Statistical analysis plan CHIDO (Child health intervention for development outcomes)

Trial design

The trial is testing a community based intervention which combines early childhood stimulation through parenting skills training, economic strengthening and case management among children infected and affected by HIV in Zimbabwe. The trial is parallel-group and cluster randomised with 1:1 allocation ratio to either the intervention package or standard of care. Dyads of 574 0-24 month old children born to HIV positive women, and their caregiver (primarily their HIV positive mother) were recruited to the trial between January and September 2016. Questionnaires were administered at baseline and 12 months after the intervention commenced when 514 dyads were followed-up.

Analysis principles

All data will be analysed based on an intention to treat adjusting for cluster and minimally for baseline prognostic factors. That is, participants will be analysed in the cluster to which they were randomised regardless of whether they received the assigned intervention. An individual-level approach to the analysis will be used for two reasons, firstly due to the fact that some clusters are half the size of others, and secondly, that it allows the joint effect of all risk factors at either individual level or cluster level to be analysed together (Hayes & Moulton, 2017). Although the number of clusters is 30, which might be considered small, Hayes & Moulton suggest 30 clusters is the minimum number required for an individual level analysis to perform reliably. Significance tests will be two-sided. Any deviations from the primary analysis approach will be explained further. Statistical methods that allow for within-cluster correlations will be used. An overall estimate of the coefficient of variation, k, will be provided for each outcome.

Lack of comparability between study arms would suggest that the outcome could differ between the arms in the absence of the intervention. Therefore, the characteristics of the clusters, women and children will be presented by study arm. All continuous variables will be summarized using the following descriptive statistics: n (non-missing sample size), mean, standard deviation, median, and inter quartile range (IQR). In the case of skewed variables, the geometric mean may be presented. The frequency and percentages (based on the non-missing sample size) of observed levels will be reported for all categorical measures. All summary tables will be structured with a column for each study arm and will be annotated with the total population size relevant to that table/trial arm, including any missing observations.

It is expected that due to the randomisation process that there will be few differences in factors between study arms. However, all factors considered important prognostic factors with the potential to be highly correlated with the outcome will be adjusted for in the analysis regardless of whether they show imbalance between the study arms.

Missing data

The amount (number, percentage) of data that is missing will be reported for each of the key outcome variables and covariates. Where a high proportion of participants lack relevant data, the possible reasons for this and any implications for interpretation of study findings will be discussed. The characteristics of participants with missing data will be compared with those participants with no missing data (complete cases) to identify systematic differences between these two groups using robust standard errors to allow for the lack of independence between participants in the same clusters. Should a high proportion of

Mebrahtu H, *et al. BMJ Global Health* 2019; 4:e001651. doi: 10.1136/bmjgh-2019-001651

participants lack data, the feasibility of imputation of missing data will be explored and/or an analysis will be carried out under the assumption of missing at random by adjusting for those characteristics found to be predictive of missingness (Sterne et al., BMJ, 2009).

Evaluation of the intervention

The cohort were followed-up 12 months after the intervention commenced in the pairs of intervention and control communities who were recruited together and in the same order as they were recruited.

Outcome measures

The primary outcome measures are:

i. Cognitive development

The age-standardized Early Learning Composite (ELC) score will be assessed for all infants at baseline and endline. Trained assessors (independent of the implementers) used the Mullen Scales of Early Learning focusing on the four cognitive scales (visual reception, fine motor, receptive language, and expressive language). The Mullen Scales of Early Learning Tool is an individually administered comprehensive measure of cognitive functioning for infants and preschool children from birth through 68 months [34]. Scores were adjusted based on the child's age. The number of assessors was kept to a minimum of 3, to maximise reliability of measurement. To increase validity, assessors were interchanged between intervention and control groups after each clinic pair, to ensure that there was no systematic bias between the groups. The Mullen assessments were video recorded. As quality control, a random 10% sample of videos were reviewed internally and another random 10% sample will be reviewed by an independent external assessor prior to the follow up survey.

ii. Retention in HIV care

The proportion of HIV exposed or infected children with full retention in care (>80%) of scheduled visits at 12 months.

Secondary outcomes (Table below) include HIV infected infants virological failure (viral load measurement > 1000 copies per ml vs virological suppression), ART adherence and retention in both the HIV positive mothers and the HIV infected infants, parental stress levels (HIV positive mothers) and mental health status of caregivers, household food security status and nutritional status of the infants.

CHIDO Trial Outcomes Outcome Measures	**Analysis Group**	**Instrument**	**Administration**	**Type of measurement**
Primary Child Development Outcomes: Mean childhood development global score	All infants	Mullen Scales for Early Learning	Child Assessment	Quantitative score, approximately normally distributed
Child HIV Outcomes: i) Retention in care	All infants	Questionnaire	Self-report	Binary, yes/no
Secondary Child HIV Outcomes HIV + infants:	HIV + infants	Viral Load Tests	Clinical/ Laboratory Tests	Binary virological suppression,

i) Viral load				<1000 copies/ml versus >=1000 copies/ml
Child Development Outcomes: i) Visual reception ii) Fine Motor iii) Receptive language iv) Expressive language	All infants	Mullen Scales for Early Learning	Child Assessment	Quantitative scores, approximately normally distributed
Nutritional Outcomes: Weight for age, height for age, weight for height (BMI) z-scores	All infants	Mid Upper Arm Circumference tape measure, height mate/board	Child Assessment	i) Quantitative scores, normally distributed ii) Binary <=-2 z-score vs >-2 z-score
Parenting Outcomes: Parenting Stress Index	All caregivers	Parental Stress Index Short Form (PSI-SF)	Interview	i) Quantitative score (percentile), approximately normally distributed ii) Binary <90% not distressed versus >=90% distressed
Adherence Outcomes: i) Retention in care ii) Viral Load	HIV +ve mothers	i) Medical Adherence Rating Scale (MARS) ii) Viral Load Tests	Interview Clinical/Laboratory Tests	i) Binary; MARS score 0-5 not adherent, score 6-10 adherent ii)Binary; virological suppression, <1000 copies/ml versus >=1000 copies/ml
Food Security Outcome	All caregivers	Household hunger (food deprivation) scale	Interview	Categorical, a) score 0-1 little to no hunger; b)2-3 moderate hunger; c)4-6 severe hunger
Mental Health Outcomes: i) Postnatal Depression ii) Common Mental Disorders	i) HIV +ve mothers ii) All Caregivers	i) Edinburgh Postnatal Depression Scale ii) Shona Symptom Questionnaire (SSQ) 8	Interview	i) Binary; EPDS 0-11 not depressed, versus EPDS>=12 depressed ii) Binary; SSQ8 score 0-5 no CMD, versus score 6-8 has CMD

Recruitment and follow-up

A CONSORT flowchart will describe recruitment, refusals and losses to follow-up.

Primary analysis

For quantitative outcomes, the effect of the intervention will be quantified by comparing the mean values between trial arms, using mixed effects linear regression, incorporating random effects for clusters, to estimate the mean difference and corresponding 95% confidence interval (CI). *A priori* adjustments will include the quantitative baseline measurement, strata in Goromonzi (two groups based on the number of children on the clinic register), district and in comparisons of the Mullen scale, the outcome assessor at baseline will also be adjusted for. Factors independently prognostic for outcome, regardless of imbalance will be adjusted for; this will be if for binary variables there are more than 20 participants who fall into each level. For assessment of the Mullen scale prognostic factors, and infant's age.

For binary outcomes, the effect of the intervention will be quantified by odds using mixed effects logistic regression to estimate the odds ratios and corresponding 95% CI. Correlation between clusters will be accounted for by incorporating random effects for clusters into all models. *A priori* adjustments will include the binary baseline measurement, strata, district and additional factors as described for the quantitative outcomes.

Pre-specified subgroup analyses will investigate whether the effect of the intervention differs by baseline early learning infant composite T-score categorised into two groups (average and above versus below average), baseline caregiver's mental health as measured by EPND, SSQ8 and PSI using the cut-offs described above to categorise the participants into two groups and infant's age group categorised into three, 0-<6m, 6-<12m and 12-24m.

Descriptive categories (Composite-scores)	***Descriptive categories (T-scores)***
Very High (130 – 155)	Very High (70 – 80)
Above Average (116 - 129)	Above Average (61 - 69)
Average (85 - 115)	Average (40 - 60)
Below Average (71 - 84)	Below Average (31 - 39)
Very Low (49 - 70)	Very Low (20 - 30)