Scaling-up the Parenting for Lifelong Health Intervention in South Sudan

Author:
Campbell McDuling

Supervisors:
Professor Francesca Little
Heiletje van Zyl

Student Number:
MCDCAM001

Declaration of Authorship

I, Campbell McDuling, declare under good faith that this report titled 'Scaling-up the Parenting for Lifelong Health Intervention in South Sudan' is solely my own work. I confirm that:

- Any contributions to the project, or references to other's work, have been appropriately credited.
- All use of other people's work and software packages have been cited and referenced using the Harvard referencing system.
- This research was completed in partial fulfillment of an Honours degree in Statistical Sciences at the University of Cape Town.

Name: Campbell McDuling

Signature: 

Date: 20 October 2022

Abstract

Background: This report analysed data from the PLH for Adolescents program delivered to 295 adolescents and 290 caregivers in South Sudan. The program aims to improve caregiver and adolescent outcomes with respect to positive parenting, child behaviour, harsh discipline, financial coping, caregiver mental health, and parental support of education.

Methods: The internal consistency of these socio-psychological indicators was assessed through a confirmatory factor analysis and computation of Cronbach's alpha. The changes over the course of the study was quantified using generalised linear and cumulative link mixed effect models. Furthermore, generalised linear models were used to identify risk factors for positive parenting, child behaviour, and harsh discipline at baseline.

Results: Significant improvements were seen among the positive parenting, child behaviour, and harsh discipline scores, as well as parental support of education items, from both adolescent and caregiver perspective. Furthermore, the data provides evidence of improvements to family financial coping scores, and depression and loneliness items, when recorded from the caregiver perspective. The item reliability analysis revealed very poor to unacceptable internal consistency for all of the scores except for the adolescents' baseline harsh discipline scores, while the caregiver mental health and parental support of education domains contained too few items to conduct this analysis.

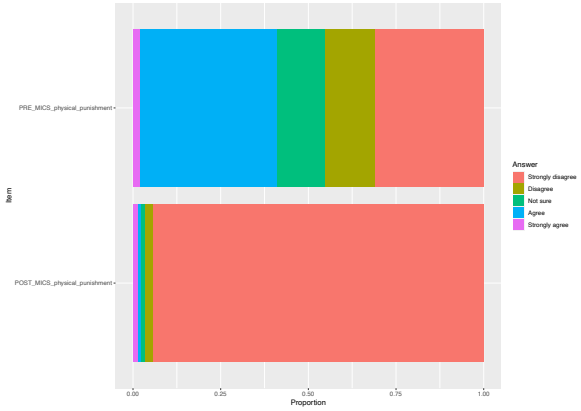
Conclusions: Analysis of the data indicates the low-cost scaled-up implementation of the PLH program may have been effective in improving positive parenting, child behaviour, and harsh discipline among this sample of South Sudanese caregivers and their adolescents. There are, however, concerns regarding the validity of the data which must be considered when interpreting the results.

South Sudan
Parenting for Lifelong Health: Promoting Positive Parenting and Preventing Violence

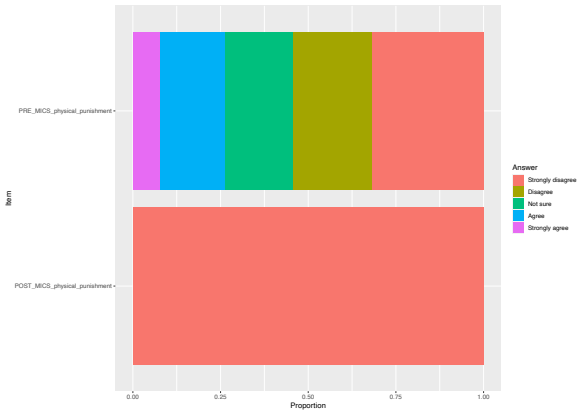
Parenting for Lifelong Health: Programmes to support parents with positive parenting, in order to prevent violence against children and promote child wellbeing

Research goal: To test whether Parenting for Lifelong Health programmes in South Sudan are achieving their goals.

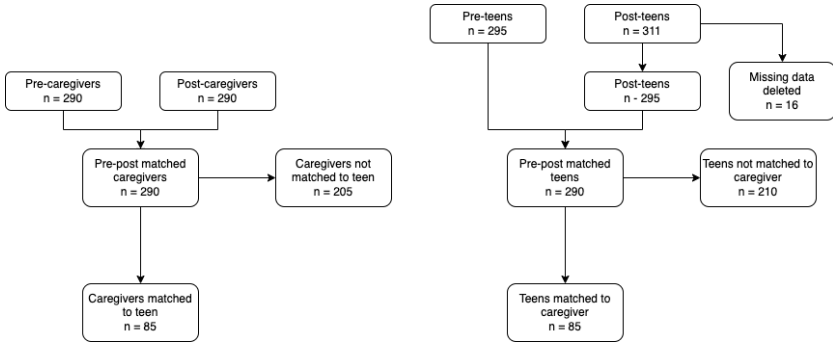
Participants: 290 parents with adolescents aged 10-17. Caregivers were matched to adolescents according to a non-unique household ID and demographic factors present in both questionnaires, resulting in a reduction in sample size to 85 caregiver-teen dyads upon which the analysis was conducted.



Caregiver agreement with corporal punishment at baseline (top) and follow-up (bottom)



Adolescent agreement with corporal punishment at baseline (top) and follow-up (bottom)



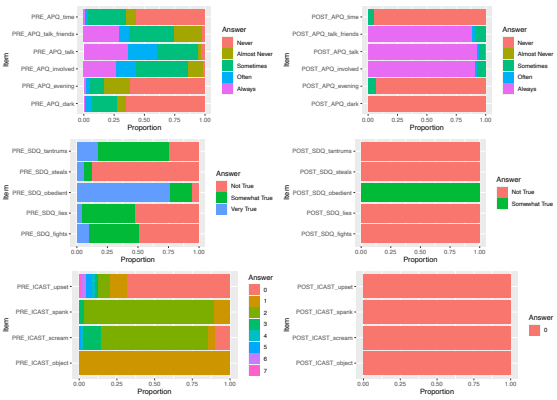
These caregivers were 79% female with a mean age of 40 years, comprised of 91% biological parents. The matched adolescents were 60% female with a mean age of 13.7 years. Around a third of the teenagers were illiterate and 11% had children of their own. Almost a third of the caregivers reported that the child's biological father was deceased, while around 15% reported that the mother was deceased. 93% of the caregivers reported having run out of money for food and essentials in the previous month. Around 74% reported living with someone who suffers from, or had succumbed to, either HIV or TB, while 21% reported caring for a severely unwell child.

Intervention: The Parenting for Lifelong Health (PLH) programme for Teens is delivered over 14 sessions in facilitated groups. Facilitators work collaboratively with parents to help them build a relationship with their children, to use positive parenting techniques (such as praise) and non-violent consequences to support their children to develop good behavioural habits. In addition, families learn to budget well, and agree on ways to keep each other safe. Manuals are available free of charge at Parenting for Lifelong Health.

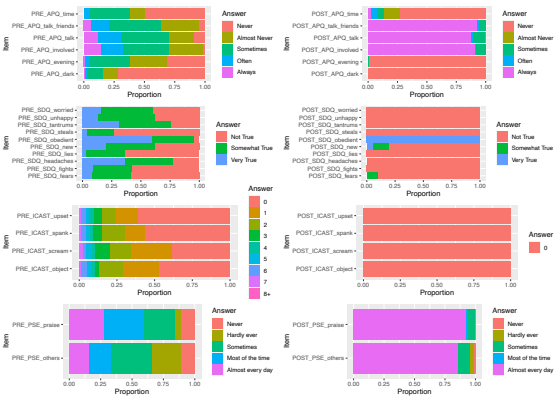
PLH was delivered to 290 caregivers and their children in South Sudan. Facilitators were trained and recruited in partnership with Clowns Without Borders SA.

Partners

Funders



Changes in selected caregiver responses from baseline (left) to follow up (right). From top to bottom: positive parenting, child behaviour, harsh discipline.



Changes in selected adolescent responses from baseline (left) to follow up (right). From top to bottom: positive parenting, child behaviour, harsh discipline, support of education.

Study design: All parents completed surveys before and after the programme. This data was analysed to determine whether the programme had an impact on harsh discipline, positive parenting, child behaviour, family financial coping and parental support of education.

Data analysis:

- Confirmatory factor analysis and Cronbach's alpha were used to assess the internal consistency of scores.
- Generalized linear models were used to assess for risk factors at baseline.
- Generalized linear and cumulative link mixed-effect models were used to quantify the changes from baseline to follow-up and identify moderators of these changes.

Key Findings:

- After the study, caregivers reported improved positive parenting (~30%), child behaviour (~59%) and reduced harsh discipline (100%).
- Caregivers also reported reduced depression, loneliness, and improved positive parenting. One negative effect was observed with respect to increased feelings of everything requiring effort.
- After their participation, adolescents reported improved positive parenting (~48%), child behaviour (~97%) and reduced harsh discipline (100%). No negative effects were observed among the teens.
- Caregivers and adolescents reported significantly more disagreement with the necessity of corporal punishment.
- The internal consistency of the above scores was mostly very poor.

Limitations of this study:

- There is no control group of parents who did not receive the programme, and so we cannot definitively state that the programme caused the changes we have observed.
- The matching procedure was flawed and greatly reduced the sample size.
- Facilitator and attendance information was not recorded.
- The validity of the data is questionable and not verified.

Implications for Policy and Practice:

- PLH for Teens, a group-based parenting programme, shows promise in reducing child maltreatment and improving child well-being in South Sudan.
- Closer supervision over the data capturing process is required for more effective future analysis.

For more information, please contact Parenting for Lifelong Health and the research team at parentinglh@gmail.com.

Partners

World Health Organization · unicef for every child · WWO · Childhood · End Violence Against Children · Together for girls · Accelerate Hope · Global Women's Institute · Parenting for Lifelong Health · IDEMS · Internet of Good Things · USAID · UNODC · ECDAN · Maestral · CDC

Funders

UK Research and Innovation · GCRF · Newton Fund · The LEGO Foundation · The Human Safety Net · University of Oxford · AK Foundation · unicef for every child · End Violence Against Children · erc

Contents

Acknowledgments

This study would not have been realized without the expertise and efforts of a number of important contributors. Firstly, my most sincere gratitude is extended towards Prof. Francesca Little - who provided me the opportunity to work on a research project of such significance - and Heiletje van Zyl - who was consistently willing to provide insight and advice. A better pair of supervisors will scarcely be found in the world of academia, and the education which has been imparted on me through the course of working on this project has been invaluable. In addition, I would like to thank my fellow Statistical Sciences Honours classmates for providing an extremely positive and supportive environment in which to work.

Secondly, I am extremely appreciative of the researchers leading the SUPER study - namely Prof. Lucie Cluver, Prof. Catherine Ward, Dr. Jamie Lachman and Dr. Yulia Shenderovich - and the investigators working along side them (Prof. F. Gardner; Prof. M. Tomlinson; Dr. I. Wessels; Dr. D. Oliver; Dr. H. Sacolo; Dr. K. Okop). The passion shown by this group, and the years of hard work which have gone into this study, is extremely inspiring. In particular, I'd like to thank Prof. Ward for her involvement, and Prof. Cluver for her enthusiasism and advice.

Thirdly, there are a number of facilitators, data capturers and other people fulfilling crucial roles on the ground in South Sudan without whom this study would never have been completed. This group included Clowns Without Borders and all who were involved in the implementation of the study. Furthermore, I'd like to recognize the work of Bojo Samuel who provided an extensive study design encapsulating the broader South Sudanese Orphaned and Vulnerable Children intervention within which this study was delivered.

Lastly, the possibility of working on this project and receiving the education I have thus far would not have been possible without the support of my father, who has been incredibly encouraging and patient, and deserves massive credit for the sacrifices he has made to this extent.

List of Figures

List of Tables

1 Introduction

Parenting for Lifelong Health (PLH) is an evidence-based parenting intervention programme developed and implemented through the collaboration and research of various academics and NGO's. The primary aim of the intervention is to improve the relationships between caregivers and their children – contributing to improvements in the socio-economic and psychological outcomes of both the individuals and the family.

The programme is mostly implemented in low to middle-income countries where child maltreatment rates are frequently unacceptably high (Ronnie, 2021). Often child maltreatment is caused by negative and inconsistent parenting, sometimes manifesting in physical and emotional abuse between caregiver and child. Child maltreatment creates significant obstacles to the social and economic progression of both child and family, causing the proliferation and propagation of anti-social behaviour. This predisposition may expose children and adolescents to a greater than average risk of poverty, sexually transmitted diseases, and substance abuse, among other serious outcomes. The Parenting for Lifelong Health (PLH) programme aims to intervene in negative relationships between caregiver and child through education in positive, consistent parenting and non-violent forms of discipline. It has been proven as an effective intervention program through randomized controlled trials in South Africa among both young children (Ward et al., 2020) and adolescents (Cluver, et al., 2018).

The PLH Scale-up of Parenting Evaluation Research (PLH SUPER) study aims to assess the effectiveness of the extension of this program to different real-world environments. The research goal of this report is thus to test the effectiveness and scalability of PLH in reducing child maltreatment, and improving caregiver-child relationships and child well-being in South Sudan. The following objectives will be used to assess the effectiveness: (1) quantify the change in socio-psychological indicators over the course of the program; (2) validate the internal consistency of the items comprising the indicators. A secondary objective is defined to identify obstacles to scalability. This will be evaluated among a sample 290 caregiver and 295 adolescent participants in the African nation of South Sudan.

2 Literature Review

Cluver et al. (2017) evaluates the impact of the PLH program on adolescents and their caregivers, among 552 families from 40 locations around the Eastern Cape Province in South Africa. This was done through the implementation and analysis of a pragmatic cluster, randomized controlled trial. The main findings include significantly greater reductions in the rates of abuse, corporal punishment and poor caregiver supervision, and improvements in positive and involved parenting among the treatment group when compared to the control group. Further to this, Ward et al. (2020), analysed the impact of the PLH program for Young Children in a randomized controlled trial among a cohort of 380 caregiver-child dyads in the Western Cape of South Africa. Key findings include significant improvements in positive parenting and child behaviour associated with the intervention. These studies provided evidence of the effectiveness of the programs among adolescents and young children in a clinical, controlled environment - however do not provide evidence of PLH's effectiveness in different real-world contexts.

The literature on child welfare specific to South Sudan is limited. Mugo, et al. (2015) provide a systematic review of the obstacles and barriers to promoting maternal, newborn and child health in the context of modern South Sudan. The article highlights the significant issues of fractured social institutions, and unstable political and macro-economic conditions, inherited from years of civil unrest and war. The authors identify a need for sustainable intervention and aid at a grass-roots level. Canavera et al, (2016), provide a qualitative analysis of the challenges to child protection services in South Sudan through key-informant and community interviews, identifying the poor conditions and harsh realities that many children in the country face. The article emphasizes the historical lack of Sudanese input and decision-making power within the humanitarian intervention structures present in the country, identifying the need to align such efforts with cultural and community-specific engagement.

3 Methods

3.1 Study Design

This study took place within a broader Orphaned and Vulnerable Children (OVC) intervention designed to provide education on positive parenting practices, household economic strengthening, and sexual health (Samuel, 2020). Participants were selected based on their vulnerability to HIV and AIDS, and their proximity to the study location. Of primary relevance to this report is the positive parenting part of the study, conducted in association with Clowns Without Borders South Africa (CWBSA). An adaptation of the PLH program was delivered to adolescents between the ages of 10 and 17 years, and their associated caregivers in weekly group sessions hosted by pairs of trained facilitators. The sessions included role-playing activities, group discussions, illustrations of positive parenting practices, and home activity assignments. Both caregiver and adolescent completed surveys at baseline and follow-up. The length of the program varied for each participant, as the end-line measurement was reached upon graduation from the program. On average, the follow-up surveys were completed around 18 months after graduation.

3.2 Data Sources and Measurement

The data used for analysis is longitudinal, primary data of demographic, and pre-post monitoring assessments collected by the data capturers in South Sudan. The assessment forms for the caregivers and adolescents can be found in Figures 24 and 25 The surveys were initially completed on paper by the participants, and then entered onto Excel sheets by the data capturers. This includes metrics constructed through the weighted sum of many self-reported Likert-scale questionnaire answers, featured below in Tables 2 and 3. The metrics are related to the primary and secondary outcome measures, and the aggregation of the individual responses to the analysed metric is done in alignment with the prescription of psychologists working with the programme. The questions posed to adolescents fall into five sections: background/demographic; relationship with caregiver (APQ); own behaviour (SDQ); discipline from caregiver (ICAST); educational support (PSE). All the responses pertain to the preceding month. The responses to the relationship with caregiver are on a discrete, ordinal scale of 0 to 5 where the options relate to frequency. An example of such an individual question is: "how often do you go out after dark without an adult present?". The responses to the behavioural questions are also ordinal and discrete, referring to agreement of the respondent to the question. For example, participants were asked to what extent they agree with the statement, "I worry a lot.". The responses pertaining to caregiver discipline relate to the frequency of occurrence of types of discipline. The questions posed to the caregiver are similar but also include additional information regarding their household financial status (FFC) and mental health (CESD). These outcome variables are presented in further detail in Section 4.5.

3.3 Data Management

The data management procedure is outlined in Table 1 and discussed below.

Table 1: Data Management Process

Data Management Process	
Step 1	Normalize variable names
Step 2	Code all responses as numeric
Step 3	Check for missing data
Step 4	Check for duplicates
Step 5	Validate variable ranges
Step 6	Reverse code identified variables
Step 7	Create composite variables
Step 8	Match adolescents with caregivers
Step 9	Spot checks to validate matching
Step 10	Investigate response patterns

The raw data received had irregular and overly complex variable labels, and thus the first step in the management process was to normalize all variable names to a universal format. In step 2, all responses were re-coded as numeric so as to make the data management and analysis easier. For example, the values for a participant's sex was changed from "female" or "male" to "0" or "1", and outcome responses entered as strings were re-coded to numeric variables reflecting the Likert scale on which they were measured. Step 3 entailed a check for observations where all, or almost all, of the responses were missing. These observations were useful for neither the exploratory nor main analysis, and were thus removed from the datasets. A check for duplicate entries was carried out in step 4, looking for any observations that were entered more than once and thus had identical values for every response. Fortunately no duplicates were found in the raw data.

The next step was to validate the ranges of each variable, and identify any impossible values. Firstly, this was done for demographic variables. Teenagers were required to be between 9 and 19 years, in accordance with the study protocol. Any recorded observations outside of these ranges were set to "NA", and assumed to be errors in data capturing. All other demographic variables were within plausible or expected ranges. The same process was done for the outcome measures. Almost all of the responses were within the expected range, barring a select few SDQ responses which were outside of the range 0-2 and were replaced with "NA".

Step 6 entailed identifying those response variables in the questionnaire which were inconsistent with the other variables in the respective domain in terms of the sentiment of the Likert scale. This was done to ensure consistency in the direction of measurement, which is necessary for step 7. In the caregiver survey, the latter three variables (APQ_evening; APQ_time; APQ_dark) relating to the "Involved parenting and parental monitoring" domain were measured in the opposite direction to the former three variables. To elaborate, consider a caregiver answers "Never" to both APQ_talk and APQ_evening. One would consider the response to signal a lack of parental involvement in the first case, but an abundance of the same underlying feature in the second. For these reasons, the latter three variables needed to be "reverse-coded" so that a response of "Never" would translate to a high score. The same process was applied to SDQ_obedient. Among the adolescent responses, the same variables as above were reverse-coded.

In step 7, composite variables were created by summing up the individual scores for each item in each domain. A score was thus created for each respondent in each domain, excluding child discipline and parental support of education - since there were only one and two measured items, respectively (refer to Tables 2 and 3). There were thus 6 composite variables created for each caregiver and 3 created for each adolescent in the study. Step 8 involved matching caregiver-teen dyads, necessary to exclude from the analysis any participants which did not have an associated 'partner'. After matching, observations were inspected at random to ensure that the matching procedure had worked. In step 10, the responses were analysed to identify any unusual response patterns which would induce bias into the analysis. For example, if a participant had provided the same response for every question, then this would most likely be due to a data capturing error or a lazy respondent and distort the results. Alternatively, if a particular question had no variation in responses, this would cast doubts over the validity of the data and also present problems in assessing internal consistency and implementing models.

3.4 Participants

Figure 1 below tracks the changes in sample size resulting from the data management process. Considering the caregiver data, two data sets were initially received containing 290 observations each. The post-intervention adolescent data initially had 6 observations more than the pre-intervention data, however these observations were identified as empty rows, and were removed. No loss to follow up was identified when matching the pre-intervention to the post-intervention data (i.e. all pre-observations were matched with a post-observation, and vice-versa).

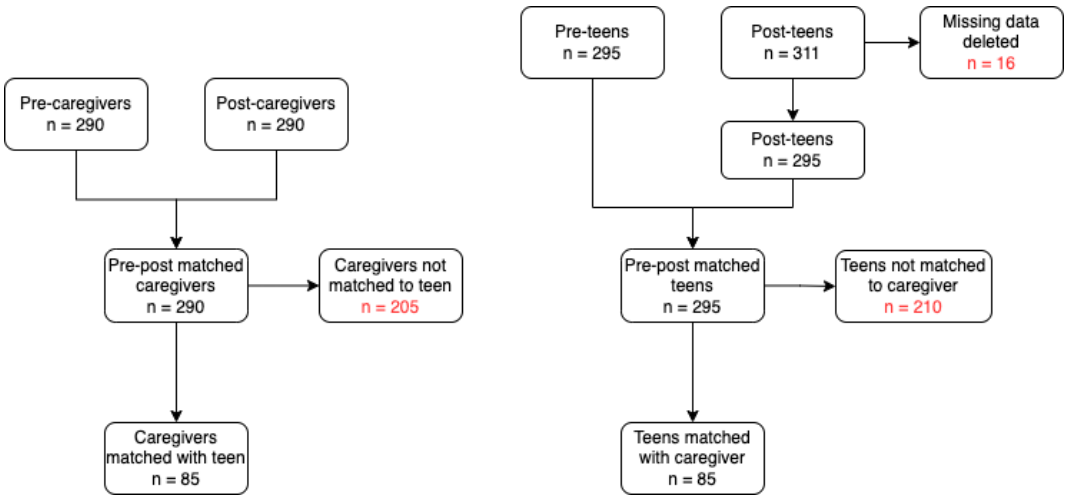


Figure 1: Participant Flowchart

The first significant problems with the data capturing process were identified when attempting to match the caregiver-teen dyads. While each participant was assigned a participant identifier, which allowed the matching of pre-intervention to post-intervention data, there was no unique identifier linking a particular caregiver with their corresponding teenager. Only a household ID was provided, which was not unique since there were multiple teens and caregivers from the same household. In addition, it was identified that caregivers with multiple teens had only completed one survey. Ideally, if a caregiver had entered the study with multiple adolescents, they would have completed a survey per adolescent. This is extremely problematic since the caregiver should have a unique relationship with each teenager under their supervision and the data does not capture this fact. In the absence of a matching ID, a strategy was devised to match caregivers to adolescents on the basis of the household ID and the child's gender - both of which were provided in every data set. Other combinations of variables present in both caregiver and teenager data were tried, for example matching on household ID and adolescent age, however there were too many inconsistencies between the caregiver and teen information for any of these combinations to be more effective than household ID and child gender. The result was far from optimal, and in the end only 85 caregiver-teen dyads were able to be matched. This meant that 205 caregivers and 210 teenagers were lost at this point of the analysis. The significance of this is that the demographic profile of the matched dyads may not be representative of the full sample, which would induce bias into the analysis. This problem is addressed Section 5, where the baseline demographic profiles of the full sample are compared to that of the matched dyads, and any emerging biases are identified.

3.5 Variables

The outcome variables included in the caregiver surveys are presented in Table 2. There were five main Likert scales relating to the following latent variables: involved parenting and parental monitoring; child behaviour; harsh discipline; parental depression; and economic strengthening. Going forward, each Likert scale is either referred to by the corresponding acronym - APQ, SDQ, ICAST, CESD, FFC - or by the domain which they represent (see Table 2). Each domain is represented by a number of questions, which supposedly together provide a representation of the underlying factor. In addition, a single question was asked relating to the participants opinion on harsh discipline (MICS), and two questions were asked relating to support of the adolescents education (PSE).

Table 2: Caregiver Outcome Measures

Domain	Variable Name	Question Posed
Involved parenting & parental monitoring		For each statement please tell us whether it is: Never, Almost never, Sometimes, Often, or Always
	APQ_talk	You have a friendly talk with your child
	APQ_involved	You get involved in activities that your child likes
	APQ_talk_friends	You talk to your child about your friends
	APQ_evening	Your child stays out in the evening past the time when he/she is supposed to be home
	APQ_time	Your child goes out without a set time to be home
	APQ_dark	Your child goes out after dark without an adult with him/her
Child behaviour		Please answer these questions about your child's behaviour in the past month: Not True, Somewhat True, or Very True
	SDQ_tantrums	Your child often has temper tantrums or hot tempers
	SDQ_obedient	Your child is generally obedient, usually does what adults request
	SDQ_fights	Your child often fights with other children or bullies them
	SDQ_lies	Your child often lies or cheats
	SDQ_steals	Your child steals from home, school or elsewhere
Harsh discipline		Can you tell me how many times in the past 4 weeks you did the following things?
	ICAST_spank	How often did you discipline your child by spanking, slapping or hitting with your hand?
	ICAST_object	How often did you discipline your child with an object like a stick or a belt?
	ICAST_scream	How often did you shout, yell or scream at your child?
	ICAST_upset	How often did you say mean things to your child that upset him/her?
Child discipline		Please tell us to what degree you agree with the following statement: Strongly disagree, disagree, not sure, agree, strongly agree
	MICS_physical_punishment	In order to bring up, raise, or educate a child properly, a child needs to be physically punished
Parental Depression		Please say how often in the past week (7 days) you have felt this way.
	CESD_depressed	How often in the past week have you felt depressed?
	CESD_effort	How often in the past week have you felt that everything you did was an effort?
	CESD_lonely	How often in the past week have you felt lonely?
Economic strengthening		Please tell us in the past four weeks:
	FFC_meat	How many times have you run out of money for two meals a day?
	FFC_electricity	How many times have you run out of money for fuel for cooking/electricity?
	FFC_transport	How many times have you run out of money for transport?
	FFC_airtime	How many times have you run out of money for airtime?
	FFC_worried	How many times were you worried about money?
Parent Support of Education	Parent Support of Education	Please tell us to what extent you do the following activities: Never, hardly ever, sometimes, most of the time, almost everyday.
	PSE_praise	I praise my child for working hard at school
	PSE_others	I support my child's schoolwork in any way that I can

The adolescent surveys were similarly designed, except they excluded the CESD and FFC items, and included additional questions relating child behaviour (Table 3). All responses are encoded as ordinal integers, except for the harsh discipline questions, which are count variables.

Table 3: Adolescent Outcome Measures

Domain	Variable Name	Question Posed
Involved parenting & parental monitoring		For each statement please tell us whether it is: Never, Almost never, Sometimes, Often, or Always
	APQ_talk	You have a friendly talk with your main caregiver
	APQ_involved	Your caregiver gets involved in activities that you like
	APQ_talk_friends	Your caregiver talks to you about your friends
	APQ_evening	You stay out in the evening past the time when you are supposed to be home
	APQ_time	You go out without a set time to be home
	APQ_dark	You go out at night without an adult present
Child behaviour		For each of the following statements please tell us if it is: Not True, Somewhat True, or Very True
	SDQ_headaches	I get a lot of headaches, stomach aches or sickness
	SDQ_tantrums	I get angry and often lose my temper
	SDQ_obedient	I usually do as I am told
	SDQ_worried	I worry a lot
	SDQ_fights	I fight a lot. I can make people do what I want
	SDQ_unhappy	I am often unhappy, downhearted or tearful
	SDQ_new	I am nervous in new situations. I easily lose confidence
	SDQ_lies	I am often accused of lying and cheating
	SDQ_steals	I take things that are not mine from home, school or elsewhere
	SDQ_fears	I have many fears, I am easily scared
Harsh discipline		Can you tell me how many times in the past 4 weeks your caregiver did the following things?
	ICAST_spank	How often did your caregiver spank, slap or hit you with his/her hand?
	ICAST_object	How often did your caregiver discipline you with an object like a stick or a belt?
	ICAST_scream	How often did your caregiver shout, yell or scream at you?
	ICAST_upset	How often did your caregiver say mean things to you?
Child discipline		Please tell us to what degree you agree with the following statement: Strongly disagree, disagree, not sure, agree, strongly agree
	MICS_physical_punishment	In order to bring up, raise, or educate a child properly, a child needs to be physically punished
Parent Support of Education	Parent Support of Education	Please tell us to what extent your caregiver does the following activities: Never, hardly ever, sometimes, most of the time, almost everyday.
	PSE_praise	Your caregiver praises you for working hard at school
	PSE_others	Your caregiver supports your schoolwork in any way that they can

The surveys were adapted from their original form. The APQ scale was adjusted from the Alabama Parenting Questionnaire; SDQ from the Strengths and Difficulties Questionaire; ICAST from the ISPCAN Child Abuse Screening Tool; CESD from the Center for Epidemiological Studies Depression scale; FFC from the Family Financial Coping scale; and PSE from the Parental Support of Education scale. Crucially, for all of these scales the adjusted versions contain far fewer items than in the original design. As a result, a confirmatory factor analysis is conducted to assess the internal consistency of the items - presented in Section 6.

3.6 Statistical Methods

The following section outlines the statistical methods used to evaluate the main objectives. All analyses were performed, and all figures generated, using the R software environment (packages cited in the Reference List).

3.6.1 Item Reliability Analysis

The internal consistency of each Likert scale will be assessed through the implementation of confirmatory factor analysis. Factor analysis will be used to deconstruct the covariance of the individual response variables and confirm that they group together to represent the underlying "factors" which they attempt to measure – such as harsh parenting. For example, a variable pertaining to how often a caregiver shouts at their child and a variable pertaining to how often the caregiver says mean things about a child are likely to be highly correlated, and thus represent an underlying factor relating to emotional abuse. The observed covariance structure, $\mathbf{\Sigma}$, is used to estimate the parameters in the model-implied covariance matrix:

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Psi}}\hat{\mathbf{\Lambda}}^T + \hat{\mathbf{\Theta}}_\epsilon$$

where $\hat{\mathbf{\Lambda}}$ is a matrix of factor loadings, $\hat{\lambda}_{ij}$, which represent how 'heavily' item i loads on the Likert scale j; $\hat{\mathbf{\Psi}}$ contains the estimated variance-covariance structure of the Likert scales; and $\hat{\mathbf{\Theta}}_\epsilon$ contains the estimated residuals. The 'marker' method is employed in this analysis, which fixes the first loading of each factor to a value of 1 - ensuring a unique solution is found (Kline, 2011). A confirmatory factor analysis has the additional step of assessing a null hypothesis specified as:

$$H_0 : \mathbf{\Sigma}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{\Sigma}}$$

$$H_1 : \mathbf{\Sigma}(\hat{\boldsymbol{\theta}}) \neq \hat{\mathbf{\Sigma}}$$

Failing to reject the null hypothesis can be loosely interpreted as failing to disprove that the model specified by the hypothesized factors and their associated items is valid. This provides ways to assess the goodness-of-fit of the factor model. In the case of ordinal data such as the Likert scales and the ordinal items comprising them here, approximate fit indices are deemed appropriate in assessing the goodness-of-fit of the above models (Hu and Bentler, 1999). Essentially, these indices represent the extent of information observed among the data which the confirmatory factor analysis models are able to capture. In this report, both the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) are used to assess goodness-of-fit, along with the Root Mean Squared Error (RMSE). In general, a CFI and a TLI close to 1 in combination with an RMSE lower than 0.05 signal an acceptable fit (Kline, 2011).

In addition, Cronbach's alpha will be used in conjunction with factor analysis to confirm the internal consistency of the individual questions that measure the latent variable of interest (Cronbach, 1951). This statistic is defined as:

$$\alpha_x = \frac{p}{p-1}\left(1 - \frac{\Sigma_{i=1}^{p}\sigma_{y_i}^2}{\sigma_x^2}\right)$$

where p is the number of items in Likert scale x; σ_x^2 is the variance Likert scale x; $\sigma_{y_i}^2$ is the variance of the i-th item in x.

An α value greater than 0.7 is usually considered to represent acceptable internal consistency; while anything lower than 0.6 is considered to represent poor consistency. A point estimate of the Cronbach's alpha is supplied for each applicable scale at baseline and at follow-up in the Results section. Further to this, a 95% bootstrapped confidence interval is computed by sampling from the data with replacement 1000 times and calculating Cronbach's alpha each time, and then taking the upper and lower 2.5% quantiles.

3.6.2 Regression Models

One way in which meaningful information can be extracted from the data is by modeling the cross-sectional pre-intervention Likert scores as functions of demographic factors, in a generalised linear modeling (GLM) framework. While this clearly does not aid in evaluating the primary objective, it is still useful in identifying significant baseline risk factors on a population level. Understanding these baseline moderators will help to

analyze those moderators which are important for explaining the change in score over time. For example, if the harsh discipline scores show variation at baseline but not at follow up, the change in an observation over time is not necessarily as meaningful as the position at baseline since all of the observations at the second time point would be homogeneous. The change over time would thus reflect the position at baseline rather than quantify the effect of the program. By regressing the pre-intervention score against a constant, one can estimate the average total harsh discipline score at baseline, and adjust this for sample strata by including a moderator fixed effect.

If x_i represents the i-th participants sex, and y_{i0} represents their ICAST score at baseline, the average effect of sex is modeled through β_1:

$$f(y_{i0}) = \beta_0 + \beta_1 x_i + \epsilon_{i0}$$

where $\epsilon_{i0} \sim D(0, \sigma_0^2)$; $f()$ is the link function appropriate for the assumed distribution D.

$\hat{\beta_1}$ thus represents the estimated average difference in score between females and males, and may indicate whether x is a risk factor for harsh discipline. In the case of a log-link function, $f()$, $e^{\hat{\beta_1}}$ reveals the estimated multiplicative factor by which the score differs between the two strata. For each total score, the effect of moderators are assessed in a simple GLM, rather than the partial effect through multiple linear regression - since the average difference between groups is of more interest that building a predictive model. The estimation method used for GLM's in this analysis is Iteratively Reweighted Least Squares (Holland & Welsch, 1977).

Of course, the assumptions underlying generalised linear regression models need to be substantiated for these estimates to be valid. The Pearson residuals of each model are thus examined against the fitted model values to assess whether they are distributed according to the assumed underlying distribution D with a constant variance σ_0^2.

Since the data consists of repeated measurements on the same participants, there exists an inherent dependence between the outcome measures. Essentially, a participants Likert score at baseline is correlated with their score after the intervention. This means that conventional least squares regression methods would result in biased coefficients, and thus the modeling procedure needs to be adjusted. This within-subject correlation is captured within a generalised mixed effect model by imposing a random effect on participant ID, which captures all of the unobservable characteristics specific to each participant. Ideally, one would also place a random effect on the facilitator to model the nested nature of multiple participants falling under one mediator, however not enough of this data was reliably obtained and thus is excluded from the analysis. The mixed-effect models allow the modelling of the subject-specific change in a Likert score over time, by including time as a fixed effect in a base model. This base model is thus specified for the i-th participant at time j as:

$$f(y_{ij}) = (\beta_0 + a_i) + \beta_1 x_{1ij} + e_{ij}$$

where y is the Likert score of a particular scale; $f()$ is the appropriate link-function; x represents time ($x = 0$ at baseline; $x = 1$ at follow-up); a_i is the subject-specific random intercept term; e_{ij} is the random error - where both a_i and e_{ij} are normally distributed: $a_i \sim N(0, \sigma_a^2)$; $e_{ij} \sim N(0, \sigma_e^2)$. σ_a^2 thus captures the covariance between the same participant's Likert score measured at different time points, while the variance in y is fully captured by $(\sigma_a^2 + \sigma_e^2)$. Independence between different participants is still considered a reasonable assumption (i.e. $cov(y_{ij}, y_{mj}) = 0$). β_0 can thus be interpreted as the average Likert score for the sample at baseline (intercept), and thus $(\beta_0 + \beta_1)$ is the average Likert score after the intervention. β_1 thus represents the average magnitude of the 'time' effect.

The distribution, D, is chosen by visually assessing the empirical distribution of y at both time points and comparing that to theoretical distributions' quantiles, probability density and cumulative probability density functions. In the case of the 'harsh discipline' scale (ICAST) - comprised of count variables - this is limited to Poisson and Negative Binomial (and zero-inflated adaptations). For the rest of the scales, the candidate distributions are Gaussian, log-Gaussian and Gamma. One significant drawback to this approach is that all of the three aforementioned theoretical distributions are continuous, while the Likert scales are actually discrete by design. This means that, especially where there is little variation in response, the data is unlikely to represent any of the candidate models and thus the model fit will be poor.

Ultimately, models are fit assuming each of the candidate distributions in turn, and the model which produces the best fit is chosen to report here. The best fit is chosen by inspecting the residuals of each fitted model. Standardized residuals are calculated using the Bayesian framework provided by the DHARMa package in R (Hartig, 2022) and represent a scaled difference between the fitted values and the observed values for each observation - which is appropriate for generalized mixed effect models. The residuals are compared to the theoretical distribution by the use of quantile-quantile (Q-Q) plots and either a scatter plot of the residuals against the fitted values with a quantile regression overlayed, or a boxplot of the residuals. A good model fit would be indicated by residuals falling along the 45°in the Q-Q plot, and either a uniformly distributed scatter plot of the residuals, or a boxplot where the empirical quartiles fall along the theoretical quartiles.

The mixed-effect regression models will be adjusted for relevant moderator variables, to assess how the impact of the program changes for different types of subjects. This will take the form of including, for example, the subject's sex as a fixed effect interaction with time:

$$f(y_{ij}) = (\beta_0 + a_i) + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{1ij} x_{2ij} + e_{ij}$$

In the case of x_{2ij} representing sex (0 = female; 1 = male): β_0 is interpreted as the average total score at baseline for females; $(\beta_0 + \beta_2)$ is the average total baseline score for males; $(\beta_0 + \beta_1)$ is the average total score for females at follow-up; and $(\beta_0 + \beta_1 + \beta_2 + \beta_3)$ is the average total score for males at follow-up. β_1 is thus the average time effect for females and $(\beta_1 + \beta_3)$ is the average time effect for males. The interpretation for continuous covariates is similar, but not identical, and will be elaborated on in Section 6. For the linear mixed-effect models, assuming a Gaussian error distribution, the coefficients are estimated using simple Restricted Maximum Likelihood (REML). For the generalised models, where a link function $f()$ other than identity is used, the evaluation of the log-likelihood function uses the adaptive Gauss-Hermite quadrature (Haberman, 2006).

For some of the domains shown in Tables 2 and 3 - namely the MICS, PSE, and CESD scales - there are too few items to conduct an item reliability analysis and thus these items cannot be confirmed to measure the underlying factor. Total scores will thus not be considered, but rather the direction of changes in the individual items over time will be quantified through the implementation of an cumulative link mixed effect model (CLMM). These models are appropriate for ordinal variables which measure an underlying continuous latent variable. This models the relative cumulative log-odds of participant i's response moving from a higher to a lower category from baseline to follow-up, while incorporating within-subject correlation by adding a random effect on participant ID:

$$logit(\frac{P[Y_{ijc} \leq c]}{1 - P[Y_{ijc} \leq c]}) = (\theta_0 - a_i) - \beta_1 x_{1j} + \epsilon_i$$

where θ_0 is the threshold parameter determining the category assignment, c_i, from the continuous estimate y_i; and a_i is the subject-specific component. In this case, e^{β_1} estimates the relative odds of moving from a higher to lower score over time (x_{1j} represents the measurement point: $x_{1j} = 0$ at baseline and $x_{1j} = 1$ at follow up).

To test for the significance of the treatment/time effect, the model specified above is compared to a null model with no x_{1j}. A resulting p-value close to zero indicates that there is not sufficient evidence that the models are identical and thus a significant treatment effect is likely (Ning, et al., 2021).

Furthermore, as will be made clear in Section 5 below, many of the responses exhibited no, or close to no, variation. In some cases, the entire domain of interest exhibited zero variation in either the pre or post-intervention data. As was illustrated above, both factor analysis and Cronbach's alpha use the covariance of the items and the variance of the factor to assess internal consistency. Where an item, or items, in a Likert scale exhibit no variation, the subset of items which do show variation are used to assess internal consistency. Where the entire scale exhibits no variation at one time point, the methods cannot be used and the internal consistency can only be assessed at the time point which does have variation. Where the internal consistency of a Likert scale is poor, the model results must be interpreted with caution since the items comprising the scale cannot be said to represent the underlying construct.

For those total scores with no variation, these responses mostly cannot be used to build reliable models in a generalized mixed-effect regression framework. In these cases, the approach is adjusted by taking the

difference between post and pre-intervention score, and modeling that as a response in a generalised linear modeling framework. The change in response for the i-th participant is thus modeled as:

$$f(y_{i1} - y_{i0}) = \beta_0 + \epsilon_i$$

where y_{ij} is the i-th participants score at time j; β_0 represents the average change in Likert score among the sample, and ϵ_i is the random error distributed according to the assumed distribution.

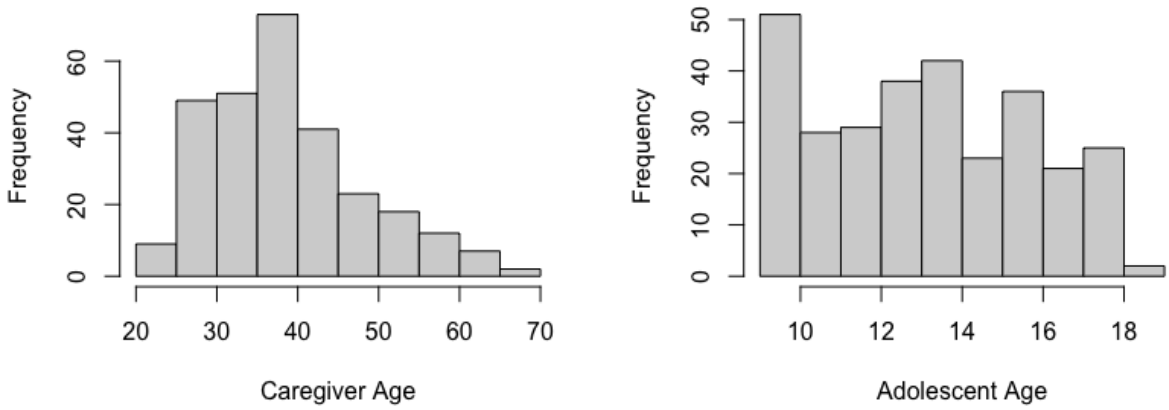
This is not optimal, since the repeated-measures component is lost and thus the within subject correlation cannot be captured. Furthermore, the 'time' effect cannot be quantified since the data is essentially reduced to cross-sectional. However, it does provide variation in the response variable, which can then be quantified on a population level and adjusted for moderators in a similar manner to the mixed-effect models.

The model fit here is assessed through plots of the Pearson residuals against fitted values. Pearson residuals are simply residuals standardized by the standard error of the observed values. In the base model (without moderators), the fitted value are simply equal to the estimated average change, β_0, and thus a box plot of the residuals is inspected. A good model fit would thus be indicated by a symmetrical box plot centered around zero.

4 Exploratory Results

4.1 Descriptive Data

The age distribution for caregivers and teens, respectively, are visualized using the histograms in Figure 2. The caregivers have a mean age of just over 29 years, with the distribution being skewed to the right. This likely reflects that there are a few adolescents that are looked after by their grandparents/other older persons. The adolescents are pretty evenly distributed between the ages of 9 and 18 years, with a few outlying 19 year old's and a mean age of 13.6 years.



(a) Caregivers

(b) Adolescents

Figure 2: Baseline Age Distributions

Descriptive statistics further outlining the demographic make-up of the participants at baseline are presented in Table 4 below. For all variables besides age, a percentage is reported along with the number of non-missing responses in square brackets. The percentage thus refers to the proportion of non-missing responses in the relevant category.

The majority of demographic questions had near full response rates (refer to Appendix B, Table 24). Five caregivers did not report their age; one caregiver did not report information regarding the whereabouts of the biological father; two caregivers did not report information regarding the whereabouts of the biological mother; one caregiver did not answer the question relating to substance abuse; and 14 caregivers did not report whether their child was enrolled in school. Six adolescents did not report the information regarding whether or not they had children of their own; six did not report their reading ability; nine did not report their school enrollment; one did not answer the question about TB/HIV; one did not report information regarding the biological father; three and four adolescents did not answer the questions regarding substance abuse and arguments in the household, respectively.

Table 4: Baseline Individual and Household Demographic Statistics

Variable	Caregiver Report (N=290)	Adolescent Report (N=295)
Sociodemographic Characteristics		
Age (mean, sd) [responses]	29.3 (9.43) [285]	13.6 (2.63) [295]
Female (%) [responses]	69 [290]	51 [295]
Adolescent has children of their own (%) [responses]	-	11 [289]
Enrolled in school (%) [responses]	-	87.8 [289]
Cannot read at all (%) [responses]	-	40.5 [289]
Can read with much difficulty (%) [responses]	-	50.0 [289]
Can read with little to no difficulty (%) [responses]	-	9.5 [289]
Caregiver is biological parent (%) [responses]	84.5 [290]	-
Caregiver is other (%) [responses]	15.5 [290]	-
Household characteristics (%) [responses]		
Father lives in household	51.2 [289]	50.7 [294]
Father deceased	27.3 [289]	23.8 [294]
Mother lives in household	71.5 [288]	70.2 [295]
Mother deceased	15.3 [288]	15.9 [295]
Ran out of money for essentials in last month	96 [290]	-
Severely unwell adult in household	37 [290]	-
Severely unwell child in household	19 [290]	-
Disabled child in household	13 [290]	-
Household member unwell/deceased from HIV/TB	64 [290]	33 [294]
Substance abuse	29 [289]	53 [292]
Physical/verbal arguments	24 [290]	44 [291]

Around 200 (69%) of the caregivers were female and, of the 290 caregivers, around 84.5% were the adolescent's biological parent. Approximately 60% of the caregivers were the biological mother while about 25% were biological fathers (not displayed in the table). The other 15.5% of caregivers consist of grandmothers/fathers, great-grandmothers/fathers, aunts/uncles, and a very small number of siblings. Around 51% reported the father living in the household, while 27% reported the father being deceased (with the remaining fathers living elsewhere). Around 71% reported the biological mother living in in the household, while 15% reported deceased mothers. 96% of the caregivers reported running out of money for essentials in the last 30 days, with only around 12 households being financially secure. 37% of the caregivers reported living with a severely unwell adult. 19% reported living with a severely unwell child and 13% with a disabled child. Around 64% of the caregivers reported having lived with a family member who suffered from HIV or Tuberculosis (TB) (hereby referred to as the HIV/TB factor). Around 29% reported substance abuse being an issue that their household battles with, while almost a quarter of the caregivers indicated that their household experiences many physical or verbal arguments.

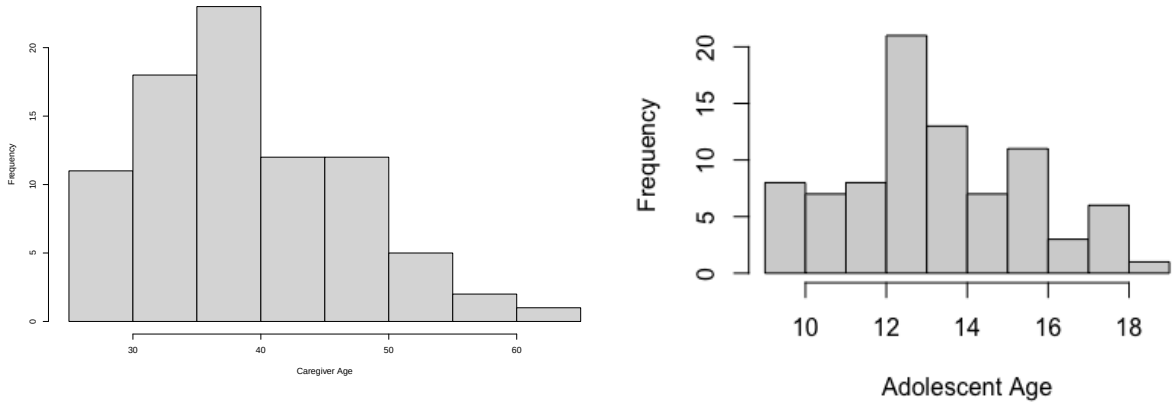
Among the adolescents, about 51% were female and approximately 11% of the teenagers had children of their own. Although almost 88% of the sample were enrolled in school, literacy seems to be a big issue, with 40.5% reporting that they cannot read at all and only 9.5% reporting that they can read with little to no difficulty. The proportions from the questions relating to the whereabouts of the biological parents differ slightly from those obtained from the caregiver report. This is likely due to the large amount of adolescents form the same household, skewing the proportions in one direction or another. For this reason, those figures obtained from the caregiver report are probably more indicative of the actual household level proportions. Interestingly, the last three household characteristics differ markedly between the two reports. Around 33% of the adolescents reported having lived with someone who suffered from HIV or TB, almost half that of the caregivers. Furthermore, significantly more adolescents reported that substance abuse and arguments were a problem in their households when compared to the caregivers. This may possibly indicate some sort of revelation bias among the caregivers, with some perhaps unwilling to divulge that information.

Table 5: Baseline Individual and Household Demographic Statistics for Matched Caregivers and Teens

Variable	Caregiver Report (N=85)	Adolescent Report (N=85)
Sociodemographic Characteristics		
Age (mean, sd) [responses]	39.8 (8.26) [84]	13.7 (2.35) [85]
Female (%) [responses]	79 [85]	60 [85]
Adolescent has children of their own (%) [responses]	-	11 [85]
Enrolled in school (%) [responses]	-	94 [80]
Cannot read at all (%) [responses]	-	33.7 [83]
Can read with much difficulty (%) [responses]	-	57.8 [83]
Can read with little to no difficulty (%) [responses]	-	8.4 [83]
Caregiver is biological parent (%) [responses]	90.6 [85]	-
Caregiver is other (%) [responses]	9.4 [85]	-
Household characteristics (%) [responses]		
Father lives in household	50.6 [85]	55.3 [85]
Father deceased	30.6 [85]	23.5 [85]
Mother lives in household	74.1 [85]	76.5 [85]
Mother deceased	15.3 [85]	16.4 [85]
Ran out of money for essentials in last month	93 [85]	-
Severely unwell adult in household	39 [85]	-
Severely unwell child in household	21 [85]	-
Disabled child in household	13 [85]	-
Household member unwell/deceased from HIV/TB	74 [85]	29 [85]
Substance abuse	34 [85]	44 [84]
Physical/verbal arguments	27 [85]	44 [84]

Considering the demographic profiles of the matched caregivers in Table 5 above, many (but not all) of the demographic variables remained largely the same after matching. The percentage of female caregivers increased by about 10%, indicating that some of the male population was lost. Similarly, after matching about 91% of the participants were biological parents - indicating that some of the other caregiver types were lost. The percentage of caregivers with unwell children and unwell adults also increased marginally, which may indicate that the matched sample comprises of a more vulnerable caregiver population than the original sample. The mean age of caregivers was increased from about 29 to 40 years. This resulted from the loss of the lower end of the age distribution, as is evident when comparing Figure 3(a) below with Figure 2(a) above. Overall, the matched caregivers contained a higher proportion of females and biological parents; an older population; a higher proportion with unwell children and adults living in their households; a slightly higher proportion with deceased fathers; a slightly higher proportion with mothers living in the household; a slightly lower proportion experiencing poverty.

Considering the baseline demographic profiles of the teenagers after matching, the mean age was almost the same but Figure 3(b) below makes it clear that most of the youngest adolescents, and some of the oldest, were lost during matching since the distribution is far less uniform after matching. The proportion of female adolescents increased from 51% to 60%, indicating a loss of male adolescents. A higher proportion of adolescents were enrolled in school, while a lower proportion reported experiencing substance abuse at home. This may suggest that a portion of the more vulnerable teenagers were lost due to the matching process. There were no other significant changes in the demographics profiles of the teenagers after matching. Overall, the matched teen sample contained a higher proportion of females; a lower proportion of participants under the age of 11; a higher enrollment rate; and a lower proportion of participants reporting substance abuse.



(a) Caregivers

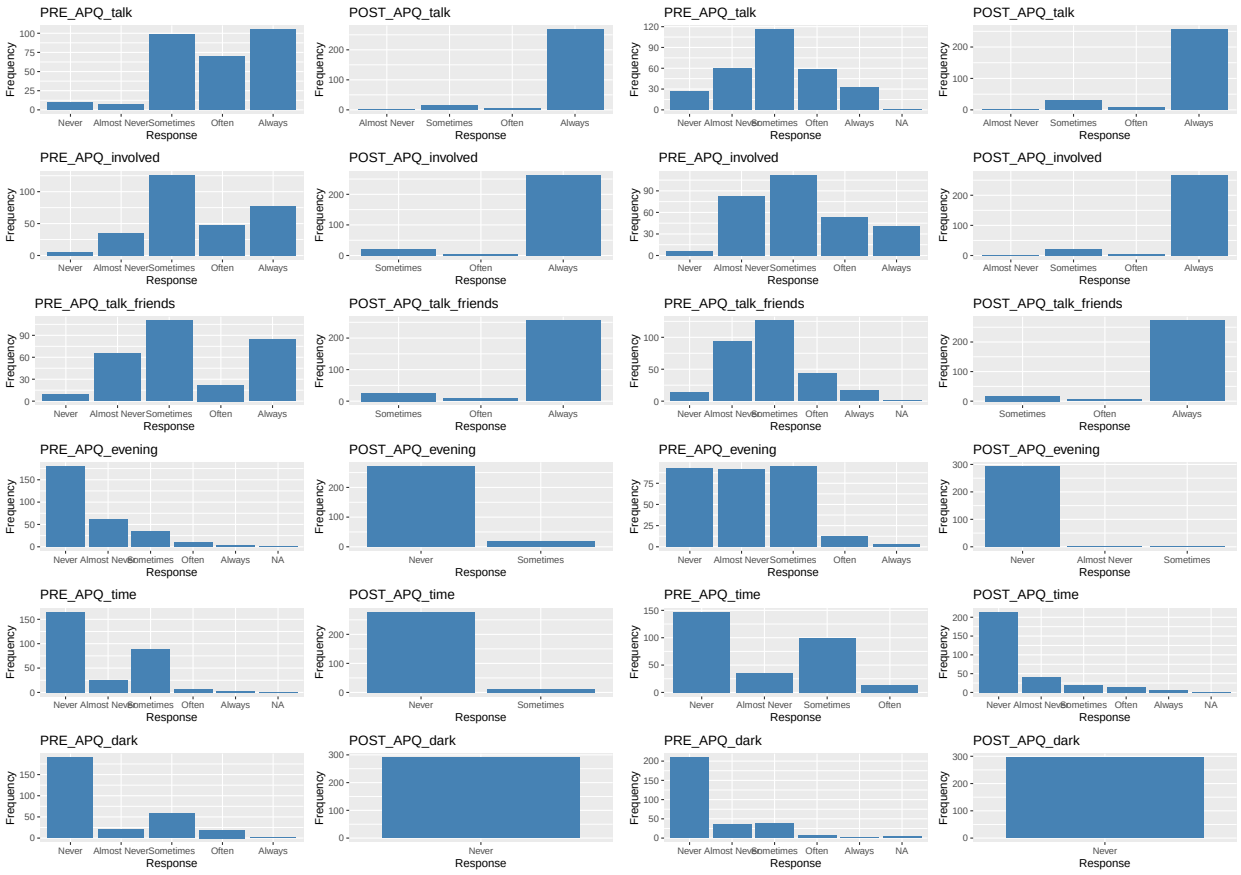
(b) Adolescents

Figure 3: Baseline Age Distributions of Matched Pairs

Some of the demographic factors - namely the relationship to adolescent and the indicator of poverty in the caregiver data, and the adolescents with own children and school enrollment in the teenagers' data - exhibit too extreme proportional imbalances to meaningfully be considered as moderators of the intervention effect (refer back to Table 5). These characteristics will thus not be considered as covariates in the regression analysis.

4.2 Outcome Data

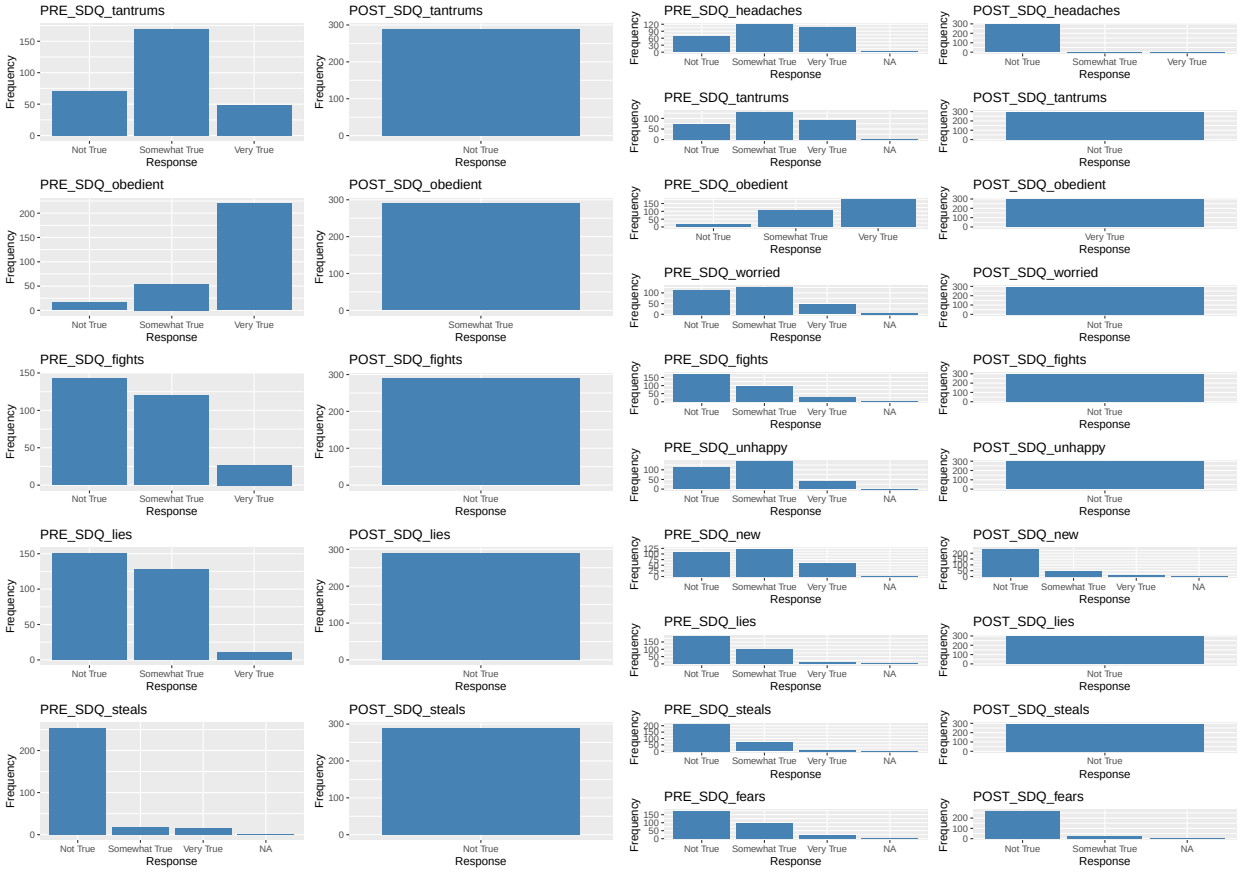
In this section, the distribution of responses for each Likert item at baseline and follow-up, across the full cohort of 290 caregivers and 295 adolescents, is investigated. In addition, the response rates to each question are provided in Appendix B in Table 25.



(a) Caregivers (b) Adolescents

Figure 4: Barplots of APQ Responses

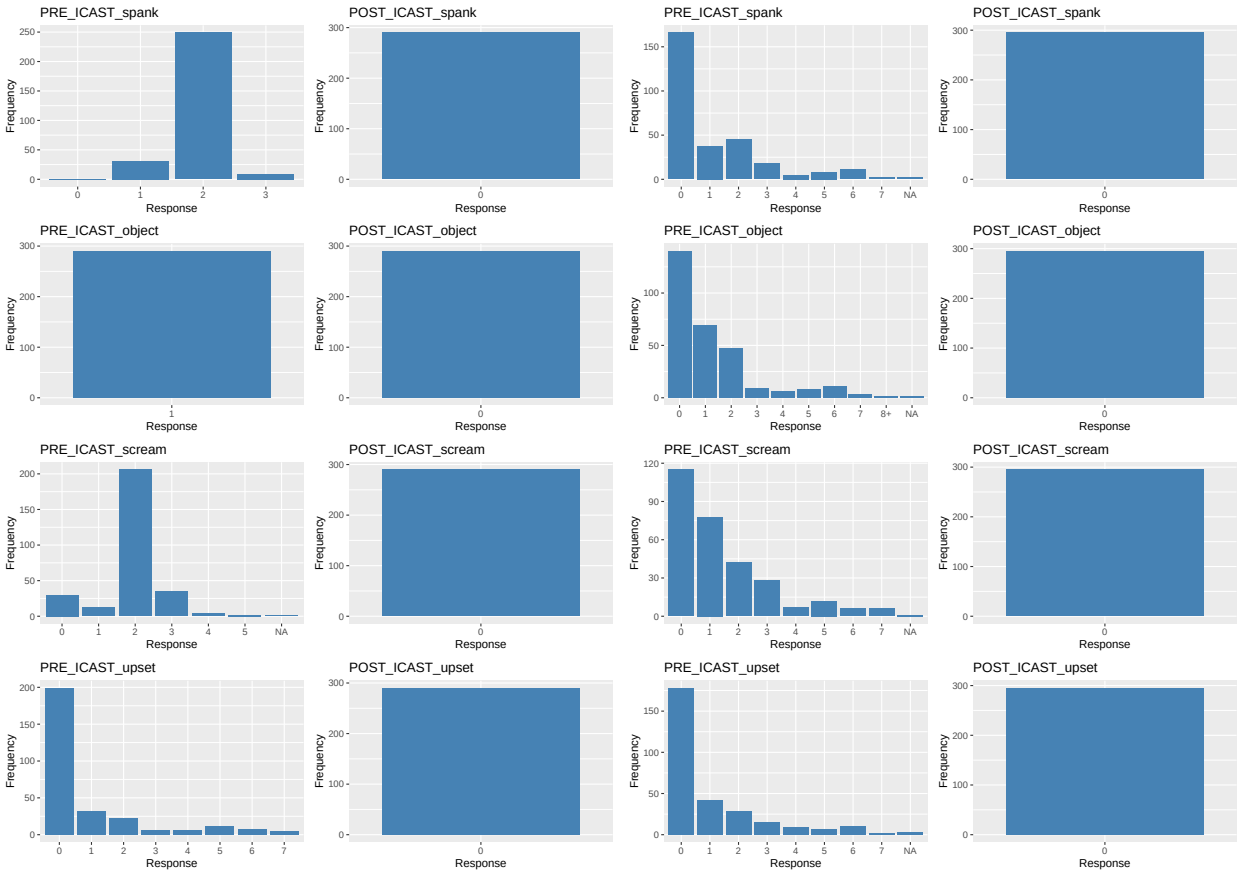
Considering the caregiver APQ responses in Figure 4(a), one can see that the responses to the first three questions were relatively left-skewed before the intervention, while the latter three were right-skewed. This reflects the need for reverse-coding as indicated previously, since the sentiment of the first three act in the opposite direction to the latter three. After the intervention, the distributions were significantly more skewed. For example, with the first three questions almost all of the caregivers answered "always" after the intervention. Similarly, with the latter three questions, almost all of the caregivers answered "never" after the intervention (literally all 290 caregivers, in the case of 'APQ_dark'). This may be the first suggestion of an effect of the intervention in improving positive parenting and parental monitoring. Similar response patterns are seen among the teen responses, with a rightward shift in the first three items and a left-ward shift in the latter three. The lack of variation in many of the items, however, does raise concerns about the ability to conduct an item reliability analysis and model the pre-post changes - since there is not much co-variation and variation to capture.



(a) Caregivers (b) Adolescents

Figure 5: Barplots of SDQ Responses

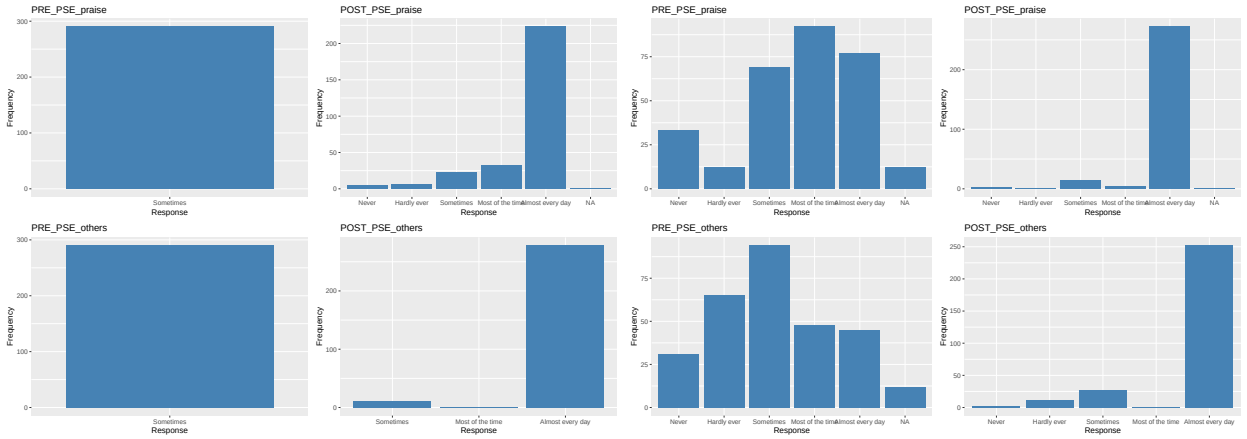
Figure 5 presents the responses to the SDQ items for caregivers and adolescents. One can see that the pre-intervention distribution of 'SDQ_obedient' for both caregiver and teenager is markedly different to the other items, illustrating the need for reverse-coding. The changes from pre to post-intervention, again suggest the program may have been effective in improving child behaviour - with all responses moving in the direction that indicates an improvement in child behaviour from both the caregiver and adolescents' perspectives. The lack of variation in the post-intervention data is even more drastic than that among the APQ responses. The responses for all of the caregiver post-intervention SDQ questions and seven of the ten adolescent post-intervention SDQ responses are completely homogeneous.



(a) Caregivers (b) Adolescents

Figure 6: Barplots of ICAST Responses

Figure 6 displays the responses for the ICAST items. The adolescent pre-intervention ICAST items, and the caregiver ICAST_upset item exhibit the patterns that one would expect of a Poisson process. One can see that the adolescent pre-intervention ICAST items exhibit a similar pattern and the ICAST scores are thus likely to be internally consistent. While the figures do indicate that the intervention reduced harsh discipline from both the caregiver and adolescents' perspectives, with all items shifting leftwards after the intervention, there is again an alarming lack of variation. In fact, all of the 290 caregivers and all of the 295 adolescents responded with '0' after the intervention to every ICAST item. This is as strong of an effect as is possible, however it is an extremely unlikely result which casts doubt over the validity of the data. Even the distribution of the caregiver's pre-intervention ICAST_object exhibits zero variation, which is of even more concern since one would not definitely not expect the caregiver sample to be homogeneous with respect to any of the responses before the intervention took place.

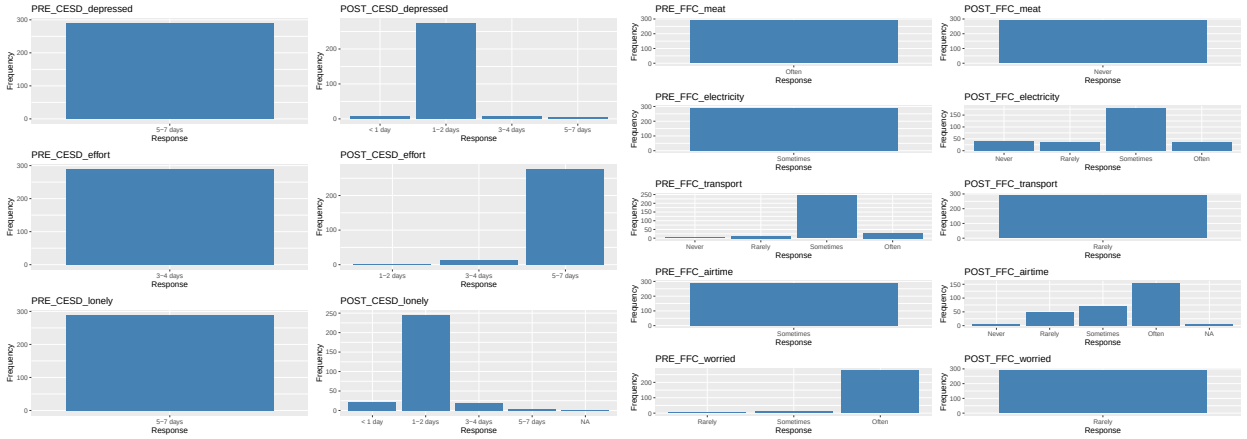


(a) Caregivers (b) Adolescents

Figure 7: Barplots of PSE Responses

Figure 7 describes the response patterns for both groups for the two questions relating to parental support of education (PSE). Similar observations can be made as with the other domains: the shift from pre to post indicate that PSE had improved after the intervention from both the caregiver and adolescent perspectives. There is again no variation in the caregiver responses prior to the intervention, with all caregivers answering "sometimes" to both questions.

Figures 8(a) and 8(b) display the caregiver responses to the CESD and FFC items, respectively. Considering 8(a), a leftward shift would indicate an improvement in the caregivers' mental health. This is evident for only the first and last item, indicating that the caregivers had experienced a reduction in the frequency of feeling depressed and lonely, but a deterioration in the feeling of 'everything requiring effort'. Similar to the caregiver PSE responses, all three of the pre-intervention CESD items were identically answered by all 290 caregivers. Figure 8(b) shows that, the items FFC_meat, FFC_transport and FFC_worried indicate improvements in the caregivers' financial strain. The second item, which pertains to worrying about money for electricity seemed to worsen on average, while that pertaining to airtime neither conclusively improved nor deteriorated. Again, there is a concerning lack of variation among most of the items.

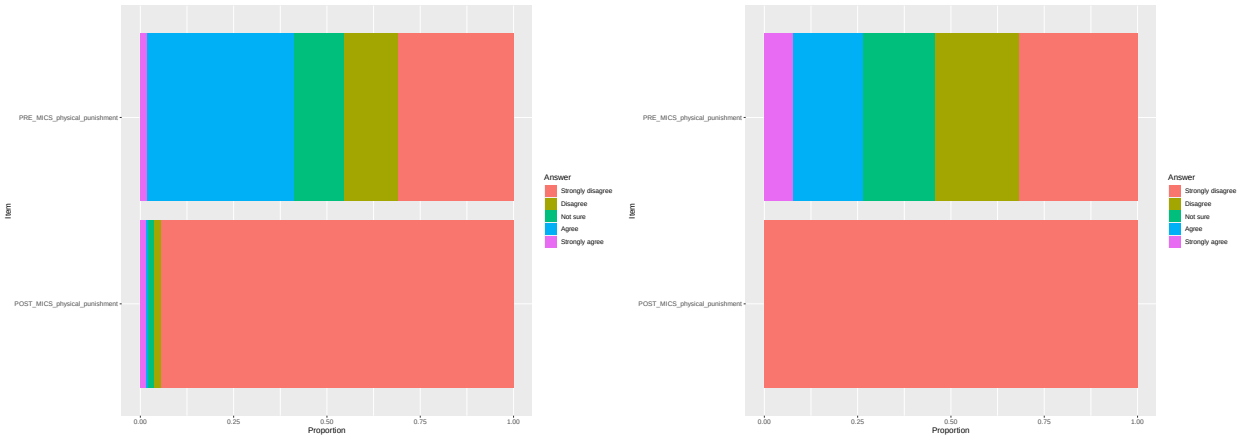


(a) Caregiver Mental Health (b) Family Financial Coping

Figure 8: Barplots of Caregiver CESD and FFC Responses

The stacked barplots in Figures 9(a) and 9(b) below displays the responses to the question regarding attitudes towards physical punishment (MICS). Before the intervention took place, there was a large diversity of

responses among both the caregivers and teenagers. While approximately a third of both groups strongly disagreed with the statement before the intervention, almost all of the participants in both groups strongly disagreed afterwards. For interest's sake, the response distributions of all items are visualised with this technique in Appendix C.



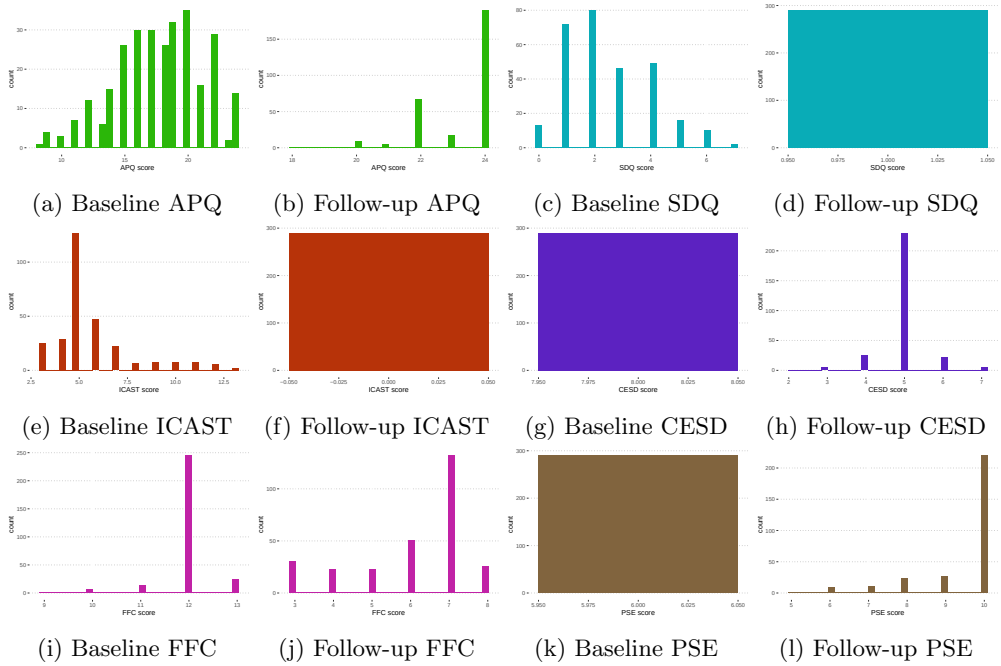
(a) Caregivers

(b) Adolescents

Figure 9: Stacked Barplots of Responses to Attitudes Towards Physical Punishment

The lack of variation seen in the majority of the items poses significant obstacles to conducting an item reliability analysis and implementing a model to quantify the changes. With respect to assessing internal consistency, the lack of variation means a lack of covariation, which renders confirmatory factor analysis and the Cronbach's alpha statistic effectively useless. In terms of building a model to quantify the changes in the Likert scales, the lack of variation carries over into the composite scores and means these scores have very unconventional distributions which are difficult to fit an appropriate model to.

This problem is illustrated, for the caregivers, through Figures 10(a) - 10(j). These bar plots display the distributions of each Likert scale before and after the intervention. One can clearly see that those scales whose items exhibited no variation, also exhibit zero variation (Figures 10(d), (f), and (g)). Those scales whose items exhibited very little variation are dominated by one score (consider Figures 10(b), (e), (h), and (i)). Ideally, for modeling purposes, one would expect to see a horizontal shift in the distribution of a scale after the intervention. This would provide a consistent distribution to quantify the changes over and infer onto the populations. This property is not observed in any of the Likert scales, except for perhaps the APQ and FFC scales. In the cases of the scales with no variation, the responses cannot even be said to follow a distribution but are just constant. In this case, the changes in scores from pre to post-intervention would be solely due to the differences at baseline. Since all the participants have the same score at follow up, the change over time is attributed to their score at baseline. As previously mentioned, modeling the effect of moderators at baseline will be more informative in these cases than the change over time.



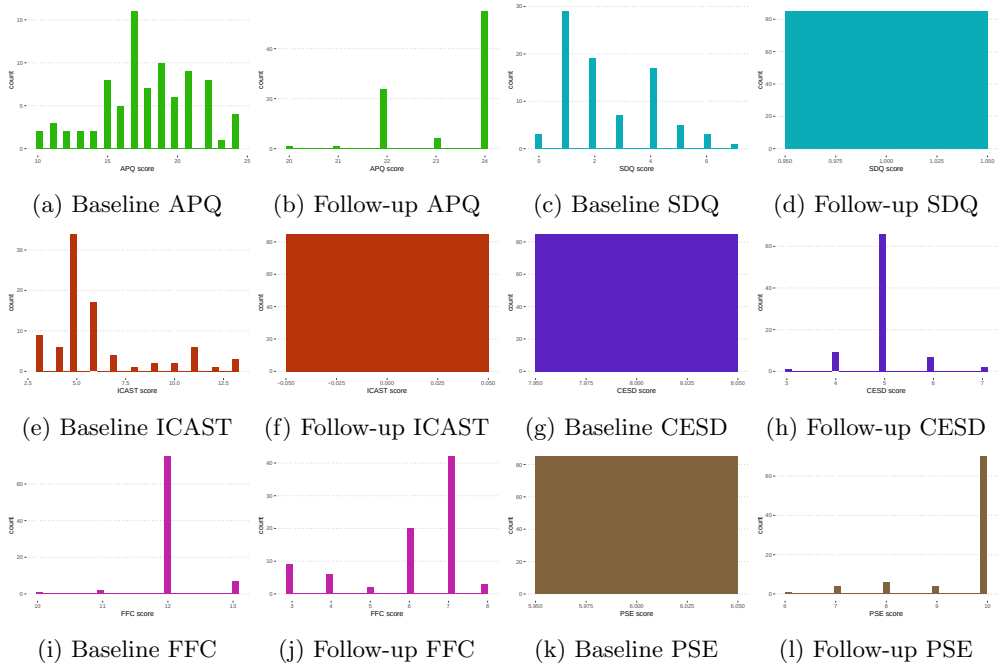
(a) Baseline APQ (b) Follow-up APQ (c) Baseline SDQ (d) Follow-up SDQ

(e) Baseline ICAST (f) Follow-up ICAST (g) Baseline CESD (h) Follow-up CESD

(i) Baseline FFC (j) Follow-up FFC (k) Baseline PSE (l) Follow-up PSE

Figure 10: Histograms of Likert Scores: Caregiver Full Sample

Another point of consideration here is whether the distribution of the matched participants' responses are similar to that of the full sample, since any departure from this would inevitably introduce bias into the analysis. Figures 11(a) to 11(l) below thus present the distributions for the matched samples.



(a) Baseline APQ (b) Follow-up APQ (c) Baseline SDQ (d) Follow-up SDQ

(e) Baseline ICAST (f) Follow-up ICAST (g) Baseline CESD (h) Follow-up CESD

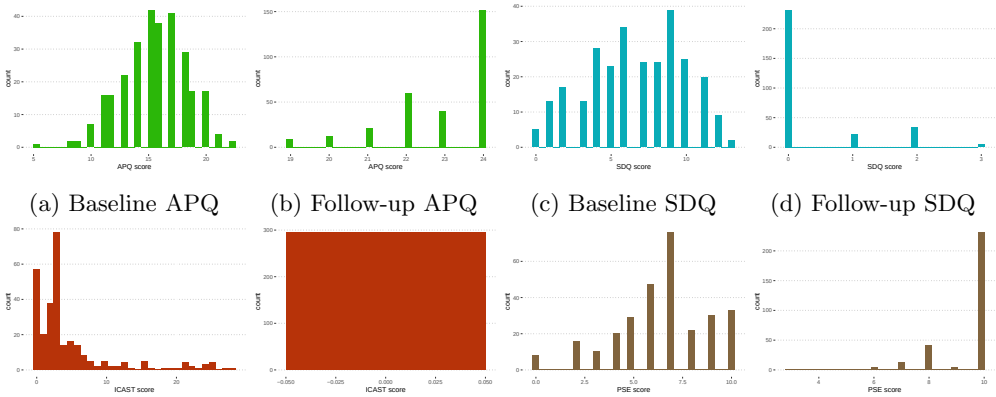
(i) Baseline FFC (j) Follow-up FFC (k) Baseline PSE (l) Follow-up PSE

Figure 11: Histograms of Likert Scores: Caregiver Matched Sample

One can see that the APQ scores generally maintained the same shape, however, are less left skewed at baseline and the scores below 20 units are lost at follow-up (Figures 11(a) and (b)). The matched samples'

SDQ scores at baseline contain proportionally less scores in the middle of the distribution (Figure 11(c)), and the distribution of the ICAST scores is slightly more dispersed towards the right tail (Figure 11(e)). The rest of the responses seem to mirror those in Figure 10 quite well. Overall, there may be a slight bias introduced into the caregiver APQ and SDQ scores, which doesn't seem to be a problem for the other scores.

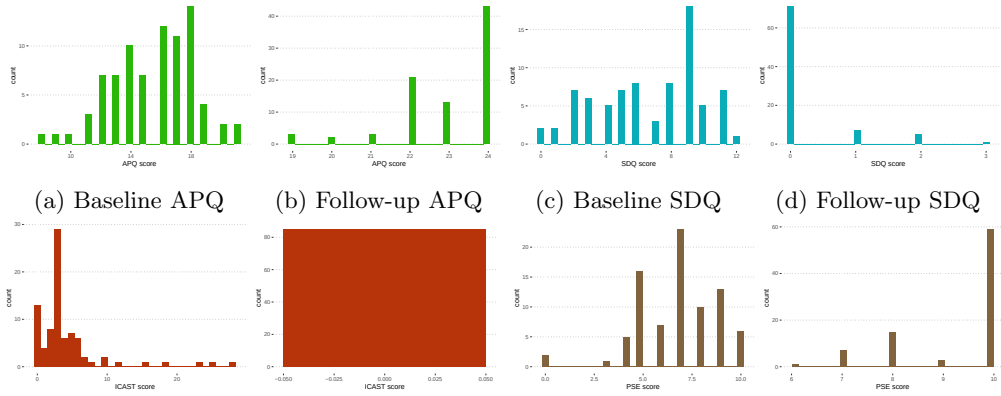
A similar investigation is conducted into the distributions of the adolescent scores, visualised for the full and matched samples in Figures 12 and 13, respectively. The distribution of the APQ scores is very similar between the cohorts at both baseline and follow up. There is, however, a loss of the median observations in the baseline scores through matching. The SDQ scores at baseline lose much of the center mass of the distribution through matching and becomes more dispersed, although the mode is maintained. The ICAST scores in the matched sample quite closely mirror that of the full sample but have proportionately less zero counts, and are more discretely distributed in the right tail. The PSE scores are very similar between the cohorts, but are a bit more sparsely dispersed in the matched sample. This is a common theme when comparing the two cohorts, which is due to the reduced sample size, and is likely to be an issue when trying the model the distributions using normal theory, or any framework which assumes a continuous random variable.



(a) Baseline APQ (b) Follow-up APQ (c) Baseline SDQ (d) Follow-up SDQ

(e) Baseline ICAST (f) Follow-up ICAST (g) Baseline PSE (h) Follow-up PSE

Figure 12: Histograms of Likert Scores: Adolescent Full Sample



(a) Baseline APQ (b) Follow-up APQ (c) Baseline SDQ (d) Follow-up SDQ

(e) Baseline ICAST (f) Follow-up ICAST (g) Baseline PSE (h) Follow-up PSE

Figure 13: Histograms of Likert Scores: Adolescent Matched Sample

Regarding the lack of variation seen in the caregiver data, Figures 12 and 13 show that similar, but less extreme problems are present among the adolescent scales. The APQ scores in Figures 13(a) and (b) are not necessarily problematic, since there is an acceptable degree of variation both before and after the intervention.

Considering the pre-intervention distribution, this scale could thus possibly be modeled with a Gaussian mixed-effect model. Similarly, the SDQ scale (Figures 13(c) and (d)) could also possibly be modeled in the mixed-effect regression framework assuming a normal or gamma distribution - although the lack of variation at follow up may inhibit a good model fit. As was seen with the individual responses, the ICAST items contained no variation post-intervention, which is carried over into the ICAST scale (Figures 13(e) and (f)).

One possible way around these problems of no variation, is by modeling the change in Likert score explicitly as a response variable in a generalised linear modeling framework. By modeling the difference between pre and post-intervention scores, one can at least obtain a response with some variation. These distributions are shown in Appendix D. In many of the cases, the differenced data needs to be transformed first so that an appropriate underlying distribution can be assumed.

Another technique which can be used to extract some insight out of the data is to model the cross-sectional data where there is variation. For example, while the post-intervention ICAST scales exhibit no variation, the pre-intervention scales do exhibit the variation one would expect from a Poisson or Negative Binomial process. These pre-intervention scales can thus be modeled in a generalised linear framework and risk factors for adolescents experiencing harsh discipline (before treatment) can be determined by including moderators.

In addition to these practical issues, the homogeneity in response patterns across participants casts some doubts over the validity of the data. A possible cause of the response patterns could be due to the participants losing interest in the questionnaire as they are completing it. In this case, one would expect the first sections to be filled out properly and thus show variation, and the latter questions to be filled out with the same answers as respondents get bored. This, however, would only explain subject-specific homogeneity, and not homogeneity across all of the respondents' answers - since one would still expect different participants to choose different answers to 'replicate', and some participants to not behave in this manner.

As mentioned above, the fact that many of the pre-intervention responses exhibit no variation is of even more concern than that seen in the follow-up data. It may be more believable that the respondents are homogeneous after taking part in the program and receiving the same advice/education. However, before this intervention took place one would certainly expect to see some differences in the caregiver's parental support of education (PSE) and emotional/psychological state (CESD), for example. Considering that this phenomenon has not been seen in any of the PLH-SUPER data obtained from other countries in which the study has been implemented, it raises alarm over whether the data has been manipulated in some way - either from an honest data capturing issue or some kind of interference in the study design. This brings about the question of whether there actually exists meaningful information representing the populations of interest which can be extracted from the data, or whether the data received from the capturers in South Sudan is not actually representative of the participants' experiences in the study.

5 Main Results

This section of the report presents the results from the item reliability analysis; mixed-effect regression models; generalised linear models; and cumulative link mixed-effect models.

5.1 Item Reliability Analysis

5.1.1 Confirmatory Factor Analysis Models

The first step in assessing the internal consistency of each score was to conduct a confirmatory factor analysis. Items were reverse coded prior to the analysis were need be. Only the factors and items which exhibited sufficient variation between observations could be used as input into a confirmatory factor analysis model. For the baseline caregiver surveys, these include APQ, SDQ and ICAST, while at follow-up only APQ could be assessed. For the adolescents these include APQ, SDQ and ICAST at baseline, and only APQ at follow-up. The numeric labels attached to the arrows from the factor into the items indicate the loadings; the bi-directional arrows from each item to itself represent the residual variances; the bi-directional arrows from each factor to itself represent the variance of the factor; and the bi-direction arrows connecting each factor represent the covariance between factors. Of primary interest are the former three metrics, while the factor covariances are of secondary importance. High magnitude factor loadings with the same sign indicate that the factor has a strong influence on the items in the same direction. A relatively high factor variance indicates that the factor explains much of the variation in the items.

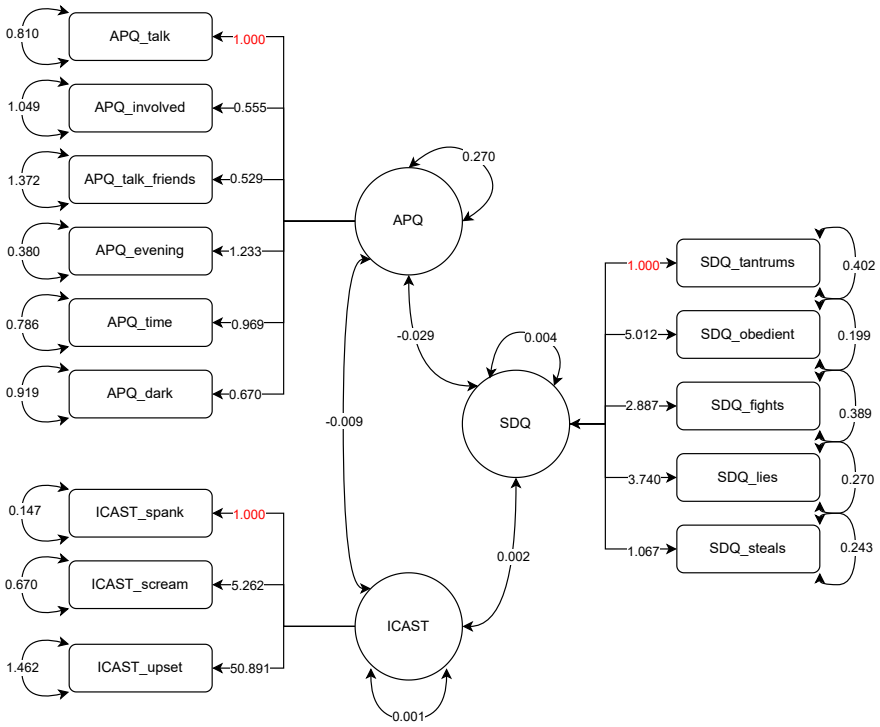


Figure 14: Confirmatory Factor Analysis Diagram: Caregivers Baseline

Figure 14 above presents the results from a 3 factor model for the caregiver data at baseline. Considering the APQ scale, the items with the suffixes 'talk', 'evening' and 'time' load highly and similarly on the factor, while the 'involved' and 'talk friends' items have the smallest loadings. The residual variances are similar across the items. Among the ICAST items, the 'upset' factor loads most highly on the factor, however the estimate is perhaps too high and may be indicative of a poor model fit. The loadings on this scale are of

significantly different magnitudes and may either be indicative of asymmetrical influence, or of a poor model fit. The SDQ items all have relatively high factor loadings in the same direction, however 'tantrums' and 'steals' are influenced least by this factor. The residual variances of each factor are relatively low, suggesting that much of the variability in these items is common and attributable to the factor. In addition, the three factors do not seem to be very closely associated.

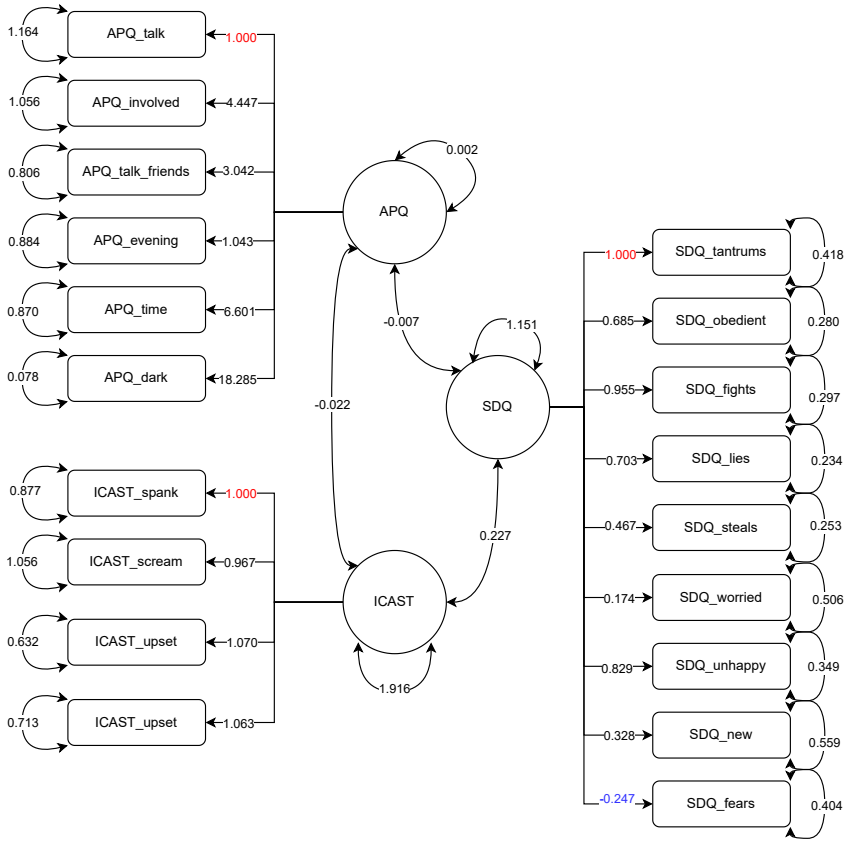
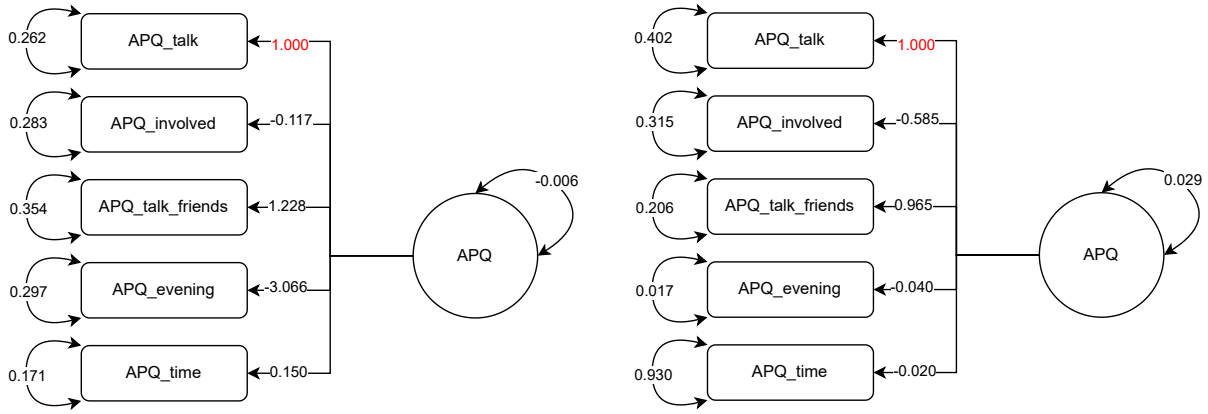


Figure 15: Confirmatory Factor Analysis Diagram: Adolescents Baseline

The 3 factor model for the adolescent baseline scores is presented in Figure 15. Considering the APQ scale, all factors load highly and in the same direction, with the scale influencing the variability in the items with suffixes: 'dark', 'time' and 'involved'. The model suggests that the factor most influences the 'dark' item, and influences the other items considerably less. The factor variance is very low. The SDQ scale perhaps provides better reliability, with most of the items loading with a similar magnitude and in the same direction. The model suggests that the factor influences the items 'worried, 'unhappy', and 'new' the least, while the 'fears' item seems to be influenced in the opposite direction. The factor seems to be most associated with the items 'tantrums', 'fights', 'lies', and 'unhappy', while the factor variance is relatively high. The ICAST item provides the most evidence in favour of good internal consistency, with all four items loading with very similar magnitudes in the same direction. The factor variance is the highest seen across all factors modeled. In addition, the APQ scale does not seem to be strongly associated with either of the other two, however there may be a positive association between ICAST and SDQ. Intuitively, this may be valid since high ICAST scores indicate harsher discipline and high SDQ scores indicate poorer child behaviour.



(a) Caregivers

(b) Adolescents

Figure 16: Confirmatory Factor Analysis Diagram: Follow up

Figure 16(a) presents the one factor model for caregivers at follow up. Only the factors 'talk' and 'talk friends' load positively, and with sufficient magnitude, on the APQ scale. Of particular interest is the item, 'evening', which loads highly but negatively on the factor despite being correctly reverse-coded. The residual variances are low, however, this is likely due to the items themselves having very little variance. The internal consistency of this scale at follow up is likely very poor, and it must be noted that the factor variance is negative - which may indicate that this model is not appropriate for the data.

Figure 16(b) shows similar results for the adolescents APQ items at follow up, with the 'talk' and 'talk friends' items loading similarly on the scale. Contrary to what one may expect to see, the 'involved' item loads in the opposite direction, while the loadings of the other two items indicate that the factor does not strongly influence their variability. In addition, the factor variance is relatively low and thus, the internal consistency of this scale at follow up is thus likely very poor too.

5.1.2 Confirmatory Factor Analysis Diagnostics

Table 6 presents three model fit criteria statistics for each of the above models. Only the caregivers baseline model indicates an appropriately fitted model, with an acceptably low RMSE, and a high CFI and TLI. The statistics associated with the adolescents' baseline model are beyond the range of acceptability, while the follow up models' TLI contradict the other two statistics and indicate an extremely poor model fit. The null hypothesis is thus rejected for the latter three CFA models and one may conclude that the model specified by the hypothesized factors and their associated items is likely invalid.

Table 6: CFA Model Fit Statistics

CFA Model	RMSE [CI]	Pr[RMSE $\leq$ 0.05]	CFI	TLI
Caregivers Baseline	0.036 [0.015, 0.052]	0.917	0.922	0.904
Caregivers Follow up	0.000 [0.000, 0.038]	0.973	1.000	-0.270
Adolescents Baseline	0.111 [0.103, 0.120]	0.000	0.654	0.603
Adolescents Follow up	0.000 [0.000, 0.037]	0.975	1.000	-0.569

5.1.3 Cronbach's Alpha Statistics

The Cronbach's alpha statistic for each of the caregiver and adolescent scales was computed for the baseline and the follow up responses. This information is displayed in Tables 7 and 8 below:

Table 7 contradicts the results of the baseline CFA model to some extent, with poor or worse internal consistency indicated by the alpha statistic for all of the scales. The highest statistic relates to the APQ scale (0.545), however this still suggests poor reliability even if the confidence interval is considered. As expected, the internal consistency of the items at follow up is even poorer for all scales except for FFC and CESD, and these results correspond with the items with no or close to no variation (see section 5.2).

Table 7: Cronbach's alpha: Caregivers

Scale	Time Point	Alpha [CI]	Interpretation
APQ	Baseline	0.545 [0.462, 0.617]	Poor
	Follow-up	0.004 [-0.188, 0.160]	Unacceptable
SDQ	Baseline	0.292 [0.140, 0.413]	Very poor
	Follow-up	NA*	
ICAST	Baseline	0.133 [0.031, 0.231]	Unacceptable
	Follow-up	NA*	
CESD	Baseline	NA*	
	Follow-up	0.074 [-0.186, 0.270]	Unacceptable
FFC	Baseline	-0.052 [-0.333, 0.312]	Unacceptable
	Follow-up	0.437 [0.388, 0.478]	Very poor

note: NA means that items comprising scale have no variance

Considering the Cronbach's alpha statistic for the adolescents (Table 8), one can see that the results for the baseline scales align with that of the CFA. The ICAST scale provides very strong internal consistency ($alpha = 0.902$), while the SDQ scale is borderline acceptable ($alpha = 0.612$) and the APQ scale's consistency is poor ($alpha = 0.341$). The scales measured at follow up have very poor internal consistency, and this can almost certainly be attributed to the lack of variation seen in the data.

Table 8: Cronbach's alpha: Adolescents

Scale	Time Point	Alpha [CI]	Interpretation
APQ	Baseline	0.341 [0.204, 0.442]	Poor
	Follow-up	-0.049 [-0.208, 0.094]	Unacceptable
SDQ	Baseline	0.612 [0.542, 0.668]	Acceptable
	Follow-up	0.313 [0.233, 0.384]	Poor
ICAST	Baseline	0.902 [0.853, 0.932]	Good
	Follow-up	NA*	

Note: NA* means that items comprising scale have no variance

Overall, the results of the CFA and Cronbach's alpha statistic indicate that only the ICAST scores (measured from the adolescent perspective at baseline) measure some underlying factor with an acceptable level of internal consistency. There is some evidence that the caregiver's APQ, and adolescent SDQ, items provide borderline acceptable internal consistency.

5.2 Baseline Regression Models

5.2.1 Results

Table 9 presents the results of univariate simple GLM's for the caregiver outcomes relating to positive parenting, child behaviour and harsh parenting measured at baseline. Each estimate is provided on the measured scale, with the exception of the SDQ scores, where the distribution was shifted up by one unit in order to model with a Gamma distribution. The intercept of that score is thus not maintained, however the effect size relative to the intercept is preserved. In addition, each estimate is accompanied by an approximate confidence interval based on the standard error (see note below table). Significance tests are also provided, however these must be interpreted with caution given the small sample size (n=85). Important to note is that each effect size is considered independently, and thus may not be truly reflective of the size of a risk factor due to variable omission. For example, some of the 'substance abuse' effect may be attributable to 'fighting at home', which is not captured in this simple approach.

Table 9: Baseline Regression Results: Caregivers

Model Index	Model Specification	Assumed Distribution	Intercept [C.I]	Effect size [C.I]	t-statistic (p-value)	AIC
null	APQ ∼1	Gaussian	17.92 [17.19, 18.65]	-	-	450.25
1	null + Caregiver sex		17.51 [16.71, 18.31]	1.94 [0.19, 3.68]	2.22 (0.29)	447.34
2	null + Substance abuse		19.34 [18.61, 20.07]	-4.17 [-5.41, -2.92]	-6.68 (<0.0001)	415.67
3	null + Fighting at home		18.89 [18.13, 19.64]	-3.58 [-5.03, -2.13]	-4.93 (<0.0001)	430.36
4	null + Sick child		18.63 [17.87, 19.38]	-3.35 [-4.99, -1.71]	-4.09 (<0.0001)	436.64
5	null + Sick adult		18.63 [17.73, 19.54]	-1.85 [-3.30, -0.40]	-2.55 (0.0126)	445.85
6	null + HIV_TB		15.50 [14.20, 16.80]	3.26 [1.75 , 4.78]	4.31 (<0.0001)	435.08
null	(SDQ+1) ∼1	Gamma (log-link)	3.45 [3.12 , 3.82]	-	-	307.40
1	null + Caregiver age		4.59 [3.79, 5.57]	0.98 [0.97, 0.99]	-3.63 (0.0005)	293.31
2	null + Substance abuse		2.75 [2.47, 3.06]	1.77 [1.47, 2.13]	6.11 (<0.0001)	272.57
3	null + Fighting at home		2.95 [2.64, 3.30]	1.62 [1.31, 2.00]	4.54 (<0.0001)	287.09
4	null + Disabled child		3.16 [2.85, 3.51]	1.69 [1.27, 2.25]	3.70 (0.0004)	293.75
5	null + Sick child		3.02 [2.71, 3.35]	1.68 [1.33, 2.11]	4.52 (<0.0001)	287.19
6	null + Sick adult		3.08 [2.71, 3.50]	1.31 [1.07, 1.60]	6.60 (0.0094)	302.09
7	null + HIV_TB		4.36 [3.60, 5.28]	0.72 [0.57, 0.90]	2.99 (0.0037)	299.96
null	ICAST ∼1	Negative binomial (theta=250) (log-link)	6.05 [5.53, 6.62]	-	-	385.99
1	null + Child sex		5.73 [5.08, 6.45]	1.14 [0.95, 1.37]	1.453 (0.15)	385.88
2	null + Caregiver sex		6.28 [5.70, 6.92]	0.82 [0.65, 1.03]	-1.72 (0.09)	385.02
3	null + Caregiver age		7.68 [6.48, 9.10]	0.98 [0.97, 0.99]	-3.19 (0.002)	374.13
4	null + Substance abuse		5.21 [4.71, 5.77]	1.47 [1.26, 1.71]	4.95 (<0.0001)	370.27
5	null + Fighting at home		5.08 [4.68, 5.52]	1.70 [1.49, 1.94]	8.04 (<0.0001)	355.95
6	null + Sick child		5.19 [4.80, 5.62]	1.78 [1.55, 2.04]	8.28 (<0.0001)	354.60
7	null + Sick adult		5.10 [4.60, 5.65]	1.48 [1.28, 1.72]	5.28 (<0.0001)	368.90
8	null + HIV_TB		7.59 [6.57, 8.78]	0.73 [0.61, 0.87]	-6.34 (0.0005)	377.13

Note: For APQ models, the effect size is additive. For SDQ and ICAST, effect sizes are multiplicative.
Note: The SDQ coefficients are relative to the observed score + 1, since Gamma distributions assume strictly positive values.
Note: CI = point-wise confidence limits, calculated by: $exp(\beta + -2 * s.e(\beta))$ for the models with a log link, and $\beta + -2 * s.e(\beta)$ otherwise.

The matched caregivers' baseline positive parenting (APQ) score on average is approximated as 17.92 ($CI : [17.19, 18.65]$). Considered independently, male caregivers and those reporting the HIV/TB factor are on average associated with higher positive parenting scores - although the sex factor does not give much evidence of significance. Caregivers reporting substance abuse, fighting at home, and living with either a sick child or adult are associated with lower APQ scores. The risk factors of greatest magnitude and significance are substance abuse ($\beta = -4.17$, $CI : [-5.14, -2.92]$, $p < 0.0001$) and fighting at home ($\beta = -3.58$, $CI : [-5.03, -2.13]$, $p < 0.0001$). The AIC statistics imply that model 2 best fits the observed data - where total APQ score is regressed against substance abuse.

The average child behaviour score among the caregivers (after shifting the distribution) is approximated as 3.45 ($CI : [3.12, 3.82]$). Those reporting substance abuse, fighting at home, living with a disabled child, sick child or sick adult are all associated with worse (higher) SDQ scores by factors of 1.77, 1.62, 1.69 and 1.68, respectively. Older caregivers are associated with a 2% reduction in SDQ score for each additional year of age, when compared to the youngest caregivers, while those reporting the HIV/TB factor are associated with

a 28% reduction in the score. Considering the AIC statistics of each model, model 2 - where SDQ score is regressed against the substance abuse factor - likely provides the best fit to the data.

Across all matched caregivers, the mean harsh discipline score is approximately 6.05 at baseline (CI : $[5.53, 6.62]$. Higher harsh discipline scores are associated with female caregivers, younger caregivers and those caring for male teens, however these factors do not provide much evidence of significance. Those caregivers reporting substance abuse are associated with a 1.47 factor increase in harsh discipline score ($CI : [1.26, 1.71]$, $p < 0.0001$); those reporting fighting at home with a 1.70 factor increase ($CI : [1.49, 1.94]$, $p < 0.0001$); and those reporting living with a sick child or sick adult are associated with a 1.78 ($CI : [1.55, 2.04]$, $p < 0.0001$) and 1.48 ($CI : [1.28, 1.72]$, $p < 0.0001$) factor increase. Caregivers reporting the HIV/TB factor are associated with an average 27% reduction in harsh discipline score (CI : $[13\%, 39\%]$, $p = 0.0005$). The AIC values indicate that models 5 and 6 provide the best fit to the data - where ICAST score is univariately regressed against fighting at home and sick child, respectively.

Table 10 presents the regression results concerning the baseline adolescent scores. The average total positive parenting score is approximated as 15.48 ($CI : [14.87, 16.08]$). Male teenagers have a mean score around 1.76 units less than females ($CI : [-2.94, -0.58]$, $p = 0.0037$), while older teens are associated with a 0.13 unit reduction in score per year when compared to the youngest ($CI : [-0.37, 0.10]$, $p = 0.26$) – an insignificant result. The risk factor for positive parenting of greatest magnitude and significance relates to those reporting fighting at home, where the average score is reduced by -2.64 units ($CI : [-3.71, -1.57]$, $p < 0.0001$). The HIV/TB factor is associated with an increased APQ score of around 1.09 units, although this estimate is likely unreliable ($CI : [-0.24, 2.43]$, $p = 0.106$). The directions in which the fighting at home and HIV/TB factors act is consistent with that seen in the previously explored caregiver results. The AIC values imply that model 4 fits the data best, where fighting at home is the moderator.

Table 10: Baseline Regression Results: Adolescents

Model Index	Model Specification	Assumed Distribution	Intercept [C.I]	Effect size [C.I]	t-statistic (p-value)	AIC
null	APQ ∼1	Gaussian	15.48 [14.87, 16.08]	-	-	401.05
1	null + Teen sex		16.18 [15.44, 16.93]	-1.76 [-2.94, -0.58]	-2.99 (0.0037)	394.38
2	null + Teen age		15.96 [14.91, 17.01]	-0.13 [-0.37, 0.10]	-1.13 (0.26)	401.75
3	null + Fighting at home		16.67 [15.95, 17.39]	-2.64 [-3.71, -1.57]	-4.92 (<0.0001)	381.35
4	null + HIV_TB		15.17 [14.46, 15.88]	1.09 [-0.24, 2.43]	1.64 (0.106)	400.35
null	SDQ+1 ∼1	Gamma (log-link)	7.62 [6.94, 8.36]	-	-	421.77
1	null + Teen age		6.71 [5.72, 7.86]	1.04 [1.00, 1.07]	1.92 (0.0589)	420.91
2	null + Substance abuse		6.32 [5.59, 7.14]	1.42 [1.19, 1.70]	3.98 (0.00016)	407.42
3	null + Fighting at home		6.81 [6.00, 7.74]	1.27 [1.05, 1.53]	2.48 (0.0153)	414.74
7	null + HIV_TB		8.04 [7.17, 9.01]	0.84 [0.68, 1.02]	-1.77 (0.0811)	421.47
null	ICAST ∼1	Poisson (log-link)	4.35 [3.91, 4.82]	-	-	597.99
1	null + Teen sex		4.04 [3.51, 4.65]	1.19 [0.96, 1.46]	1.62 (0.104)	597.36
2	null + Child age		3.67 [3.04, 4.45]	1.05 [1.00, 1.09]	2.91 (0.0285)	595.18
3	null + Caregiver sex		4.65 [4.15, 5.21]	0.69 [0.52, 0.92]	-2.564 (0.0103)	592.85
4	null + Substance abuse		3.76 [3.23, 4.38]	1.34 [1.08, 1.65]	2.75 (0.006)	588.28
5	null + Fighting at home		2.93 [2.47, 3.49]	2.06 [1.66, 2.57]	6.65 (<0.00001)	549.84
6	null + HIV_TB		3.47 [3.02, 3.98]	1.89 [1.53, 2.33]	6.01 (<0.00001)	565.50

Note: For APQ models, the effect size is additive. For SDQ and ICAST, effect sizes are multiplicative.
Note: The SDQ coefficients are relative to the observed score + 1, since Gamma distributions assume strictly positive values.
Note: CI = point-wise confidence limits, calculated by: $exp(\beta \pm 2 * s.e(\beta))$ for the models with a log link, and $\beta \pm 2 * s.e(\beta)$ otherwise.

Considering the SDQ models, the reader is reminded here that the adolescent SDQ scale contains many more items than the caregiver scale, and thus the magnitude of both intercept and effect should not be compared between the two. One can see that the average adolescent score at baseline is approximated as 7.62 at baseline ($CI : [6.94, 8.36]$). An older teen is associated with an average 4% increase in score per year from the youngest, but this estimate is not significantly different from no change ($CI : [0\%, 7\%]$, $p = 0.0589$). Those reporting substance abuse and fighting in the household on average have higher SDQ scores of factors 1.42 ($CI : [1.19, 1.70]$, $p = 0.00016$) and 1.27 ($CI : [1.05, 1.53]$, $p = 0.0153$), respectively. Lastly, the HIV/TB factor is associated with a 0.84 change in baseline SDQ score compared to those not reporting this factor, however this is not significantly different to no effect ($CI : [0.68, 1.02]$, $p = 0.0811$). The best model fit, according to AIC, is achieved by model number 2.

Lastly, the average adolescent ICAST score at baseline is around 4.34 ($CI : [3.91, 4.82]$). Those adolescents with male caregivers are associated with an average 31% reduction in ICAST score ($CI : [8\%, 48\%]$, $p = 0.0103$), and male teenagers are associated with a non-significant 19% increase ($CI : [-4\%, 46\%]$, $p = 0.104$), when compared to females. Those reporting substance abuse in the household are associated with an average 34% increased scored when compared to those not reporting this ($CI : [8\%, 65\%]$, $p = 0.006$). Of both significance and large magnitude are the fighting at home and HIV/TB factors, which are associated with an 2.06 ($CI : [1.66, 2.57]$, $p < 0.0001$) and 1.89 ($CI : [1.53, 2.33]$, $p < 0.0001$) factor increase in score, respectively, when compared to their counterparts. The AIC statistics infer that model 5, where fighting at home is the covariate, provides the best fit to the observed data.

5.2.2 Model Diagnostics

Residual plots associated with each of the above models are supplied in the Appendix E (Figures 20 - 25), in order to assess the assumptions underlying the above models - that the residuals should be normally distributed with a constant variance. Considering that the data is not continuous, while the assumed underlying distribution is, one should not be surprised to see deviations from these assumptions.

Among the caregiver APQ models, the residuals approximate a normal distribution with some undesirable over-dispersion at the tails in all cases. Although all of the models' residuals seem to exhibit constant variance between the levels of the covariate, the deviation of the residuals from the theoretical distribution is perhaps too extreme to ignore among models null, 1, and 5. The coefficient estimates associated with models 2, 3, 4 and 6 are thus likely to be biased to a negligible degree, and considered acceptable.

The residuals associated with the caregiver SDQ models indicate a much poorer fit than the former. Again one can see the discrete nature of the data is problematic, with bands of residual points shown in the Q-Q plots. Models 1, 4 and 7 seem to provide residuals which most closely mirror the expected distribution, however these are still not very good, while models 1, 4, and 6 exhibit heteroskedasticity. The coefficients associated with all of these models are thus likely to be biased to some degree.

Lastly, the residuals associated with the caregiver ICAST models deviate significantly from the expected distribution. While none of these models truly satisfy the underlying assumptions, model 5 produces residuals which most closely mirror the expected distribution and variance, however the variance between the two levels of this covariate is not constant. The estimates associated with these ICAST models are likely to be biased to some degree.

The adolescent APQ models provide satisfactorily distributed residuals for models 3 and 4, where the variance is approximately constant between the two levels of the explanatory variables. Those residuals associated with models null, 1 and 2 deviate too excessively from the theoretical quantiles to satisfy the assumptions. Models 3 and 4 thus likely produce negligibly biased coefficient estimates.

Among the adolescent SDQ models, the residuals are over-dispersed at the tails for all models excepting model 4. Model 4, however deviates from the expected distribution around the center of the Q-Q plot. None of these models can be said to truly satisfy the underlying assumptions, and the associated estimates are thus likely to be biased.

The adolescent ICAST models produce residuals that certainly do not satisfy the underlying assumptions, which can be attributed to the lack of a zero-inflated model in the face of data dominated by scores of zero. There are thus many observations with residuals of high magnitude, which correspond with the 'zero' observations, causing a distribution of residuals which deviates significantly from the theoretical distribution at the upper quantiles. The coefficient estimates associated with these models are thus certainly biased to a non-negligible degree.

5.3 Mixed Effect Regression Models

The results from the generalised mixed-effect regression models are presented in section 6.3.1 below. Since the baseline scores were analysed separately, this section will focus on the time effect, and the moderator-time interaction. An important note here is that the mixed effect model coefficients are estimated with a different

method than the GLM's, and thus the baseline coefficients are expected to vary marginally between the two models. Where there is a binary moderator, the time effect is interpreted as the average change for the one group from baseline to follow up, while the interaction represents how this time effect is altered for the other group. The residuals resulting from the best fitting models are then analysed in section 6.3.2 to validate the assumptions underlying these models.

5.3.1 Results

Examining Table 11, one can see that - across the caregiver responses - the average change in APQ score over time is estimated as 5.39 ($CI : [4.70, 6.08]$), suggesting an improvement in positive parenting. The residual variance is around 4.37 times the variance of the random effect. The factors which moderate this time effect most are substance abuse, fighting at home, those living with with a sick child and those living with some with HIV/TB. Those caregivers who report substance abuse at home experienced an average improvement in APQ score around 3.96 units greater than those who did not ($CI : [2.75, 5.18]$). Similarly, those caregivers who reported fighting at home showed an average improved APQ score around 3.76 units greater than those who did not ($CI : [2.37, 5.14]$), while those who reported living with a sick child are associated with an improved score 3.64 units greater than otherwise ($CI : [2.08, 5.19]$). The caregivers who reported the HIV/TB factor are associated with a change in APQ score 3.4 units lower than their counterparts ($CI : [1.96, 4.84]$). The AIC statistics suggest that model 2, with substance abuse as a fixed effect, provides the best fit when compared to the other models.

Table 11: Mixed Effect Results: Caregiver APQ

Model Index	Fixed Effects Included	Fixed effect size:	Intercept [CI]	Moderator [CI]	Time [CI]	Time*Moderator [CI]	Random effects:	$\sigma^2_{subject}$	$\sigma^2_{residual}$	AIC
Base	Time		17.89 [17.33, 18.45]	-	5.39 [4.70, 6.08]	-		1.15	5.03	794.22
1	Time + Caregiver sex + Time*Caregiver sex		17.47 [16.85, 18.09]	2.02 [0.65, 3.38]	5.93 [5.18, 6.67]	-2.54 [-4.15, -0.92]		1.31	4.61	784.82
2	Time + Substance abuse + Time*Substance abuse		19.34 [18.77, 19.91]	-4.27 [-5.25, -3.30]	4.04 [3.33, 4.75]	3.96 [2.75, 5.18]		0.68	3.53	732.44
3	Time + Fighting at home + Time*Fighting at home		18.84 [18.25, 19.43]	-3.52 [-4.66, -2.39]	4.37 [3.65, 5.09]	3.76 [2.37, 5.14]		0.96	4.02	758.39
4	Time + Sick child + Time*Sick child		18.58 [17.99, 19.16]	-3.32 [-4.60, -2.04]	4.62 [3.91, 5.33]	3.64 [2.08, 5.19]		1.08	4.23	768.18
5	Time + Sick adult + Time*Sick adult		18.57 [17.87, 19.26]	-1.76 [-2.89, -0.64]	4.52 [3.67, 5.36]	2.25 [0.89, 3.61]		1.23	4.62	784.39
6	Time + HIV_TB + Time*HIV_TB		15.57 [14.58, 16.56]	3.18 [2.02, 4.34]	7.91 [6.67, 9.15]	-3.40 [-4.84, -1.96]		1.00	4.22	766.35

Note: Effect sizes are additive.
Note: CI = point-wise confidence limits, calculated by: $\beta \pm 2 * s.e(\beta)$.
Note: APQ modeled assuming Gaussian distribution.

Considering the adolescent responses (Table 12), the average change in APQ score is estimated as 7.50 ($CI : [6.98, 8.01]$), a greater change than that seen from the caregiver perspective. The overall variability is less captured by the random effect here than in the caregiver responses. The greatest time-effect moderators in this case relate to the adolescents' sex, fighting at home, and the HIV/TB factor. Male teenagers, on average, are associated with a change in score 1.53 units greater than their female counterparts ($CI : [0.52, 2.54]$), while adolescents who reported fighting at home are associated with a change in score 1.95 units greater than those who did not ($CI : [0.99, 2.92]$). Lastly, adolescents who reported living with someone who suffers from HIV/TB experienced an average change in APQ score around 1.41 units less than those who did not ($CI : [0.29, 2.52]$). Considering the AIC values, the best fitting model here is model 5, where child behaviour scores are regressed against the fighting at home factor.

Table 12: Mixed Effect Results: Adolescent APQ

Model Index	Fixed Effects Included	Fixed effect size:	Intercept [CI]	Moderator [CI]	Time [CI]	Time*Moderator [CI]	Random effects:	$\sigma^2_{subject}$	$\sigma^2_{residual}$	AIC
Base	Time		15.56 [15.09, 16.04]	-	7.50 [6.98, 8.01]	-		1.72	2.76	717.39
1	Time + Teen sex + Time*Teen sex		16.19 [15.60, 16.78]	-1.60 [-2.54, -0.66]	6.88 [6.24, 7.53]	1.53 [0.52, 2.54]		1.65	2.55	707.91
2	Time + Teen age + Time*Teen age		15.90 [15.09, 16.71]	-0.09 [-0.28, 0.09]	6.92 [6.04, 7.80]	0.16 [-0.04, 0.36]		1.77	2.72	724.84
3	Time + Caregiver sex + Time*Caregiver sex		15.48 [14.95, 16.02]	0.37 [-0.75, 1.49]	7.68 [7.10, 8.26]	-0.85 [-2.10, 0.40]		1.76	2.74	718.33
4	Time + Substance abuse + Time*Substance abuse		15.70 [15.05, 16.35]	-0.23 [-1.19, 0.74]	7.25 [6.55, 7.96]	0.50 [-0.54, 1.55]		1.74	2.78	711.88
5	Time + Fighting at home + Time*Fighting at home		16.66 [16.09, 17.24]	-2.54 [-3.41, -1.66]	6.61 [5.97, 7.26]	1.95 [0.99, 2.92]		1.33	2.38	686.13
6	Time + HIV_TB + Time*HIV_TB		15.29 [14.73, 15.85]	0.98 [-0.05, 2.02]	7.90 [7.30, 8.49]	-1.41 [-2.52, -0.29]		1.81	2.61	714.01

Note: Effect sizes are additive.
Note: CI = point-wise confidence limits, calculated by: $\beta \pm 2 * s.e(\beta)$.
Note: APQ modeled assuming Gaussian distribution.

Table 13 reveals that the average change in total caregiver SDQ score is estimated as a reduction of 1.45 units ($CI : [1.14, 1.77]$), while the random effect variance is around 4 times the residual variance. The main factors which modify this effect include substance abuse, fighting at home, living with a disabled child, a sick child, a sick adult, or someone with HIV/TB. Caregivers who reported substance abuse are associated with a change in score 2.11 units greater than those who do not ($CI : [1.58, 2.63]$), while reporting fighting at home was associated with a change in score 1.83 units less than otherwise ($CI : [1.22, 2.63]$). The corollaries for those reporting living with a disabled child, sick child, sick adult or person with HIV/TB are: -2.20 ($CI : [-3.04, -1.37]$); -2.04 ($[-2.70, -1.38]$; -0.94 $CI : [-1.56, -0.32]$; and 1.23 $CI : [0.56, 1.91]$, respectively. Model 3, where substance abuse is the moderator, seems to provide the best fit among these models ($AIC = 454.77$).

Table 13: Mixed Effect Results: Caregiver SDQ

Model Index	Fixed Effects Included	Fixed effect size:	Intercept [CI]	Moderator [CI]	Time [CI]	Time*Moderator [CI]	Random effects:	$\sigma^2_{subject}$	$\sigma^2_{residual}$	AIC
Base	Time		2.46 [2.20, 2.72]	-	-1.45 [-1.77, -1.14]	-		0.26	1.04	528.48
1	Time + Caregiver age + Time*Caregiver age		3.40 [2.90, 3.91]	-0.07 [-0.10, -0.04]	-2.43 [-3.04, -1.81]	0.07 [0.03, 0.10]		0.23	0.94	520.00
2	Time + Teen sex + Time*Teen sex		2.21 [1.88, 2.54]	0.62 [0.10, 1.15]	-1.20 [-1.60, -0.80]	-0.62 [-1.25, 0.01]		0.26	1.01	528.17
3	Time + Substance abuse + Time*Substance abuse		1.76 [1.51, 2.01]	2.13 [1.69, 2.57]	-0.75 [-1.05, -0.45]	-2.11 [-2.63, -1.58]		0.17	0.65	454.77
4	Time + Fighting at home + Time*Fighting at home		1.96 [1.70, 2.23]	1.83 [1.32, 2.34]	-0.95 [-1.27, -0.63]	-1.83 [-2.45, -1.22]		0.19	0.79	485.34
5	Time + Disabled child + Time*Disabled child		2.17 [1.92, 2.41]	2.21 [1.54, 2.89]	-1.16 [-1.46, -0.86]	-2.20 [-3.04, -1.37]		0.21	0.82	492.67
6	Time + Sick child + Time*Sick child		2.03 [1.78, 2.28]	2.06 [1.52, 2.61]	-1.02 [-1.32, -0.71]	-2.04 [-2.70, -1.38]		0.19	0.77	481.67
7	Time + Sick adult + Time*Sick adult		2.10 [1.78, 2.41]	0.95 [0.44, 1.46]	-1.08 [-1.47, -0.69]	-0.94 [-1.56, -0.32]		0.24	0.97	520.24
8	Time + HIV_TB + Time*HIV_TB		3.35 [2.88, 3.82]	-1.22 [-1.77, -0.67]	-2.36 [-2.95, -1.78]	1.23 [0.56, 1.91]		0.24	0.93	513.98

Note: Effect sizes are additive.
Note: Note: CI = point-wise confidence limits, calculated by: $\beta \pm 2 * s.e(\beta)$.
Note: SDQ modeled assuming Gaussian distribution.

Regarding the adolescent's perspective of child behaviour (Table 14), the average change in SDQ score across the participants is estimated as -6.38 units ($CI : [6.07, 7.08]$). In this model, the ratio between the residual variance and the random intercept variance is about 5.5. An additional year of age is associated with a reduction in score 0.23 units more than that of the youngest teens ($CI : [0.02, 0.47]$). The adolescents reporting substance abuse, and those reporting fighting at home, were associated with reductions in scores about 2.49 units ($CI : [1.31, 3.68]$) and 1.81 units ($CI : [0.53, 3.08]$) greater than their counterparts, respectively. Those who reported living with someone affected by HIV/TB generally had reductions in their scores about 1.31 units smaller than the change in scores for those who did not report his factor ($CI : [-0.07, 2.70]$). Of these models, the AIC indicates that model 3 best fits the data - which is consistent with the results of the caregiver SDQ models.

Table 14: Mixed Effect Results: Adolescent SDQ

Model Index	Fixed Effects Included	Fixed effect size:	Intercept [CI]	Moderator [CI]	Time [CI]	Time*Moderator [CI]	Random effects:	$\sigma^2_{subject}$	$\sigma^2_{residual}$	AIC
Base	Time		6.58 [6.07, 7.08]	-	-6.38 [-7.02, -5.73]	-		0.76	4.22	730.05
1	Time + Teen age + Time*Teen age		5.63 [4.78, 6.47]	0.27 [0.08, 0.46]	-5.57 [-6.67, -4.47]	-0.23 [-0.47, 0.02]		0.61	4.18	732.19
2	Time + Caregiver age + Time*Caregiver age		5.96 [4.92, 7.00]	0.04 [-0.02, 0.10]	-5.70 [-7.03, -4.36]	-0.05 [-0.13, 0.03]		0.80	4.24	734.88
3	Time + Substance abuse + Time*Substance abuse		5.23 [4.59, 5.88]	2.75 [1.81, 3.68]	-5.15 [-5.96, -4.35]	-2.49 [-3.68, -1.31]		0.67	3.49	693.48
4	Time + Fighting at home + Time*Fighting at home		5.81 [5.15, 6.47]	1.76 [0.76, 2.76]	-5.57 [-6.42, -4.72]	-1.81 [-3.08, -0.53]		0.71	4.02	713.79
5	Time + HIV_TB + Time*HIV_TB		6.98 [6.37, 7.59]	-1.26 [-2.34, -0.19]	-6.79 [-7.56, -6.02]	1.31 [-0.07, 2.70]		0.75	4.14	727.19

Note: Effect sizes are additive.
Note: CI = point-wise confidence limits, calculated by: $\beta \pm 2 * s.e(\beta)$.
Note: SDQ modeled assuming Gaussian distribution.

Tables 15 and 16 present the results for the mixed effect models for the caregiver and adolescent harsh discipline score, respectively. Since, among both groups, the responses are homogeneously '0' at follow-up, the ICAST scores over time are estimated to change by a factor of 0. Essentially, no matter the baseline score, all participants scores are zero at follow up and this means that the magnitude of the change over time can only possibly be attributed to a participant's position at baseline. Examining Table 15, one can thus see that the time-effect moderators are essentially just the complement of the baseline moderator effect. For example, as seen previously, those caregivers reporting substance abuse were associated with a 147% increase in the baseline score. The effect on time for this factor is thus a change in score 0.64 times that of those who did not report substance abuse. The moderator-time interaction is effectively meaningless then, since the interpretation is a 0.64 factor modification on 0.00. Among these caregiver models, the AIC indicates that the best fit is seen in models 5 and 6 - where harsh discipline scores are regressed against fighting at home and sick child, respectively.

Table 15: Mixed Effect Results: Caregiver ICAST

Model Index	Fixed Effects Included	Fixed effect size:	Intercept [CI]	Moderator [CI]	Time [CI]	Time*Moderator [CI]	Random effects:	$\sigma^2_{subject}$	AIC
Base	Time		5.93 [5.32, 6.60]	-	0.00 [0.00, Inf]	-		0.0453	384.99
1	Time + Teen sex + Time*Teen sex		5.65 [4.92, 6.48]	1.13 [0.91, 1.39]	0.00 [0.00, Inf]	0.86 [0.00, Inf]		0.0420	387.71
2	Time + Caregiver sex + Time*Caregiver sex		6.16 [5.48, 6.92]	0.83 [0.64, 1.08]	0.00 [0.00, Inf]	1.25 [0.00, Inf]		0.0400	387.05
3	Time + Caregiver age + Time*Caregiver age		7.49 [6.11, 9.18]	0.98 [0.97, 1.00]	0.00 [0.00, Inf]	1.02 [0.00, Inf]		0.0282	378.00
4	Time + Substance abuse + Time*Substance abuse		5.19 [4.59, 5.87]	1.47 [1.21, 1.77]	0.00 [0.00, Inf]	0.64 [0.00, Inf]		0.0130	375.17
5	Time + Fighting at home + Time*Fighting at home		5.08 [4.53, 5.69]	1.70 [1.42, 2.05]	0.00 [0.00, Inf]	0.54 [0.00, Inf]		0.00203	361.26
6	Time + Sick child + Time*Sick child		5.19 [4.66, 5.78]	1.78 [1.47, 2.15]	0.00 [0.00, Inf]	0.51 [0.00, Inf]		0.00157	359.83
7	Time + Sick adult + Time*Sick adult		5.06 [4.44, 5.77]	1.49 [1.23, 1.80]	0.00 [0.00, Inf]	0.63 [0.00, Inf]		0.0169	373.40
8	Time + HIV_TB + Time*HIV_TB		7.40 [6.20, 8.82]	0.74 [0.60, 0.91]	0.00 [0.00, Inf]	1.45 [0.00, Inf]		0.0221	381.66

Note: Effect sizes are multiplicative.
Note: CI = point-wise confidence limits, calculated by: $exp(\beta \pm 2 * s.e(\beta))$.
Note: ICAST modeled assuming Poisson distribution with log-link.

Table 16 presents similar results for the adolescent score, however the model is adjusted by using a zero-inflated negative Binomial distribution, with a zero-part fixed effect on the intercept and the moderator. This is done because the baseline adolescent scores are also dominated by zeros, and thus these need to be modeled explicitly. Unlike the other models in this section, the intercept and baseline moderator effect are actually more informative here than the model presented in Section 6.2.1 (Table 10). The zero-part intercept effect gives the probability of a response being zero at baseline. One can see from the base model's zero-part intercept, that the probability of an adolescent's ICAST score being zero at baseline is estimated as 0.92 ($CI : [0.64, 0.98]$).

As with the caregiver ICAST scores, the observations among teens are homogeneously zero at the second time point and thus there is clearly no point in putting a zero-part effect on time. This also means that the time effect is a factor change of 0.00, and the moderator-time interaction cannot be estimated to any meaningful degree of precision and is fully dependant on a participant's position at baseline. The zero-part

moderator effect shows how the probability of a zero-observation at baseline differs between sample strata (multiplicatively). Nonetheless, all of these estimates are also accompanied by extremely wide confidence intervals (containing 1) which mean that the coefficient cannot be interpreted with any degree of certainty. Given the data, attempting to model the change over time in ICAST scores, and the effect that moderators have on this, is thus not really possible in this framework.

Table 16: Mixed Effect Results: Adolescent ICAST

Model Index	Fixed Effects Included	Fixed effect size:	Intercept [CI]	Moderator [CI]	Time [CI]	Time*Moderator [CI]	Zero-part Intercept [CI]	Zero-part Moderator [CI]	Random effects:	$\sigma^2_{subject}$	AIC
Base	Time		3.53 [2.66, 4.69]	-	0.00 [0.00, Inf]	-	0.08 [0.02, 0.36]			0.531	427.21
1	Time + Teen sex + Time*Teen sex		3.24 [2.32, 4.52]	1.29 [0.79, 2.09]	0.00 [0.00, Inf]	0.84 [0, Inf]	0.05 [0.00, 0.77]	2.65 [0.14, 48.90]		0.509	431.93
2	Time + Teen age + Time*Teen age		3.53 [2.29, 5.41]	1.01 [0.92, 1.10]	0.00 [0.00, Inf]	0.95 [0, Inf]	0.36 [0.11, 1.23]	0.59 [0.32, 1.11]		0.505	427.73
3	Time + Caregiver sex + Time*Caregiver sex		3.65 [2.68, 4.95]	0.85 [0.48, 1.52]	0.00 [0.00, Inf]	1.44 [0, Inf]	0.06 [0.01, 0.44]	2.35 [0.16, 34.80]		0.528	432.07
4	Time + Substance abuse + Time*Substance abuse		3.39 [2.33, 4.92]	1.08 [0.67, 1.75]	0.00 [0.00, Inf]	0.75 [0, Inf]	0.16 [0.05, 0.56]	0.08 [0.00, 79.11]		0.533	424.52
5	Time + Fighting at home + Time*Fighting at home		2.75 [1.93, 3.91]	1.77 [1.12, 2.80]	0.00 [0.00, Inf]	0.48 [0, Inf]	0.13 [0.03, 0.61]	0.34 [0.02, 6.01]		0.449	418.65
6	Time + HIV_TB + Time*HIV_TB		3.09 [2.44, 3.90]	2.40 [1.59, 3.63]	0.00 [0.00, Inf]	0.53 [0, Inf]	0.01 [0.00, 2.08]	37.25 [0.22, 6443.84]		0.304	411.96

Note: Effect sizes are multiplicative.
Note: CI = point-wise confidence limits, calculated by: $exp(\beta \pm 2 * s.e(\beta))$.
Note: ICAST modeled assuming zero-inflated negative Binomial distribution with log-link.

Mixed effect regression models were also fit to the FFC scale, however the model fit was so poor no matter the transformation applied or underlying distribution assumed, and thus is excluded from the analysis. The residual plots associated with this model are provided in the Appendix F and discussed in the section below.

5.3.2 Model Diagnostics

In this section, the distribution of (scaled) DHARMa residuals associated with each of the mixed effect models are analyzed to determine whether the models satisfy the underlying assumptions and provide a good fit to the data. Q-Q plots are again used to determine whether the residuals conform to the expected uniform distribution, while box plots (or scatter plots) of predictor against residual are used to determine whether there is a significant deviation of homoskedasticity. For the sake of brevity, only residual plots pertaining to the best fitting model for each Likert score is displayed here, while the rest are discussed briefly and relegated to Appendix F.

Figure 17 displays these plots for the caregiver APQ model with substance abuse as a fixed effect.

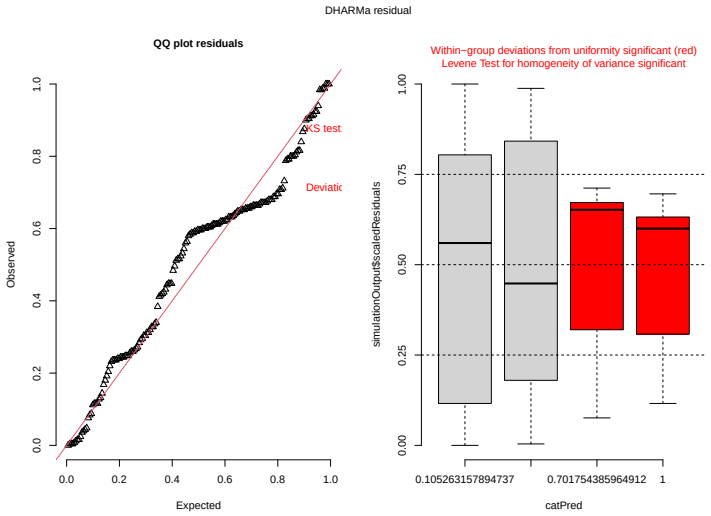


Figure 17: Residual plot: APQ Caregivers - Mixed Model 2

Considering the Q-Q plot above, once can see that there is significant deviation from the expected distribution,

while the box plot indicates that the variance between the sample strata (substance abuse reporters and non-reporters) is significantly different. Considering Figure 37 in Appendix F, one can see that these same issues are present in all of the caregiver APQ models - with a visually identifiable dispersion from uniformity around the center of the expected distribution. One must then conclude that the coefficient estimates displayed in Table 11 are biased.

In Figure 18, the residuals are displayed for the adolescent APQ model with fighting at home included as a fixed effect. These residuals approximate a uniform distribution far better those resulting from the caregiver model, which one can visually see by the points lying close to the 45°line. There is, however, evidence to suggest that the variance in residuals between the two groups is not constant, and thus the coefficient estimates resulting from this model are likely to be biased. Considering Figure 38 in Appendix F, one can see that the residuals from the base model and models 2, 3, and 4 exhibit similar behaviour, while the deviation from uniformity is slightly more pronounced for the residuals associated with models 1 and 6.

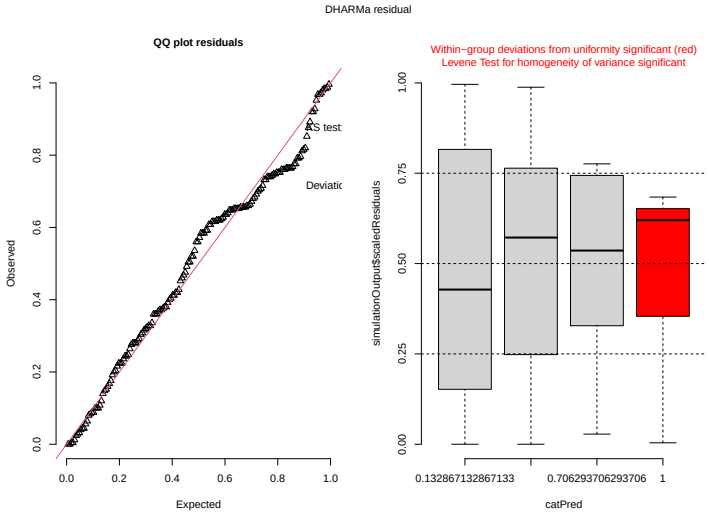


Figure 18: Residual plot: APQ Adolescents - Mixed Model 5

The residuals for the caregiver SDQ model with substance abuse as a moderator are illustrated in Figure 19. It is immediately apparent in this case that the model does not fit the data well, with the residuals barely resembling uniformity. There is a band of observed residuals which indicates a disproportionate number of observations with DHARMa residuals between 0.4 and 0.6, while the variance between groups is far from homogeneous. In Appendix F, Figure 39, one can see that this is similar for all of the caregiver SDQ models. The coefficient estimates associated with these models are thus likely to be significantly biased.

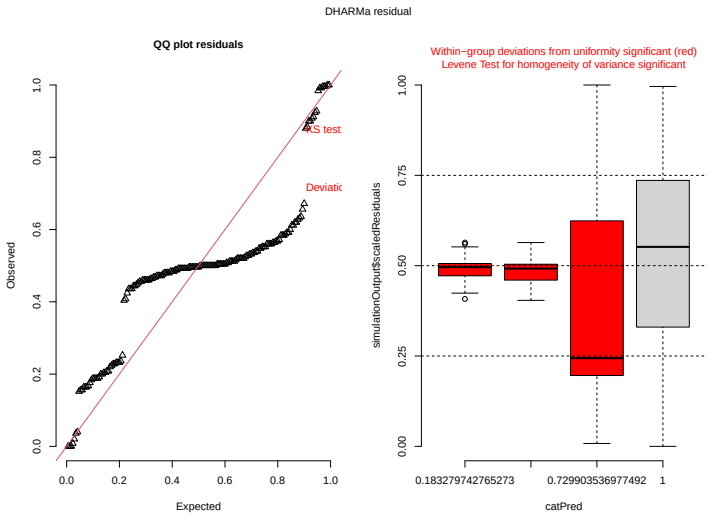


Figure 19: Residual plot: SDQ Caregivers - Mixed Model 3

Figure 20 shows similar, but less prominent, patterns in the adolescent SDQ model with substance abuse as a moderator. There is again a concentration of residuals between 0.4 and 0.6, deviating significantly from uniformity, and the box plots indicate that there is not homogeneity between the groups with respect to residual variance. Figure 40, Appendix F, reveals that this pattern is also seen in the other adolescent SDQ models, and thus the associated coefficient estimates are certainly biased in this case too.

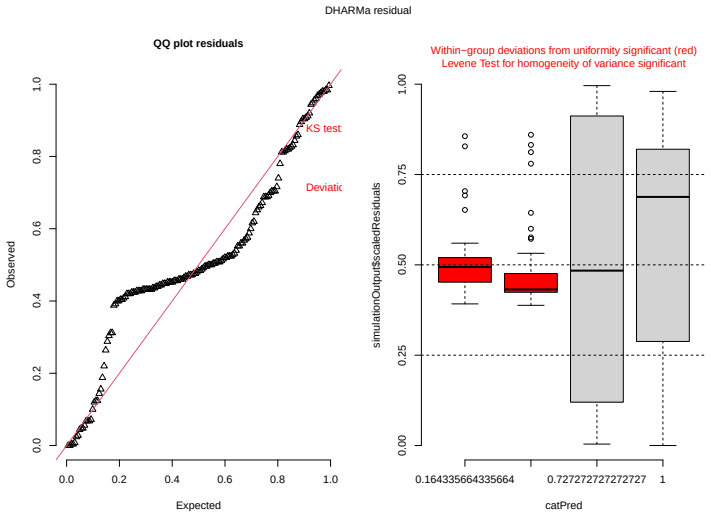


Figure 20: Residual plot: SDQ Adolescents - Mixed Model 3

The residuals generated from the caregiver ICAST model with sick child as a moderator is shown in Figure 21, below. In this case, the residuals very closely mirror the expected distribution, with the most deviation occurring at the upper tail. The right panel shows a scatter plot of the DHARMa residuals against the transformed fitted values, with outlying residuals indicated by red stars. In this case, the residuals seem to exhibit relative uniformity in both directions, and the variance is relatively constant - barring a lower variance among the largest fitted values. These residuals thus provide an acceptable fit, and the associated coefficients can be (cautiously) considered as unbiased. The residuals associated with the rest of the models

in Table 15 are displayed in Figure 41, Appendix F, where one can see very similar patterns. The coefficient estimates displayed in Table 15 can thus all be considered to be unbiased.

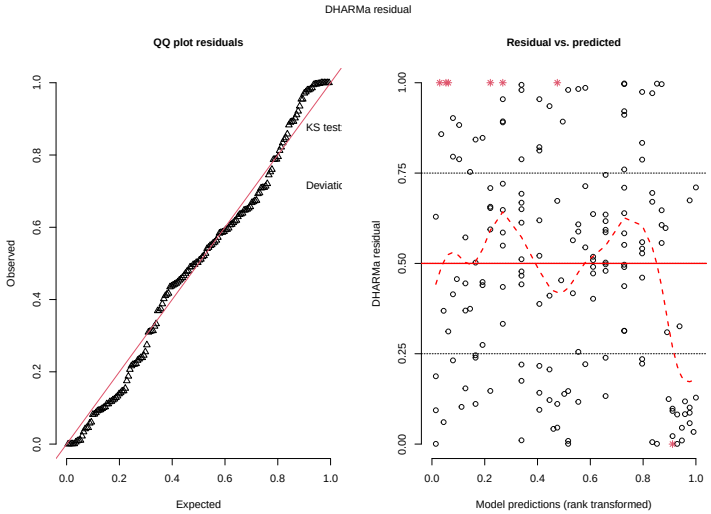


Figure 21: Residual plot: ICAST Caregivers - Mixed Model 6

Figure 22 displays the residuals associated with the adolescent ICAST model with the HIV/TB factor as a moderator. While the variance of the residuals is relatively constant across the fitted values, their distribution deviates from uniformity to a significant degree at both the lower and upper tails. This 'S' shape can be seen across all of the adolescent ICAST models' residuals in Appendix F, Figure 42. One must then conclude that the coefficient estimates provided in Table 16 are biased to some degree, and that the models do not fit the data well.

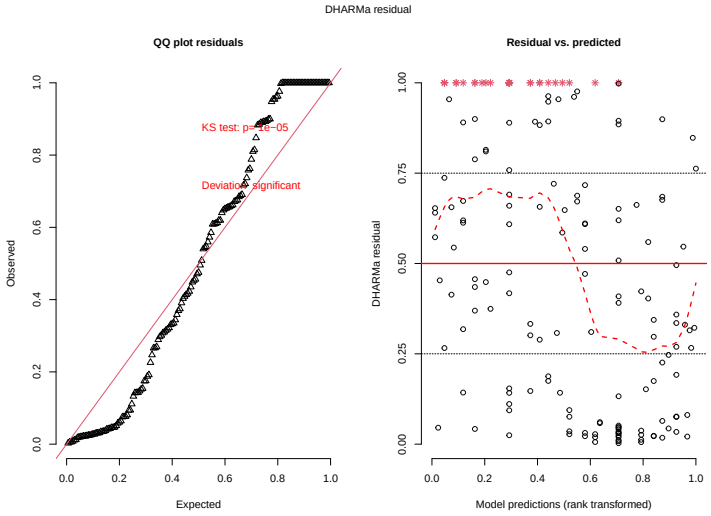
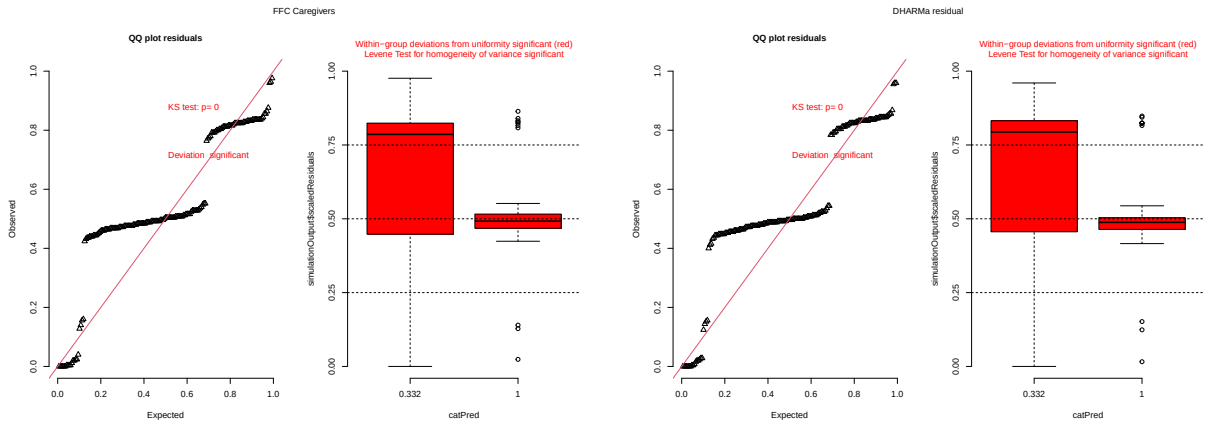


Figure 22: Residual plot: ICAST Adolescents - Mixed Model 6

Lastly, Figures 23a and 23b display the residuals associated with the FFC mixed models assuming a different underlying distributions, with only time as a fixed effect. Once can see that, due to the extremely discrete nature of the data, an appropriate model fit cannot be obtained regardless of the assumed distribution. Further transformations and distributions were used in order to attempt to model this data in a generalised

linear mixed effect framework, however only marginal improvements were obtained, and thus the modeling framework was abandoned with respect to this score.



(a) Base model assuming Gaussian

(b) Base model assuming log-Gaussian

Figure 23: Residual plots for FFC Mixed Models

Overall, the model diagnostics discussed above suggest that the mixed effect models pertaining to the adolescent APQ and caregiver ICAST scores fit the data well, and the associated results can be considered as unbiased. The residuals generated from the caregiver APQ models, while not strictly following the expected form, indicate that the model fit is not extremely poor and one can accept the associated coefficient estimates with a healthy but measured amount of skepticism. However, the residuals associated with the adolescent ICAST models, the caregiver FFC models, and both the caregiver and adolescent SDQ models, deviate too significantly from the expected form for any of the models to be considered as fitting the data acceptably. These scores are thus taken forward into the next section, where the difference in a participant's score over time is modeled explicitly as a response variable. This is not done, however, for the adolescent ICAST scores as the homogeneity at follow-up has the consequence of the difference in a participant's score simply reflecting their baseline score, which is essentially meaningless in this context.

5.4 Generalised Linear Models

5.4.1 Results

The distributions of the differenced data is presented through Figures 29 and 30 in Appendix D. Considering the SDQ responses, the change over time is almost exclusively negative, with one of two outlying observations among the caregiver data (29(b)). While one could certainly attempt to transform these variables such that they approximate a normal distribution, any effective transformation in this regard results either in parameter estimates which cannot be back transformed to the original scale or in an extremely discrete distribution of responses. The changes on the observed scale are thus modeled with a normal distribution, despite their shape. The FFC scores, shown in Figure 29(e), are transformed through the logarithm of it's reflection, where the parameters can be back transformed to the original scale, and modeled assuming a Gaussian distribution.

The results from the caregiver models are presented in Table 17. Given that the modeling of the SDQ scores is practically the same as with the mixed effect models above, the intercept and moderator estimates are very similar to those shown in Table 13. One can again see here that the difference in the change in Likert score between sample strata is mostly due to the differences seen at baseline (Section 6.2.1). Consistent with this, and the results presented in Table 13, the AIC indicates that the best fitting model seems to be the one where the substance abuse factor is included as a covariate.

The average change in FFC score is estimated as -5.83 units ($CI : [-6.09, -5.57]$) across the caregivers, signaling an improvement in family financial coping. The factors which moderate this change to the greatest extent are substance abuse and HIV/TB. Those caregivers who reported substance abuse in the household are associated with a reduction in FFC score around 13% greater than those who did not ($CI : [3\%, 23\%]$), while those living with someone with HIV/TB generally had a reduction in FFC score around 0.83 times that of the caregivers who did not report this factor ($CI : [0.76, 0.91]$). It cannot be said that any of the other demographic factors moderate the change in either direction with an acceptable degree of certainty.

Table 17: Differenced Scores Regression Results: Caregivers

Model Index	Model Specification	Assumed Distribution	Intercept [CI]	Effect size [CI]	p-value	AIC
Null	Δ SDQ $\sim$1	Gaussian	-1.45 [-1.81, -1.10]	-	-	322.03
1	null + Caregiver age		-2.43 [-3.12, -1.74]	0.07 [0.03, 0.11]	0.0013	309.69
2	null + Teen sex		-1.20 [-1.65, -0.75]	-0.62 [-1.33, 0.09]	0.082	320.93
3	null + Substance abuse		-0.75 [-1.09, -0.41]	-2.11 [-2.70, -1.52]	<0.00001	283.57
4	null + Fighting at home		-0.95 [-1.31, -0.59]	-1.83 [-2.52, -1.15]	<0.00001	298.91
5	null + Disabled child		-1.16 [-1.50, -0.83]	-2.20 [-3.13, -1.27]	<0.00001	303.88
6	null + Sick child		-1.02 [-1.36, -0.67]	-2.04 [-2.78, -1.30]	<0.00001	297.41
7	null + Sick adult		-1.08 [-1.51, -0.64]	-0.95 [-1.65, -0.26]	0.0076	316.69
8	null + HIV/TB		-2.36 [-3.02, -1.71]	1.23 [0.48, 1.99]	0.0017	313.85
Null	- Δ FFC $\sim$1	Gaussian (log-link)	-5.83 [-5.57, -6.09]	-	-	-26.05
1	null + Caregiver age		-6.20 [-6.80, -5.65]	1.00 [0.99, 1.00]	0.10	-30.28
2	null + Teen sex		-5.65 [-5.98, -5.33]	1.08 [0.99, 1.18]	0.0921	-26.98
3	null + Substance abuse		-5.59 [-5.89, -5.29]	1.13 [1.03, 1.23]	0.0099	-30.91
4	null + Fighting at home		-5.71 [-6.02, -5.42]	1.07 [0.97, 1.18]	0.1610	-26.08
5	null + Sick adult		-6.06 [-6.41, -5.73]	0.91 [0.83, 0.99]	0.032	-28.80
6	null + HIV/TB		-6.69 [-7.26, -6.17]	0.83 [0.76, 0.91]	0.00017	-38.677

Note: Intercepts and effect sizes presented on observed scale.
Note: Effect sizes are additive for SDQ models and multiplicative for FFC models
Note: CI = point-wise confidence limits, calculated by: $exp(\beta \pm 2 * s.e(\beta))$ for the FFC models, and $\beta \pm 2 * s.e(\beta)$ otherwise.

The results of the adolescent SDQ models are presented in Table 18, where one can see similar consequences as with the caregiver SDQ models. The effect sizes of the moderators are very similar to that seen with the mixed effect models, and the AIC indicates that model 3 (with substance abuse as a covariate) fits the data best.

Table 18: Differenced Scores Regression Results: Adolescents

Model Index	Model Specification	Assumed Distribution	Intercept [CI]	Effect size [CI]	p-value	AIC
Null	Δ SDQ $\sim$1	Gaussian	-6.37 [-7.05, -5.68]	-	-	403.24
1	null + Caregiver age		-5.69 [-7.10, -4.27]	-0.05 [-0.13, 0.04]	0.262	399.68
2	null + Teen age		-5.57 [-6.73, -4.40]	-0.22 [-0.49, 0.04]	0.0931	402.33
3	null + Substance abuse		-5.15 [-6.02, -4.27]	-2.50 [-3.77, -1.23]	0.000178	385.91
4	null + Fighting at home		-5.56 [-6.46, -4.66]	-1.81 [-3.16, -0.47]	0.00851	394.02
5	null + HIV/TB		-6.78 [-7.60, -5.96]	1.30 [-0.16, 2.75]	0.0782	402.04

Note: Intercepts and effect sizes presented on the observed scale.
Note: Effect sizes are additive.
Note: CI = point-wise confidence limits, calculated by: $\beta \pm 2 * s.e(\beta)$.

5.4.2 Model Diagnostics

The residual plots associated with each of these models are provided in Appendix G.

Considering Figure 43, the residuals do seem to approximate a normal distribution relatively well for models 1, 4, and 8. Model 1, however exhibits heteroskedasticity as the residuals are not evenly distributed across the fitted values. One can conclude that the parameter estimates provided by models 4 and 8 are negligibly biased.

Figure 44 shows that the FFC models still fit the data poorly, where the residuals significantly deviate from the expected distribution at the upper tail in all models. None of these models thus satisfy the assumptions underlying GLM's and the estimates are considered biased.

Lastly, the residuals associated with the adolescent SDQ models in Table 18 are presented in Figures 45(a) - (f). Across all of these residual plots, there is significant deviation from the theoretical distribution at both the upper and lower tails, and thus the parameter estimates associated with these models are likely biased.

5.5 Cumulative Link Mixed-Effect Models

This section presents the results of cumulative link mixed-models fit to the response items pertaining to caregiver depression, parental support of education, and attitudes towards physical punishment. These items are measured on Likert scales, and thus the models allow the estimation of the average relative odds of participants' responses moving from a higher to a lower category over time. Furthermore, this model is applied to the total scores to estimate the relative odds of a participant's score being smaller at follow up when compared to baseline.

5.5.1 Results

Table 19 presents the estimated average odds of a participant's response for a question moving from a higher score to a lower score (such as 'agree' to 'disagree' for MICS). The relative odds of a caregiver's response for the first two CESD questions moving to a lower category after the intervention are estimated to be close to certainty, while the corollary for the question relating to 'effort' is close to zero. This is an intuitive result since, if one considers the distribution of responses previously explored in Section 5.2, Figure 8(a), it can be seen that the responses for the 'effort' and 'lonely' questions are homogeneously '5-7' days at baseline, and almost all '1-2' days at follow up. Similarly, the responses for the 'effort' question are all '3-4 days' at baseline and dominated by the '5-7 days' category at follow up.

Similarly, the average odds of a caregiver disagreeing more with the MICS question at follow up compared to at baseline is around 55.56 ($CI : [45.441, 58.490]$). This suggests that it is almost certain that a caregiver will disagree with physical punishment being a necessary tool for child discipline more after participating in the study. One can again compare this interpretation with Figure 9(a).

Lastly, the average odds of a caregiver praising their child less and supporting their child's schoolwork less at the second time point are both relatively close to zero. Again, the responses to this question were homogeneously 'sometimes' at baseline and mostly 'almost every day' at follow up, and thus this interpretation makes logical sense.

Table 19: Cumulative Link Model Results: Caregiver Items

Domain	Model Specification	$\hat{Odds}(Score_2 < Score_1)$ [CI]	p-value	$\sigma^2_{subject}$
CESD	*Depressed ∼Time	5.96e+28 [3.58e+16, 9.92e+40]	<0.00001	1252.03
	*Lonely ∼Time	3.85e+26	**NA	3652.45
	*Effort ∼Time	2.19e-20 [0, Inf]	0.914	1009.21
MICS	Opinion ∼Time	51.555 [45.441, 58.490]	<0.00001	1.16
PSE	Praise ∼Time	0.00152 [8.87e-5, 0.02606]	<0.00001	0.83
	*Others ∼Time	7.19e-14 [7.82e-14, 8.01e-14]	<0.00001	2148.33

Note: CI refers to two standard errors above and below estimate.
Note: '*' - model did not converge.
Note: '**' - standard errors could not be estimated.

Table 20 presents similar results among the relevant teen items. The odds of an adolescent disagreeing more with the MICS question after the intervention is extremely high. This reflects the fact that the respondents homogeneously 'disagreed strongly' with this statement in the follow up survey, while only about a third responded in this manner at baseline (Figure 9(b)). The movement of response for the PSE questions are less certain than that of MICS, but only relatively so. The relative odds of a teen reporting that their caregiver praises them for doing homework (Praise) and supports their schoolwork (Others) less after the intervention than before are less than one and close to zero. This indicates that it is very likely that an adolescent feels more supported with regards to their education after the intervention when compared to before.

Table 20: Cumulative Link Model Results: Adolescent Items

Domain	Model Specification	$\hat{Odds}(Score_2 < Score_1)$ [CI]	p-value	$\sigma^2_{subject}$
MICS	Opinion ∼Time	1.43e+8	**NA	0.73
PSE	Praise ∼Time	0.03195 [0.01107, 0.09222]	<0.00001	0.35
	Others ∼Time	0.05511 [0.02427, 0.07267]	<0.00001	0.84

Note: CI refers to two standard errors above and below estimate.
Note: '**' - standards errors could not be estimated.

The results relating to the cumulative link mixed models fit to the caregiver and adolescent scores are presented in Table 21.

Table 21: Cumulative Link Model Results: Scores

Cohort	Model Specification	$\hat{Odds}(Score_2 < Score_1)$ [CI]	p-value	$\sigma^2_{subject}$
Caregivers	APQ ∼Time	0.0059 [0.0017, 0.0209]	<0.00001	0.94
	SDQ ∼Time	167.32 [20.80, 1345.78]	<0.00001	3.31
	*ICAST ∼Time	5.07e+124	**NA	16754.68
	FFC ∼Time	6.47e+36	**NA	854.12
Adolescents	APQ ∼Time	5.77e-5 [5.36e-6, 6.22e-4]	<0.00001	3.19
	SDQ ∼Time	1700.85 [210.50, 13743.00]	<0.00001	2.21
	ICAST ∼Time	1.16e+13	**NA	0.93

Note: CI refers to two standard errors above and below estimate.
Note: '*' - model did not converge.
Note: '**' - standards errors could not be estimated

One can see that, on average, the relative odds of a caregiver's APQ score being lower at follow up when compare to baseline is close to zero. In addition, caregivers are extremely likely to have lower SDQ, ICAST and FFC scores at follow up than at baseline. The parameter estimation procedure did not converge for the ICAST model, while the standard errors associated with the ICAST and FFC models' coefficient estimates

could not be computed. The subject-specific random effect explains a non-negligible proportion of the variance in all of the models, however the latter two estimates are disproportionately high, perhaps indicating an error in the estimation procedure.

The results are similar for the adolescent scores: the odds of recording a lower APQ at follow-up than at at baseline are also extremely small, while the odds of recording a smaller SDQ or ICAST score at follow-up are extremely large.

While these models could be adjusted for moderators to compare these odds estimates for different sample strata, given that the estimates are either very low or very high, any adjustment likely be negligible.

5.5.2 Model Diagnostics

As outlined in Section 4.6, the significance of the direction of treatment effect can be assessed by conducting a likelihood ratio test between a null model and the model with time as a fixed effect. The resulting likelihood ratio statistics and associated p-values are thus presented in Table 22 below.

Table 22: Significance of Treatment Effect: Items

Domain	Model Comparison	Likelihood Ratio Statistic	p-value
CESD Caregiver	Depressed ∼Time vs. Depressed ∼1	262.87	<0.00001
	Lonely ∼Time vs. Lonely ∼1	282.19	<0.00001
	Effort ∼Time vs. Effort ∼1	222.08	<0.00001
MICS Caregiver	Opinion ∼Time vs. Opinion ∼1	79.18	<0.00001
PSE Caregiver	Praise ∼Time vs. Praise ∼1	158.40	<0.00001
	Others ∼Time vs. Others ∼1	225.91	<0.00001
MICS Adolescent	Opinion ∼Time vs. Opinion ∼1	82.31	<0.00001
PSE Adolescent	Praise ∼Time vs. Praise ∼1	77.84	<0.00001
	Others ∼Time vs. Others ∼1	67.72	<0.00001

The results indicate that the likely movement of scores from baseline to follow up, quantified through the models above, indicate that a significant effect of reduced caregiver depression and loneliness, and both caregiver and teen agreement with corporal punishment, are likely correlated with the program. The results also suggest that the increased parental support of education is associated with the program from both the caregiver and teen perspective. Interestingly, it is also suggested that there is a significant association of caregivers feeling that everything requires effort more frequently after the program than before.

Table 23 presents the same information for the score models. The results suggest that the treatment effect was significant with respect to: a reduction in both caregiver and adolescent APQ scores; an increase in both caregiver and adolescent SDQ and ICAST scores; a reduction in caregiver FFC scores.

Table 23: Significance of Treatment Effect: Scores

Cohort	Model Comparison	Likelihood Ratio Statistic	p-value
Caregiver	APQ ∼Time vs. APQ ∼1	148.00	<0.00001
	SDQ ∼Time vs. SDQ ∼1	83.05	<0.00001
	ICAST ∼Time vs. ICAST ∼1	374.28	<0.00001
	FFC ∼Time vs. FFC ∼1	304.70	<0.00001
Adolescent	APQ ∼Time vs APQ ∼1	224.83	<0.00001
	SDQ ∼Time vs. SDQ ∼1	181.29	<0.00001
	ICAST ∼Time vs. ICAST ∼1	158.22	<0.00001

6 Discussion

6.1 Key Results

The analysis of the data provides strong evidence in favour of the scaled-up implementation of the PLH for Adolescents program's effectiveness in improving socio-psychological outcomes for this sample of South-Sudanese adolescents and their caregivers. The results of both the mixed effect generalised linear regression models and the cumulative link mixed effect models indicate that there was an improvement in positive parenting, and a reduction in poor child behaviour and harsh discipline, from both the caregiver and adolescent perspective. In addition, there seem to have been significant improvements in caregiver mental health, family financial coping, and parental support of education.

Caregiver positive parenting scores after the intervention were, on average, between 26.3% and 34.0% higher than at baseline, while the adolescent scores were improved by between 44.9% and 51.5%. Child behaviour scores were improved by between 28.0% and 53.7% from the caregivers' perspectives, and by between 81.6% and 106.7% from the adolescents' perspectives. Harsh discipline was reduced entirely from both the caregiver and adolescent perspectives. Family financial coping was also shown to have improved over the course of the study.

Caregiver depression and loneliness was significantly reduced between the two time points, however the frequency at which caregivers felt that everything required effort was increased. The amount which caregivers felt they supported their adolescent's schoolwork increased between the two measurement points, and this is consistent with the responses from the adolescents' perspectives. Lastly, the proportion of both caregivers and adolescents who agreed with the necessity of corporal punishment as a discipline device was significantly reduced.

From the caregivers' perspectives, the improvements to positive parenting and child behaviour were more extreme for those caregivers who reported that the following factors were present in their household: substance abuse; physical or verbal arguments; living with a sick child or sick adult. These improvements were less drastic for those who reported living with someone who suffered from, or had succumbed to, HIV or TB. From the adolescents' perspectives, the improvements in positive parenting were relatively larger for those who reported physical or verbal arguments at home. The improvements in child behaviour were larger for those who reported substance abuse or physical/verbal arguments at home. For both outcomes, the changes were less for those who reported living with someone who suffered from, or had succumbed to, HIV or TB.

Given the homogeneity in response at follow-up, moderators for the change in harsh discipline score could not be estimated reliably. However, risk factors for harsh discipline at baseline were assessed. Harsh discipline, from the caregiver perspective, was generally higher for caregivers who reported substance abuse, fighting at home, and living with a sick child or adult, and generally lower for those who reported the HIV/TB factor. From the adolescent perspective, baseline harsh discipline was generally higher for those with female caregivers, and those who reported substance abuse or fighting at home.

Overall, these results imply that caregivers/adolescents who report substance abuse, fighting at home, or living with a sick child or adult, generally experienced greater improvements with respect to positive parenting and child behaviour. These results, however, may be misleading. Both of these scores exhibited very little variation at follow-up and, as was seen in the baseline regression results, the moderators of the change in scores over time mostly reflect the differences between sample strata at baseline.

Many of the models produced residuals which suggested that the underlying assumptions were not satisfied, and thus the coefficient estimates were likely biased. In this study, however, the direction of change may be considered more important than the magnitude of change. The cumulative link models thus confirmed the significance, and direction, of change in scores over time.

With respect to the secondary objective, the item reliability analysis suggested that the only scale which exhibited satisfactory internal consistency was the adolescent harsh discipline scale when measured at baseline. For the scales that had sufficient items, this mostly reflects the lack of variation rather than the design of the surveys and one may expect to obtain different results if the same surveys are used in a different study region.

6.2 Limitations

Interpretation of the results in this report must be done with caution. Firstly, the scaled-up study design results in certain inherent limitations. The study is conducted with no control group, and thus any attempt to interpret the results with respect to a causal relationship would be ill-advised. Another concern relates to the design of the assessment questionnaires, which were significantly reduced from their original form. Some of the sections had too few items to conduct an item reliability analysis. The item reliability analysis revealed that, even where there were a sufficient number of questions, none of the scales were internally consistent, excepting the adolescent harsh discipline scale at baseline. These scales thus cannot be said to represent the underlying factor which they attempt to measure. Beyond this, other structural changes were implemented such as condensing positive parenting and parental monitoring/supervision into one category.

In terms of the moderator effects, confounding was not explored and the effects were considered independently. This means that the size of effect of a moderator reported in this paper is likely not representative of it's partial effect. Furthermore, some of these demographic factors must be interpreted with caution, since they reflect self-reporting of a factor and thus revelation bias may be a cause for concern. This limitation could be addressed in future research by clustering the participants according to demographic factors, and using these clusters as moderators instead to more broadly stratify the sample.

There were also several limitations imposed by the nature of the data received. The matching procedure was impeded by the lack of a unique identifier linking caregiver to adolescent, which resulted in a dramatic reduction in sample size. Furthermore, where a caregiver attended the program with more than one adolescent, they had identical responses for each adolescent. This suggests that only one survey form had been completed by these caregivers, where ideally they would have completed one per adolescent. This would certainly be another source of bias in the study. Furthermore, neither facilitator nor attendance rates were recorded. The effects of facilitator clusters could thus not be isolated, and the relationship between attendance and the program's impact could not be assessed. These problems were worsened by the lack of reliable communication lines between the data capturers and the analysts. This may be the nature of scaled-up, low-cost studies, however the analysis could be significantly improved by maintaining open communication lines.

Another limitation to the analysis here was that assessment forms were only completed at baseline and at one follow-up point. Including another measurement point, perhaps another 12 months after the follow-up assessment, would provide an opportunity to assess the long-term effectiveness of the program. This is, however, subject to additional costs which is perhaps not realistic given the low-cost study design. Furthermore, the follow-up assessments were generally completed at the same time by caregiver and adolescent. If the assessments were staggered, then the analysis could incorporate a structural equation model to assess a mediation effect (i.e. whether the program affects caregiver outcomes, which then affect the adolescent outcomes).

Lastly, and perhaps most significantly, as revealed in the exploratory analysis the distribution of responses for both cohorts casts some doubt over the validity of the data. While the modeling results reflect what is seen in the data accurately, if the data does not mirror the subjects' experiences through their participation in the study then neither does the analysis.

Given the inclusion criteria, the main findings are at most generalisable to adolescents vulnerable to HIV and AIDS, however - given the problems experienced with matching and the concerns regarding the validity of the data - the results are likely not generalisable to any broader population.

6.3 Conclusion

This report aimed to analyse data from the implementation of PLH for Adolescents among a cohort of 295 South Sudanese adolescents and their caregivers. This was approached with primary aims of: (1) quantifying the changes in socio-psychological indicators; and (2) assessing the internal consistency of these psychometric scales. A secondary aim was defined to identify obstacles to the program's scalability.

The first primary aim was evaluated through the use of generalised linear and cumulative link mixed effect models. The cumulative link mixed models revealed that there was a significant improvement in positive parenting, child behaviour, harsh discipline, and family financial coping scores over time. In addition, items relating to caregiver depression and loneliness, opinions towards corporal punishment, and parental support of education were shown to have significantly improved from baseline to follow-up. The generalised linear mixed effect models quantified the magnitude of these changes, and assessed whether they differed between sample strata. The changes in positive parenting, child behaviour and harsh discipline were generally more extreme for caregivers/adolescents who reported: substance abuse or fighting in the household; or living with a severely sick adult or child. On the contrary, the changes were generally less extreme for those who reported living with someone who has suffered from, or succumbed to, HIV or TB. It was shown, however, that the moderators of the change in scores were mostly due to differences in scores at baseline rather than differences in the program's effectiveness between sample strata.

The second primary aim was assessed through a confirmatory factor analysis, and calculation of Cronbach's alpha. These methods showed that only the harsh discipline score, when measured from the adolescents' perspectives at baseline, exhibited acceptable internal consistency. To a lesser extent, there was some evidence that the caregivers' positive parenting score and the adolescents' child behaviour scores exhibited borderline acceptable internal consistency. Mostly the poor results from the item reliability analysis are attributable to the lack of variation seen in the responses.

Throughout the analysis, attention was brought to the questionable response patterns seen in the data. The homogeneity in responses, particularly among the caregiver data at both baseline and follow-up, cast significant doubts over the validity of the data. Further to this, the lack of a unique identifier linking caregiver to adolescent resulted in a reduction in sample size. The lack of communication lines between data capturers and researchers meant that these problems could not be elaborated on, and is one of the obstacles to scalability identified in this report.

Overall, the patterns seen in the data give evidence to the program's effectiveness in improving adolescent-caregiver outcomes relating to positive parenting, child behaviour, harsh discipline, and family financial coping. It is possible, however, that the data does not reflect the outcomes of the participants. This provides a significant limitation assessing the program's impact. Results must be interpreted with caution and are certainly not generalised to a broader population.

References

Bartoń, K. (2022). MuMIn: Multi-Model Inference. R package version 1.46.0. https://CRAN.R-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Canavera, M., Lanning, K., Polin, K. & Stark, L. (2016). 'And then they left': Challenges to child protection systems strengthening in South Sudan. Child Soc, 30: 356-368. https://doi.org/10.1111/chso.12172

Christensen, R. H. B. (2019). ordinal - Regression Models for Ordinal Data. R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal.

Cluver, L.D., Meinck, F., Steinert, J.I., et al. (2018). Parenting for Lifelong Health: a pragmatic cluster randomised controlled trial of a non-commercialised parenting programme for adolescents and their families in South AfricaBMJ Global Health ;3:e000539.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. psychometrika, 16(3), 297-334.

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software, 64(4), 1-34. DOI 10.18637/jss.v064.i04.

Fox, J. & Weisberg, S. (2019). An R Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Haberman, S. (2006). Adaptive quadrature for item response models. Educational Testing Service, Princeton, New Jersey.

Hartig, F. (2022). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.6 ¡https://CRAN.R-project.org/package=DHARMa¿.

Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. Communications in Statistics-theory and Methods, 6(9), 813-827.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal, 6(1), 1-55.

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0, https://CRAN.R-project.org/package=ggpubr.

Kline, R. B. (2011). Convergence of structural equation modeling and multilevel modeling.

Mugo, N., Zwi, A. B., Botfield, J. R., & Steiner, C. (2015). Maternal and Child Health in South Sudan: Priorities for the Post-2015 Agenda. SAGE Open, 5(2). https://doi.org/10.1177/2158244015581190

Neuwirth, E. (2022). RColorBrewer: ColorBrewer Palettes. R package version 1.1-3, https://CRAN.R-project.org/package=RColorBrewer.

Ning, Y., Ho, P.J., Støer, N.C., Lim, K.K., Wee, H.L., Hartman, M., Reilly, M., Tan, C.S. (2021). A New Procedure to Assess When Estimates from the Cumulative Link Model Can Be Interpreted as Differences for Ordinal Scales in Quality of Life Studies. Clin Epidemiol ;13:53-65. https://doi.org/10.2147/CLEP.S288801

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Revelle, W. (2022) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version 2.2.5.

Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses, Journal of Statistical Software, 17 (5), 1-25. https://doi.org/10.18637/jss.v017.i05

Rizopoulos, D. (2022). GLMMadaptive: Generalized Linear Mixed Models using Adaptive Gaussian Quadrature. R package version 0.8-5, https://CRAN.R-project.org/package=GLMMadaptive.

Ronnie, G. (2021). Parenting for Lifelong Health – Scale-up of Parenting Evaluation Research [SUPER]: Background and Overview.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Samuel, B. (2020). Evaluating the effect of Orphans and Vulnerable Children (OVC) program accelerators across multiple Sustainable Development Goals (SDGs) -related survival outcomes for adolescents impacted by HIV and AIDS in Nimule, Eastern Equatoria State, South Sudan - Research concept note.

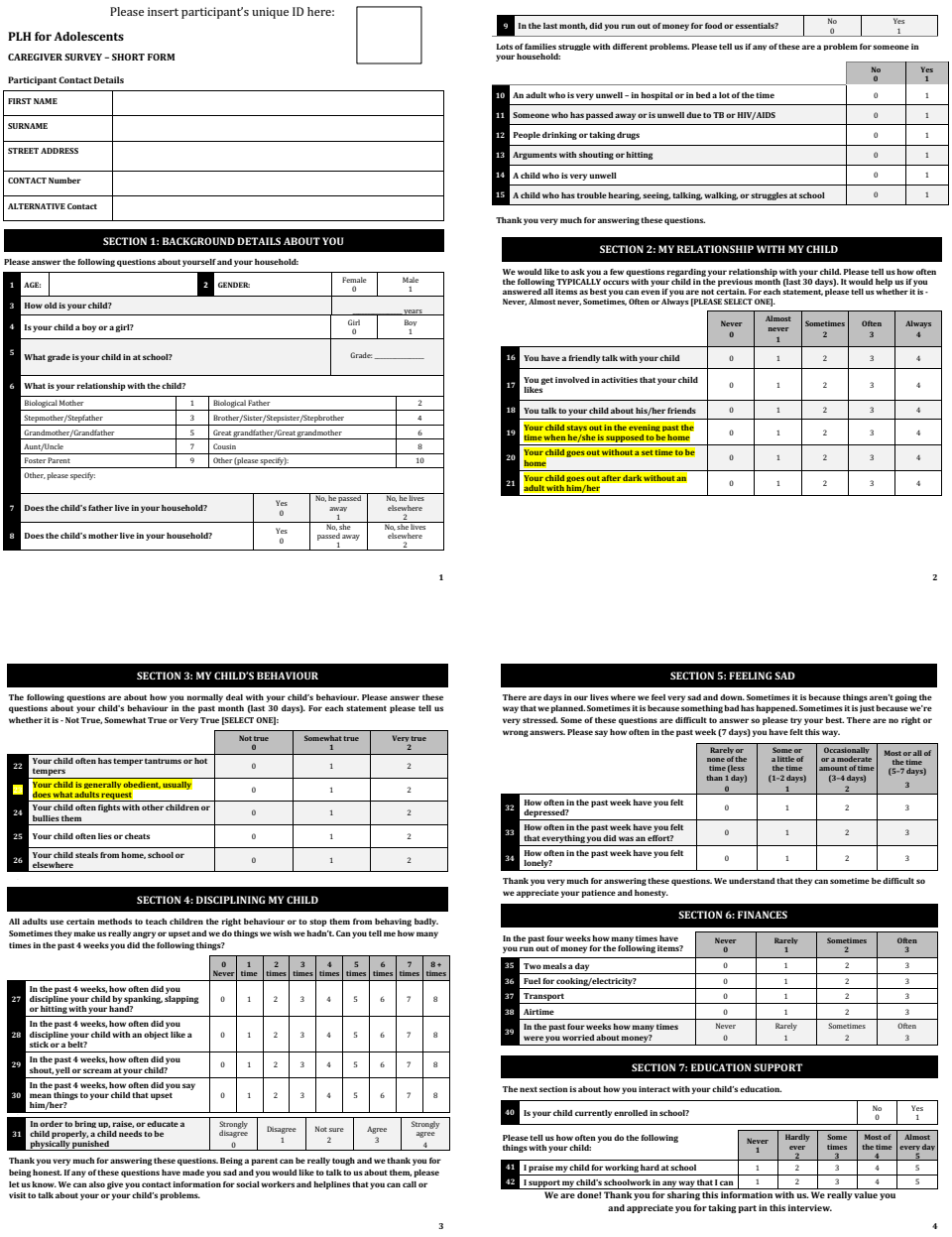
Ward, C. L., Wessels, I. M., Lachman, J. M., Hutchings, J., Cluver, L. D., Kassanjee, R., Nhapi, R., Little, F., & Gardner, F. (2020). Parenting for Lifelong Health for Young Children: a randomized controlled trial of a parenting program in South Africa to prevent harsh parenting and child conduct problems. Journal of child psychology and psychiatry, and allied disciplines, 61(4), 503–512. https://doi.org/10.1111/jcpp.13129

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. https://ggplot2.tidyverse.org

Wickham, H., François, R. & Henry, L. (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.8. https://CRAN.R-project.org/package=dplyr

Appendix

Appendix A: Survey Forms



Please insert participant's unique ID here:

PLH for Adolescents

CAREGIVER SURVEY – SHORT FORM

Participant Contact Details

FIRST NAME	
SURNAME	
STREET ADDRESS	
CONTACT Number	
ALTERNATIVE Contact	

SECTION 1: BACKGROUND DETAILS ABOUT YOU

Please answer the following questions about yourself and your household:

1	AGE:		2	GENDER:	Female 0 / Male 1
3	How old is your child?				_____ years
4	Is your child a boy or a girl?				Girl 0 / Boy 1
5	What grade is your child in at school?				Grade: _____
6	What is your relationship with the child?				
	Biological Mother	1	Biological Father		2
	Stepmother/Stepfather	3	Brother/Sister/Stepsister/Stepbrother		4
	Grandmother/Grandfather	5	Great grandfather/Great grandmother		6
	Aunt/Uncle	7	Cousin		8
	Foster Parent	9	Other (please specify):		10
	Other, please specify:				
7	Does the child's father live in your household?	Yes 0	No, he passed away 1	No, he lives elsewhere 2	
8	Does the child's mother live in your household?	Yes 0	No, she passed away 1	No, she lives elsewhere 2	

1

9	In the last month, did you run out of money for food or essentials?	No 0	Yes 1

Lots of families struggle with different problems. Please tell us if any of these are a problem for someone in your household:

		No 0	Yes 1
10	An adult who is very unwell – in hospital or in bed a lot of the time	0	1
11	Someone who has passed away or is unwell due to TB or HIV/AIDS	0	1
12	People drinking or taking drugs	0	1
13	Arguments with shouting or hitting	0	1
14	A child who is very unwell	0	1
15	A child who has trouble hearing, seeing, talking, walking, or struggles at school	0	1

Thank you very much for answering these questions.

SECTION 2: MY RELATIONSHIP WITH MY CHILD

We would like to ask you a few questions regarding your relationship with your child. Please tell us how often the following TYPICALLY occurs with your child in the previous month (last 30 days). It would help us if you answered all items as best you can even if you are not certain. For each statement, please tell us whether it is - Never, Almost never, Sometimes, Often or Always [PLEASE SELECT ONE].

		Never 0	Almost never 1	Sometimes 2	Often 3	Always 4
16	You have a friendly talk with your child	0	1	2	3	4
17	You get involved in activities that your child likes	0	1	2	3	4
18	You talk to your child about his/her friends	0	1	2	3	4
19	Your child stays out in the evening past the time when he/she is supposed to be home	0	1	2	3	4
20	Your child goes out without a set time to be home	0	1	2	3	4
21	Your child goes out after dark without an adult with him/her	0	1	2	3	4

2

SECTION 3: MY CHILD'S BEHAVIOUR

The following questions are about how you normally deal with your child's behaviour. Please answer these questions about your child's behaviour in the past month (last 30 days). For each statement please tell us whether it is - Not True, Somewhat True or Very True [SELECT ONE]:

		Not true 0	Somewhat true 1	Very true 2
22	Your child often has temper tantrums or hot tempers	0	1	2
23	Your child is generally obedient, usually does what adults request	0	1	2
24	Your child often fights with other children or bullies them	0	1	2
25	Your child often lies or cheats	0	1	2
26	Your child steals from home, school or elsewhere	0	1	2

SECTION 4: DISCIPLINING MY CHILD

All adults use certain methods to teach children the right behaviour or to stop them from behaving badly. Sometimes they make us really angry or upset and we do things we wish we hadn't. Can you tell me how many times in the past 4 weeks you did the following things?

		0 Never	1 time	2 times	3 times	4 times	5 times	6 times	7 times	8+ times
27	In the past 4 weeks, how often did you discipline your child by spanking, slapping or hitting with your hand?	0	1	2	3	4	5	6	7	8
28	In the past 4 weeks, how often did you discipline your child with an object like a stick or a belt?	0	1	2	3	4	5	6	7	8
29	In the past 4 weeks, how often did you shout, yell or scream at your child?	0	1	2	3	4	5	6	7	8
30	In the past 4 weeks, how often did you say mean things to your child that upset him/her?	0	1	2	3	4	5	6	7	8

31	In order to bring up, raise, or educate a child properly, a child needs to be physically punished	Strongly disagree 0	Disagree 1	Not sure 2	Agree 3	Strongly agree 4

Thank you very much for answering these questions. Being a parent can be really tough and we thank you for being honest. If any of these questions have made you sad and you would like to talk to us about them, please let us know. We can also give you contact information for social workers and helplines that you can call or visit to talk about your or your child's problems.

3

SECTION 5: FEELING SAD

There are days in our lives where we feel very sad and down. Sometimes it is because things aren't going the way that we planned. Sometimes it is because something bad has happened. Sometimes it is just because we're very stressed. Some of these questions are difficult to answer so please try your best. There are no right or wrong answers. Please say how often in the past week (7 days) you have felt this way.

		Rarely or none of the time (less than 1 day) 0	Some or a little of the time (1–2 days) 1	Occasionally or a moderate amount of time (3–4 days) 2	Most or all of the time (5–7 days) 3
32	How often in the past week have you felt depressed?	0	1	2	3
33	How often in the past week have you felt that everything you did was an effort?	0	1	2	3
34	How often in the past week have you felt lonely?	0	1	2	3

Thank you very much for answering these questions. We understand that they can sometime be difficult so we appreciate your patience and honesty.

SECTION 6: FINANCES

In the past four weeks how many times have you run out of money for the following items?		Never 0	Rarely 1	Sometimes 2	Often 3
35	Two meals a day	0	1	2	3
36	Fuel for cooking/electricity?	0	1	2	3
37	Transport	0	1	2	3
38	Airtime	0	1	2	3
39	In the past four weeks how many times were you worried about money?	Never 0	Rarely 1	Sometimes 2	Often 3

SECTION 7: EDUCATION SUPPORT

The next section is about how you interact with your child's education.

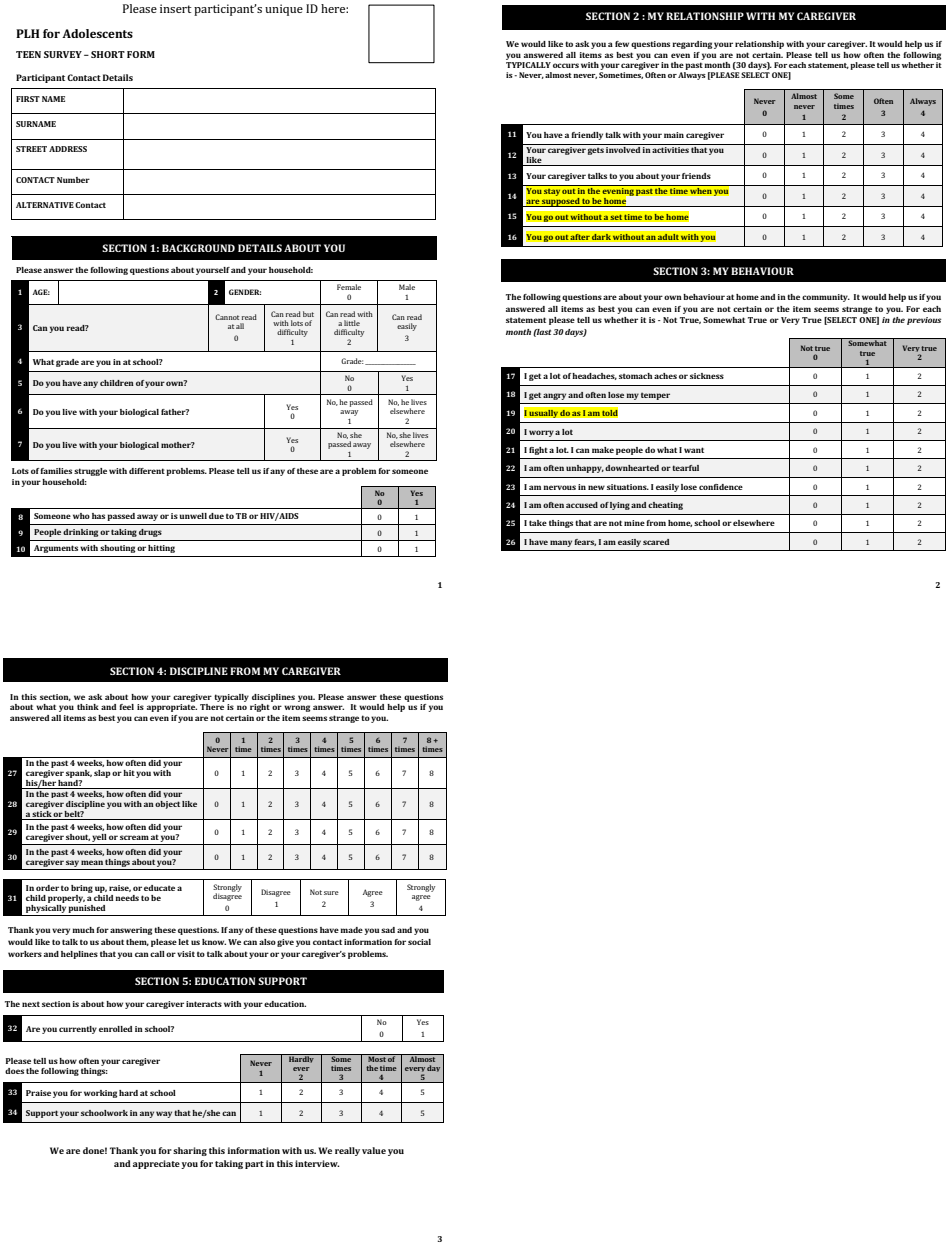
		No 0	Yes 1
40	Is your child currently enrolled in school?		

Please tell us how often you do the following things with your child:		Never 1	Hardly ever 2	Some times 3	Most of the time 4	Almost every day 5
41	I praise my child for working hard at school	1	2	3	4	5
42	I support my child's schoolwork in any way that I can	1	2	3	4	5

We are done! Thank you for sharing this information with us. We really value you and appreciate you for taking part in this interview.

4

Figure 24: Caregiver Assessment Forms



Please insert participant's unique ID here:

PLH for Adolescents

TEEN SURVEY – SHORT FORM

Participant Contact Details

FIRST NAME	
SURNAME	
STREET ADDRESS	
CONTACT Number	
ALTERNATIVE Contact	

SECTION 1: BACKGROUND DETAILS ABOUT YOU

Please answer the following questions about yourself and your household:

1	AGE:	2	GENDER:	Female 0	Male 1
3	Can you read?	Cannot read at all 0	Can read but with lots of difficulty 1	Can read with a little difficulty 2	Can read easily 3
4	What grade are you in at school?			Grade: _____	
5	Do you have any children of your own?			No 0	Yes 1
6	Do you live with your biological father?		Yes 0	No, he passed away 1	No, he lives elsewhere 2
7	Do you live with your biological mother?		Yes 0	No, she passed away 1	No, she lives elsewhere 2

Lots of families struggle with different problems. Please tell us if any of these are a problem for someone in your household:

		No 0	Yes 1
8	Someone who has passed away or is unwell due to TB or HIV/AIDS	0	1
9	People drinking or taking drugs	0	1
10	Arguments with shouting or hitting	0	1

1

SECTION 2 : MY RELATIONSHIP WITH MY CAREGIVER

We would like to ask you a few questions regarding your relationship with your caregiver. It would help us if you answered all items as best you can even if you are not certain. Please tell us how often the following TYPICALLY occurs with your caregiver in the past month (30 days). For each statement, please tell us whether it is - Never, almost never, Sometimes, Often or Always [PLEASE SELECT ONE]

		Never 0	Almost never 1	Some times 2	Often 3	Always 4
11	You have a friendly talk with your main caregiver	0	1	2	3	4
12	Your caregiver gets involved in activities that you like	0	1	2	3	4
13	Your caregiver talks to you about your friends	0	1	2	3	4
14	You stay out in the evening past the time when you are supposed to be home	0	1	2	3	4
15	You go out without a set time to be home	0	1	2	3	4
16	You go out after dark without an adult with you	0	1	2	3	4

SECTION 3: MY BEHAVIOUR

The following questions are about your own behaviour at home and in the community. It would help us if you answered all items as best you can even if you are not certain or the item seems strange to you. For each statement please tell us whether it is - Not True, Somewhat True or Very True [SELECT ONE] ***in the previous month (last 30 days)***

		Not true 0	Somewhat true 1	Very true 2
17	I get a lot of headaches, stomach aches or sickness	0	1	2
18	I get angry and often lose my temper	0	1	2
19	I usually do as I am told	0	1	2
20	I worry a lot	0	1	2
21	I fight a lot. I can make people do what I want	0	1	2
22	I am often unhappy, downhearted or tearful	0	1	2
23	I am nervous in new situations. I easily lose confidence	0	1	2
24	I am often accused of lying and cheating	0	1	2
25	I take things that are not mine from home, school or elsewhere	0	1	2
26	I have many fears, I am easily scared	0	1	2

2

SECTION 4: DISCIPLINE FROM MY CAREGIVER

In this section, we ask about how your caregiver typically disciplines you. Please answer these questions about what you think and feel is appropriate. There is no right or wrong answer. It would help us if you answered all items as best you can even if you are not certain or the item seems strange to you.

		0 Never	1 time	2 times	3 times	4 times	5 times	6 times	7 times	8+ times
27	In the past 4 weeks, how often did your caregiver spank, slap or hit you with his/her hand?	0	1	2	3	4	5	6	7	8
28	In the past 4 weeks, how often did your caregiver discipline you with an object like a stick or belt?	0	1	2	3	4	5	6	7	8
29	In the past 4 weeks, how often did your caregiver shout, yell or scream at you?	0	1	2	3	4	5	6	7	8
30	In the past 4 weeks, how often did your caregiver say mean things about you?	0	1	2	3	4	5	6	7	8

31	In order to bring up, raise, or educate a child properly, a child needs to be physically punished	Strongly disagree 0	Disagree 1	Not sure 2	Agree 3	Strongly agree 4

Thank you very much for answering these questions. If any of these questions have made you sad and you would like to talk to us about them, please let us know. We can also give you contact information for social workers and helplines that you can call or visit to talk about your or your caregiver's problems.

SECTION 5: EDUCATION SUPPORT

The next section is about how your caregiver interacts with your education.

		No 0	Yes 1
32	Are you currently enrolled in school?		

	Please tell us how often your caregiver does the following things:	Never 1	Hardly ever 2	Some times 3	Most of the time 4	Almost every day 5
33	Praise you for working hard at school	1	2	3	4	5
34	Support your schoolwork in any way that he/she can	1	2	3	4	5

We are done! Thank you for sharing this information with us. We really value you and appreciate you for taking part in this interview.

3

Figure 25: Adolescent Assessment Forms

Appendix B: Response Rates

Table 24: Demographic Characteristics Response Rates (full sample)

Variable	**Caregiver Report**	**Adolescent Report**
	Responses rate % (N=290)	Response rate % (N=295)
Sociodemographic Characteristics		
Age	98.3	100
Female	100	100
Adolescent has children of their own	-	98.0
Enrolled in school	95.2	96.9
Adolescent in high school	99.7	96.6
Can read with little to no difficulty	-	98.0
Caregiver is biological parent	100	-
Caregiver is grandparent	100	-
Caregiver is other	100	-
Household characteristics		
Father lives in household	99.7	99.7
Father deceased	99.7	99.7
Mother lives in household	99.3	100
Mother deceased	99.3	100
Ran out of money for essentials in last month	100	-
Severely unwell adult in household	100	-
Severely unwell child in household	100	-
Disabled child in household	100	-
Household member unwell/deceased from HIV/TB	100	99.7
Substance abuse	99.7	99.0
Physical/verbal arguments	100	98.6

Table 25: Outcome Measures Response Rates (full sample)

Domain	Variable Name	Caregiver Response Rate % (N=290)		Adolescent Response Rate % (N=295)	
		Baseline	Follow-up	Baseline	Follow-up
Involved parenting & parental monitoring					
	APQ_talk	100	100	99.7	100
	APQ_involved	100	100	100	100
	APQ_talk_friends	100	100	99.7	100
	APQ_evening	99.7	100	100	100
	APQ_time	99.7	100	100	99.7
	APQ_dark	100	100	98.3	100
Child behaviour					
	SDQ_tantrums	100	100	99.0	100
	SDQ_obedient	100	100	100	100
	SDQ_fights	100	100	99.3	100
	SDQ_lies	100	100	99.3	100
	SDQ_steals	99.3	100	99.7	100
	SDQ_headaches	-	-	99.0	100
	SDQ_worried	-	-	97.6	100
	SDQ_unhappy	-	-	99.7	100
	SDQ_new	-	-	99.3	99.7
	SDQ_fears	-	-	99.3	99.7
Harsh discipline					
	ICAST_spank	100	100	99.3	100
	ICAST_object	100	100	99.7	100
	ICAST_scream	99.7	100	99.7	100
	ICAST_upset	100	100	99.0	100
Child discipline					
	MICS_physical_punishment	100	100	98.0	100
Parental Depression					
	CESD_depressed	100	100	-	-
	CESD_effort	100	100	-	-
	CESD_lonely	100	99.7	-	-
Economic strengthening					
	FFC_meat	100	100	-	-
	FFC_electricity	100	100	-	-
	FFC_transport	100	100	-	-
	FFC_airtime	100	97.6	-	-
	FFC_worried	100	100	-	-
Parent Support of Education	Parent Support of Education				
	PSE_praise	100	100	98.6	100
	PSE_others	100	100	98.6	100

Appendix C: Additional Visualisation of Survey Responses

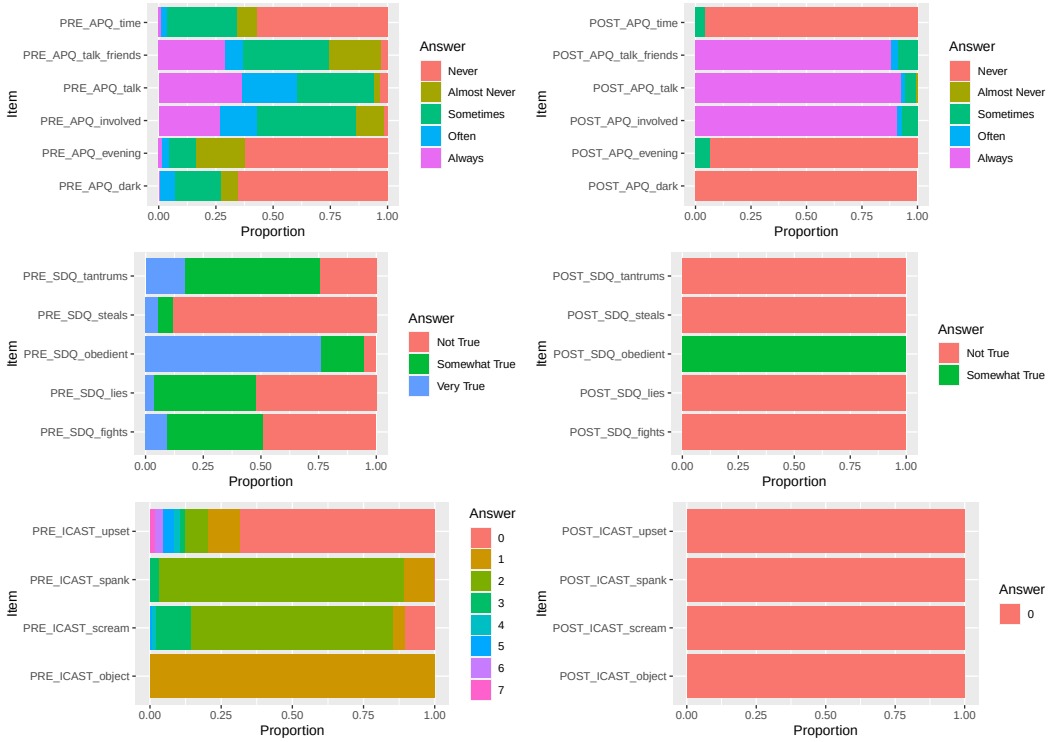


Figure 26: Stacked Barplots of Caregiver Likert Responses: APQ-ICAST

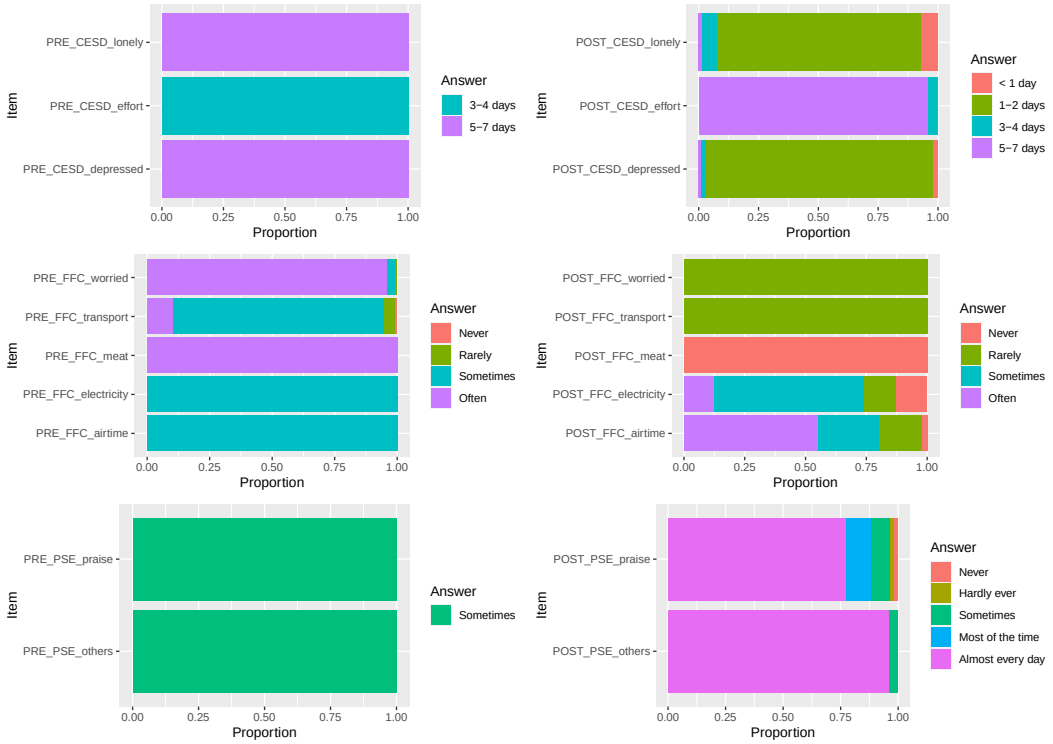


Figure 27: Stacked Barplots of Caregiver Likert Responses: CESD-PSE

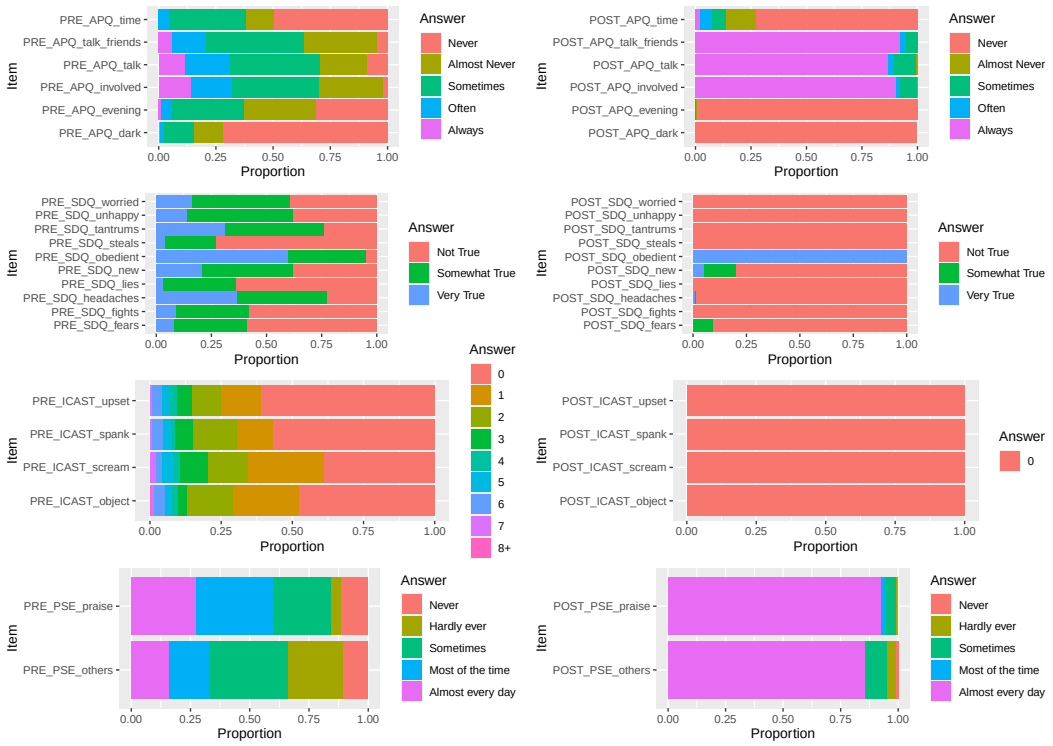
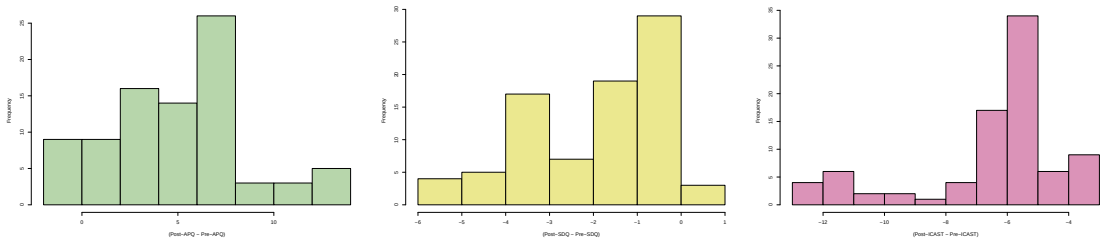
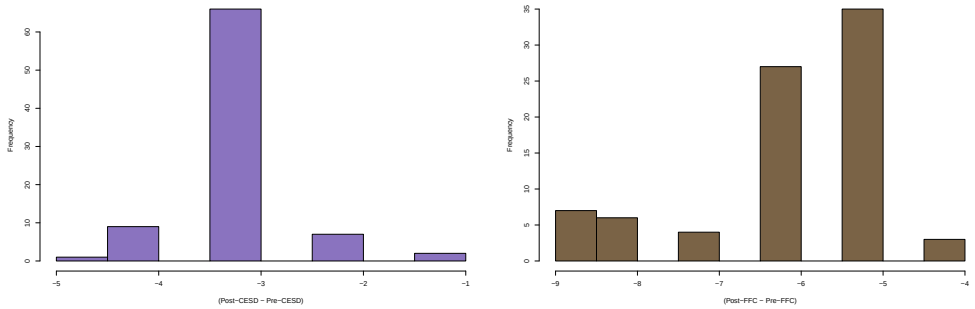


Figure 28: Stacked Barplots of Adolescent Likert Responses

Appendix D: Distribution of Scales after Differencing

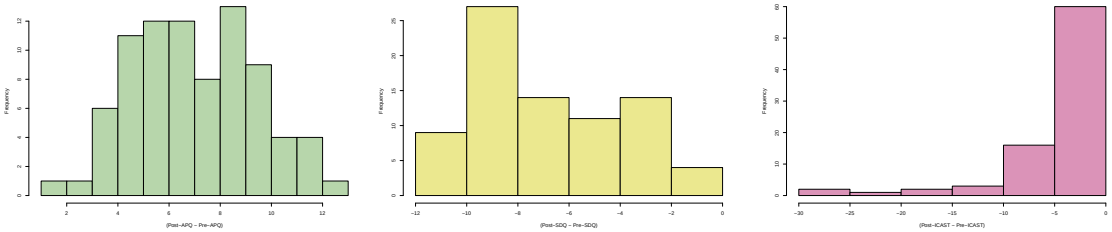


(a) APQ (b) SDQ (c) ICAST



(d) CESD (e) FFC

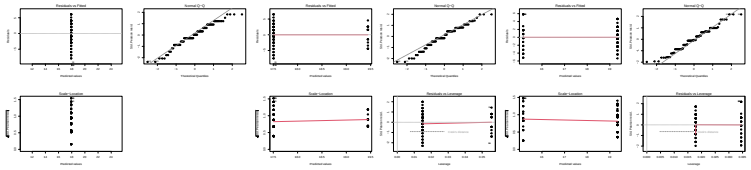
Figure 29: Histograms of Differenced Caregiver Likert Scales



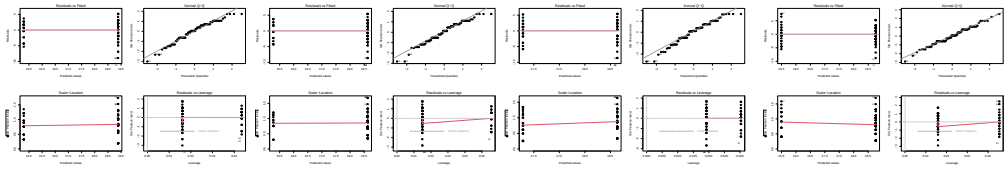
(a) APQ (b) SDQ (c) ICAST

Figure 30: Histograms of Differenced Adolescent Likert Scales

Appendix E: Residual Plots for Baseline Regression Models

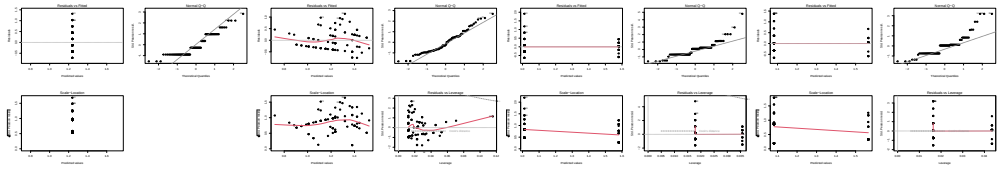


(a) Null (b) Model 1 (c) Model 2

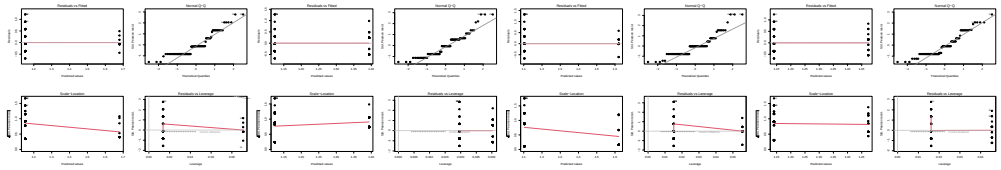


(d) Model 3 (e) Model 4 (f) Model 5 (g) Model 6

Figure 31: Residual Plots of Baseline APQ Regression Models: Caregivers

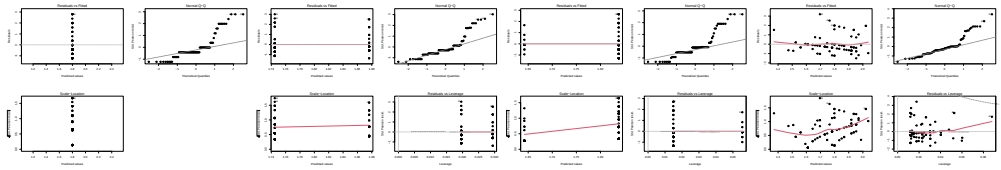


(a) Null (b) Model 1 (c) Model 2 (d) Model 3

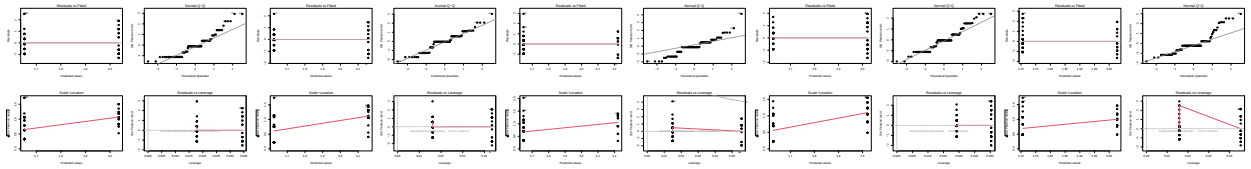


(e) Model 4 (f) Model 5 (g) Model 6 (h) Model 7

Figure 32: Residual Plots of Baseline SDQ Regression Models: Caregivers

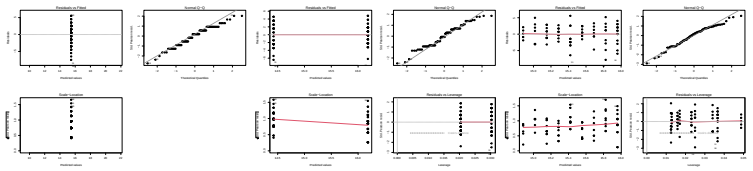


(a) Null (b) Model 1 (c) Model 2 (d) Model 3

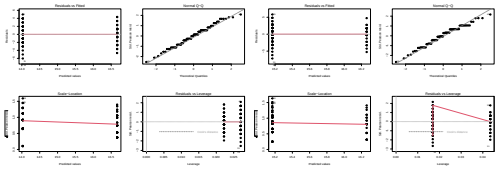


(e) Model 4 (f) Model 5 (g) Model 6 (h) Model 7 (i) Model 8

Figure 33: Residual Plots of Baseline ICAST Regression Models: Caregivers

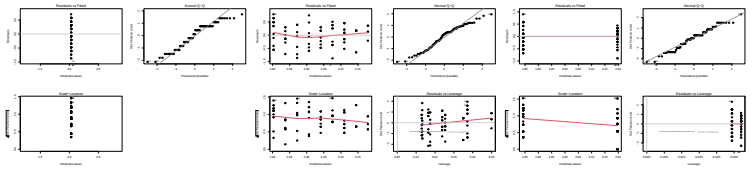


(a) Null (b) Model 1 (c) Model 2

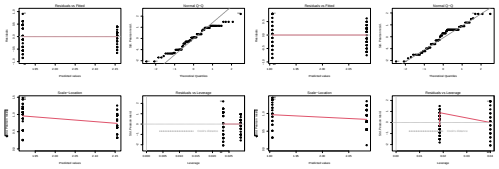


(d) Model 3 (e) Model 4

Figure 34: Residual Plots of Baseline APQ Regression Models: Adolescents

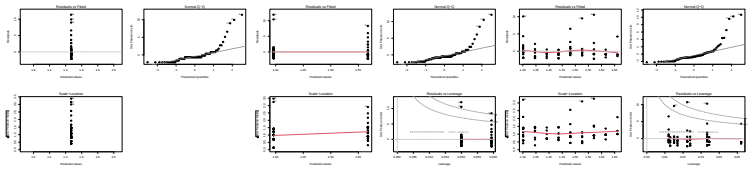


(a) Null (b) Model 1 (c) Model 2

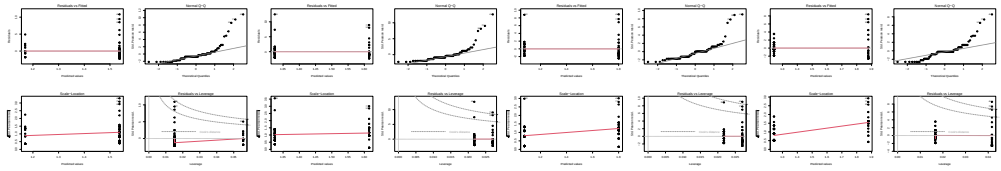


(d) Model 3 (e) Model 4

Figure 35: Residual Plots of Baseline SDQ Regression Models: Adolescents



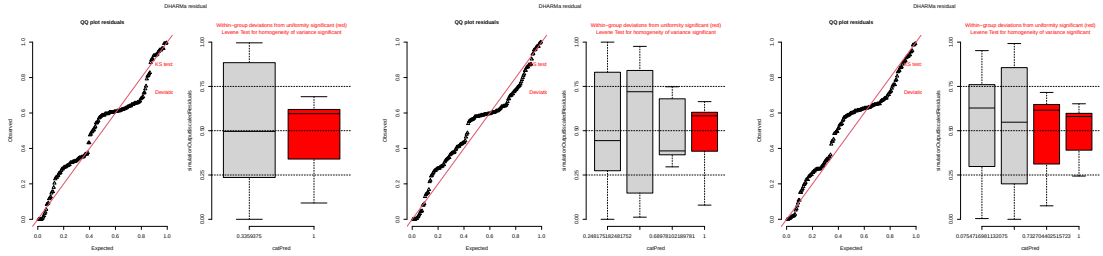
(a) Null (b) Model 1 (c) Model 2



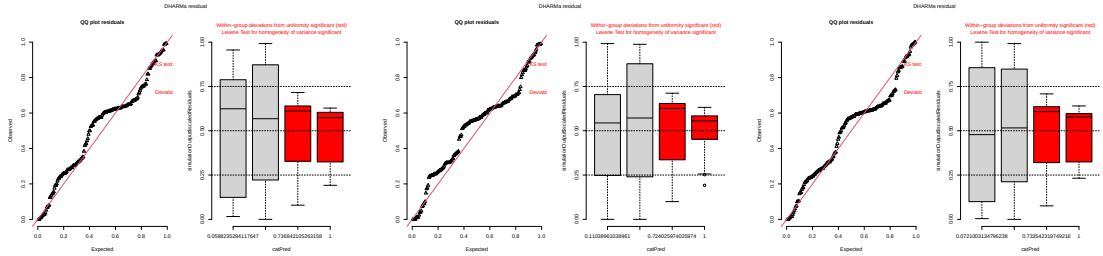
(d) Model 3 (e) Model 4 (f) Model 5 (g) Model 6

Figure 36: Residual Plots of Baseline ICAST Regression Models: Adolescents

Appendix F: Residual Plots for Mixed Effect Models

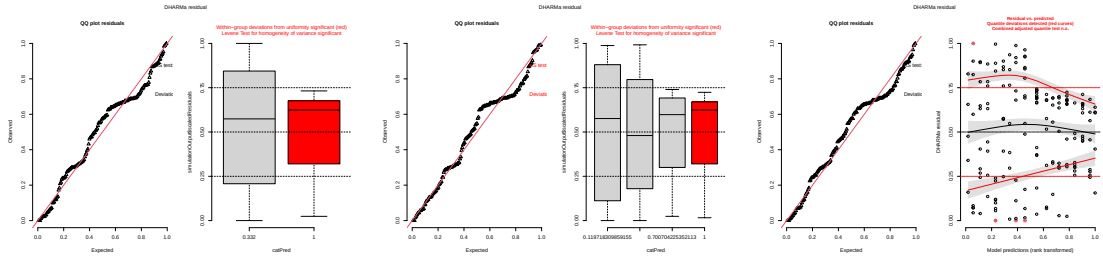


(a) Base model (b) Model 1 (c) Model 3

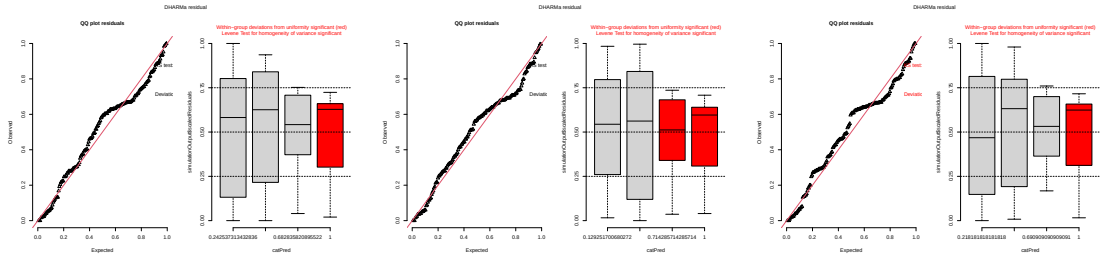


(d) Model 4 (e) Model 5 (f) Model 6

Figure 37: DHARMa Residuals: Caregiver APQ

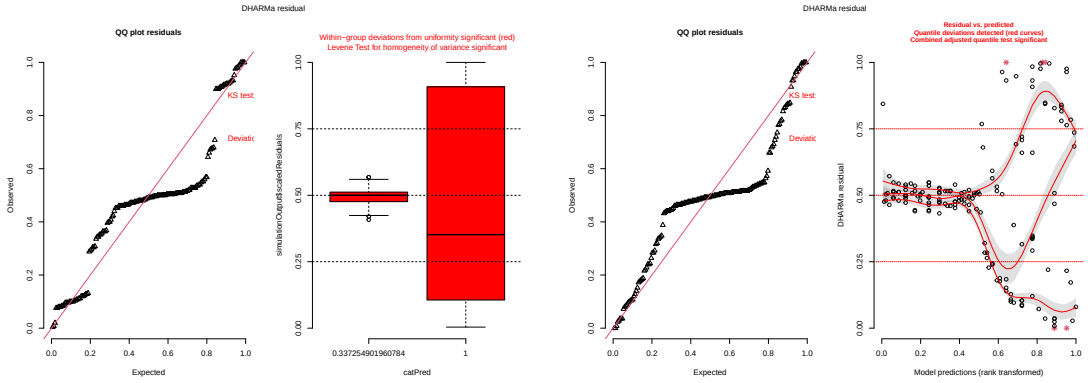


(a) Base model (b) Model 1 (c) Model 2



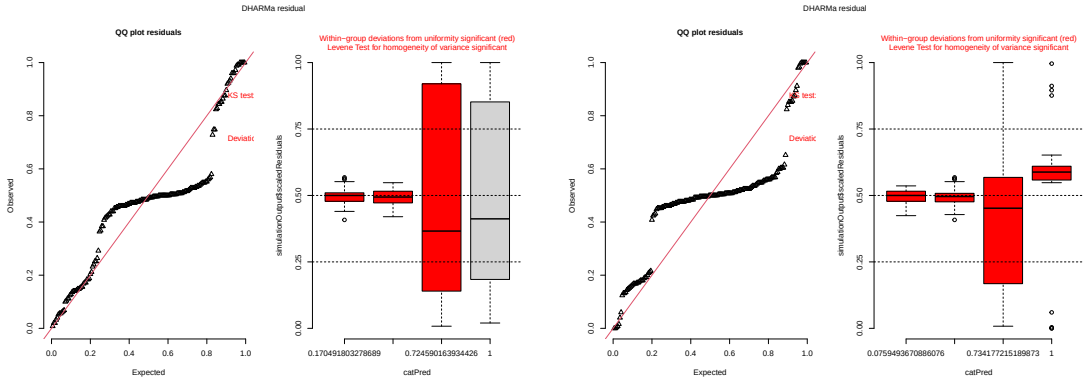
(d) Model 3 (e) Model 4 (f) Model 6

Figure 38: DHARMa Residuals: Adolescent APQ



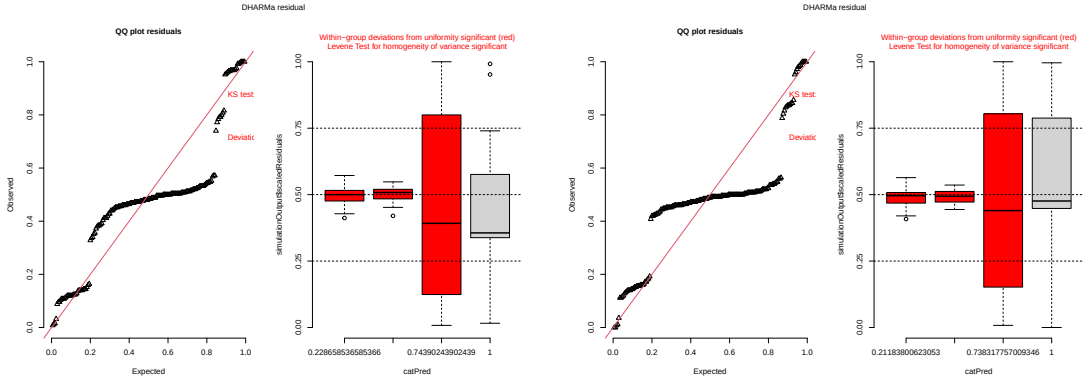
(a) Base model

(b) Model 1



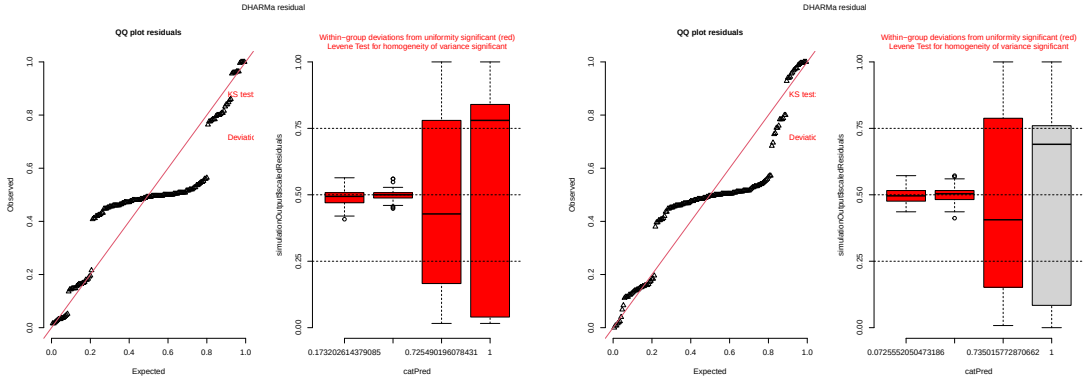
(c) Model 2

(d) Model 4



(e) Model 5

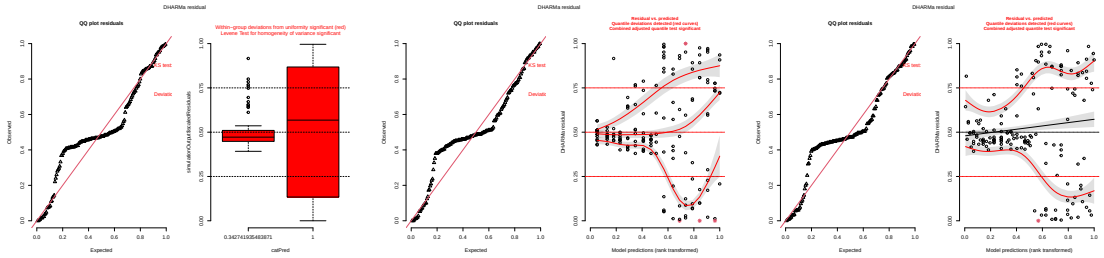
(f) Model 6



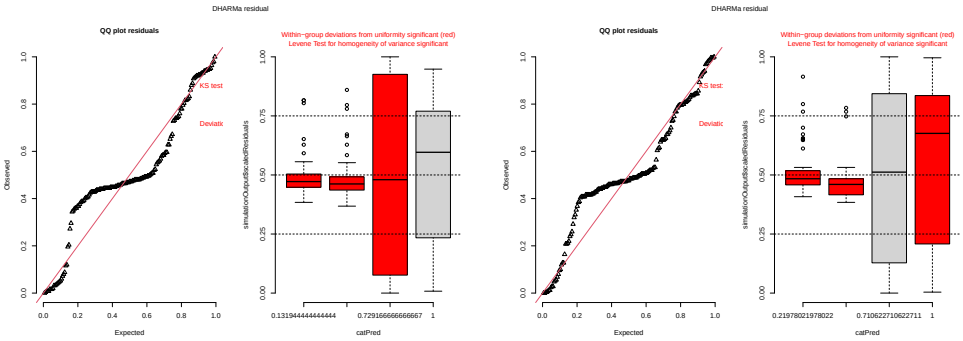
(g) Model 7

(h) Model 8

Figure 39: DHARMa Residuals: Caregiver SDQ

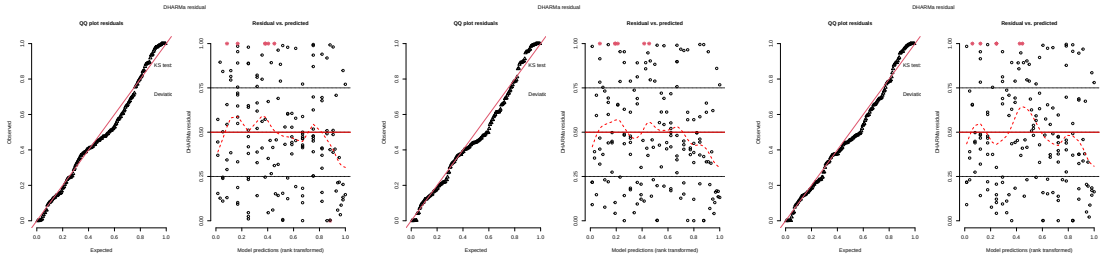


(a) Base model (b) Model 1 (c) Model 2

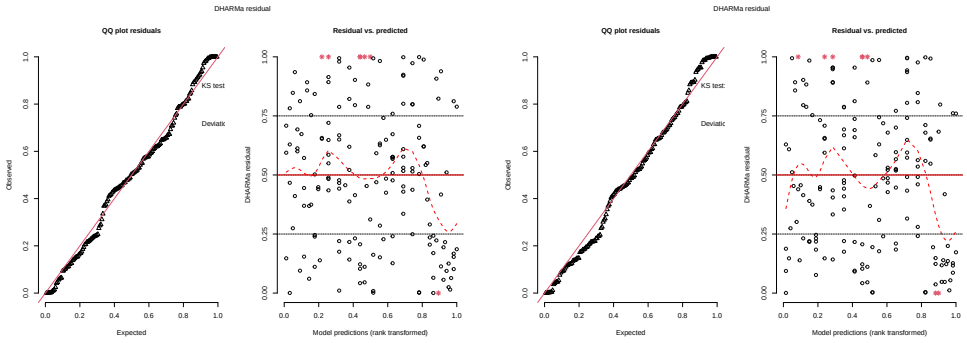


(d) Model 4 (e) Model 5

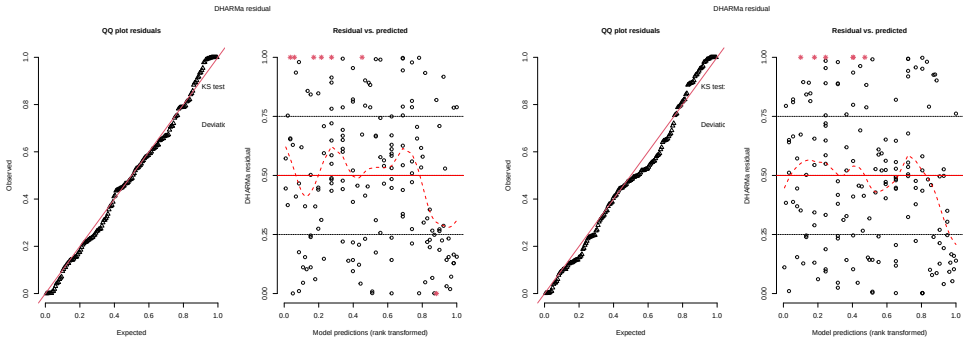
Figure 40: DHARMa Residuals: Adolescent SDQ



(a) Base model (b) Model 1 (c) Model 2

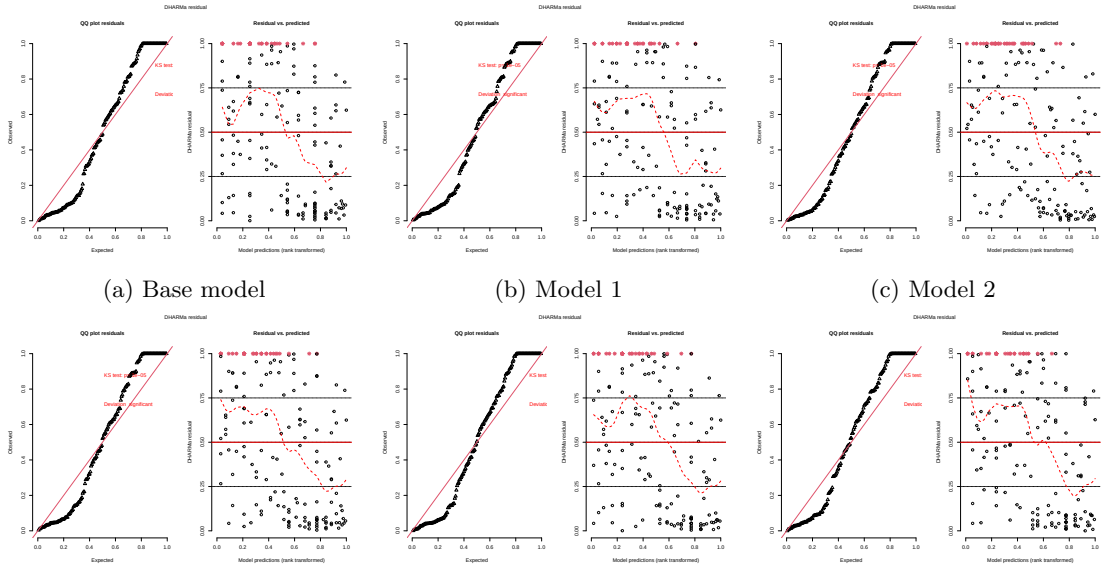


(d) Model 4 (e) Model 5



(f) Model 7 (g) Model 8

Figure 41: DHARMa Residuals: Caregiver ICAST

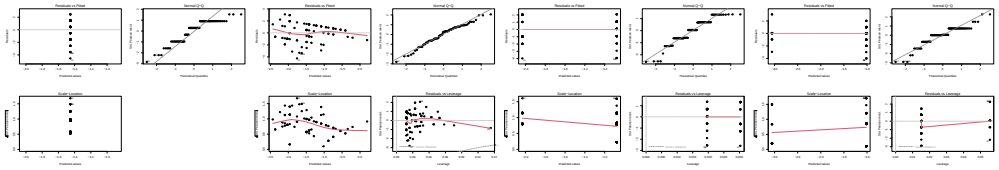


(a) Base model (b) Model 1 (c) Model 2

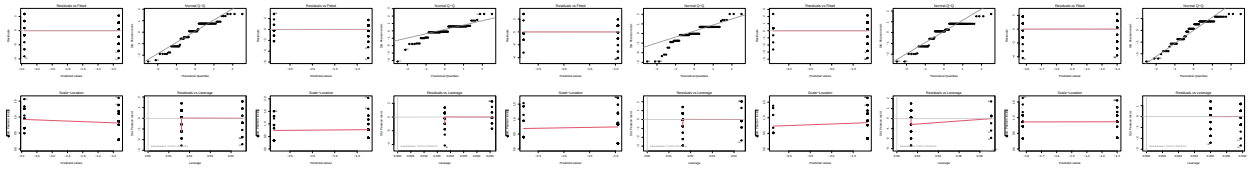
(d) Model 3 (e) Model 4 (f) Model 5

Figure 42: DHARMa Residuals: Adolescent ICAST

Appendix G: Residual Plots: Differenced Data Models

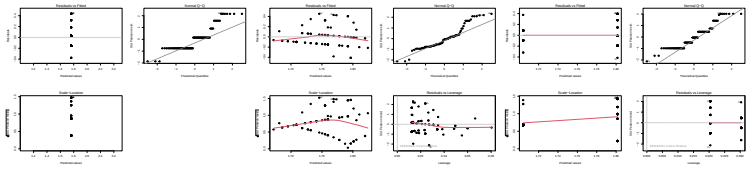


(a) Null (b) Model 1 (c) Model 2 (d) Model 3

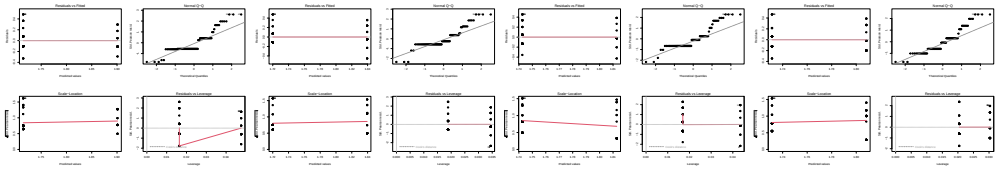


(e) Model 4 (f) Model 5 (g) Model 6 (h) Model 7 (i) Model 8

Figure 43: Residual Plots of Differenced SDQ GLMs : Caregivers

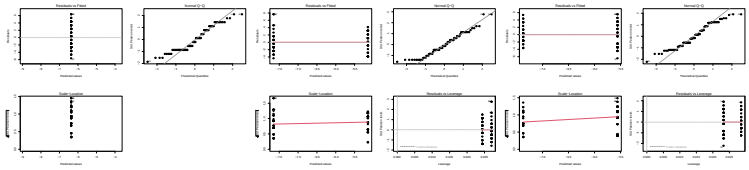


(a) Null (b) Model 1 (c) Model 2

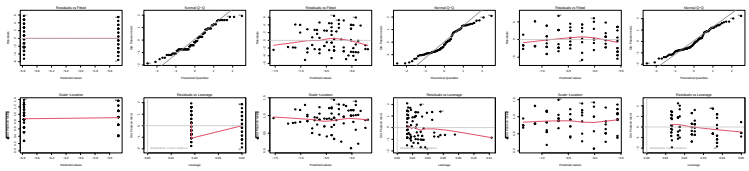


(d) Model 3 (e) Model 4 (f) Model 5 (g) Model 6

Figure 44: Residual Plots of Differenced FFC Models



(a) Null (b) Model 1 (c) Model 2



(d) Model 3 (e) Model 4 (f) Model 5

Figure 45: Residual Plots of Differenced SDQ Models: Adolescents

Appendix H: Link to Code and Datasets

https://drive.google.com/drive/folders/1OOSDYs_J3XJ1t02nmlpDFeRquC9L5D5u?usp=sharing