with no empty cells for the selected fixed factors).

NOTE: In model 1 the crude values were adjusted for baseline differences.

NOTE: In model 2 the crude values were adjusted for baseline, age, division, school and history of abduction (i.e. fully adjusted for identified factors and covariates).

8.4c) Impact of the Gum Marom Kids League (randomised control trial)

i) Descriptive statistics

The mean and 95% CI of all outcome variables were calculated for the boys in groups [T], [W] and [O] (TABLES 8.11 & 8.12).

TABLE 8.11: Crude mean and 95% CI of the PF outcomes for the boys in the Gulu sample at M1 and M2 - stratified by time and RCT study group

Group	Time	BFA (z score)	SBJ (cm)	MSFT (level complete)
[T] (n=73)	M1	**-0.663** (-0.849↔-0.477)	**187.89** (182.93↔192.85)	**6.34** (5.89↔6.79)
	M2	**-0.759** (-0.954↔-0.565)	**181.67** (176.63↔186.71)	**7.00** (6.56↔7.44)
		n=72	n=70	n=70
[W] (n=71)	M1	**-0.650** (-0.852↔-0.449)	**183.78** (179.07↔188.49)	**6.17** (5.72↔6.62)
	M2	**-0.726** (-0.940↔-0.513)	**186.22** (181.27↔191.17)	**7.16** (6.70↔7.62)
		n=69	n=68	n=69
[O] (n=450)	M1	**-0.678** (-0.761↔-0.594)	**181.03** (179.11↔182.95)	**5.92** (5.73↔6.11)
	M2	**-0.681** (-0.764↔-0.599)	**180.90** (178.84↔182.96)	**6.62** (6.42↔6.82)
		n=436	n=429	n=429

TABLE 8.12: Crude mean and 95% CI of the MH outcomes for the boys in the Gulu sample at M1 and M2 - stratified by time and RCT study group

Group	Time	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
[T] (n=73)	M1	**9.95** (8.61↔11.29)	**7.49** (6.35↔8.63)	**10.03** (8.74↔11.32)	**8.19** (7.16↔9.22)	**21.21** (18.53↔23.89)	**27.33** (24.16↔30.50)
	M2	**11.58** (9.98↔13.18)	**8.79** (7.40↔10.18)	**11.34** (9.84↔12.84)	**8.79** (7.66↔9.92)	**24.40** (21.19↔27.61)	**31.30** (27.46↔35.14)
		n=73	n=73	n=73	n=73	n=73	n=73
[W] (n=71)	M1	**11.37** (9.84↔12.90)	**8.54** (7.36↔9.72)	**11.70** (10.28↔13.12)	**8.84** (7.67↔10.01)	**24.53** (21.47↔27.59)	**30.86** (27.28↔34.44)
	M2	**8.94** (7.81↔10.07)	**6.06** (5.17↔6.95)	**8.27** (7.07↔9.47)	**7.11** (6.33↔7.89)	**18.29** (15.96↔20.62)	**23.97** (21.33↔26.61)
		n=70	n=70	n=70	n=70	n=70	n=70
[O] (n=450)	M1	**11.59** (11.01↔12.17)	**8.76** (8.28↔9.24)	**11.77** (11.20↔12.34)	**8.77** (8.35↔9.19)	**24.80** (23.66↔25.94)	**31.25** (29.91↔32.59)
	M2	**10.98** (10.40↔11.56)	**7.93** (7.47↔8.39)	**10.69** (10.11↔11.27)	**7.83** (7.43↔8.23)	**22.75** (21.61↔23.89)	**28.65** (27.32↔29.98)
		n=448	n=448	n=448	n=448	n=448	n=448

ii) Statistical findings

I found that the GMKL had a negative effect on the crude MH and SBJ scores of the boys in group [T] when compared to their wait-listed peers. There was also no "disappointment effect" evident for the boys in group [W] for all of the outcome measures. The underlying assumptions for the statistical analyses of within-group changes in each of the RCT treatment groups were satisfied (APPENDIX A18).

When examining the crude means, the boys in group [T] improved significantly between M1 and M2 for the MSFT [p=0.035]. There was also a highly significant improvement in performance over time by the boys in groups [W] and [O] for the MSFT [p<0.001]. A significant deterioration in performance was noted for the boys in group [T] for the SBJ [p=0.011], but no change for the wait-listed group [p=0.333] or the boys who did not register for the GMKL [p=0.882]. No significant change was noted for BFA in the intervention group [p=0.169], wait-list boys [p=0.239] or those not registered [p=0.872] (FIGURES 8.22 – 8.24).

FIGURE 8.22: Crude change from M1 to M2 in the mean and 95% CI of BFA for the boys in group [T], [W] and [O] of the Gulu sample

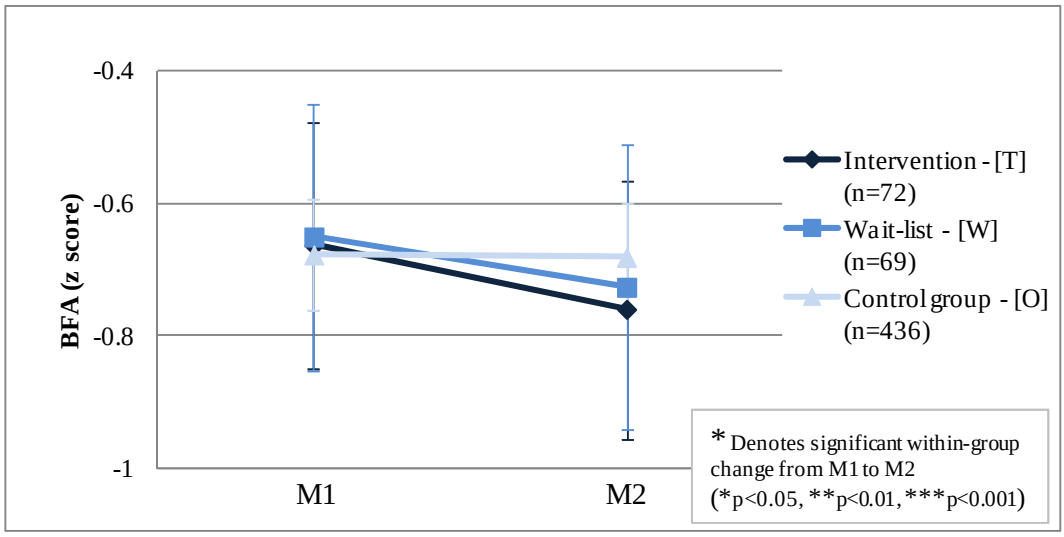


FIGURE 8.23: Crude change from M1 to M2 in the mean and 95% CI of the SBJ for the boys in group [T], [W] and [O] of the Gulu sample

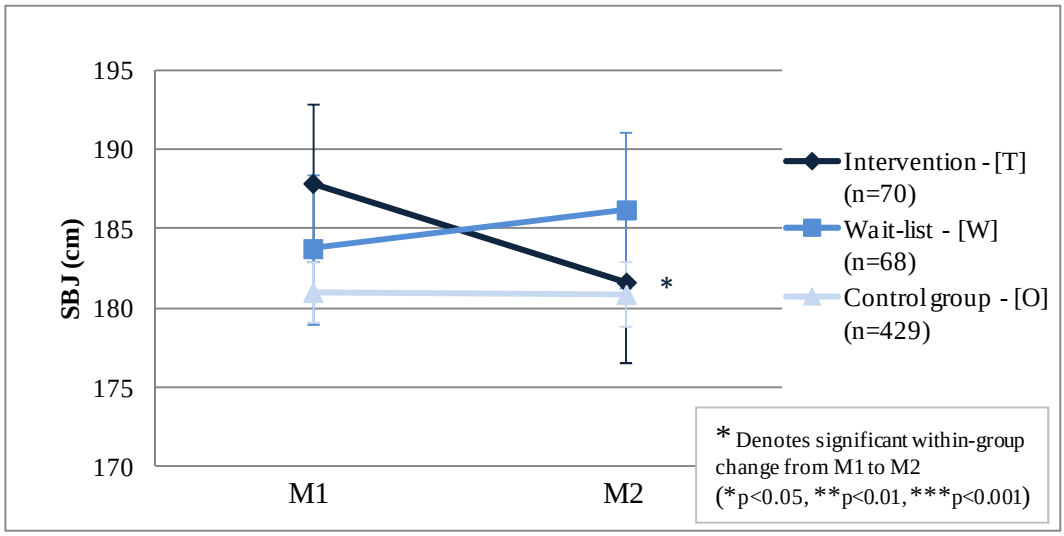
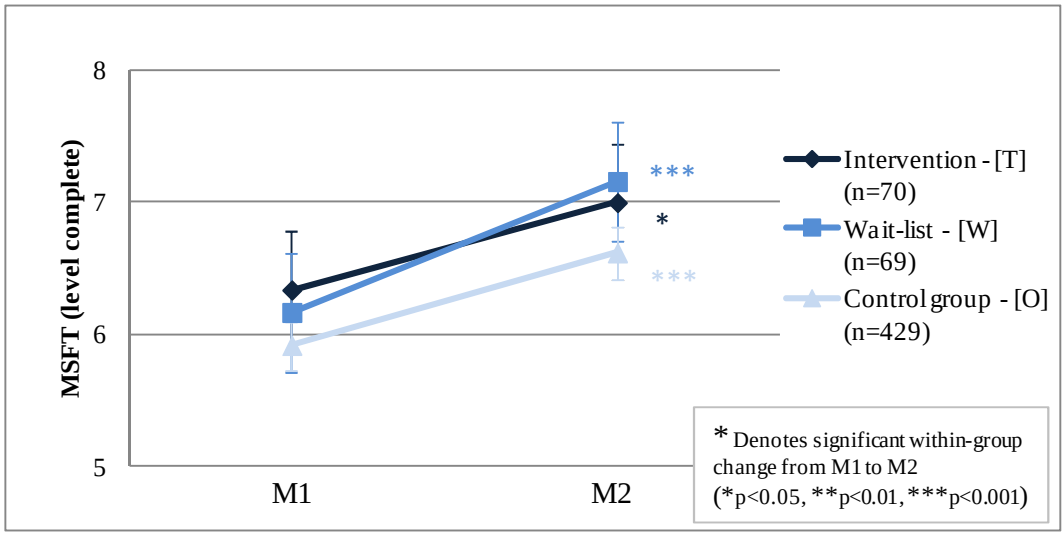


FIGURE 8.24: Crude change from M1 to M2 in the mean and 95% CI of the MSFT for the boys in group [T], [W] and [O] of the Gulu sample



The deterioration in the crude means for the boys in group [T] was not statistically significant for TT [p=0.085], K [p=0.078], P [p=0.109], ML [p=0.344], DT [p=0.076] and AT [p=0.067]. The boys in group [W] experienced a highly significant improvement for P, DT and AT [p<0.001]. The wait-listed boys also scored significantly better at M2 than M1 for TT [p=0.005], K [p=0.001] and ML [p=0.005]. A highly significant improvement was noted for boys in group [O] for ML and AT [p<0.001]. The improvement for boys in group [O] was also significant for K [p=0.002], P [p=0.001] and DT [p=0.001]. There was no significant change in group [O] for TT [p=0.061] (FIGURES 8.25 – 8.30).

FIGURE 8.25: Crude change from M1 to M2 in the mean and 95% CI of TT for the boys in group [T], [W] and [O] of the Gulu sample

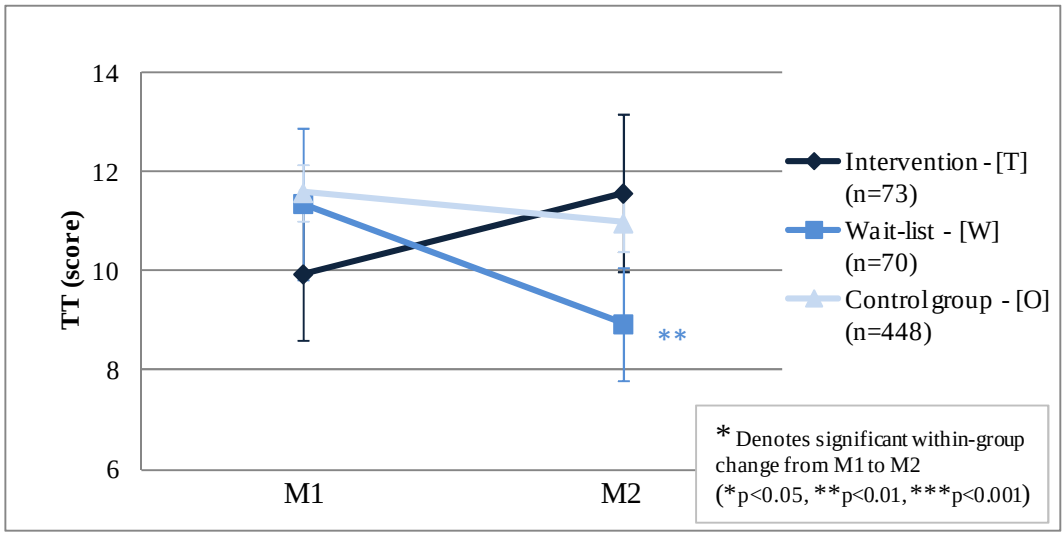


FIGURE 8.26: Crude change from M1 to M2 in the mean and 95% CI of K for the boys in group [T], [W] and [O] of the Gulu sample

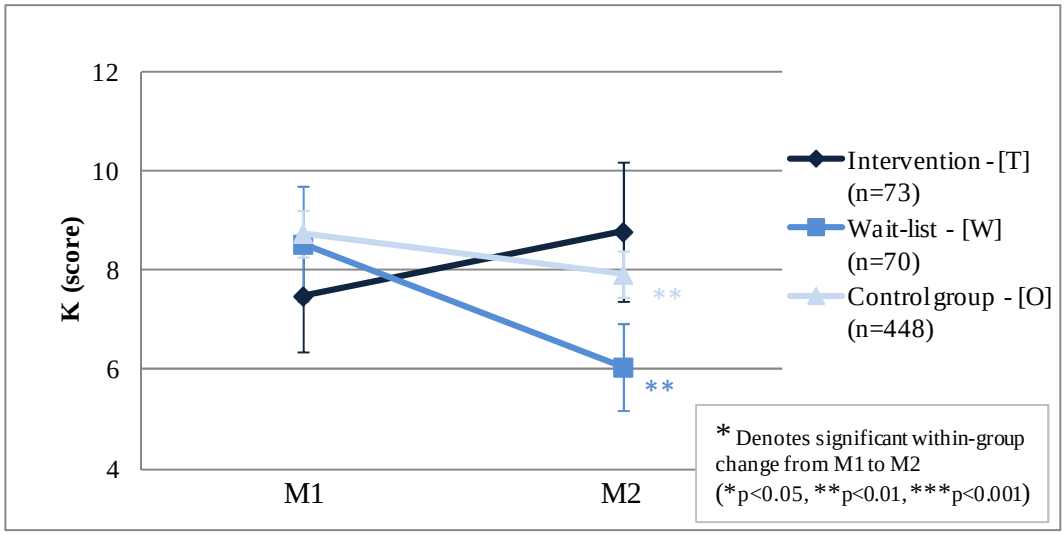


FIGURE 8.27: Crude change from M1 to M2 in the mean and 95% CI of P for the boys in group [T], [W] and [O] of the Gulu sample

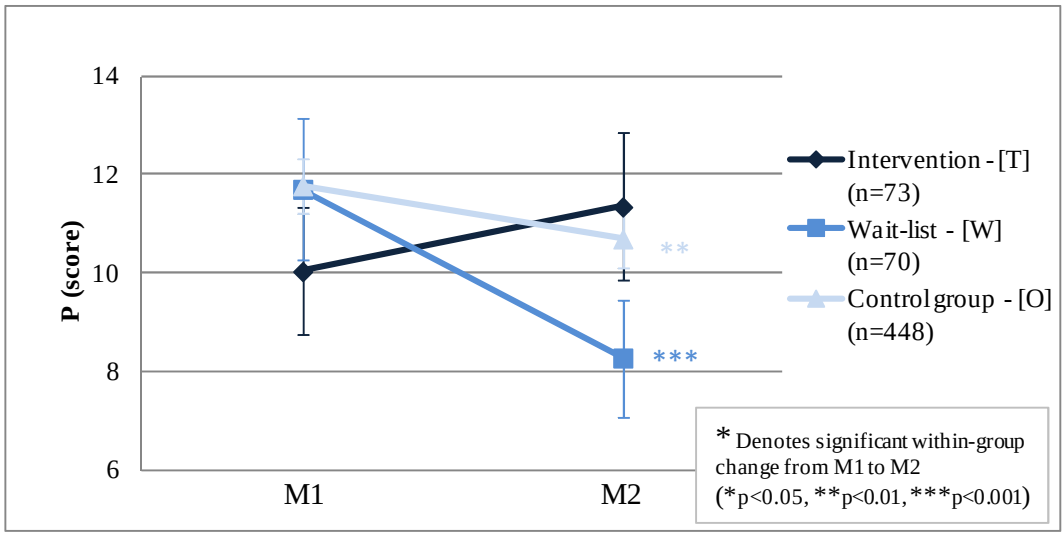


FIGURE 8.28: Crude change from M1 to M2 in the mean and 95% CI of ML for the boys in group [T], [W] and [O] of the Gulu sample

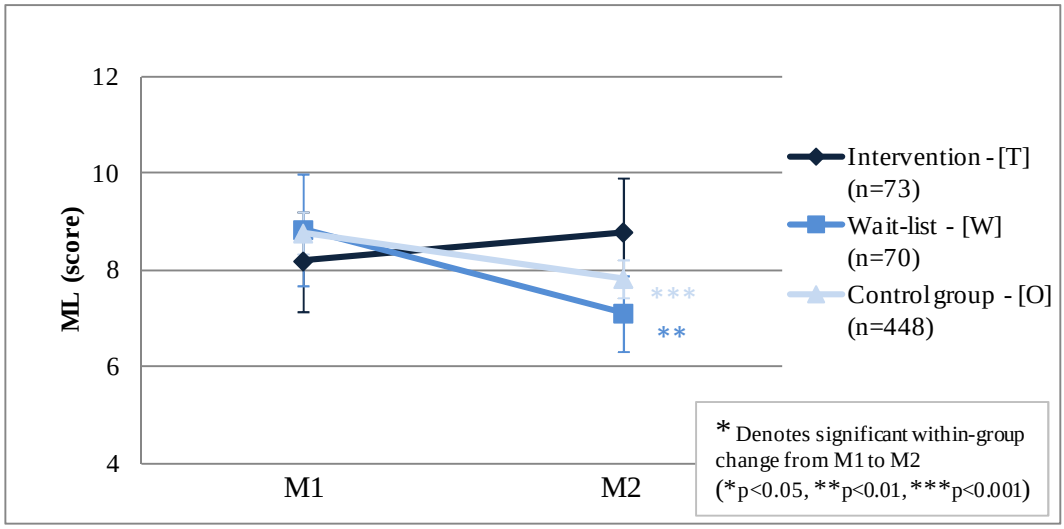


FIGURE 8.29: Crude change from M1 to M2 in the mean and 95% CI of DT for the boys in group [T], [W] and [O] of the Gulu sample

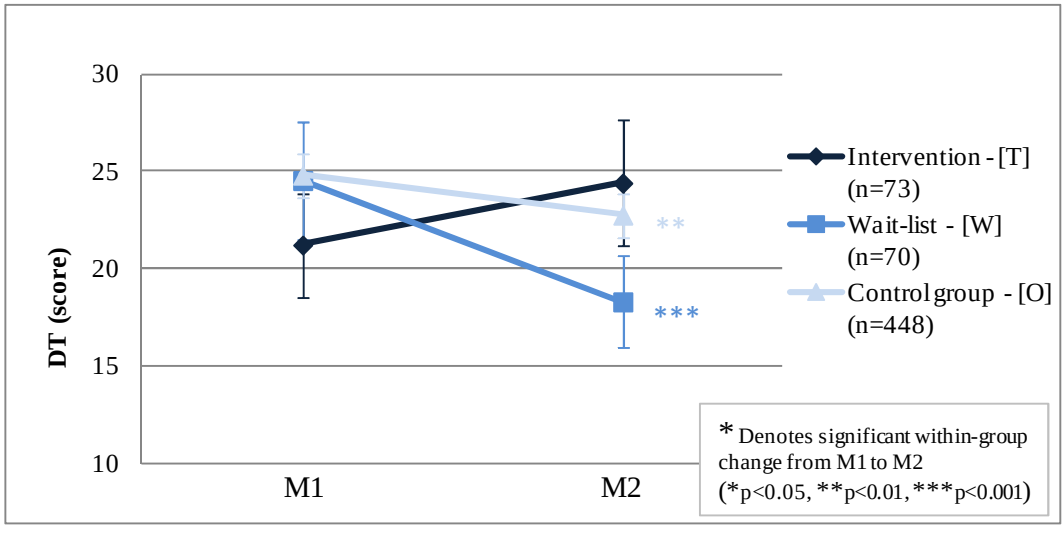
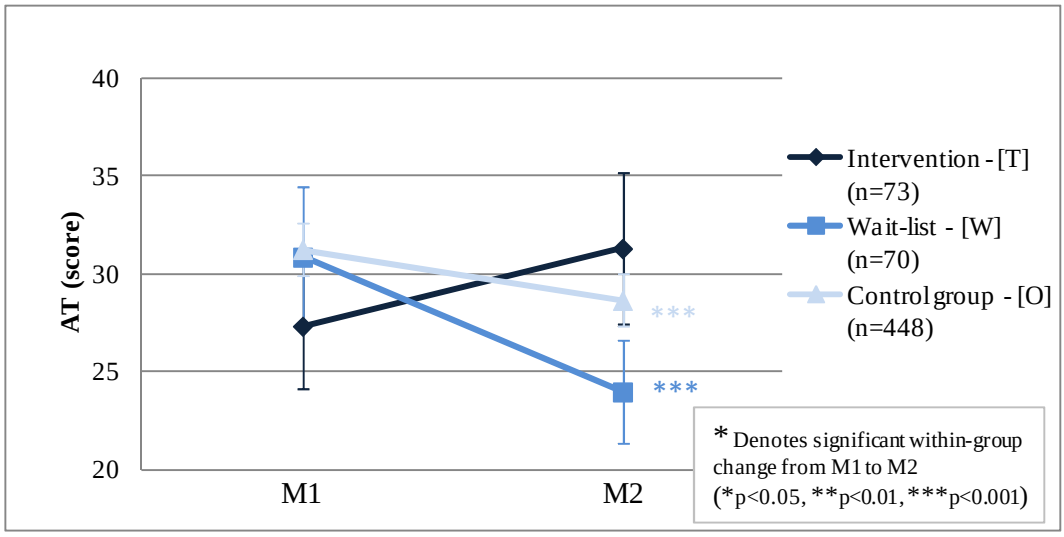


FIGURE 8.30: Crude change from M1 to M2 in the mean and 95% CI of AT for the boys in group [T], [W] and [O] of the Gulu sample



The underlying assumptions for the between-group statistical analyses were satisfied (APPENDIX A19).

When considering the effect of the intervention on the boys crude means, the improvement in group [W] was significantly better than the changes in group [T] for the SBJ [p=0.013], TT [p=0.002] and ML [p=0.008]. There was also a similar highly significant between-group difference for K, P, DT and AT [p<0.001]. The changes that occurred in the wait-list and intervention groups were not significantly different for BFA [p=0.828] or MSFT [p=0.421].

There was no indication of a "disappointment effect" when analysing the crude means that compared the changes in group [W] vs. [O]. The between-group differences were not significant for any of the PF outcomes (BFA [p=0.259], SBJ [p=0.297], MSFT [p=0.346]) or ML [p=0.242]. There was a significantly greater improvement in group [W] than group [O] for TT [p=0.041], K [p=0.022], P [p=0.009], DT [p=0.015] or AT [p=0.034].

Similar results were noted for both the intervention and "disappointment effect" when the crude values were adjusted for baseline in Model 1. There continued to be a highly

significant between-group difference when comparing changes in group [T] vs. [W] for K, P, DT and AT [p<0.001]. The improvement of the boys wait-listed was still significantly better than the changes seen in group [T] for the SBJ [p=0.029], TT [p=0.003] and ML [p=0.006]. No significant intervention effect was evident for BFA [p=0.804] or MSFT [p=0.541].

The between-group differences for group [W] vs. [O] when applying Model 1 continued to be insignificant for all of the PF outcomes (BFA [p=0.269], SBJ [p=0.118], MSFT [p=0.087]) and ML [p=0.157]. The statistical certainty that there was no "disappointment effect" was strengthened when adjusting for baseline values for TT [p=0.007], K [p=0.002], P [p=0.001], DT [p=0.002] and AT [p=0.006].

When the Model 2 analysis was applied, there was no between-group intervention effect evident for any of the PF outcomes (BFA [p=0.263], SBJ [p=0.392], MSFT [p=0.141]). The statistical certainty for K was weakened, but continued to indicate better results for group [W] than for the boys exposed to the intervention [p=0.024]. The between-group differences comparing group [T] vs. [W] for all of the other MH outcomes became statistically insignificant (TT [p=0.207], P [p=0.113], ML [p=0.786], DT [p=0.084], AT [p=0.119]).

Analysis of the data using Model 2 indicated that the change in BFA for the boys in group [O] was significantly different to the decrease seen in those wait-listed [p=0.013]. The between-group differences remained statistically insignificant for the SBJ [p=0.288] and MSFT [p=0.142]. The Model 2 analysis also indicated that there was no significant "disappointment effect" for any of the MH outcomes (TT [p=0.337], K [p=0.202], P [p=0.523], ML [p=0.953], DT [p=0.396], AT [p=0.417]).

A summary of these results was tabulated (TABLES 8.13 & 8.14).

TABLE 8.13: Repeated measures ANOVA comparing the change in the mean from M1 to M2 between boys in groups [T] and [W] of the Gulu sample for each of the outcomes

Data analysed	BFA (z score)	SBJ (cm)	MSFT (level complete)	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Crude data	F=0.048	F=6.319 #	F=0.651	F=10.344 ##	F=13.425 ###	F=17.334 ###	F=7.200 ##	F=14.770 ###	F=14.329 ###
Model 1	F=0.062	F=4.854 #	F=0.376	F=8.947 ##	F=13.018 ###	F=14.698 ###	F=7.846 ##	F=12.694 ###	F=13.010 ###
Model 2	F=1.282	F=0.746	F=2.232	F=1.636	F=5.372 #	F=2.593	F=0.074	F=3.101	F=2.510

NOTE: # Denotes a significant difference between "change in [T]" vs. "change in [W]" (# p<0.05, ## p<0.01, ### p<0.001)

TABLE 8.14: Repeated measures ANOVA comparing the change in the mean from M1 to M2 between boys in groups [W] and [O] of the Gulu sample for each of the outcomes

Data analysed	BFA (z score)	SBJ (cm)	MSFT (level complete)	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Crude data	F=1.275	F=1.089	F=0.892	F=4.185 #	F=5.285 #	F=6.814 ##	F=1.375	F=5.968 #	F=4.538 #
Model 1	F=1.224	F=2.458	F=2.946	F=7.307 ##	F=9.672 ##	F=10.547 ##	F=2.008	F=9.453 ##	F=7.529 ##
Model 2	F=6.277 #	F=1.134	F=2.171	F=0.926	F=1.635	F=0.409	F=0.003	F=0.723	F=0.660

NOTE: # Denotes a significant difference between "change in [W]" vs. "change in [O]" (# p<0.05, ## p<0.01, ### p<0.001)

NOTE: Comparisons completed with a split-plot design using a model III sum of squares analysis (unbalanced groups with no empty cells for the selected fixed factors).

NOTE: In model 1 the crude values were adjusted for baseline differences.

NOTE: In model 2 the crude values were adjusted for baseline, age, division, school and history of abduction (i.e. fully adjusted for identified factors and covariates).

8.4d) Impact of the Gum Marom Kids League (accelerometry)

i) Descriptive statistics

The mean and 95% CI of MVPA and VPA were calculated for the group [T] and [W] accelerometer sub-samples (TABLE 8.15).

TABLE 8.15: Crude mean and 95% CI for the volume of MVPA and VPA for the boys in the accelerometer study in Gulu - stratified by intervention group

Intervention group	Daily average on valid days (minutes)		Daily average during the GMKL training times - 4-7pm (minutes)		Average on the GMKL activity day (minutes)	
	MVPA	VPA	MVPA	VPA	MVPA	VPA
[T] n=30	**116.48** (103.35↔129.60)	**6.19** (2.77↔9.60)	**45.09** (38.72↔51.47)	**5.09** (2.71↔7.47)	**142.37** (122.32↔162.41)	**16.87** (8.05↔25.68)
[W] n=24	**114.09** (104.51↔123.68)	**4.32** (2.71↔5.94)	**40.53** (35.41↔45.65)	**4.10** (2.13↔6.08)	**119.25** (97.24↔141.26)	**8.77** (4.20↔13.33)

ii) Statistical findings

I found that the boys who registered for the GMKL engaged in high volumes of MVPA, but very little of this was sustained at a vigorous intensity. It was also not possible to identify a "dose" of the GMKL intervention using only the accelerometer outputs. The underlying assumptions for the statistical analyses of between-group differences were satisfied (APPENDIX A20 – A21).

All of the subjects in both groups averaged more than 60 minutes of MVPA per day. The "heat maps" for bouts of MVPA of at least 10 minutes duration demonstrated consistent patterns for several of the participants. Bouts of MVPA immediately before school (7-8am) and after school (4-5pm) were common in both groups. There was also a high prevalence of MVPA bouts during school lunch break (1-2pm). Although it is likely that these were

episodes of active transport, it is not possible to identify the "type" of PA from the accelerometer readings (APPENDIX A22).

The "heat maps" for VPA demonstrated that only nine boys in group [T] and six in group [W] had one or more bouts of at least 10 minutes duration. Only one boy in each group had VPA bouts on at least three days during the week (APPENDIX A23).

It was not possible to identify a "dose" of the GMKL intervention using only the accelerometer data. There were no consistent patterns on the MVPA or VPA "heat maps" for group [T] when examining the after school training and GMKL activity days (APPENDIX A22-A23). The times for the GMKL activities could not be matched with the accelerometer results to identify the "dose" because of the haphazard nature in which the scheduled programme was delivered (APPENDIX A16).

There was no significant difference between groups [T] vs. [W] when comparing overall mean volumes of MVPA [p=0.789] or VPA [p=0.377]. Similar results were observed during the after-school GMKL training times for both MVPA [p=0.296] and VPA [p=0.549]. Despite a trend suggesting that group [T] was more active on the GMKL activity days, the differences were statistically insignificant for MVPA [p=0.135] and VPA [p=0.143] (FIGURES 8.31 & 8.32).

FIGURE 8.31: Comparison of the crude mean and 95% CI of the MVPA for the boys in groups [T] and [W] participating in the accelerometer study in Gulu

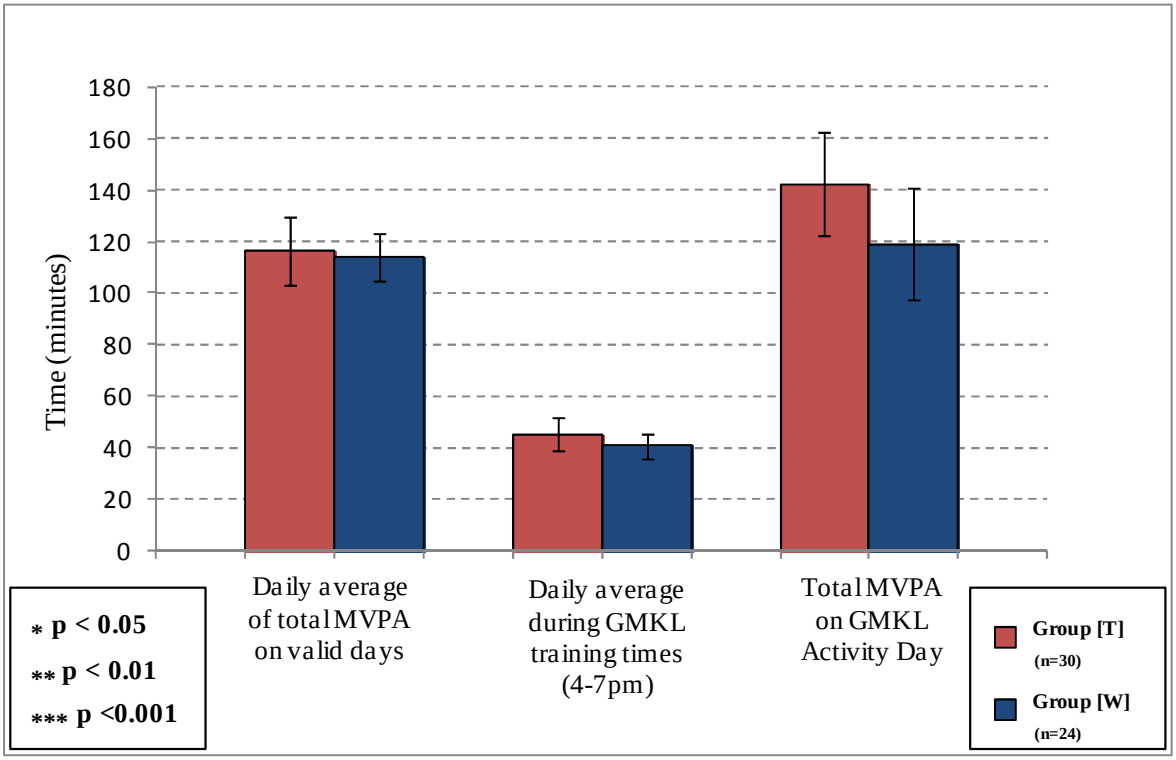
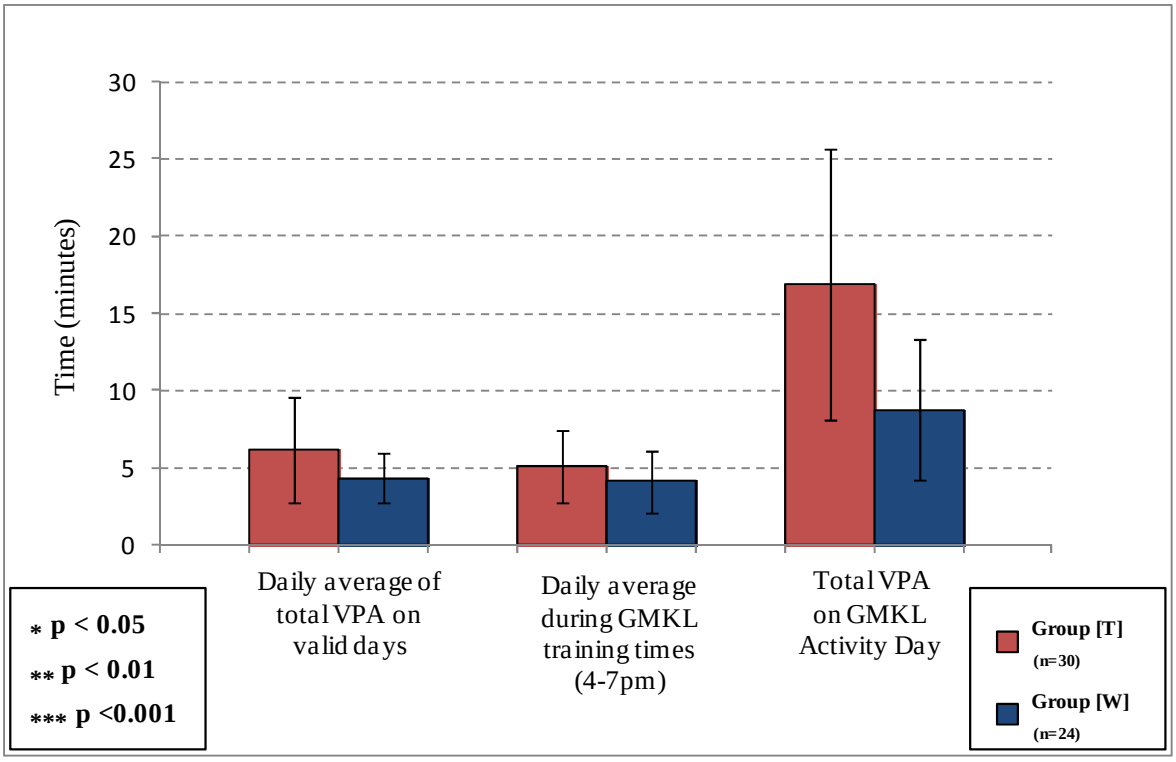


FIGURE 8.32: Comparison of the crude mean and 95% CI of the VPA for the boys in groups [T] and [W] participating in the accelerometer study in Gulu



8.5) Preliminary discussion

8.5a) Principal findings

There were differences in the distribution of the sample in each intervention group that varied according to gender, age, division, school and history of abduction.

There was some gender-based variation in the within-group changes for the PF outcomes. The decrease in BFA was statistically insignificant for all groups except the girls control in the cohort study. The within-group changes in the SBJ were also statistically insignificant except for the deterioration in performance seen in the boys intervention group. There were significant improvements in the MSFT for all groups in both studies.

The significant within-group changes were more consistent for the MH outcomes. There was a statistically insignificant trend suggesting that the MH of the boys in group [T] deteriorated. The small improvement seen for the girls in group [T] was also insignificant. In contrast, the control groups for both genders had significant improvements in their MH.

There were no significant differences for any of the outcomes in the girls sample when comparing the changes in groups [T] vs. [O]. After adjusting the data of the boys sample for baseline, age, division, school and history of abduction, the only significant intervention effect in the cohort study was in the MSFT. This finding was not sustained when comparing boys in groups [T] vs. [W]. However, the negative influence of the intervention on K was significantly different to the improvement of the wait-listed boys in the RCT.

There was no sign of a negative "disappointment effect" on the MH or PF of the boys who registered for the GMKL, but were not selected to participate. The only significant difference between groups [W] vs. [O] was the decrease in mean BFA for the wait-listed boys.

The young adolescents measured in this study appeared to be engaging in sufficient volumes of MVPA to satisfy the WHO recommendations. However, only a small proportion and insufficient bouts of this were of vigorous intensity. There was no significant difference between groups [T] vs. [W] for MVPA or VPA during the GMKL intervention period.

8.5b) Limitations and strengths

The current study built on the data collected and analysed in Chapters 6 and 7 of this thesis. Therefore, it shared several of the same limitations and strengths (SECTION 6.5b). The effect size of clinical relevance was still unknown and continued to cloud the interpretation of the results [179]. However, the concerns about measurement validity were attenuated because of the strengths of the longitudinal design and the reliability of the metrics.

There were additional concerns about the sample distribution that were related to the intervention group allocation. These included age-, division- and school-level disparities that varied according to gender and group. The differences between intervention groups were also evident for several of the outcome variables and the randomisation process did not appear to divide the boys groups evenly. Although the disproportionate sample distribution was a weakness in the current research, this was adjusted for statistically [2,179,190].

The data adjustments had implications on the statistical power of the analysis [179]. The study design meant that the intervention group size and allocation were heavily influenced by the GMKL registration process (SECTION 7.4a). The large number of P6 pupils absent from school at M1 and ineligible for the GMKL also impacted on the sample size in the current study (SECTION 6.4a). Although low levels of loss to follow-up attenuated these recruitment shortfalls, there was an inadequate sample size for the comparisons between the boys in

groups [T] vs. [W] in the RCT. Therefore, measuring the P6 classes from 10 schools in Gulu municipality was not sufficient to achieve the target ***n*** for the boys PF and MH (SECTION 6.3c). Failure to account for potential confounders when calculating the sample size further compromised the RCT study power [179,201]. However, all of the other longitudinal analyses in this study appeared to have an adequate sample size.

The sample size for the PA assessment was even smaller because there were only seven accelerometers. This prevented the PA assessment of the boys in group [O] and the entire girls sample. The measures were cross-sectional, which limited statements about causation. Although the accelerometry output trends did triangulate the results of the PF measurements, this was limited to the boys in groups [T] and [W]. The results for the girls and the "disappointment effect" for the boys could not be cross-validated.

The design of the study also meant that the group sizes were uneven for the PF and MH outcomes. This was an artefact of embedding measurement in the school system to facilitate subject tracking. However, the evaluation design was also one of the greatest strengths of this study. Accessing the subjects through schools contributed to the low levels of loss to follow-up at M2. It also enabled the successful inclusion of a randomly allocated control group for the first time in SfD research (SECTION 2.9). Adapting the methods to the specifics of the intervention optimised the level of evidence attained within the capacity and resource constraints of Gulu. This included the application of locally adapted metrics rather than a generic evaluation toolkit that was not sensitive to the intricacies of local programming. The accelerometers were the only equipment "carried" in by the researcher. This was a methodological weakness in the study that was otherwise locally feasible and sustainable (SECTION 5.4d).

The accelerometry measurements also highlighted several of the biggest challenges for conducting research in the Gulu context. Local assistance was integral to overcoming all of these difficulties. Despite thorough planning and regular monitoring, on several occasions the target subjects were not at school at the time when they were supposed to receive and/or return the accelerometers. The distribution of the accelerometers was most severely affected in the second week of measurement when school attendance after lunch was affected by severe wet weather. Consequently, PA measures were unevenly balanced between groups [T] vs. [W]. The disparate sample sizes were exacerbated when an accelerometer became faulty later in the study unbeknownst to the research team.

Only four students did not return their accelerometer as planned during the study. On all occasions this was because they were attending off-site school activities that were not previously planned. In each case, the accelerometers were retrieved by the research team with the assistance of the class teachers and peers. Although this had no effect on the measurement outcomes, poor local time management hindered another intended use of the accelerometers in this study. It was not possible to identify a "dose" of the intervention by retrospectively matching the accelerometer output to the GMKL schedule because of the disorganized way it was implemented. This also demonstrated the inability of accelerometry to identify PA "type" as has been clearly demonstrated by previous studies [213,214].

Local norms in planning and time management also presented various challenges when collecting the PF and MH measurements at M2. Haphazard school attendance and timetabling was overcome by revisiting several schools during the assessment period. This replicated the process and potential threats to measurement validity that occurred at M1 and were identified in Chapter 6 (SECTION 6.5b).

The measurements at M2 revealed that numerous children were missed when completing the baseline assessments at M1. This was primarily due to a change of school term dates at the time of M1 that caused confusion in class timetabling and absenteeism due to other school-related activities (SECTION 6.5b). This may have contributed to the under-powered sample size. It was also possible that the students who remained in class and were captured in this study at M1 were a biased sample.

There was a self-selection bias at baseline in the cohort study due to the voluntary subscription for the GMKL (CHAPTER 7) [2,179,190]. This was adjusted for in the statistical analysis. Selection bias was also likely when using the accelerometer outputs as an indicator for PA levels in Gulu [2,179,190]. Only including the boys in groups [T] and [W] selected the subjects who were the most enthusiastic and available for a PA intervention. This was exacerbated by completing the accelerometry assessment in the schools that had the highest registration rate for the GMKL. Therefore, the sample comprised boys who were the most likely to play sport and attended schools that supported this behaviour. However, there was no inherent bias when comparing the accelerometer outputs between groups [T] vs. [W].

The subject tracking system provided a reliable means to identify subjects with the same name and to maintain fidelity in the treatment groups. This presented multiple challenges throughout the intervention as subjects attempted to “jump the queue” from the wait-list by being dishonest about their names. Several coaches advocated this behaviour with a maligned focus on winning more football games. They also kept poor records of training attendance and occasionally attempted to exclude participants by putting them on the bench for extended periods during games.

However, the research team and programme personnel were vigilant about supervising these issues in the interests of fairness and to preserve the study design. No known group contamination occurred, but the attendance records kept by the coaches were not adequate for inclusion in the data analysis. Therefore, a uniform "dose" of the programme was assumed for each participant.

The subsequent intention to treat analysis of subjects who completed both M1 and M2 was limited by the inability to adjust for actual participation in the intervention [179]. Although this enabled a true evaluation of programme impact on the intended beneficiaries, the generalisability of the results outside of this population was limited. Both the biased sampling process and the specificity of the evaluation design to the intervention that was delivered contributed to this weakness in the current study [2,179,190].

There was also no systematic assessment of other known potential adverse effects of the intervention. This included officially recording the occurrence of injury [107]. Although there were several minor incidents brought to the attention of the programme implementers, only one of these was a major injury that required medical attention (i.e. a fractured ankle).

8.5c) Mechanisms and implications

There are several potential explanations for the results of the longitudinal studies. When interpreting these outcomes, the group sizes and potentially underpowered results are taken into account [179,199]. Consequently, several statistically insignificant results that demonstrated consistent trends and were theoretically reasonable were included in the discussion below.

The results of the PF and PA variables suggested that the GMKL had an impact on group [T] that was more pronounced in the boys sample. The statistically significant self-selection bias

for the girls who performed better in the SBJ and MSFT may have contributed to this gender-based disparity (CHAPTER 7). Alternatively, differences in programme implementation and uptake may be responsible. The GMKL also appeared to have an effect on the PF of the wait-listed boys, but only a marginal impact on the subjects in group [O]. It is likely that this was mediated by facets of the GMKL programme that improved local capacity for sporting activities through:

- Developing infrastructure;
- Providing football equipment and facilities;
- Training volunteer coaches, many of whom were primary school teachers.

The results of the MSFT fitted with these assertions. Although test familiarity and musculoskeletal maturation probably contributed to the improvements observed for all of the groups, there may have also been a community-wide increase in cardiorespiratory fitness [55,169,170,215]. It was not possible to tease out the relative contribution of these factors and repeat measures prior to the intervention may be indicated in future research. However, the between-group comparisons did provide some clues in the current study.

The difference between the groups in the cohort study indicated that the intervention had an impact on the MSFT that was additional to the changes seen throughout the community. Although this was only statistically significant for the boys sample, a similar trend was observed for the adjusted data of the girls groups. Therefore, it appeared that the duration, frequency and intensity of the GMKL intervention was sufficient to have an effect on aerobic capacity that was more pronounced in the boys sample [54,216].

There also appeared to be an improvement in the MSFT of the wait-listed boys that was greater than the community-wide trend. The absence of a statistical difference between

groups [T] vs. [W] fitted with this hypothesis. However, no strong assertions could be made because the improvement in the wait-listed boys was not significantly different to the change in group [O] for the MSFT.

The results for BFA also supported the existence of a direct and community-wide effect of the GMKL. Only marginal changes were expected due to the high proportion of subjects who were in the "healthy" range at baseline (SECTION 6.4b) [21,25,26,62]. However, there was a notable downward trend for BFA in all groups.

The crude decline in mean BFA for both of the girls groups was of equivalent magnitude. The statistical significance of the change in the girls from group [O] was an artefact of its larger sample size [199]. Although it was statistically insignificant, the between-group differences of the adjusted data in the cohort studies suggested a greater decrease for group [T] than the control groups for both genders. A similar trend was noted for the boys RCT results comparing groups [T] vs. [W]. This suggested that the GMKL had an effect that was additional to the community-wide trend.

The community-wide effect on BFA appeared to be more pronounced in the wait-listed boys. This was evidenced by the significant difference between the boys in groups [W] vs. [O]. The statistically insignificant difference between groups [T] vs. [W] could not be used to indicate that the groups were the same, but it did fit with these assertions [199].

The changes in the SBJ were more diverse. The stability of the SBJ results in group [O] in both genders suggested that the community-wide effect in this group was marginal. Although the results for the other groups were more disparate, they were interpreted in accordance with the direct and community-wide effects described above.

The performances in the SBJ and MSFT compared to global norms at baseline suggested that the sample was anaerobically adapted at the outset (SECTION 6.4b) [54,168,170]. Therefore, there was a large scope for aerobic improvement from the relatively low "dosage" of the GMKL [27]. Aerobic exercise of sufficient intensity causes physiological adaptation of muscular endurance at the expense of muscular power and results in the reduction of fat-free mass (i.e. muscle bulk) [54,55,217]. Although the review of Matos et al (2007) suggested that children may respond differently, the studies cited had small samples of subject who were younger and not anaerobically "pre-trained" [28].

The outcomes of this study suggested that compensation in muscular strength for cardiorespiratory gains did occur for the boys in group [T]. They experienced a significant improvement in aerobic capacity as evidenced by the results of the MSFT. This was matched by what appeared to be the largest decrease in BFA. If the decline in BFA was due to a loss in fat-free mass, the decrease in muscular power and performance in the SBJ was to be expected [54,55,217].

However, the opposite trend was observed for the mean SBJ of the girls in group [T]. This suggested that the girls in the GMKL adapted differently to the boys. Adolescent males typically experience greater hypertrophy of muscle tissue than females in response to strength training [25,28]. Therefore, it is feasible that in an anaerobically adapted sample, the boys were predisposed to losing more fat-free mass in response to aerobic training [54,55,217].

Alternatively, it is possible that the girls were exposed to a "dosage" of the GMKL that had a different effect on their SBJ performance when compared to the boys sample [54,217]. It was not possible to assess if there was a gender-based difference in PA "dosage" because

accelerometer data were not collected from the girls sample. Comparing programme engagement was also limited by the poor attendance records kept by the coaches. However, anecdotal evidence suggested that the girls teams organised fewer additional training sessions during the intervention. This was due to cultural norms that discouraged girls from playing sport and burdened them with more home duties [130]. They also played their games on a smaller playing field suggesting that they experienced less cardiorespiratory exertion on the GMKL activity days. Playing an 11-a-side football game on a full-size field may have been a critical component of the aerobic adaptation experienced by the boys in the GMKL.

The wait-listed boys demonstrated similar trends for the PF outcomes to the girls in group [T]. This suggested that the diversity in the SBJ results was related to "dosage" rather than gender. The results of the PA assessment indicated that the community-wide effect of the GMKL on the boys in group [W] was at a lower "dosage" than for those in group [T] [54,217]. Therefore, the PA associated with the GMKL had a different effect on the PF of groups [T] vs. [W] that may have been of clinical importance.

All subjects who underwent PA assessment in both groups reached the basic WHO recommendation of 60 minutes of MVPA per day [17,26]. Examination of the "heat maps" indicated that MVPA was dispersed throughout the day (APPENDIX A22). In contrast, the descriptive statistics for VPA suggested that it primarily occurred after school. The "heat maps" showed that these tended to be in bouts that were shorter than 10 minutes and the recommendations for VPA were not being met by either group [17,26]. Therefore, extended periods of aerobic exertion were not common. Considering that the measured sample was biased towards the fittest boys, these results may explain the relatively poor performance in the MSFT of the Gulu sample when compared to global norms (SECTION 6.4b) [43,168,170].

Although it was not statistically significant, the difference between groups [T] vs. [W] for MVPA and VPA was particularly marked for the GMKL activity days. This indicated that participating in organised games of football was a critical component of the PF adaptations in this study. However, the football games did not consistently translate to bouts of VPA on the accelerometer "heat maps" (APPENDIX A23). Therefore, the value of assessing PA bouts of 10 minute duration in future studies of young adolescents in Gulu is questionable.

To summarise the PF outcomes, there was a marginal community-wide effect on group [O]. This was more pronounced in the wait-listed boys and was probably because they had a genuine interest in engaging in sport. However, the boys in group [T] appeared to receive a higher "dosage" of aerobic PA than those wait-listed and the girls in group [T]. Although these groups did improve in the MSFT, the "dosage" was not sufficient to induce the physiological adaptation seen in the boys exposed to the GMKL that compromised muscle bulk and power. In contrast, the impact on PF was more general and the modest improvements in aerobic capacity were accompanied by a small increase in the SBJ. However, considering the needs of the target population (CHAPTER 6), it appeared that the PF of the boys in group [T] benefited the most from the intervention.

The MH outcomes were reasonably consistent across syndromes and have been interpreted uniformly for the longitudinal analysis. There was a community-wide improvement in the MH scores. This was evidenced by the significant declines in group [O] for both genders.

The magnitude of the crude within-group decline was similar for both girls groups and the statistical significance of the changes in group [O] was an artefact of its larger sample size [199]. The between-group comparisons suggested that the GMKL had no additional effect on the MH scores for the girls in group [T].

In the boys sample there was a consistent trend suggesting that those in group [W] improved even more than those in group [O]. This indicated that being on the wait-list for the GMKL had a positive influence on MH that was in addition to the community-wide changes. It was likely that the exposure to the GMKL mediated impact on local capacity and the prospect of playing in the next league contributed to these outcomes.

However, participating in the GMKL adversely affected the scores of the boys for all of the MH syndromes. After full adjustment of the data there was a persistent trend suggesting that the MH of the boys in group [T] had deteriorated compared to the rest of the community.

The obvious conclusion is that the GMKL made the MH of the boys in group [T] worse. However, an alternative viewpoint suggested by one GMKL coordinator was that exposure to the programme may have made the boys more comfortable to express their MH problems and created a response bias [2]. This interpretation cannot be discounted because the validity of the tool in questionnaire format was not established (SECTION 5.5b).

However, the APAI had been used in one previous longitudinal study [126]. Although it was delivered in interview format, the changes in the scores for the MH syndromes were interpreted as equally valid before and after the intervention for both genders [126]. In the current study, the intra-tester reliability of the tool in questionnaire format also suggested that there was a genuine deterioration in the boys MH (SECTION 5.5b).

There were several potential reasons that explained why the GMKL programme had a negative impact on MH in the boys sample. The league points system was designed to focus on community- and peace-building initiatives (APPENDIX A16). However, the majority of

coaches failed to recognise this and placed a large emphasis on football performance. Coalter (2010) emphasised that the "peer leaders" were important role models in SfD interventions and that they were critical mediators for achieving positive results in Gulu [130]. This view has been supported in the sociology literature [121,218]. In the GMKL the volunteer coaches filled this role and may have contributed to creating a competitive dynamic that was particularly evident in the boys leagues. Burnett et al (2001) described a similar gender-biased focus on competition for a league-based SfD intervention in South Africa [219].

It is likely that the focus on winning increased the pressure on the players to perform and caused stress [220]. This was amplified by the absence of organised sports leagues for younger children in Gulu. Consequently, the competitive nature of sport had not been gradually introduced from a very young age as is often the case in many HICs [37]. Instead, many of the participants saw the GMKL as a once-off opportunity to play football like their premier league heroes. Often this was the only reference for organised sport that was available for the participants other than their coach.

Therefore, it is possible that the competitive nature of the GMKL coupled with the pressure and expectation from coaches manifested as a stressful emotional experience for the participants [220]. This was compounded by the identification of football skill deficiencies that was often realised by coach reprimand within an environment focused on performance. It is important to recognise that in league-based SfD interventions there can only be one "winner". It appears that in the current study these factors combined to have a negative impact on the MH variables that were assessed [220].

Although this anecdotal evidence did not definitively identify the cause of the MH changes in this study, it did provide the most plausible explanation. The mediating factors were apparent

in the boys teams and their coaches more so than in the girls leagues. As expected, they were also isolated to the GMKL participants and did not affect the boys in group [W]. However, programme implementation indicators and delivery processes were not formally assessed. Therefore, it is not possible to draw strong conclusions about these potential mechanisms.

To summarise the MH outcomes, the subjects in group [O] had a significant improvement in MH during the intervention period. This may have been caused by a community-wide effect of the GMKL or due to another unknown factor. There was no additional benefit for the girls in group [T]. However, the MH of the boys in group [T] became significantly worse. Perhaps this was due to the competitive nature and pressure of the GMKL programme. The wait-listed boys experienced greater MH benefits than observed in group [O]. This may have been because the GMKL improved the local capacity for sporting activities, but the boys in group [W] were not exposed to the detrimental experiences of those in the competitive league.

Future studies should be more vigilant in collecting data about injuries, intervention compliance and the process of programme delivery. Accurate records of subject participation in the GMKL would indicate how well the intervention actually reached the intended adolescents. It could also be used to assess if the level of compliance moderated the PF, PA and MH results. Further investigation that includes more schools from under-represented divisions and/or multiple year levels is also indicated. Sample size calculations should be adjusted according to the school attendance, GMKL registration rates and the loss to follow-up proportions identified in this thesis. They should also take into account adjustment for potential confounding factors in the statistical analysis. Future studies could also match subjects prior to allocation to group [T] vs. [W] and collect more accelerometry measures from all groups. These suggestions would improve the size and distribution of the sample when repeating the same evaluation design.

8.6) Summary box

LONGITUDINAL STUDY – IMPACT EVALUATION

- The concerns about measurement validity and sample distribution identified in Chapters 5, 6 and 7 persisted in the current study. Additional age-, division- and school-level anomalies in the group allocation process were adjusted for statistically.

- This study was the first time a SfD programme evaluation had included a randomly allocated control group.

- The challenges of conducting research in the Gulu were overcome pragmatically with local support, but the effect on the final results was unknown.

- The evaluation design failed to recruit sufficient sample sizes and components of the study were underpowered. Therefore, some of the consistent statistically insignificant trends that had theoretical justification contributed to the data interpretation.

- Boys in Gulu who registered for the GMKL engaged in high volumes of MVPA, but very little of this was sustained at a vigorous intensity. This explained why the Gulu sample performed poorly on the MSFT at baseline and meant that there was a large capacity for aerobic improvement.

- The GMKL appeared to have a direct positive effect on the aerobic PF of its participants that was more pronounced in the boys sample. The deterioration in the mean SBJ of the boys in group [T] was probably due to physiological adaptation to a higher "dosage" of aerobic PA and was of minimal concern when considering the local needs.

- Participation in an organised football game on a full-sized field appeared to be a key contributing factor to higher levels of MVPA and VPA for the boys in group [T]. This may also explain some of the PF outcomes in this study.

- Development of local capacity appeared to have a positive community-wide impact on PF that was most evident in the wait-listed boys who were more keen and available to play sport.

- The negative effect of the GMKL on the MH of the boys in group [T] may have been caused by exposure to new emotional pressures and stress associated with competition. This was not the case for the girls due to an apparently less intense focus on winning.

- The MH benefits experienced by the wait-listed boys may have been because they were exposed to the improved local capacity for sport without the negative effects of the competitive league.

- It was not possible to identify a "dose" of the GMKL using only accelerometer outputs.

- Future SfD programmes in Gulu should:
  - Train coaches as "peer leaders" who are responsible for focusing on participation;
  - Reconsider the league structure to reduce the focus on competition and winning;
  - Ensure all games are played on a full-sized playing field for both genders.

- The evaluation design of future studies assessing the impact of SfD programmes should:
  - Attain the highest level of evidence with methods that are adapted to the programme model and local context;
  - Assess appropriately matched and representative groups that are of sufficient size;
  - Ensure that indicators of programme engagement, uptake and injuries are included;
  - Use statistical methods that adjust for disparities in subject distribution;
  - Include a formal process evaluation to help explain the programme impacts.

PART VI

RESEARCH FINDINGS AND IMPLICATIONS

CHAPTER 9

Final Discussion and Conclusions

9.1) Purpose of chapter

The purpose of this chapter is to collate the common themes of this thesis to address the research objectives. I identify the principal outcomes and present them according to the theoretical framework for evaluation that forms the backbone of this research. I also consider the common limitations and strengths of the studies completed. The interpretation of the final results is described in the context of established SfD conceptual models and draws upon my fieldwork observations from Gulu. The implications of these studies are used to make recommendations for future programmes and research. I conclude by summarising my contribution to the evidence base and the implementation of SfD programmes.

9.2) Principal findings

The sequence of studies in this thesis did not follow the same order that was described by Nutbeam et al (2006) in the Stages of Evaluation for Health Promotion Interventions (FIGURE 1.1) [2]. However, all of the key research questions in the first three stages of this model were addressed (SECTION 1.3). I report the principal findings, according to the order described in the original theoretical framework, to provide an initial insight into how the chronology of programme development impacted on the final outcomes.

9.2a) Problem definition – what is the problem?

Defining the problems to be addressed is the first stage of evaluation. This process "identifies the major health problems experienced within a defined population, the known causes of these problems, and the scope for change in those causal and contributory factors" [2]. The results should guide the initial formulation of programme objectives for health interventions. A summary for this phase of the SfD evaluation that I completed in Gulu is presented below according to the components described by Nutbeam et al (2006) [2]:

i) Epidemiology and demographics (CHAPTER 2)

- NCDs and MH disorders are significant problems throughout the world that are worsening with urbanisation and contributing to the development of the double burden of disease in many LMICs;
- The diverse health needs of post-conflict communities vary throughout the world, but there are several common characteristics that may suggest an increased risk for NCDs.

ii) Social, behavioural and organisational research (CHAPTER 3)

- Gulu is a post-conflict zone in northern Uganda that is likely to experience the double burden of disease if current patterns in health care, PA behaviour, socio-economic development and urbanisation continue.

iii) Community needs analysis (CHAPTER 6)

- Young adolescents in Gulu engage in high volumes of MVPA, but very little of this is sustained at a vigorous intensity;
- A high proportion of the young adolescents in Gulu are in the "healthy" weight range, which suggests that there is not widespread acute malnourishment in the community;
- The young adolescents in Gulu perform better than global norms in the SBJ, which suggests the development of good muscular power;
- The young adolescents in Gulu perform worse than global norms in the MSFT, particularly the girls, which suggests poor cardiorespiratory fitness;
- A history of abduction is strongly associated with persistent poor MH and is particularly pronounced in girls;
- There is no association between PF and MH for young adolescents in Gulu, which suggests that a SfD intervention that depends on PA and/or PF mediators to improve MH is unlikely to be successful.

9.2b) Solution generation – how might the problem be solved?

The second stage of evaluation comprises generating potential solutions for the identified problems. This process "clarifies the potential content and methods for intervention, and further defines the different needs of populations" [2]. The results should contribute to refining the programme objectives and developing the theoretical rationale for their achievement. A summary for this phase of the SfD evaluation that I completed in Gulu is presented below according to the components described by Nutbeam et al (2006) [2]:

i) Literature review (CHAPTER 2)

- Increasing PA and PF have been found to positively affect adolescent RFs for NCDS (including cardiorespiratory fitness and musculoskeletal health) and MH disorders that may track into adulthood;

- The SfD sector is rapidly expanding and claims to have the same health benefits that are mediated by increasing PA and PF in the target population;
- There is a paucity of published evidence supporting the claimed impact of SfD programmes on health-related outcomes.

ii) Theory development (CHAPTER 2)

- Existing conceptual models that explain the link between SfD programmes and health-related outcomes are scarce and rudimentary;
- I have formulated a theoretical framework describing the processes that link the needs of the community and intended reach of the programme with the actions, outputs and claimed outcomes of a SfD programme;
- The model described by Bouchard (2001) is cited in a literature review of the SfD sector as the mechanism explaining the impact of sustainable sport programmes on the health-related outcomes assessed in this thesis [107,138].

iii) Pretesting methods and materials (CHAPTER 5)

- It is pragmatically and ethically feasible to assess the PA, PF and MH of young adolescents in Gulu using measurement methods adapted to the local context;
- All of the modified methods for PF assessed in this thesis have good intra- and inter-tester reliability;
- The MH measurement protocol assessed in this thesis has good intra-tester reliability and internal consistency, but the results from APAI questionnaires and interviews are not interchangeable;
- There appears to be good compliance with PA measurement even though the accelerometers were “carried” in to Gulu for this component of the assessment;
- The validity of the adapted measurement methods used in this thesis is not known, but the modifications from previously established protocols are minimal;

- Local support in Gulu is critical for the effective implementation of the measurement methods developed in this thesis.

9.2c) Innovation testing – was the solution effective?

Testing whether the proposed solution was effective in addressing the identified problem is the third stage of evaluation. This process "assesses the extent to which defined objectives can be achieved operating in the best possible conditions for success" [2]. The focus should be on the impact of the programme and indicate if it "worked" [2]. In this case, Gulu was the most accessible and well resourced municipality in northern Uganda. A summary for this phase of the SfD evaluation that I completed in Gulu is presented below according to the impact and process components described by Nutbeam et al (2006) [2]:

i) Impact and outcome assessment (CHAPTERS 7 & 8)

- Voluntary registration for the GMKL results in a self-selection bias, particularly pronounced in girls, for subjects who are physically fitter (i.e. the subjects in greatest need are not reached by the programme);
- A history of abduction is not a barrier to registering for the GMKL;
- The GMKL appears to have a direct positive effect on the aerobic PF of its participants;
- The GMKL contributed to developing local capacity and resources that appear to have a community-wide positive impact on PF, which is most evident in the wait-listed boys who are more keen and available to play sport;
- The boys in the GMKL intervention group appear to be exposed to a higher "dosage" of aerobic PA, which may be due to better programme uptake than the girls and regular participation in an organised football game on a full-sized field;

- A higher “dosage” of aerobic PA may explain why the boys in group [T] experienced large gains in cardiorespiratory fitness (MSFT) and deterioration in muscular power (SBJ);
- It is not possible to identify a “dose” of the GMKL using only accelerometers;
- The current GMKL model may introduce new emotional pressures that have a negative impact on participants who are focused on competition and winning;
- The competitive focus appears to be more pronounced in the boys than the girls when comparing the GMKL participants;
- Wait-listed boys appear to experience large MH benefits because they are exposed to improved local capacity for sport programmes, but do not endure the negative effects of a competitive league.

ii) Process evaluation (CHAPTERS 7 & 8)

- Although there are no formal process evaluation results presented in this thesis, the delivery was monitored in accordance with the theoretical model that I established in Chapter 2 (SECTION 2.10) and I elaborate on this insight in the discussion below.

9.3) Limitations and strengths of thesis

The limitations and strengths of the studies that I have conducted were described in detail in the previous chapters. Most of these were caused by forced adaptations to the challenges of conducting research and implementing a SfD programme in Gulu. This included the local pragmatic difficulties related to:

- Planning and time management;
- Infrastructure and facilities;
- Resources and capacity;
- Community access and communication;
- Cultural norms and expectations.

These challenges were compounded by ethical concerns about conducting research in Gulu. There was a clear tension between the agendas of development agencies and academia [205]. A culturally sensitive approach that was sustainable and locally reproducible was not always synonymous with obtaining the optimal level of evidence using rigorous scientific methods. Although all of the challenges faced in Gulu did force compromises in the research methods, they did not prevent any of the studies being completed. Perhaps the greatest strength in this thesis was demonstrating that it is possible to conduct rigorous and ethically responsible research in post-conflict and LMIC settings. Building the capacity of locally established systems was integral to achieving these objectives [114]. I believe that this enabled me to make a contribution to the evidence base for SfD interventions that was contextually relevant and was sensitive to the development agenda.

There were several pervasive themes regarding the limitations and strengths of the studies that I conducted in Gulu. These can be categorised according to the measurement methods,

evaluation design, subject recruitment and data interpretation. It is important to note that often the biggest limitations of the research were also related to its greatest strengths.

9.3a) Using modified measurement methods

The adaptation of the measurement methods to the local context was practically and ethically necessary for this research. However, it increased the emic characteristics of the metrics and made the results less generalisable to other settings and studies. Stringent application of scientific rigour would have required the application of gold standard metrics using calibrated equipment in a controlled environment. This was not possible for assessing PF and I used field tests instead. The methods chosen were the most commonly applied when assessing large cohorts and there were sufficient bodies of evidence to justify this approach [55,168-170].

Standardised methods have been described for the BFA, HFA, SBJ and MSFT field tests [44,51,55,62,162,164,165,168-170]. It was possible to "carry" in the equipment and personnel required to conduct the PF assessments according to previously validated protocols. However, it was not logistically feasible to implement the standardised methods that were developed in HICs. Perhaps more importantly, the "carrier" approach is contrary to the priorities of development agencies that focus on building local capacity and utilising sustainable resources available in Gulu [112]. The only exception to this was the PA measurements, which adapted the use of accelerometers to the local context.

The modification of the field tests for PF meant that the validity of the measurement methods was uncertain (SECTION 5.5b). There were similar concerns for the modified MH measures. Although the APAI had been validated in interview form, delivering it as a questionnaire had poor inter-method agreement and its validity was not assessed [128].

The lack of metric validation meant that the extent of measurement error and bias was unknown in all of the studies. This warranted caution when interpreting the results and compromised the certainty of the conclusions drawn from the PF and MH data [2]. It was particularly pertinent in the cross-sectional analyses when assessing the community needs and the GMKL programme reach (SECTIONS 6.4c & 7.4b). Satisfactory inter-tester reliability and consistency in the research team meant that the measurement of change over time was less susceptible to the use of unvalidated metrics. Therefore, the results of the longitudinal studies were more robust to this limitation in the research methods (SECTIONS 8.4b & 8.4c).

However, it is important to recognise that the modification for all of the metrics was minimal. Similar adaptations have been made in other studies when applying field tests in LMICs with little consideration for the effect on the validity of the results [55,130,168-170,183-185]. Although the deterioration of the quality of the measures was uncertain, it was likely to be small and successful adaptation to the local context was actually a strength of the research.

9.3b) Developing appropriate evaluation designs

The evaluations implemented in this thesis were designed to optimise the level of evidence obtained within the pragmatic and ethical constraints of the local context. It was important to find a balance between generating evidence for intervention effectiveness and collecting data for programme development [2]. Consideration of local ownership and capacity building compromised the scientific rigour of the research methods [114]. It was also not logistically possible to apply an independent and blinded approach to the group allocation or measurement procedures in any of the studies in this thesis.

The feasibility and reliability assessment of measurement methods in Chapter 5 highlighted that effective evaluation design needed to be sensitive and adaptable to the local environment

(SECTION 5.5b). The research methods had to be initially adjusted to the Gulu context and allow for its unpredictable nature by being flexible [121]. The subsequent loss in evidence quality was a pragmatic trade-off that enabled the studies to take place.

The evaluation of the programme needs, reach and impact was specifically devised for the GMKL intervention in Gulu (SECTIONS 6.3a, 7.3a & 8.3a). The adaptation of the evaluation design to the local context and systems increased the specificity of the research. Although this optimised the quality of evidence, it limited the generalisability of the methods and results to other SfD interventions [2]. This was also evident in the specific search terms that I used in my systematic literature review (SECTION 2.9). However, the findings of Cronin (2011) supported this approach and suggested that more focused literature reviews were needed [114]. The specificity of my research has also been justified by a growing recognition that the post-conflict context presents unique health issues that are worthy of investigation [139-142].

The results of my literature review indicated the novelty of the evaluation design developed in this thesis for the outcomes and population of interest (SECTION 2.9). The majority of the published articles had a sociology focus and were descriptive [71,115-125]. There was only one SfD literature review that focused on health-related outcomes and it cited papers that were not contextually relevant [107]. The programme evaluations that had been published either did not focus specifically on SfD interventions or did not report empirical data relevant to the outcomes of interest [126,130-133]. None of the longitudinal studies identified in the literature review included a randomly sampled control group.

However, there is growing support for the implementation of strong evaluation design in the SfD sector [114,120,135-137]. As an example, Zak Kaufmann from the London School of Hygiene and Tropical Medicine is currently conducting an RCT assessing the impact of Grassroots

Soccer programmes on HIV/AIDS transmission in South Africa [221]. This work and the research that I completed in Gulu are seminal in the SfD sector. They are the first studies to use an experimental design to assess causation for the health outcomes of SfD interventions.

As previously described, I had planned to be an independent evaluator of a TKL intervention in Gulu. However, due to poor programme management I was required to coordinate the intervention and broker partnerships with OAP and YCP. Although I ensured that I remained blinded when initial group allocation was completed for the GMKL, I was not masked to programme implementation. This was a weakness in the evaluation design that was a direct result of the pragmatic challenges that I faced in Gulu. Managing the intervention was also time-consuming and compromised my evaluation of programme processes. My initial intention was to use the theoretical framework that I had developed to assess process indicators for programme delivery (SECTION 2.10). Reliable attendance records for each team would have also assisted in assessing participation levels and intervention "dose". Assessing the knowledge and attitudes of the coaches using semi-structured interviews or surveys could have provided an insight into training quality and their focus within the programme. Focus groups with the beneficiaries could have also identified differences in programme delivery and exposure. However, my process evaluation was limited to assessing programme reach (CHAPTER 7) and informal observations recorded in my fieldwork diary. These compromises have limited the interpretation of my impact focussed results [2].

9.3c) Recruiting and retaining subjects

The subject recruitment for all of the studies in this thesis was successfully conducted through primary schools. This approach effectively harnessed the most robust system that existed in Gulu for accessing a large sample of young adolescents. Embedding the assessment in the school PES classes facilitated subject tracking and retention.

An alternative approach that had been previously implemented in Gulu was to recruit subjects for evaluation through the community and conduct the assessments at an independent testing venue [126,130]. Although this option was considered for all of the studies in this thesis, it is likely that it would have caused a self-selection bias in the sample measured [2,179,190]. Assessing the reach of the GMKL programme demonstrated that this was the case for voluntary engagement in physical exertion (SECTION 7.4b). Prior to conducting the needs analysis, I hypothesised that young adolescents willing to engage in PA would also have relatively good MH and be more inclined to attend an independent testing venue for PF [130]. Therefore, it was likely that using a community-based voluntary approach to subject recruitment for evaluation would not have engaged a representative sample for PF and/or MH levels in Gulu. This would have limited the assessment of the metric quality to a narrower range in the pilot study (CHAPTER 5). It also would not have provided a true indication of the broader community needs and the reach of the GMKL programme (CHAPTERS 6 & 7).

However, using an independent testing venue would have assessed the metrics and needs in the population that the GMKL programme was most likely to reach. Therefore, these concerns about self-selection bias during measurement probably would not have affected the RCT that I conducted [2,179,190]. Previous studies identified in the literature review integrated the baseline assessment into the programme registration [130,131,133]. I could have used a similar approach, but it was not known if this would measure a sufficient number of subjects to form a wait-list control group. Therefore, nesting the RCT within an existing cohort study that accessed subjects through schools was the only sure way to recruit a control group.

Tracking and retaining the subjects for follow-up testing would have also presented logistical and ethical concerns if the measurement procedures were community-based. There were

pragmatic difficulties related to local transport and time management. This was the primary reason for conducting the pilot testing in a boarding school setting (SECTION 5.3c). These challenges could have also been overcome by providing an incentive to participate in the studies. The provision of a transport "allowance" was a widely accepted practice in Gulu. However, it was seen locally as a "payment" to augment income and was rarely actually used for transport. Therefore, its potential coercive influence was accentuated and there was limited logistical benefit [161]. The ethical concerns of development agencies regarding aid dependency also precluded the use of incentives for participation in research in Gulu [209].

Although the subject recruitment methods used in this thesis accounted for all of these concerns, they did cause uneven group sizes for various studies. They also contributed to the disproportionate distribution of potential confounding factors between groups (SECTIONS 6.4a, 7.4a & 8.4a). However, statistical methods were used to address these issues.

The sample size calculations for the evaluation of programme needs, reach and impact took into account various factors that influenced subject recruitment (SECTIONS 6.3b, 7.3b & 8.3b). This included adjusting for uneven group sizes by using estimates of programme registration and the incidence of history of abduction in the target population. The calculations also accounted for the anticipated levels of school attendance, subject eligibility and loss to follow-up. Although this appeared to recruit a sufficient number of subjects for most of the studies, the results of the RCT were underpowered [179]. This may have been further compromised by failing to account for statistical adjustments for potential confounders in the sample size calculation [179,201]. An increase of more than 10% is unlikely to be needed for each confounding variable [201].

The absence of sufficient data to conduct a power calculation for the pilot study may have also contributed to insufficient subject recruitment when assessing feasibility and reliability (SECTION 5.3b). These were notable limitations when interpreting the results.

9.3d) Interpreting the data appropriately

A common assertion in the SfD sector is that the purpose of M&E is to "prove success" [135]. This approach stems from the rhetoric that promotes sport as an essentially positive intervention [121]. It implies that the role of research is to provide data that confirm the claimed outcomes of SfD programmes [135]. Evaluators with this mindset are at risk of ignoring important assumptions and using retrospective "data dredging" to identify the necessary evidence for the relevant stakeholders. In the systematic literature review completed for this thesis, poorly defined analytical approaches were more prevalent in the SfD reports from the grey literature that were not peer-reviewed [130-133].

However, rigorous evaluation should objectively assess the assumptions of SfD interventions rather than just seek to find proof of effect [2]. This requires a clearly defined analytical approach prior to commencing the data collection to avoid "data dredging" and ensure the results are not over-interpreted. The number of outcome measures and statistical tests performed in this thesis also introduced the problem of multiplicity and increased the risk of a type I error [222]. A Bonferroni or Holm adjustment is often applied to the p-value in this context. However, I applied an approach advocated by Zhang et al (1997) in selecting a primary endpoint (i.e. MSFT) and designating the other outcomes as secondary endpoints [223].

It is also important to note that there was a specific purpose for these analyses in all of the studies. Each PF outcome variable measured a distinct characteristic and aggregating the results into a composite score was not appropriate [55,222]. The MH outcomes were assessed

individually to recognise the distinction between local syndromes [127]. Following the precedent set by Bolton et al (2007), the APAI scores were also combined to provide an indication of disorders according to a "Western" construct of MH [126]. The targeted approach to analysis was a strength of the studies in this thesis.

In Chapters 6-8, the statistical tests included adjusting the crude data for the discrepancies in sample distribution and the differences at baseline (SECTIONS 6.4, 7.4 & 8.4) [2,179,190]. This increased the number of analyses and the likelihood that I would identify statistically significant results by chance (i.e. a type I error) [179]. Therefore, the magnitude of the statistical findings was taken into account and the theoretical rationale for the results was also considered. On several occasions, I applied my field observations to interpret various study outcomes. Although this provided some insight into the programme delivery, the lack of formal process evaluation was a weakness in this thesis. Consequently, it was not possible to adjust the data for actual programme uptake and an equal "dosage" of the GMKL was assumed for all of the participants in Chapter 8 (SECTIONS 8.4b & 8.4c).

A consistent problem faced when interpreting the data throughout this thesis was the lack of clinically relevant parameters of effect size for the SBJ, MSFT and APAI. Although the analyses that I performed identified statistical significance, they did not indicate whether there were discernible health implications for these findings. This highlighted the paucity of evidence linking performance in the SBJ and MSFT field tests with criterion-referenced recommendations for health (SECTION 2.6a). It also reflected the relative infancy of the APAI as a measurement tool for MH in northern Uganda [127]. The inability to interpret whether the health impact of the GMKL was clinically relevant was a limitation of this thesis.

9.4) Mechanisms and implications

The mechanisms and implications for the various components of this research were considered in detail in the previous chapters. Each of these studies added to the evidence base for the health and SfD sectors in post-conflict settings. However, perhaps the most important findings of this research were the broader themes related to SfD programme development that became evident when applying a structured evaluation process [2].

As previously discussed, I completed the first three stages of the evaluation framework described by Nutbeam et al (2006) for health programmes (FIGURE 1.1) [2]. Their dispersion throughout the thesis was due to a complex dynamic between the funders, implementers and evaluators that is common in the SfD sector [71,115,121,125,135,136]. Perhaps this haphazard approach to the development of the GMKL contributed to several discrepancies between the programme aims and delivery that may explain some of the study results.

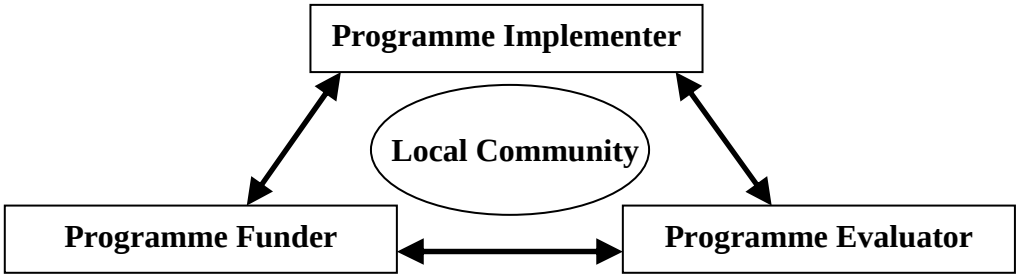
The relationships between the objectives, actions and intermediary outputs of the relevant stakeholders for a SfD intervention are depicted in my theoretical framework (FIGURE 2.5). It also describes the link to the final health-related outcomes via the mechanisms described by Bouchard (2001) [138]. These theoretical models form the basis for this final discussion which also draws on informal process observations contained in my fieldwork diary.

9.4a) Programme objectives

At the centre of my theoretical framework are the "aims and objectives formed in response to local needs and the intended reach of the SfD programme" (FIGURE 2.5). The funders, implementers and evaluators for the intervention should all contribute to this process. Consultation with the community is also critical for identifying their priorities and promoting sustainability through local ownership [2,131,133,135].

Ideally, programme objectives are informed by the planning and development phases described by Nutbeam et al (2006) [2]. Clearly defining the problem and conducting a formative evaluation of potential solutions should direct all of the relevant stakeholders towards concerted goals. Therefore, evaluators can play a key role in identifying the local issues that SfD programme implementers ought to address and donors need to fund. The local community should remain at the centre of these relationships throughout the development and implementation of the intervention (FIGURE 9.1) [125,131,133].

FIGURE 9.1: The relationships between the different types of stakeholders for developing and delivering a SfD programme



However, in reality the synchronous development of organisational relationships and programme objectives rarely occurs in the SfD sector. Numerous authors have referred to the tensions that exist among stakeholders [70,71,115,120,121,125,135-137]. Funders in the SfD sector are demanding increasing accountability from implementers and see evaluators as a means to enforce these sentiments [125]. This means that implementers often perceive evaluators as potential threats who divert resources from interventions and are only necessary for satisfying donor demands or acquiring more funds [120,125]. Therefore, evaluators are not typically integrated into programme design and are pressured to produce results on tight budgets that satisfy implementers and/or funders rather than report objective outcomes [135,136].

The dysfunctional relationships between stakeholders have several practical manifestations. Perhaps most importantly, the concerns of the local community are often displaced by the priorities of the other stakeholders [115]. This means that implementers are typically granted funding based on overly ambitious objectives that have been developed without conducting an appropriate formative evaluation or assessing the local needs [125].

However, this was not the case for all of the interventions identified in my literature review (SECTON 2.9). The programmes evaluated by the SAD in post-conflict regions had conducted a "situational assessment" [131,133]. It appears that the objectives of these programmes were at least partly informed by this process and included the development of theoretical "problem trees" by the participants [133]. As previously discussed, Coalter (2006 & 2010) developed a logic model for the interventions that he evaluated, but this was not informed by a comprehensive needs analysis in Gulu [129,130]. The evaluation completed by Betancourt et al (2007) "piggy-backed" on existing interventions. Although the results of the study indicated a local need, this was retrospective and it is not known whether this was assessed by the implementing organisations prior to delivery [126].

The current GMKL in Gulu was not informed by an extensive planning and development phase as described by Nutbeam et al (2006) [2]. This was evidenced by the order in which the components of problem definition and solution generation were presented in this thesis (SECTION 9.2). However, perhaps most importantly, none of these components of the evaluation were completed before the programme was planned by the implementers and the necessary funds were acquired.

The disconnect that exists between the stakeholders highlights an important "chicken and egg" scenario in the SfD sector. It is very difficult for implementers to work with evaluators

to plan SfD programmes without sufficient resources. However, donors typically require clearly articulated objectives from implementing partners prior to providing them with funding. This does not acknowledge the contribution that evaluators can make to the formative stages of programme development [2]. There is clearly an indication to increase resources for prospective evaluation as part of SfD intervention planning [120,125,137]. This should focus on assessing specific community needs and establishing locally relevant theoretical frameworks for programmes prior to implementation [2,121,135,136].

The objectives of the GMKL were further complicated by the partnership of two previously independent implementing organisations (i.e. YCP and TKL). This avoided programme replication in Gulu and was the most economical use of resources to increase the size of the intervention. However, there was some diversity in the objectives of the organisations that may have been diluted when the partnership was formed.

The YCP was a "plus sport" organisation that primarily focused on peace-building initiatives [129]. Since 2008, the support of OAP had shifted YCP activities towards using football as a vehicle to achieve their peace-building objectives. It also assured the YCP of medium-term funding despite the lack of previous evaluation or the development of a theoretical framework for its impact.

TKL was established in Gulu in 2003 and came from the other end of the SfD spectrum. It was a "sport plus" subsidiary of the KKL, which focused on developing sport for young people [129]. The initial objective of TKL was to assist with the reintegration of previously abducted children in Gulu. However, the aims and donors of TKL quickly diversified. Funding for the TKL activities in this thesis was provided by the ASC and was based on a proposal that outlined multiple objectives. These included health-related outcomes for IDPs

in Gulu, but were formulated without effectively assessing the relevant needs or priorities of the local community.

It is also important to recognise my role in formulating the objectives of the GMKL. Prior to arriving in Gulu, I was not expecting to contribute to managing the intervention or to evaluate a programme that had a strong peace-building agenda. However, I did bring a health focus to the partnership that influenced the overall programme objectives.

A series of stakeholder meetings were also conducted in Gulu prior to commencing the GMKL activities. Representatives from the local government, schools, sports clubs and NGOs were consulted. However, the purpose of these meetings was primarily to gather the support of the community for the GMKL intervention that had already been designed. These local stakeholders did confirm the importance of several of the planned programme activities (e.g. football rather than netball interventions for girls). However, several of the issues raised during these meetings also highlighted a lack of awareness of the local needs and demonstrated poor theoretical understanding of sustainable SfD programmes (e.g. a request for an electronic scoreboard when most playing fields in Gulu did not even have goalposts).

The programme objectives that were finally agreed upon by all stakeholders for the GMKL fused all of these influences. The aim was to implement an intervention that provided a vehicle for delivering peace-building initiatives and recognised the broader benefits of sport and PA on physical fitness and health. There was a particular focus on MH as it bridged the peace-building and health agendas of all stakeholders [142].

It is important to recognise the repercussions of formulating the GMKL objectives without a true community needs analysis or formative evaluation [2]. The initial stages of the evaluation

that I conducted were retrospective in terms of the GMKL programme planning and development. Therefore, it was purely fortuitous that several of the needs identified in my evaluation matched the objectives of the programme. This included the review of the limited and dubious data that suggested a health-related need for PA interventions in Gulu (CHAPTER 3). Establishing field tests that reliably assessed health parameters associated with PA in Gulu enabled me to confirm that the adolescents had poor cardiorespiratory fitness (CHAPTERS 5 & 6). Despite the paucity of evidence and theory for SfD interventions in post-conflict settings, the deficit in the MH of previously abducted adolescents also suggested that a PA intervention may be of benefit (CHAPTERS 2 & 6).

However, the formative evaluation completed in this thesis also indicated that it was unlikely that MH would be improved by simply increasing PF or PA (SECTION 6.4d). Therefore, there was a mismatch in the MH objective of the GMKL and the theoretical understanding of how this would be achieved. If the programme planning and development had been adequately informed by this formative evaluation, the stakeholders could have reassessed their objectives and/or delivery methods. This would have prompted the development of an alternative theoretical model for MH to the one described by Bouchard (2001) [138]. However, this did not occur and the repercussions for the GMKL implementation and outcomes are considered further below.

9.4b) Programme implementation

The actions and intermediary outputs of my theoretical framework describe the processes of delivering a sustainable SfD programme that can be tailored to the outcomes of interest (FIGURE 2.5). As previously discussed, I was not able to conduct a formal process evaluation of the GMKL. However, as a member of the programme coordinating team, I did have some insight into the implementation of each action and the realisation of associated

intermediary outputs. These observations corresponded to the process evaluation of innovation testing in the model described by Nutbeam et al (2006) [2]. In some instances, I was also able to refer to the success of specific outcomes to provide an indication of upstream delivery of actions and outputs.

The personnel and volunteers of the GMKL appeared to implement all of the actions described in my theoretical framework. However, the delivery of each of these varied in both quantity and quality according to the outcome of interest.

The GMKL funders and implementers did effectively "mobilise and distribute resources" for the football component of the SfD programme in Gulu. This was evidenced by the community-wide effect seen for the PF outcomes (SECTIONS 8.4b & 8.4c). Several of the volunteer coaches were primary school teachers in Gulu. Therefore, the provision of sports equipment and training in football coaching were transferred to the target schools in the evaluation. The rehabilitation of several football facilities in Gulu as part of the programme may have also contributed to the community-wide changes observed.

The coordinators of the GMKL placed a high priority on "teaching and training local employees, volunteers and leaders" before the football season commenced. However, its implementation appeared to be misdirected and was poorly reinforced during the programme. The impact of the intervention on the MH of the participants supported the observation that the coaches and players were focusing on winning the football games rather than participation (SECTION 8.5c) [220]. This problem was avoided in the post-conflict programmes previously evaluated by the SAD and Betancourt et al (2007) because they focused on play rather than competitive sport [126,131,133]. Although it appeared that most of the GMKL participants were

still included in the activities, the focus on winning football games was a clear departure from the original objectives of the programme.

Word of mouth is the most powerful medium to "advocate and disseminate information" in Gulu. This was augmented by the GMKL coordinators through weekly newsletters and regular radio announcements. Various stakeholder meetings also contributed to informing key members of the community about the programme. This appeared to circulate information about the football league successfully, but failed to have the same effect for the associated peace-building agenda. This may reflect the universal appeal of sport that is often described in the rhetoric of the SfD sector [68]. However, it is also likely that the differences in information dissemination were influenced by the varying levels of local engagement of the GMKL football logistics and peace-building coordinators.

The "completion of research and programme M&E" had a strong focus on the health objectives of the programme (CHAPTERS 5 – 8). Although there were some efforts to assess the peace-building components, unvalidated metrics in a pre-experimental design were used and were poorly coordinated by the programme personnel. Therefore, a large proportion of the objectives were not effectively evaluated.

The "participation in conferences and international conventions" was also limited. Only two of the local personnel in Gulu regularly attended workshops and these were focused on football coaching. Although several staff members in TKL headquarters in Kampala had broader experiences, they were not based in Gulu during the intervention and had minimal input. The coordinators of OAP were in a similar situation. Therefore, I provided the main contribution to this action for the GMKL. This demonstrates the dearth of opportunities available for local evaluators to share their experiences and is reflected in the paucity of

evidence in the SfD sector (SECTION 2.9). It also highlights the importance of feedback to local communities from events where relevant information has been shared.

As indicated in the theoretical framework, achieving the primary intermediary outputs was related to the implementation of the associated actions (FIGURE 2.5). Therefore, it follows that the GMKL was effective in developing "local infrastructure, facilities and equipment" for football. It also contributed to "local knowledge and skills" that were specific to coaching sport and the evaluated health-related outcomes. The peace-building objectives were notably neglected for all of the primary intermediary outputs and there was consequently a strong "sport plus" focus [129]. This was particularly evident in the development of "promotional materials using various media", "guidelines for best practice" and "local community, government and international awareness".

The initial stages of programme implementation had subsequent downstream repercussions for the secondary intermediary outputs (FIGURE 2.5). These can also be considered short-term impacts that are directly related to intervention delivery and the outcomes of interest [2]. As expected, the "local capacity to implement sustainable SfD programmes" that had a "sport plus" and health focus appeared to become well established in Gulu [129]. "Understanding the programme effect" on health-related outcomes improved and the "local support" for the football component of the intervention also developed. Although the influence on "policy development" and "information sharing between stakeholders" was marginal at the time of implementation, it is intended that the outcomes of this research will contribute to these processes. This highlights the importance of local and broader dissemination of evaluation findings in the establishment of effective SfD programmes [135]. However, it is important to recognise the health-related focus of these outputs and continued neglect of the peace-building agenda.

Therefore, it appears that all of the actions and intermediary outputs described in the theoretical framework were accomplished to varying degrees for the health-related objectives of the GMKL (FIGURE 2.5). As previously discussed, their implementation was not guided by a comprehensive formative evaluation and there was a mismatch between the MH objectives and the theory for programme delivery (SECTION 9.4a). The retrospective nature of the formative analysis meant that the programme implementation was not adapted accordingly. Although MH was prioritised by all stakeholders, alternative mechanisms to address it were not identified or delivered within the programme.

Coalter (2011) suggested that sport is only part of the greater social phenomenon that surrounds it in SfD interventions [135]. This indicates that the psychosocial impact may be a separate "plus" component of SfD programmes that is not mediated by PA. If the formative evaluation had informed the planning and development of the GMKL, the implemented actions would have been adapted according to a theoretical model developed specifically to achieve the desired MH outcomes. Although separating MH from PA and PF would have further diluted the actions of programme delivery, it would have addressed all of the objectives of the stakeholders more effectively.

A prospective community needs analysis could have also informed the delivery of the GMKL programme. Identifying the populations most at risk for the outcomes of interest would have enabled targeted implementation of actions [137]. This may have influenced who the intervention reached and improved the registration rates of those most in need. The self-selection bias for the GMKL was not surprising considering it was a health-focused and voluntary "sport plus" programme that did not target any of its actions towards at-risk groups in the community (SECTION 7.5c).

9.4c) Programme outcomes

This section refers to the impact and outcome assessment component of innovation testing described by Nutbeam et al (2006) [2]. The programme outcomes are depicted in the outer ring of my theoretical framework (FIGURE 2.5). I have previously discussed the need for specific conceptual models for the various claimed impacts of sustainable SfD programmes (SECTION 2.10). The framework described by Bouchard (2001) has been cited in the SfD literature and bridges the gap to the health-related outcomes assessed in this thesis (FIGURE 2.6) [107,138].

Despite assessing only boys in the accelerometry study, the intervention group did appear to engage in more PA than their wait-listed peers (SECTION 8.4d). Although the results were statistically insignificant, the sample size was small and the biggest difference was seen on the days when the GMKL activities took place. These changes were observed in a sample of young adolescents who were most likely to be physically active in Gulu. They demonstrated high levels of MVPA, but low levels of sustained VPA. This fitted with the cross-sectional results for PF compared to global norms at baseline and suggested that the adolescents in Gulu were anaerobically adapted at the outset (SECTION 6.4b) [54,168,170].

Therefore, it was not surprising that the GMKL appeared to have a positive impact on the aerobic capacity of its participants [27]. For the boys, these changes were significantly greater than those observed in the rest of the community (SECTION 8.4b). Although there appeared to be no additional benefit in cardiorespiratory fitness when compared to those wait-listed, the boys exposed to the GMKL did experience a loss in muscle power (SECTION 8.4c). This was probably a compensatory adaptation to higher levels of sustained aerobic PA [54,55,217].

The self-selection bias for fitter adolescents may have limited the potential change in cardiorespiratory fitness for those who registered for the GMKL (SECTION 7.4b) [27,54]. This was particularly pertinent for the girls exposed to the intervention. They were significantly fitter than their peers at baseline and experienced similar changes to the control group for each of the outcome variables (SECTION 8.4b). However, it is also likely that lower levels of programme engagement and PA participation contributed to this finding (SECTION 8.5c).

The self-selection bias also meant that the intervention did not reach those most in need for the PF outcomes assessed in this thesis. As previously discussed, this was an inherent risk of conducting a voluntary and community-based "sport plus" programme that focused on health-related outcomes (SECTION 9.4b) [129]. However, no effort was made to mitigate this effect by targeting the programme delivery to the needs of the community.

The changes in PA and PF may have also been affected by the dilution of programme objectives (SECTION 9.4a). Despite the poor delivery of the peace-building agenda, it was allocated a large proportion of the programme resources and intervention time (SECTION 9.4b). This did divert the programme implementation from its health objectives and may have compromised its impact on these outcomes.

As expected from the results of the retrospective formative analysis, the improvement in PF did not translate into an improvement in MH (SECTION 6.4d). However, the deterioration in MH for the boys in the GMKL when compared to the rest of the community was also not anticipated (i.e. there was not a negative correlation between PF and MH at baseline).

The wait-listed boys experienced an improvement in MH that exceeded the rest of the community (SECTION 8.4c). This suggested that there were inherent components of the

GMKL that had a negative impact on the MH of the boys in the intervention group, but positively influenced those wait-listed. As previously discussed, it is likely that the improved capacity in the community for sporting activities and the competitive nature of the boys league contributed to these observations (SECTION 8.5c) [219,220].

The difference in the health-related outcomes between the boys groups in the RCT brings into question the merits of a league-based SfD programme in Gulu. Although there were some additional gains in PA for the boys participating in the GMKL, these did not translate into significant differences in PF. However, more importantly, the negative effect of the competitive league on MH may have long-term health implications. Tracking of PA behaviour from adolescence into adulthood is influenced by the quality of the experience (SECTION 2.3) [21,23,34-37]. The deterioration in MH suggested that the GMKL was a negative experience for the participants. The potential long-term deterrent for participation in PA may actually mitigate the short-term benefits in PF experienced by the boys in the intervention group. In contrast, this was not likely to be the case in the wait-listed boys whose increase in PF was accompanied by improvements in MH.

Therefore, simply improving local resources and facilities for sporting activities may be a more appropriate use of SfD resources in Gulu. As previously described, it is likely that the health-related improvements of the wait-listed boys were mediated by the volunteer coaches who were also primary school teachers. The training and sports equipment they received were transferred to the school setting for use by the students. This suggests that the development of capacity and infrastructure within schools may be the most effective venue for SfD interventions to improve PA, PF and MH in Gulu. There are numerous examples from other LMICs where school-based interventions have improved health-related outcomes, but these are contingent on conducting inclusive and scheduled PES classes [181-184,186,187].

However, the results presented in this thesis cannot be used to conclude that community-based SfD leagues are intrinsically deleterious for MH. Although Coalter (2010) assessed different psychosocial outcomes for a similarly structured league in Gulu, a negative impact was not apparent [130]. The results of Burnett et al (2006) and (2009) also suggested a positive impact on MH from SfD leagues in South Africa [188,189]. It should be noted that none of these studies had a control group and definitive conclusions cannot be drawn. However, the girls intervention group in this thesis did not experience negative changes in MH and their results were equivalent to the rest of the community (SECTION 8.4b). Therefore, the discrepancy between programme objectives and implementation for non-competitive play may be a risk that is particularly relevant for boys football leagues in Gulu.

An effective formative evaluation may have mitigated the negative impact that the GMKL had on the MH of the boys who participated. As previously discussed, the development of a separate theoretical model for MH that was not mediated by PA would have prompted an adaptation of the actions delivered (SECTION 9.4b). This may have included an ongoing focus on the "peer leader" approach and the "plus" component of SfD participation [121,129,130,218].

9.5) Recommendations

Effective evaluation should inform the delivery of future interventions [2,129]. Nutbeam et al (2006) describe three stages that contribute to this process [2]. In this theoretical model, there is an initial focus on replicating and disseminating the intervention. This moves towards ongoing monitoring as the programme becomes more established (FIGURE 1.1). I have referred to these phases to formulate recommendations for the implementation and evaluation of future SfD interventions in Gulu and beyond.

9.5a) Intervention demonstration – can the programme be replicated?

Intervention demonstration involves replication of the programme with the intention of reproducing the desired outcomes. A second season of the GMKL was completed early in 2011. However, there have since been significant changes in the implementing organisations that have forced modification of the GMKL programme delivery.

The funding provided by the ASC for TKL was not renewed. In response to the demands of a new donor, TKL has turned its focus in Gulu towards SfD programmes for disabled children.

YCP has experienced significant internal management problems that have also prompted OAP to reduce funding. Their focus in Gulu has shifted back towards school holiday training camps for football and peace-building.

Despite these changes, both TKL and OAP remain interested in implementing SfD leagues for able-bodied adolescents in Gulu. If this does occur, it is recommended that the needs of the community are reassessed prior to implementation to inform the programme aims. If future interventions have the same objectives and replicate the community-based model, they should modify recruitment to improve reach by:

- Introducing more registration sites in all divisions of the municipality;
- Engaging schools in the intervention planning process;
- Building local support for the participation of girls in sport;
- Promoting the programme prior to registration to at-risk groups identified during the needs assessment.

Future programme implementation must also be more targeted in response to the formative evaluation. If the MH of the participants remains an objective for SfD interventions in Gulu, a new theoretical framework should be developed. This may shift the priorities of the programme towards non-competitive PA that is not focused on winning a sports league. It is possible that developing local capacity and infrastructure for football participation will have the greatest overall effect on health in Gulu. This may be particularly valuable and resource-efficient within schools. The role of the coaches as "peer leaders" is likely to be critical in redirecting the attention of the participants to engage in enjoyable play [121,130,218]. However, regular football games on full-sized fields may still be necessary if PA and cardiorespiratory fitness also remain health priorities of the programme.

The evaluation of intervention dissemination should assess "effectiveness" and "indicate the essential conditions that need to be established for the programme to be successfully implemented" [2]. This involves a shift towards assessing the process of programme delivery [2]. The theoretical framework that I developed in this thesis could be used for this purpose (SECTION 2.10). This would necessitate further articulation of the link between sustainable SfD programmes and each of the intended outcomes prior to commencing future interventions.

However, if there are any major changes in the format of intervention implementation, a full repeat of the impact evaluation is also warranted for all of the programme objectives. The design of this should:

- Attain the highest level of evidence with methods that are adapted to the programme model and local context;
- Assess appropriately matched groups that are representative of the target population and of sufficient size;
- Ensure indicators of the programme engagement, uptake and injuries are included;
- Use statistical methods that adjust for disparities in subject distribution;
- Include a formal process evaluation to help explain the programme impacts.

All future GMKL programmes should also evaluate who was actually reached in the context of the local needs and the programme objectives.

9.5b) Intervention dissemination – can the programme be disseminated?

Intervention dissemination involves delivering the programme in other locations. Only programmes that have proven impact should reach this stage of implementation and evaluation. However, programme dissemination is already occurring for several derivatives of the GMKL model. TKL has received funding to conduct SfD leagues in other parts of Uganda and Rwanda. They are also interested in expanding into southern Sudan. OAP has funded the implementation of a similar programme in Anaka, which is located 54km from Gulu in northern Uganda.

The evaluation of intervention dissemination focuses on "identifying the ways in which successful programs can be widely disseminated" [2]. Nutbeam et al (2006) advocate a further shift towards process evaluation at this stage [2]. However, this assumes that a

relatively consistent model of the intervention is replicated at each location and that its impacts have been established. In reality, SfD programmes continue to be disseminated despite the paucity of evidence supporting their efficacy. Although the sector has not yet progressed to this stage of implementation and evaluation, its donor driven nature has caused premature dissemination. This further demonstrates the disconnect that exists between funders, implementers and evaluators.

PA and PF are the only outcomes for the current GMKL model that have progressed to the dissemination stage of implementation and evaluation. Therefore, it is recommended that when delivering the GMKL model in a new location the implementers should:

- Assess the needs of the community using metrics that are appropriate to the local context and have been quality assessed;
- Formulate objectives that are responsive to the local needs;
- Develop programme theory for each of the programme objectives based on the results of a formative evaluation;
- Establish whether the intended outcomes and theoretical framework for delivery are similar to previous successful programmes. If they are (i.e. for PA and PF objectives), the intervention should be implemented with a process focused evaluation to assess the quality of delivery;
- Establish if any of the objectives are new or have not been realised in previous programmes. If this is the case, the implementation and evaluation procedures described for programme replication should be applied (SECTION 9.5a).

Whether the population that the intervention is reaching matches the community needs and programme objectives should continue to be assessed during dissemination.

9.5c) Programme monitoring – can the programme be sustained?

Programme monitoring is "directed towards supporting programme management" and occurs when an intervention is readily "implemented according to established methods" [2]. At this stage the delivery should be sustainable. The inability for local personnel to maintain the GMKL in Gulu without the support of OAP and the ASC demonstrates that it has not reached this phase. Therefore, it is recommended that future implementation focuses on building local capacity for the independent mobilisation of programme resources. This should promote community ownership and may be best achieved by embedding programmes within existing systems of local government and/or schools.

Evaluation at the programme monitoring stage is almost entirely focused on the process of intervention delivery [2]. An indication of effect can also be provided by concurrent "population-level monitoring of the outcome indicators of interest" [2]. I will be establishing an annual surveillance of PA, PF and MH for young adolescents in the P6 classes of 12 Primary Schools in Gulu during a follow-up fieldtrip in March 2012. The purpose is to identify trends that might demonstrate changes in local needs and/or the impact of future interventions in Gulu.

9.6) Conclusion – what does this research contribute?

The studies that I have completed for this DPhil thesis make several important and novel contributions to current knowledge in the public health and SfD paradigms. The modified measurement methods that I developed for PA, PF and MH enabled me to describe these health parameters in young adolescents living in a post-conflict and resource-poor setting. My approach to assessing PF is transferable to other resource-limited settings and may be useful for assessing risk factors and indicators of the epidemiological transition. I also identified persistent MH deficits in adolescents previously abducted during the war. However, it is evident that voluntary competitive sport-for-development programmes may not be an effective intervention for this sub-population. Similar interventions in post-conflict settings are likely to reach and improve the fitness of adolescents who are least in need and may have a detrimental impact on their MH. These are important considerations for policy makers in the health and sport-for-development sectors.

I have also identified several areas specific to my research that require further investigation and outlined recommendations for future evaluations. There is an indication to assess the impact of SfD on a broader scale. This should be accompanied by process evaluations that improve our understanding of effective programme delivery and participation. It is important that future research recognises the unique characteristics of different target populations and is applied to different models of implementation. Theoretical frameworks specific to each of these conditions and the various claimed outcomes also need further development.

My research also has practical implications for the SfD sector. It informs funders and implementers about the role of evaluators in programme planning and development. Applying experimental methods to assess the intervention reach and outcomes was a novel process for SfD programmes. Therefore, this thesis demonstrates several principles of

rigorous impact evaluation that may be transferrable to future programmes in the field. Although I did not formally assess the process of programme delivery, I have also made a practical contribution to this component of SfD evaluation. The theoretical framework that I developed for the implementation of sustainable SfD interventions is a tool that should be widely disseminated for testing and improvement.

I have also made a contribution to understanding the health priorities for young adolescents and improving SfD programme capacity in Gulu. I worked closely with the local personnel and volunteers of the GMKL to build their capacity for implementing SfD programmes. I also trained the research team to assess the components of PA, PF and MH evaluated in this study. The results of my research could influence future programme implementation in Gulu and similar settings to improve effectiveness and optimise outcomes. At this stage, I have provided verbal feedback based on my research to the personnel of the implementing organisations. I will be disseminating a report for local stakeholders during a follow-up fieldtrip in March 2012. This report will also be made available through existing platforms for the SfD grey literature.

There are a growing number of scholars becoming interested in SfD and it is anticipated that research in this sector will continue to expand. I endeavour to distribute my research findings within the academic community according to my publication and conference plan (APPENDIX A24). Some of my results have already been disseminated in one academic paper (APPENDIX A25) and two conference presentations (APPENDIX A26) [153,224,225]. However, there are currently no academic journals that specifically cater for SfD research. Consequently, I have contributed to establishing the Journal of Sport for Development (JSFD) for the publication of peer-reviewed articles [226]. This will provide a space for open dialogue between stakeholders and the dissemination of best practices from case-studies in the field.

On a personal level, this DPhil has contributed to expanding my research skills in evaluation design and statistical methods. I have also had the opportunity to improve my proficiency in academic writing and presentation. My field experience enabled me to learn a great deal about managing programmes and research in resource-poor settings. This included working with people from diverse backgrounds to implement rigorous research in the face of several pragmatic and ethical constraints.

Above all else, I believe that my work drives forward a streamlining process for the SfD sector. I think that there is little value in conducting further desk research about the broad claims of these programmes from offices in developed countries. The SfD movement does not need more descriptive articles outlining the paucity of evidence or to establish another "M&E toolkit" based on "Western" constructs. It is time for academics to get their hands dirty by contributing to building capacity for independent evaluations in the field rather than from their armchairs at home. Only then will we begin to understand the true impact of SfD programmes on the broad array of its claimed outcomes. Although the SfD sector has rapidly expanded since the UN International Year of Sport and Physical Education in 2005, I believe that improved evaluation will lead to a "healthy" contraction. I adapt a common mantra from academia in suggesting that SfD organisations must "evaluate or perish". Only those that rigorously assess their impact and appropriately target their programmes will actually survive. This should weed out the "briefcase NGOs" that move from one trend to another at the whim of potential funders, but to the detriment of the supposed beneficiaries. I hope that it will also ensure that the SfD sector does not become superseded by the next fad in development - I suspect that the "International Year of Music and Dance" is just around the corner...

9.7) Summary box

FINAL DISCUSSION AND CONCLUSIONS

- The chronology of programme development for the GMKL indicated that planning was not informed by a community needs assessment or a formative evaluation.

- There were several common limitations that pervaded the methods of this thesis. Most of the weaknesses were pragmatic trade-offs that enabled the studies to take place. Therefore, they were also some of the greatest strengths in the research. These were primarily related to the measurement methods, evaluation design, subject recruitment and data interpretation.

- It appears that a mismatch between programme objectives and implementation influenced the final outcomes of the GMKL.

- There is a tension between programme funders, implementers and evaluators in the SfD sector. This may cause the priorities of the local community to be neglected and the asynchronous development of programme objectives among stakeholders.

- The objectives of the GMKL fused the aims of the contributing stakeholders and may have diluted the programme implementation for each intended outcome. Although these were not based on a comprehensive community needs analysis, there were local deficits in PA, PF and MH that matched the objectives of the programme.

- The actions of the GMKL were not responsive to the formative evaluation that indicated an alternative theoretical framework for the MH objectives may have been necessary.

- It appears that the GMKL did not reach those most in need when considering the outcomes of interest in this research. This may have been exacerbated by the voluntary and community-based registration process for the intervention.

- The intention to implement a programme which focused on peace-building, PA participation and MH rather than football results was not realised. The delivery of a structured "sport plus" league appeared to be sufficient to have a positive influence on PA and PF, but had a detrimental effect on the MH of boys who participated. This may have been caused by the competitive nature of the league.

- The findings reported in this thesis can inform the design and evaluation of future SfD interventions. It is necessary to develop theoretical frameworks for each programme objective prior to implementation, replication and dissemination. Assessing the impact of the intervention and evaluating the process of delivery both remain important for GMKL programmes at this stage.

- This research contributed to:
  - Expanding the evidence base for the public health and the SfD sectors;
  - Outlining the future direction of SfD research and evaluation;
  - Promoting the role of evaluators in SfD programme planning and development;
  - Understanding the health priorities for young adolescents in Gulu;
  - Building local capacity to implement SfD programmes in Gulu;
  - Developing my skills in research and programme management.

REFERENCES (Thesis)

1. E. Stanley. *The Conduct of Life*. 1873, Addres at Liverpool College.
2. D. Nutbeam, A. Bauman. *Evaluation in a Nutshell*. 2006, Sydney: McGraw-Hill.
3. R. Beaglehole, J. Epping-Jordan, eds. *Preventing Chronic Diseases: A Vital Investment*. 2005, WHO: Geneva.
4. S. Yusuf, S. Reddy, S. Ounpuu, S. Anand. *Global Burden of Cardiovascular Diseases: Part I: General Considerations, the Epidemiologic Transition, Risk Factors, and Impact of Urbanization.* Circulation, 2001. **104**(22): p. 2746-53.
5. S. Yusuf, S. Reddy, S. Ounpuu, S. Anand. *Global Burden of Cardiovascular Diseases: Part Ii: Variations in Cardiovascular Disease by Specific Ethnic Groups and Geographic Regions and Prevention Strategies.* Circulation, 2001. **104**(23): p. 2855-64.
6. S. Dalal, J.J. Beunza, J. Volmink, C. Adebamowo, F. Bajunirwe, M. Njelekela, D. Mozaffarian, W. Fawzi, W. Willett, H.O. Adami, M.D. Holmes. *Non-Communicable Diseases in Sub-Saharan Africa: What We Know Now.* Int J Epidemiol, 2011. **40**(4): p. 885-901.
7. M. Prince, V. Patel, S. Saxena, M. Maj, J. Maselko, M. R. Phillips, A. Rahman. *No Health without Mental Health.* Lancet, 2007. **370**(9590): p. 859-77.
8. World Health Organisation (WHO). *Mental Health: A State of Well-Being*. 2009, WHO: Geneva.
9. World Health Organisation (WHO). *The World Health Report 2001 - Mental Health: New Understanding, New Hope*. 2001, WHO: Geneva.
10. V. Patel, A. J. Flisher, S. Hetrick, P. McGorry. *Mental Health of Young People: A Global Public-Health Challenge.* Lancet, 2007. **369**(9569): p. 1302-13.
11. World Health Organisation (WHO). *Mental Health Atlas 2005*. 2005, WHO: Geneva.
12. S. Saxena, G. Thornicroft, M. Knapp, H. Whiteford. *Resources for Mental Health: Scarcity, Inequity, and Inefficiency.* Lancet, 2007. **370**(9590): p. 878-89.
13. D. Bhugra, I.H. Minas. *Mental Health and Global Movement of People.* The Lancet, 2007. **370**(9593): p. 1109-1111.
14. S. Allender, B. Lacey, P. Webster, M. Rayner, M. Deepa, P. Scarborough, C. Arambepola, M. Datta, V. Mohan. *Level of Urbanisation and Noncommunicable Disease Risk Factors in Tamil Nadu, India.* Bulletin of the World Health Organization, 2009. **87**.
15. Physical Activity and Health. Resolution WHA57.17. Global Strategy on Diet. *In: Fiftyseventh World Health Assembly, Geneva, 17–22 May 2004. Resolutions and Decisions, Annexes.* 2004, WHO: Geneva.
16. World Health Organisation (WHO). *The World Health Report 2002: Reducing Risks, Promoting Healthy Life*. 2002, WHO: Geneva.
17. World Health Organisation (WHO). *Global Recommendations on Physical Activity for Health*. 2010, WHO: Geneva.
18. D.E. Warburton, C.W. Nicol, S.S. Bredin. *Health Benefits of Physical Activity: The Evidence.* CMAJ, 2006. **174**(6): p. 801-9.
19. W.L. Haskell, I.M. Lee, R.R. Pate, K.E. Powell, S.N. Blair, B.A. Franklin, C.A. Macera, G.W. Heath, P.D. Thompson, A. Bauman. *Physical Activity and Public Health:*

Updated Recommendation for Adults from the American College of Sports Medicine and the American Heart Association. Circulation, 2007. **116**(9): p. 1081-93.

20. D.E.R. Warburton, S. Charlesworth, A. Ivey, L. Nettlefold, S.S.D. SD Bredin. *A Systematic Review of the Evidence for Canada's Physical Activity Guidelines for Adults.* International Journal of Behavioral Nutrition and Physical Activity, 2010. **7**(39).
21. NICE Public Health Collaborating Centre. *Physical Activity and Children: Review 1 - Descriptive Epidemiology*. 2008, NICE: London.
22. S. Cumming, C. Riddoch, J.W. Twisk, I. Ferreira, P.W. Franks, H.C. Looker, H.C. Kemper, J.C. Garza, N.G. Murray, S. Sharma, K.L. Drenner, S.R. Tortolero, W.C. Taylor, S. Trost, S.P.J. Kremers, H. Schaalma, R.M. Meertens, W. van Mechelen, G.J. Kok. *Physical Activity, Physical Fitness and Health*, in *Paediatric Exercise Science and Medicine*, N. Armstrong, W. van Mechelen, Editors. 2008, Oxford University Press: Oxford.
23. S.N. Blair, D.G. Clark, K.J. Cureton, K.E. Powell. *Exercise and Fitness in Childhood: Implications for a Lifetime of Health*, in *Perspectives in Exercise Science and Sports Medicine*, C.V. Gisolfi, D.R. Lamb, Editors. 1989, McGraw Hill: New York. p. 401-430.
24. S. J. Biddle, T. Gorely, D. J. Stensel. *Health-Enhancing Physical Activity and Sedentary Behaviour in Children and Adolescents.* J Sports Sci, 2004. **22**(8): p. 679-701.
25. W.B. Strong, R.M. Malina, C.J. Blimkie, S.R. Daniels, R.K. Dishman, B. Gutin, A.C. Hergenroeder, A. Must, P.A. Nixon, J.M. Pivarnik, T. Rowland, S. Trost, F. Trudeau. *Evidence Based Physical Activity for School-Age Youth.* J Pediatr, 2005. **146**(6): p. 732-7.
26. I. Janssen, A.G. LeBlanc. *Systematic Review of the Health Benefits of Physical Activity and Fitness in School-Aged Children and Youth.* International Journal of Behavioural Nutrition and Physical Activity, 2010. **7**(40).
27. P.L. Kristensen, N.C. Moeller, L. Korsholm, E. Kolle, N. Wedderkopp, K. Froberg, L.B. Andersen. *The Association between Aerobic Fitness and Physical Activity in Children and Adolescents: The European Youth Heart Study.* Eur J Appl Physiol, 2010. **110**(2): p. 267-75.
28. N. Matos, R.J. Winsley. *Trainability of Young Athletes and Overtraining.* Journal of Sports Science and Medicine, 2007. **6**: p. 353-367.
29. S.A. Paluska, T.L. Schwenk. *Physical Activity and Mental Health: Current Concepts.* Sports Med, 2000. **29**(3): p. 167-80.
30. R.K. Dishman, P.J. O'Connor. *Lessons in Exercise Neurobiology: The Case of Endorphins.* Mental Health and Physical Activity, 2009. **2**: p. 4-9.
31. N. Armstrong, W. van Mechelen, eds. *Paediatric Exercise Science and Medicine*. 2 ed. 2008, Oxford University Press: Oxford.
32. L.B. Andersen, H. Hasselstrøm, V. Grønfeldt, S.E. Hansen, F. Karsten. *The Relationship between Physical Fitness and Clustered Risk, and Tracking of Clustered Risk from Adolescence to Young Adulthood: Eight Years Follow-up in the Danish Youth and Sport Study.* International Journal of Behavioral Nutrition and Physical Activity, 2004. **1**(6).
33. I. Ferreira, J.W. Twisk, W. van Mechelen, H.C. Kemper, C.D.A Stehouwer. *Development of Fatness, Fitness, and Lifestyle from Adolescence to the Age of 36*

Years: Determinants of the Metabolic Syndrome in Young Adults: The Amsterdam Growth and Health Longitudinal Study. Arch Intern Med., 2005. **165**(1): p. 42-48.

34. P.C. Hallal, C.G. Victora, M.R. Azevedo, J.C.K. Wells. *Adolescent Physical Activity and Health: A Systematic Review.* Sports Med, 2006. **36**(12): p. 1019-1030.
35. R.M. Malina. *Tracking of Physical Activity and Physical Fitness across the Lifespan.* Res Q Exerc Sport., 1996. **67**(3): p. 48-57.
36. R. Telama. *Tracking of Physical Activity from Childhood to Adulthood: A Review.* Obesity Facts, 2009. **2**(3): p. 187-95.
37. D. Kirk. *Physical Education, Youth Sport and Lifelong Participation: The Importance of Early Learning Experiences* European Physical Education Review, 2005. **11**(3): p. 239-55.
38. D.E. Warburton, C.W. Nicol, S.S. Bredin. *Prescribing Exercise as Preventative Therapy.* CMAJ, 2006. **174**(7): p. 961-974.
39. A. Hurtig-Wennlof, J. R. Ruiz, M. Harro, M. Sjostrom. *Cardiorespiratory Fitness Relates More Strongly Than Physical Activity to Cardiovascular Disease Risk Factors in Healthy Children and Adolescents: The European Youth Heart Study.* Eur J Cardiovasc Prev Rehabil, 2007. **14**(4): p. 575-81.
40. S.N. Blair, Y. Cheng, J.S. Holder. *Is Physical Activity or Physical Fitness More Important in Defining Health Benefits?* Med Sci Sports Exerc, 2001. **33**(6 Suppl): p. S379-99; discussion S419-20.
41. J. Myers, A. Kaykha, S. George, J. Abella, N. Zaheer, S. Lear, T. Yamazaki, V. Froelicher. *Fitness Versus Physical Activity Patterns in Predicting Mortality in Men.* Am J Med, 2004. **117**(12): p. 912-8.
42. P.T. Williams. *Physical Fitness and Activity as Separate Heart Disease Risk Factors: A Meta-Analysis.* Med Sci Sports Exerc, 2001. **33**(5): p. 754-61.
43. F.B. Ortega, J.R. Ruiz, M.J. Castillo, M. Sjostrom. *Physical Fitness in Childhood and Adolescence: A Powerful Marker of Health.* Int J Obes (Lond), 2008. **32**(1): p. 1-11.
44. J.R. Ruiz, J. Castro-Pinero, E.G. Artero, F.B. Ortega, M. Sjostrom, J. Suni, M.J. Castillo. *Predictive Validity of Health-Related Fitness in Youth: A Systematic Review.* British Journal of Sports Medicine, 2009. **43**(12): p. 909-23.
45. J.C. Eisenmann, G.J. Welk, E.E. Wickel, S.N. Blair. *Combined Influence of Cardiorespiratory Fitness and Body Mass Index on Cardiovascular Disease Risk Factors among 8-18 Year Old Youth: The Aerobics Center Longitudinal Study.* Int J Pediatr Obes, 2007. **2**(2): p. 66-72.
46. M.H. Rice, C.C. Howell. *Measurement of Physical Activity, Exercise, and Physical Fitness in Children: Issues and Concerns.* J Pediatr Nurs, 2000. **15**(3): p. 148-56.
47. A.L. Claessens, G. Beunen, R.M. Malina. *Anthropometry, Physique, Body Composition, and Maturity,* in *Paediatric Exercise Science and Medicine*, N. Armstrong, W. van Mechelen, Editors. 2008, Oxford University Press: Oxford.
48. R.M. Malina. *Body Composition in Athletes: Assessment and Estimated Fatness.* Clin Sport Med, 2007. **26**: p. 37-68.
49. J.C.K. Wells. *A Critique of the Expression of Paediatric Body Composition Data.* Arch Dis Child, 2001. **85**: p. 67-72.

50. A. Sopher, W. Shen, A. Pietrobelli. *Paediatric Body Composition Methods.*, in *Human Body Composition*, S.B. Heymsfield, T.G. Lohman, Z.M. Wang, S.B. Going, Editors. 2005, Human Kinetics: Champaign.

51. J. Castro-Pinero, E.G. Artero, V. Espana-Romero, F.B. Ortega, M. Sjostrom, J. Suni, J.R. Ruiz. *Criterion-Related Validity of Field-Based Fitness Tests in Youth: A Systematic Review.* Br J Sports Med, 2010. **44**(13): p. 934-43.

52. M.B.A. De Ste Croix. *Muscle Strength*, in *Paediatric Exercise Science and Medicine*, N. Armstrong, W. Van Mechelen, Editors. 2008, Oxford University Press: Oxford.

53. N.J. Farpour-Lambert, C.J.R. Blimkie. *Muscle Strength*, in *Paediatric Exercise Science and Medicine*, N. Armstrong, W. van Mechelen, Editors. 2008, Oxford University Press: Oxford.

54. R. Maughan, M. Gleeson. *The Biochemical Basis of Sports Performance*. 2004, Oxford University Press: Oxford.

55. G.R. Tomkinson, T.S. Olds. *Field Tests of Fitness*, in *Paediatric Exercise Science and Medicine*, N. Armstrong, W. van Mechelen, Editors. 2008, Oxford University Press: Oxford.

56. Y.K. Kesaniemi, E. Danforth Jr, M.D. Jensen, P.G. Kopelman, P. Lefebvre, B.A. Reeder. *Dose-Response Issues Concerning Physical Activity and Health: An Evidence-Based Symposium.* Med Sci Sports Exerc, 2001. **33**(6 Suppl): p. S351-8.

57. M.H. Whaley, P.H. Brubaker, R.M. Otto, L.E. Armstrong. *Acsm's Guidelines for Exercise Testing and Prescription* 7ed. 2005, Baltimore: Lippincott Williams & Wilkins.

58. A.E. Mark, I. Janssen. *Influence of Bouts of Physical Activity on Overweight in Youth.* Am J Prev Med, 2009. **36**(5): p. 416-21.

59. K.J. Mackelvie, H.A. McKay, K.M. Khan, P.R. Crocker. *A School-Based Exercise Intervention Augments Bone Mineral Accrual in Early Pubertal Girls.* J Pediatr, 2001. **139**(4): p. 501-8.

60. B.E. Ainsworth, W.L. Haskell, M.C. Whitt, M.L. Irwin, A.M. Swartz, S.J. Strath, W.L. O'Brien, D.R. Bassett Jr, K.H. Schmitz, P.O. Emplaincourt, D.R. Jacobs Jr, A.S. Leon. *Compendium of Physical Activities: An Update of Activity Codes and Met Intensities.* Med Sci Sports Exerc, 2000. **32**(9 Suppl): p. S498-504.

61. World Health Organisation (WHO). *Global Physical Activity Questionaire (Gpaq) Analysis Guide*. 2008, WHO: Geneva.

62. M. de Onis, A. Onyango, E. Borghi, A. Siyam, C. Nishida, J. Siekmann. *Development of a Who Growth Reference for School-Aged Children and Adolescents.* Bulletin of the World Health Organization, 2007. **85**: p. 661-668.

63. The Cooper Institute. *Fitnessgram: Standards for Healthy Fitness Zone (Revision 8.6 and 9.X).* 2010, The Cooper Institute: Dallas.

64. A.R. Adegboye, S.A. Anderssen, K. Froberg, L.B. Sardinha, B.L. Heitmann, J. Steene-Johannessen, E. Kolle, L.B. Andersen. *Recommended Aerobic Fitness Level for Metabolic Health in Children and Adolescents: A Study of Diagnostic Accuracy.* Br J Sports Med., 2011. **45**(9): p. 722-8.

65. F. Lobelo, R.R. Pate, M. Dowda, A.D. Liese, J.R. Ruiz. *Validity of Cardiorespiratory Fitness Criterion-Referenced Standards for Adolescents.* Med Sci Sports Exerc., 2009. **41**(6): p. 1222-9.

66. J.R. Ruiz, F.B. Ortega, N.S. Rizzo, I. Villa, A. Hurtig-Wennlöf, L. Oja, M. Sjöström. *High Cardiovascular Fitness Is Associated with Low Metabolic Risk Score in Children: The European Youth Heart Study.* Pediatr Res., 2007. **61**(3): p. 350-5.

67. P. Freedson, K. J. Cureton, G. W. Heath. *Status of Field-Based Fitness Testing in Children.* Preventative Medicine, 2000. **31**: p. 77-85.

68. Sport for Development and Peace International Working Group (SDPIWG). *Harnessing the Power of Sport for Development and Peace: Recommendations to Governments*. 2008, SDPIWG: Toronto.

69. International Olympic Committee (IOC). *International Relations Commission - Cooperation with Un Agencies*. 2009; Available from: http://www.olympic.org/international-relations-commission.

70. P.B. Massao, S. Straume. *Urban Youth and Sport for Development*. 2011, UN-HABITAT: Nairobi.

71. T. Kay. *Sport and International Development: The Unproven Remedy*, in *Routledge Handbook of Sports Development*, B. Houlihan, M. Green, Editors. 2011, Routledge: New York.

72. Sport for Development and Peace International Working Group (SDPIWG). *Sport for Development and Peace: From Practice to Policy Preliminary Report*. 2006, SDPIWG: Toronto.

73. Sport for Development and Peace International Working Group (SDPIWG). *Sport for Development and Peace: Governments in Action*. 2008, SDPIWG: Toronto.

74. Swiss Academy for Development (SAD). *The International Platform on Sport and Development*. 2010; Available from: http://www.sportanddev.org/.

75. United Nations Educational Scientific and Cultural Organisation (UNESCO). *First International Conference of Ministers and Senior Officials Responsible for Physical Education and Sport (Mineps I) Final Report*. 1976, UNESCO: Paris.

76. United Nations Educational Scientific and Cultural Organisation (UNESCO). *Second International Conference of Ministers and Senior Officials Responsible for Physical Education and Sport (Mineps Ii) Final Report*. 1988, UNESCO: Paris.

77. United Nations Educational Scientific and Cultural Organisation (UNESCO). *Third International Conference of Ministers and Senior Officials Responsible for Physical Education and Sport (Mineps Iii) Final Report*. 1999, UNESCO: Paris.

78. United Nations Educational Scientific and Cultural Organisation (UNESCO). *Fourth International Conference of Ministers and Senior Officials Responsible for Physical Education and Sport (Mineps Iv) Final Report*. 2004, UNESCO: Paris.

79. UN Office on Sport for Development and Peace (UNOSDP). *The Magglingen Declaration and Recommendations*. 2003: Magglingen.

80. United Nations Educational Scientific and Cultural Organisation (UNESCO). *International Charter of Physical Education and Sport*. 1978, UNESCO: Paris.

81. I. Beutler. *Report on the Year of Sport and Physical Education (Iyspe 2005) - Sport for a Better World*. 2006, UN Office on Sport for Development and Peace (UNOSDP): Geneva.

82. United Nations General Assembly. *Declaration on the Rights of the Child*. 1959, United Nations: New York.

83. Committe on the Rights of the Child (CRC). *Convention on the Rights of the Child*. 1989, United Nations: New York.

84. United Nations General Assembly. *Resolution Adopted by the General Assembly - 48/10 International Year of Sport and the Olympic Ideal*. 1993, United Nations: New York.

85. United Nations General Assembly. *Resolution Adopted by the General Assembly - 48/11 Observance of the Olympic Truce*. 1993, United Nations: New York.

86. United Nations General Assembly. *Resolution Adopted by the General Assembly - 52/21 Building a Peaceful and Better World through Sport and the Olympic Ideal*. 1997, United Nations: New York.

87. United Nations General Assembly. *Resolution Adopted by the General Assembly - 59/10 Sport as a Means to Promote Education, Health, Development and Peace*. 2004, United Nations: New York.

88. United Nations General Assembly. *Resolution Adopted by the General Assembly - 63/135 Sport as a Means to Promote Education, Health, Development and Peace*. 2008, United Nations: New York.

89. United Nations General Assembly. *Resolution Adopted by the General Assembly - 64/3 Observer Status for the International Olympic Committee in the General Assembly*. 2009, United Nations: New York.

90. Commission of the European Communities. *White Paper on Sport*. 2007, Commission of the European Communities: Brussels.

91. The European Parliament. *European Parliament Resolution on Development and Sport.* Official Journal of the European Union, 2006. **C 285 E**: p. 135-137.

92. UN Office on Sport for Development and Peace (UNOSDP). *Report on the Second Magglingen Conference - Development through Sport: Moving to the Next Stage*. 2005, UNOSDP: Magglingen.

93. Africa International Sport Convention (CISA). *Report on the First Cisa - Sport for Education and Development*. 2007, CISA: Dakar.

94. African Union (AU). *Decision on the International Year of African Football (Doc.Assembly/Au.8 (Vi)) Add.11)*. 2006, AU: Addis Ababa.

95. United Nations General Assembly. *Report of the Secretary-General - Sport for Development and Peace: The Way Forward*. 2006, United Nations: New York.

96. International Olympic Committee (IOC). *Report: First International Forum on Sport, Peace and Development*, International Cooperation and Development Department, Editor. 2009, IOC: Lausanne.

97. Beyond Sport. *Beyond Sport - Summit and Awards 2009*. 2011; Available from: http://www.beyondsport.org/.

98. Jacobs Foundation. *Report: Forum on Productive Youth Development through Sport in Africa*. 2009, Jacobs Foundation: Zurich.

99. United Nations General Assembly. *Resolution Adopted by the General Assembly - 65/1 Keeping the Promise: United to Achieve the Millennium Development Goals*. 2010, United Nations: New York.

100. Interdisciplinary Centre of Excellence for Sports Science and Development (ICESSD). *International Sport and Development Conference - Beyond 2010*. 2010, University of the Western Cape: Cape Town.

101. Sport for Development and Peacec International Working Group (SDPIWG). *Annotated Agenda: Plenary Session of the Sdpiwg*. 2011, United Nations Office on Sport for Development and Peace (UNOSDP): Geneva.

102. United Nations Office on Sport for Development and Peace (UNOSDP). *Recommendations: Second International Forum on Sport for Peace & Development*. 2011, UNOSDP: Geneva.

103. R. Giulianotti. *Sport, Peacemaking and Conflict Resolution: A Contextual Analysis and Modelling of the Sport, Development and Peace Sector.* Ethnic and Racial Studies, 2010. **34**(2): p. 207-228.

104. R. Giulianotti. *The Sport, Development and Peace Sector: A Model of Four Social Policy Domains.* Jnl Soc. Pol, 2011. **40**(4): p. 757-776.

105. United Nations (UN). *Sport for Development and Peace: Towards Achieving the Millennium Development Goals*. 2003, UN: New York. p. 2.

106. P. Donnelly, S. Darnell, S. Wells, J. Coakley. *The Use of Sport to Foster Child and Youth Development and Education* in *Literature Reviews on Sport for Development and Peace*. 2007, Sport for Development & Peace International Working Group (SDPIWG): Toronto.

107. D. Zakus, D. Njelesani, S. Darnell. *The Use of Sport and Physical Activity to Achieve Health Objectives,* in *Literature Reviews on Sport for Development and Peace*. 2007, Sport for Development & Peace International Working Group (SDPIWG): Toronto.

108. J. Larkin, S. Razack, F. Moola. *Gender, Sport and Development,* in *Literature Reviews on Sport for Development and Peace*. 2007, Sport for Development & Peace International Working Group (SDPIWG): Toronto.

109. P. Parnes, G. Hashemi. *Sport as a Means to Foster Inclusion, Health and Well-Being of People with Disabilities*, in *Literature Reviews on Sport for Development and Peace*. 2007, Sport for Development & Peace International Working Group (SDPIWG): Toronto.

110. B. Kidd, M. MacDonnell. *Peace, Sport and Development*, in *Literature Reviews on Sport for Development and Peace*. 2007, Sport for Development & Peace International Working Group (SDPIWG): Toronto.

111. United Nations (UN). *Post-Conflict Stabilization, Peacebuilding and Recovery Frameworks*, Policy and Strategy Concepts, Editor. 2006, UN Disarmament, Demobilization and Reintegration Resource Centre Geneva.

112. Center for Strategic and International Studies (CSIS), Association of te United States Army (AUSA). *Post-Conflict Reconstruction - Task Framework*. 2002, Center for Strategic and International Studies: Washington DC.

113. United Nations Development Group (UNDG). *United Nations Development Assistance Framework [Undaf] - Post-Conflict Transition*. 2005, UNDG: New York. p. 1.

114. O. Cronin. *Comic Relief Review: Mapping the Research on the Impact of Sport and Development Interventions*. 2011, Orla Cronin Research: Manchester.

115. B. Kidd. *A New Social Movement: Sport for Development and Peace.* Sport in Society, 2008. **11**(4): p. 370-80.

116. S. Darnell, D. Black. *Mainstreaming Sport into International Development Studies.* Third World Quarterly, 2011. **32**(3): p. 367-78.

117. I. Beutler. *Sport Serving Development and Peace: Achieving the Goals of the United Nations through Sport.* Sport in Society, 2008. **11**(4): p. 359-69.

118. J.O. Koss, A. Alexandrova. *Essay: Hiv/Aids Prevention and Peace through Sport.* Lancet, 2005. **366 Suppl 1**: p. S3-4.

119. L. Micheli, M. Mountjoy, L. Engebretsen, K. Hardman, S. Kahlmeier, E. Lambert, A. Ljungqvist, V. Matsudo, H. McKay, C. J. Sundberg. *Fitness and Health of Children through Sport: The Context for Action.* Br J Sports Med, 2011. **45**(11): p. 931-6.

120. R. Levermore. *Sport: A New Engine of Development?* Progress in Development Studies, 2008. **8**(2): p. 183-90.

121. J. Coakley. *Youth Sports: What Counts as "Positive Development?".* Journal of Sport & Social Issues, 2011. **35**: p. 306-324.

122. F. Coalter. *A Wider Social Role for Sport: Who's Keeping the Score?* 2007, London: Routledge.

123. F. Coalter. *Sport-in-Development: Development for and through Sport?*, in *Sport and Social Capital*, M. Nicholson, R. Hoye, Editors. 2008, Elsevier: Oxford.

124. R. Levermore, A. Beacom. *Sport and International Development.* 2009, New York: Palgrave Macmillan.

125. R. Levermore. *Sport in International Development*, in *Routledge Handbook of Sports Development*, B. Houlihan, M. Green, Editors. 2011, Routledge: New York.

126. P. Bolton, J. Bass, T. Betancourt, L. Speelman, G. Onyango, K.F. Clougherty, R. Neugebauer, L. Murray, H. Verdeli. *Interventions for Depression Symptoms among Adolescent Survivors of War and Displacement in Northern Uganda: A Randomized Controlled Trial.* Journal of the American Medical Association, 2007. **298**(5): p. 519-527.

127. T.S. Betancourt, L. Speelman, G. Onyango, P. Bolton. *A Qualitative Study of Mental Health Problems among Children Displaced by War in Northern Uganda.* Transcult Psychiatry, 2009. **46**(2): p. 238-56.

128. T.S. Betancourt, J. Bass, I. Borisova, R. Neugebauer, L. Speelman, G. Onyango, P. Bolton. *Assessing Local Instrument Reliability and Validity: A Field-Based Example from Northern Uganda.* Soc Psychiatry Psychiatr Epidemiol, 2009. **44**(8): p. 685-92.

129. F. Coalter. *Sport-in-Development: A Monitoring and Evaluation Manual*. 2006, UK Sport: London.

130. F. Coalter. *Sport-for-Development Impact Study - a Research Initiative Funded by Comic Relief and Uk Sport and Managed by International Development through Sport*, Department of Sports Studies, Editor. 2010, University of Stirling: Stirling.

131. M. Akhbari, V. Kunz, U. Selvaraju, R. Wijesekera. *Sport & Development - a Summary of Sad's Experiences and Good Practices*. 2010, Swiss Academy for Development (SAD): Bienne.

132. M. Akhbari. *A Pilot Study Using Photo Monitoring in the Context of the Project "Sport and Play for Lebanese Children and Youth Affected by Conflict"*. 2009, Swiss Academy for Development (SAD): Bienne.

133. V. Kunz. *Move 4 New Horizons - a Holistic Educational Programme for Disadvantaged Children in Nepal: Interim Evaluation Report*. 2010, Swiss Academy for Development (SAD): Bienne.

134. D. Nutbeam, E. Harris, M. Wise. *Theory in a Nutshell*. 3 ed. 2010, Sydney: McGraw-Hill.

135. F. Coalter. *Sports Development's Contribution to Social Policy Objectives,* in *Routledge Handbook of Sports Development*, B. Houlihan, M. Green, Editors. 2011, Routledge: New York.

136. F. Coalter. *The Politics of Sport-for-Development: Limited Focus Programmes and Broad Gauge Problems?* International Review for the Sociology of Sport, 2010. **45**(3): p. 295-314.

137. D. Hartman, C. Kwauk. *Sport and Development: An Overview, Critique, and Reconstruction.* Journal of Sport and Social Issues, 2011. **35**(3): p. 284-305.

138. C. Bouchard. *Physical Activity and Health: Introduction to the Dose-Response Symposium.* Med Sci Sports Exerc, 2001. **33**(6 Suppl): p. S347-50.

139. D. Pedersen. *Political Violence, Ethnic Conflict, and Contemporary Wars: Broad Implications for Health and Social Well-Being.* Soc Sci Med, 2002. **55**(2): p. 175-90.

140. H.A. Ghobarah, P. Huth, B. Russett. *The Post-War Public Health Effects of Civil Conflict.* Soc Sci Med, 2004. **59**(4): p. 869-84.

141. P. B. Spiegel, F. Checchi, S. Colombo, E. Paik. *Health-Care Needs of People Affected by Conflict: Future Trends and Changing Frameworks.* Lancet, 2010. **375**(9711): p. 341-5.

142. B. Roberts, P. Patel, M. McKee. *Noncommunicable Diseases and Post-Conflict Countries.* Bull World Health Organ, 2012. **90**: p. 2-2A.

143. AlertNet - Alerting Humanitarians about Emergencies. *Factsheet: Alertnet Top 10 'Forgotten' Emergencies*. 2005, Thomson Reuters Foundation - AlertNet: London.

144. World Health Organisation (WHO). *Annual Report: Promoting Effective Coordination of Emergency Health, Nutrition and Hiv/Aids Response and Recovery in Uganda*. 2007, WHO: Geneva.

145. Women's Commission for Refugee Women and Children. *Listening to Youth: The Experiences of Young People in Northern Uganda* 2007, WCRWC: New York.

146. A. Branch. *Gulu Town in War... And Peace? Displacement, Humanitarianism and Postwar Crisis.*, in *Crisis States Working Papers*, Crisis States Research Centre, Editor. 2008, LSE Destin Development Studies Initiative.

147. J. Annan, C. Blattman, R. Horton. *The State of Youth and Youth Protection in Northern Uganda - Findings from the Survey for War Affected Youth*. 2006, UNICEF: Kampala.

148. J. Annan, C. Blattman, K. Carlson, D. Mazurana. *The State of Female Youth in Northern Uganda: Findings from the Survey of War-Affected Youth (Sway)*. 2008, UNICEF: Kampala.

149. Gulu District Health Office (DHO). *Gulu Dho Annual Workplan 2008-2009*. 2008, Gulu DHO: Gulu.

150. Gulu District Health Office (DHO). *Gulu Dho Annual Workplan 2009-2010*. 2009, Gulu DHO: Gulu.

151. Uganda Bureau of Statistics (UBS). *Uganda Demographic and Health Survey 2006*. 2007, Marco International Inc: Maryland.

152. World Health Organisation (WHO). *Mental Health: A State of Well-Being Fact File*. 2009; Available from: http://www.who.int/features/factfiles/mental_health/mental_health_facts/en/index.html.

153. J.D. McMullen, P.S. O'Callaghan, J.A. Richards, J.G. Eakin, H. Rafferty. *Screening for Traumatic Exposure and Psychological Distress among War-Affected Adolescents in Post-Conflict Northern Uganda.* Soc Psychiatry Psychiatr Epidemiol, 2011.

154. World Health Organisation (WHO). *Global School-Based Student Health Survey (Gshs): Uganda*. 2003, WHO: Geneva.

155. J. Marshall, K. Hardman. *The State and Status of Physical Education in Schools in International Context.* European Physical Education Review, 2000. **6**(3): p. 203-229.

156. K. Hardman. *An Update on the Status of Physical Education in Schools Worldwide: Technical Report for the World Health Organisation (Who)*. 2005, WHO: Geneva.

157. Uganda Ministry of Education and Sports. *Uganda Educational Statistics Abstract 2007*. 2007, The Republic Of Uganda Government: Kampala.

158. K. Deininger. *Does Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda.* Economics of Education Review, 2003. **22**: p. 291–305.

159. K. Hardman, J. Marshall. *Update on the State and Status of Physical Education World-Wide*. 2005, University of Worcester: Worcester.

160. C. Mackenzie, C. McDowell, E. Pittaway. *Beyond 'Do No Harm': The Challenge of Constructing Ethical Relationships in Refugee Research.* Journal of Refugee Studies, 2007. **20**(2): p. 299-319.

161. E. Pittaway, L. Bartolomei, R. Richard Hugman. *'Stop Stealing Our Stories': The Ethics of Research with Vulnerable Groups.* J Human Rights Practice, 2010. **2**(2): p. 229-251.

162. V. Espana-Romero, E.G. Artero, D. Jimenez-Pavon, M. Cuenca-Garcia, F.B. Ortega, J. Castro-Pinero, M. Sjostrom, M.J. Castillo-Garzon, J.R. Ruiz. *Assessing Health-Related Fitness Tests in the School Setting: Reliability, Feasibility and Safety; the Alpha Study.* International Journal of Sports Medicine, 2010. **31**(7): p. 490-7.

163. E.G. Artero, V. Espana-Romero, J. Castro-Pinero. *Reliability of Field-Based Fitness Tests in Youth: A Systematic Review.* Int J Sport Med, 2011. **In press**.

164. F.B. Ortega, E.G. Artero, J.R. Ruiz, G. Vicente-Rodriguez, P. Bergman, M. Hagstromer, C. Ottevaere, E. Nagy, O. Konsta, J.P. Rey-Lopez, A. Polito, S. Dietrich, M. Plada, L. Beghin, Y. Manios, M. Sjostrom, M.J. Castillo. *Reliability of Health-Related Physical Fitness Tests in European Adolescents. The Helena Study.* International Journal of Obesity, 2008. **32 Suppl 5**: p. S49-57.

165. J. Castro-Pinero, F.B. Ortega, E.G. Artero, M.J. Girela-Rejon, J. Mora, M. Sjostrom, J.R. Ruiz. *Assessing Muscular Strength in Youth: Usefulness of Standing Long Jump as a General Index of Muscular Fitness.* Journal of Strength & Conditioning Research, 2010. **24**(7): p. 1810-7.

166. WHO Multicentre Growth Reference Study Group. *Who Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age: Methods and Development.* 2006, World Health Organization: Geneva.

167. M.J. LaMonte, B.E. Ainsworth. *Quantifying Energy Expenditure and Physical Activity in the Context of Dose Response.* Medicine & Science in Sports & Exercise, 2001. **33**(6): p. 370-378.

168. T.S. Olds, G.R. Tomkinson, L. Leger, G Cazorla. *Worldwide Variation in the Performance of Children and Adolescents: An Analysis of 109 Studies of the 20-M Shuttle Run Test in 37 Countries.* Journal of Sports Sciences, 2006. **24**(10): p. 1024-1038.

169. G.R. Tomkinson, L.A. Leger, T.S. Olds, G. Cazorla. *Secular Trends in the Performance of Children and Adolescents (1980-2000): An Analysis of 55 Studies of the 20m Shuttle Run Test in 11 Countries.* Sports Med, 2003. **33**(4): p. 285-300.

170. G.R. Tomkinson, T.S. Olds, J. Borms. *Who Are the Eurofittest?* Med Sport Sci, 2007. **50**: p. 104--28.

171. J. Lee, D. Koh, C.N. Ong. *Statistical Evaluation of Agreement between Two Methods for Measuring a Quantitative Variable.* Computers in Biology & Medicine, 1989. **19**(1): p. 61-70.

172. J.M. Bland, D.G. Altman. *Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement.* Lancet, 1986. **1**(8476): p. 307-10.

173. J.M. Bland, D.G. Altman. *A Note on the Use of the Intraclass Correlation Coefficient in the Evaluation of Agreement between Two Methods of Measurement.* Computers in Biology & Medicine, 1990. **20**(5): p. 337-40.

174. K. Corder, U. Ekelund, R.M. Steele, N.J. Wareham, S. Brage. *Assessment of Physical Activity in Youth.* J Appl Physiol, 2008. **105**(3): p. 977-87.

175. A. Agresti, B.A. Coull. *Approximate Is Better Than Exact for Interval Estimation of Binomial Proportions.* The American Statistician, 1998. **52**(2): p. 119-126.

176. R.A. Petersen. *A Meta-Analysis of Cronbach's Coefficient Alpha.* Journal of Consumer Research, 1994. **21**(2): p. 381-91.

177. K.R. Murphy, C.O. Davisdshofer. *Psychological Testing: Principles and Applications*. 1988, Eglewood Cliffs: Prentice-Hall.

178. N. J. Wareham. *Commentary: Measuring Physical Activity in Sub-Saharan Africa.* Int J Epidemiol, 2001. **30**(6): p. 1369-70.

179. D.G. Altman. *Practical Statistics for Medical Research*. 1991, London: Chapman & Hall.

180. A. Strohle. *Physical Activity, Exercise, Depression and Anxiety Disorders.* Journal of Neural Transmission, 2009. **116**(6): p. 777-84.

181. A. Bauman, S. Schoeppe, M. Lewicka, T. Armstrong, V. Candeias, J. Richards. *Review of Best Practice in Interventions to Promote Physical Activity in Developing Countries*. 2005, (WHO): Geneva.

182. C. M. Hoehner, J. Soares, D. Parra Perez, I. C. Ribeiro, C. E. Joshu, M. Pratt, B. D. Legetic, D. C. Malta, V. R. Matsudo, L. R. Ramos, E. J. Simoes, R. C. Brownson. *Physical Activity Interventions in Latin America: A Systematic Review.* Am J Prev Med, 2008. **34**(3): p. 224-233.

183. M. Bonhauser, G. Fernandez, K. Puschel, F. Yanez, J. Montero, B. Thompson, G. Coronado. *Improving Physical Fitness and Emotional Well-Being in Adolescents of Low Socioeconomic Status in Chile: Results of a School-Based Controlled Trial.* Health Promot Int, 2005. **20**(2): p. 113-22.

184. C.E. Draper, L. De Kock, A.T. Grimsrud, M. Rudolph, S.M. Nemutandani, T.L. Kolbe-Alexander, E. Lambert. *Evaluation of a School-Based Physical Activity Intervention in Alexandra Township.* SAJSM, 2010. **22**(1): p. 12-19.

185. M. Armstrong. *Youth Fitness Testing in South African Primary School Children: National Normative Data, Fitness and Fatness, and Effects of Socioeconomic Status*, in *Department of Human Biology*. 2009, University of Cape Town: Cape Town.

186. C. W. Fuller, A. Junge, J. DeCelles, J. Donald, R. Jankelowitz, J. Dvorak. *'Football for Health'--a Football-Based Health-Promotion Programme for Children in South Africa: A Parallel Cohort Study.* Br J Sports Med, 2010. **44**(8): p. 546-54.

187. C. W. Fuller, A. Junge, C. Dorasami, J. DeCelles, J. Dvorak. *'11 for Health', a Football-Based Health Education Programme for Children: A Two-Cohort Study in Mauritius and Zimbabwe.* Br J Sports Med, 2011. **45**(8): p. 612-8.

188. C. Burnett, W. Hollander. *Monitoring and Evaluation of Four Active Community Clubs in the Eastern Cape Province, Republic of South Africa*. 2009, Australian Sports Commission (ASC): Canberra.

189. C. Burnett, W. Hollander, R. Lombard, N. Bews, P. Singh, L. Badenhorst. *Australia - Africa 2006 Sport Development Programme: An Impact Study on the Active Community Clubs Initiative*. 2006, Australian Sports Commission (ASC): Canberra.

190. R. Beaglehole, R. Bonita, T. Kjellstrom. *Basic Epidemiology*. 1993, Geneva: World Health Organisation (WHO).

191. Gulu District Health Office (DHO). *Gulu Dho Annual Workplan 2006-2007*. 2006, Gulu DHO: Gulu.

192. Gulu District Health Office (DHO). *Gulu Dho Annual Workplan 2007-2008*. 2007, Gulu DHO: Gulu.

193. P. Onek. *Gulu District Health Sector Development Plan - 2010/2011 - 2012/13*. 2010, Gulu District Health Office: Gulu.

194. World Health Organisation (WHO). *A Prioritized Research Agenda for Prevention and Control of Noncommunicable Diseases*. 2011, WHO: Geneva.

195. E. Whitley, J. Ball. *Statistics Review 4: Sample Size Calculations.* Crit Care, 2002. **6**(4): p. 335-41.

196. G. Van Belle. *Statistical Rules of Thumb*. 2002, New York: John Wiley & Sons.

197. StatsDirect. *Sample Size Estimation*. 2011; Available from: https://statsdirect.com/help/statsdirect.htm#sample_size/sscorr.htm.

198. J.R. Thomas, J.K. Nelson. *Research Methods in Physical Activity*. 2001, Champaign: Human Kinetics.

199. G. Keppel. *Design and Analysis: A Researcher's Handbook*. 4th ed. 2004, Englewood Cliffs, N.J: Prentice-Hall. xii, 611 p.

200. D.C. Montgomery. *Design and Analysis of Experiments*. 7 ed. 2009, Hoboken: John Wiley and Sons.

201. B.R. Kirkwood, J.A.C. Sterne. *Medical Statistics*. 2006, Oxford: Blackwell Science Ltd.

202. W. Chen, C.C. Lin, C.T. Peng, C.I. Li, H.C. Wu, J. Chiang, J.Y. Wu, P.C. Huang. *Approaching Healthy Body Mass Index Norms for Children and Adolescents from Health-Related Physical Fitness.* Obes Rev, 2002. **3**(3): p. 225-32.

203. T.S. Olds, K. Ridley, G.R. Tomkinson. *Declines in Aerobic Fitness: Are They Only Due to Increasing Fatness?* Med Sport Sci, 2007. **50**: p. 226-40.

204. C. Bouchard, R.M. Malina, L. Perruse. *Genetics of Fitness and Physical Performance*. 1997, Champaign: Human Kinetics.

205. K. Hughes, C. Hutchings. *Can We Obtain the Required Rigour without Randomisation? Oxfam Gb's Non-Experimental Global Performance Framework.* 2011, The International Initiative for Impact Evaluation (3ie): New Delhi.

206. S. Allender, G. Cowburn, C. Foster. *Understanding Participation in Sport and Physical Activity among Children and Adults: A Review of Qualitative Studies.* Health Educ Res, 2006. **21**(6): p. 826-35.

207. NICE Public Health Collaborating Centre. *Promoting Physical Activity for Children: Review 3 – Qualitative Correlates*. 2007, NICE Public Health Collaborating Centre: London.

208. G. Brunton, A. Harden, R. Rees, J. Kavanagh, S. Oliver, A. Oakley. *Children and Physical Activity: A Systematic Review of Barriers and Facilitators*, Institute of Social Science Research Unit, University of London Education, Editors. 2003, Evidence for Policy and Practice Information and Co-ordinating Centre: London.

209. F.A. Swarts. *Ethics and Global Peace: Ngo Perspectives*, World Congress of NGOs, Editor. 2007, World Association of Non-Government Orgaisations (WANGO): Toronto.

210. P.S. Freedson, E. Melanson, J. Sirard. *Calibration of the Computer Science and Applications, Inc. Accelerometer.* Med Sci Sports Exerc, 1998. **30**(5): p. 777-81.

211. N. Owen, G. Healy, D. Dunstan. *The Science of Sedentary Behaviour*, in *Scottish Physical Activity Research Collaboration (SPARColl)*. 2010, SPARColl: Glasgow.

212. M. Hagstromer, R.P. Troiano, M. Sjostrom, D. Berrigan. *Levels and Patterns of Objectively Assessed Physical Activity--a Comparison between Sweden and the United States.* Am J Epidemiol, 2010. **171**(10): p. 1055-64.

213. J. Staudenmayer, D. Pober, S. Crouter, D. Bassett, P. Freedson. *An Artificial Neural Network to Estimate Physical Activity Energy Expenditure and Identify Physical Activity Type from an Accelerometer.* J Appl Physiol, 2009. **107**(4): p. 1300-7.

214. A. Doherty, P. Kelly, O. Mytton, A. Hamilton, J. Kerr, S. Marshall, H. Badland, C. Foster. *Making Counts, Count More? Using Wearable Cameras to Contextualise Accelerometer Data.* Med Sci Sports Exerc, 2012. **under review**.

215. A. Nedeljkovic, D.M. Mirkov, M. Kukolj, D. Ugarkovic, S. Jaric. *Effect of Maturation on the Relationship between Physical Performance and Body Size.* J Strength Cond Res, 2007. **21**(1): p. 245-50.

216. A.D. Mahon. *Aerobic Training*, in *Paediatric Exercise Science and Medicine*, N. Armstrong, W. van Mechelen, Editors. 2008, Oxford Uninversity Press: Oxford

217. W.J. Kraemer, J.F. Patton, S.E. Gordon, E.A. Harman, M.R. Deschenes, K. Reynolds, R.U. Newton, N.T. Triplett, J.E. Dziados. *Compatibility of High-Intensity Strength and Endurance Training on Hormonal and Skeletal Muscle Adaptations.* J Appl Physiol, 1995. **78**(3): p. 976-89.

218. N. Schulenkorf, J. Sugden. *Sport for Development and Peace in Divided Societies: Cooperating for Inter-Community Empowerment in Israel.* European Journal for Sport and Society, 2011. **8**(4): p. 235-256.

219. C. Burnett. *Social Impact Assessment and Sport Development: Social Spin-Offs of the Australia-South Africa Junior Sport Programme.* International Review for the Sociology of Sport, 2001. **36**: p. 41-57.

220. K.L. Siegenthaler, G.L. Gonzalez. *Youth Sports as Serious Leisure: A Critique.* Journal of Sport & Social Issues, 1997. **21**(3): p. 298-314.

221. Z. Kaufman. *Transfer Document: Assessing the Effectiveness of Sports-Based Hiv Prevention: A Systematic Review of the Evidence and Protocol for a Cluster-Randomised Trial*. 2011, London School of Hygiene and Tropical Medicine: London.

222. R. J. Feise. *Do Multiple Outcome Measures Require P-Value Adjustment?* BMC Med Res Methodol, 2002. **2**: p. 8.

223. J. Zhang, H. Quan, J. Ng, M. E. Stepanavage. *Some Statistical Methods for Multiple Endpoints in Clinical Trials.* Control Clin Trials, 1997. **18**(3): p. 204-21.

224. J.A. Richards, K. Siefken, A. McManus, C. Foster. *Physical Activity Research in Developing Countries – What Are the Opportunities and Challenges?* , in *Annual meeting for the International Society for Behavioural Nutrition (ISBNPA)*. 2011, ISBNPA: Melbourne.

225. J.A. Richards, C. Foster. *The Reliability of Physical Fitness Testing for Children in Gulu, Northern Uganda – a Post-Conflict Setting in a Low-Income Country*, in *Annual meeting for the International Society for Behavioural Nutrition (ISBNPA)*. 2011, ISBNPA: Melbourne.

226. Z. Kaufman, J. Richards, E. Wolff, K. Gannett. *Journal of Sport for Development (Jsfd)*. 2011; Available from: http://jsfd.org/.