**Supplementary materials**

Table S1. Baseline characteristics of study participants by loss to follow up

| Baseline characteristic | Retained(n=969) | Lost to follow-up, passed away, or missing data (n=77) |  |
| --- | --- | --- | --- |
| n (%) | p-value |
| ART adherent in the past week  | 642 (66) | 39 (51) | 0.006 |
| **Healthcare factors** |  |  |  |
| No medication stock-outs | 918 (95) | 71 (92) | 0.350 |
| Confidentiality at the clinic | 740 (76) | 52 (68) | 0.082 |
| Travel to the clinic below 1 hour | 859 (89) | 67 (87) | 0.660 |
| Wait time in the clinic below 1 hour | 503 (52) | 38 (49) | 0.670 |
| **Caregiving factors** |  |  |  |
| Good monitoring and supervision  | 332 (34) | 20 (26) | 0.140 |
| Good caregiver-teen communication | 258 (27) | 22 (29) | 0.710 |
| No emotional or physical violence  | 680 (70) | 48 (62) | 0.150 |
| Not witnessing arguments or fights between adults in the home  | 859 (89) | 64 (83) | 0.150 |
| **Household structure** |  |
| Orphan (maternal, paternal, or both) | 578 (60) | 38 (49) | 0.077 |
| Number of caregivers with whom the teen has ever lived (m, SD) | 1.89 (1.04) | 2.10 (1.55) | 0.093 |
| Biological parent is primary caregiver | 425 (44) | 43 (56) | 0.042 |
| Household size (m, SD) | 6.78 (2.94) | 7.06 (5.93) | 0.460 |
| **Socio-economic factors** |  |
| Girls  | 532 (55) | 44 (57) | 0.700 |
| Mean age (m, SD) | 13.59 (2.87) | 14.65 (3.06) | 0.002 |
| Rural area residence | 258 (27) | 13 (17) | 0.059 |
| Informal housing | 180 (19) | 16 (21) | 0.590 |
| Can afford 8 basic necessities | 310 (32)  | 28 (36) | 0.430 |
| **HIV factors** |
| Horizontal mode of HIV infection  | 201 (21) | 21 (32) | 0.034 |
| Recent ARV initiation (under 2 years prior to baseline)  | 255 (26) | 18 (26) | 0.970 |

Table S2. Single-variable probit regressions testing associations of adolescent experiences and their adherence (n=969 adolescents).

|  | Coefficient | Lower 95% CI | Higher 95% CI | p-value |
| --- | --- | --- | --- | --- |
| Explanatory variables |  |  |  |  |
| *Healthcare factors* |  |  |  |  |
| No medication stock outs - between | 0.027 | 0.208 | 0.130 | 0.897 |
| No medication stock outs - within | 0.384 | 0.221 | 1.740 | 0.082 |
| Confidentiality - between | 0.114 | 0.101 | 1.130 | 0.259 |
| Confidentiality - within  | 0.221 | 0.096 | 2.300 | 0.021 |
| Travel to the clinic below 1 hour - between | 0.280 | 0.153 | 1.830 | 0.067 |
| Travel to the clinic below 1 hour - within  | 0.384 | 0.172 | 2.230 | 0.026 |
| Wait time in the clinic below 1 hour - between | 0.072 | 0.092 | 0.780 | 0.438 |
| Wait time in the clinic below 1 hour - within | -0.151 | 0.090 | -1.680 | 0.093 |
| *Caregiving factors* |   |  |  |  |
| Good monitoring and supervision - between | 0.221 | 0.094 | 2.340 | 0.019 |
| Good monitoring and supervision - within | -0.031 | 0.100 | -0.320 | 0.752 |
| Good caregiver-teen communication - between | 0.165 | 0.105 | 1.570 | 0.116 |
| Good caregiver-teen communication - within | -0.039 | 0.095 | -0.410 | 0.684 |
| No emotional or physical violence victimisation - between | 0.466 | 0.097 | 4.790 | <0.001 |
| No emotional or physical violence victimisation - within | 0.410 | 0.094 | 4.350 | <0.001 |
| Not witnessing any arguments or fights between adults at home - between | 0.521 | 0.143 | 3.650 | <0.001 |
| Not witnessing any arguments or fights between adults at home - within | 0.251 | 0.134 | 1.870 | 0.062 |
| Control factors |  |  |  |  |
| *Household structure* |  |  |  |  |
| Orphan - between | 0.047 | 0.075 | 0.620 | 0.533 |
| Orphan - within | -0.045 | 0.161 | -0.280 | 0.782 |
| Caregiver is a biological parent - between | -0.108 | 0.077 | -1.390 | 0.163 |
| Caregiver is a biological parent - within | -0.144 | 0.149 | -0.970 | 0.335 |
| Household size - between | -0.003 | 0.011 | -0.270 | 0.785 |
| Household size - within | 0.009 | 0.016 | 0.560 | 0.579 |
| The number of previous caregivers - between | 0.038 | -1.840 | 0.066 | -0.144 |
| The number of previous caregivers - within | -0.042 | 0.043 | -0.980 | 0.329 |
| *Socio-economic factors* |  |  |  |  |
| Age - between | -0.036 | 0.012 | -3.100 | 0.002 |
| Age- within | -0.048 | 0.037 | -1.290 | 0.196 |
| Rural area residence - between | -0.097 | 0.082 | -1.180 | 0.237 |
| Rural area residence - within \* | -0.065 | 0.211 | -0.310 | 0.758 |
| Informal home - between | 0.069 | 0.100 | 0.690 | 0.493 |
| Informal home - within | -0.035 | 0.217 | -0.160 | 0.873 |
| Can afford 8 basic necessities - between \* | 0.235 | 0.027 | 0.443 | 0.027 |
| Can afford 8 basic necessities - within | -0.173 | -0.372 | 0.027 | 0.099 |
| Timepoint (follow-up) | -0.043 | 0.061 | -0.710 | 0.480 |
| Sex (girl)- between | -0.114 | 0.068 | -1.670 | 0.094 |
| *HIV factors* |  |  |  |  |
| Mode of infection (horizonal) - between | -0.364 | 0.081 | -4.480 | <0.001 |
| Recent ART initiation (under 2 years prior to baseline) - between | -0.252 | 0.074 | -3.420 | 0.001 |

*95%CI - confidence interval*

Table S3. Full results from the multivariable within-between probit regressions testing associations of adolescent experiences and their adherence (n=969 people, 1,928 observations)

|  | *Coefficient* | *Lower CI* | *Higher CI* | p-value |
| --- | --- | --- | --- | --- |
| *Explanatory variables* |  |  |  |  |
| *Healthcare factors* |  |  |  |  |
| No medication stock outs - between | *-0.032* | *-0.437* | *0.373* | 0.877 |
| No medication stock outs - within | *0.335* | *-0.116* | *0.786* | 0.146 |
| Confidentiality - between | *0.112* | *-0.093* | *0.317* | 0.284 |
| Confidentiality - within  | *0.240\** | *0.042* | *0.438* | 0.018 |
| Travel to the clinic below 1 hour - between | *0.267* | *-0.055* | *0.590* | 0.105 |
| Travel to the clinic below 1 hour - within  | *0.451\** | *0.105* | *0.796* | 0.011 |
| Wait time in the clinic below 1 hour - between | *0.048* | *-0.142* | *0.238* | 0.621 |
| Wait time in the clinic below 1 hour - within | *-0.170* | *-0.352* | *0.012* | 0.068 |
| *Caregiving factors* |  |  |  |  |
| Good monitoring and supervision - between | *0.056* | *-0.149* | *0.260* | 0.594 |
| Good monitoring and supervision - within | *-0.082* | *-0.287* | *0.124* | 0.438 |
| Good caregiver-teen communication - between | *0.166* | *-0.049* | *0.382* | 0.130 |
| Good caregiver-teen communication - within | *-0.003* | *-0.196* | *0.189* | 0.972 |
| No emotional or physical violence victimisation - between | *0.434 \*\*\** | *0.228* | *0.639* | <0.001 |
| No emotional or physical violence victimisation - within | *0.378 \*\*\** | *0.183* | *0.574* | <0.001 |
| Not witnessing any arguments or fights between adults at home - between | *0.280* | *-0.017* | *0.578* | 0.065 |
| Not witnessing any arguments or fights between adults at home - within | *0.191* | *-0.083* | *0.464* | 0.171 |
| Control factors |  |  |  |  |
| Household structure |  |  |  |  |
| Orphan - between | *-0.003* | *-0.192* | *0.186* | 0.974 |
| Orphan - within | *-0.061* | *-0.400* | *0.278* | 0.724 |
| Caregiver is a biological parent - between | *-0.144* | *-0.338* | *0.050* | 0.147 |
| Caregiver is a biological parent - within | *-0.169* | *-0.479* | *0.142* | 0.288 |
| Household size - between | *-0.004* | *-0.028* | *0.021* | 0.769 |
| Household size - within | *0.015* | *-0.017* | *0.047* | 0.368 |
| The number of previous caregivers - between | *-0.059* | *-0.137* | *0.020* | 0.141 |
| The number of previous caregivers - within | *-0.037* | *-0.124* | *0.049* | 0.397 |
| Socio-economic factors |  |  |  |  |
| Age - between | -0.011 | -0.042 | 0.019 | 0.467 |
| Age- within | -0.157 | -0.344 | 0.030 | 0.100 |
| Rural area - between | -0.110 | -0.287 | 0.068 | 0.225 |
| Rural area - within | 0.010 | -0.419 | 0.439 | 0.963 |
| Informal home - between | 0.145 | -0.062 | 0.352 | 0.171 |
| Informal home - within | 0.032 | -0.410 | 0.474 | 0.887 |
| Can afford 8 basic necessities - between  | 0.143 | -0.076 | 0.363 | 0.200 |
| Can afford 8 basic necessities - within | -0.184 | -0.391 | 0.023 | 0.082 |
| Timepoint (follow-up) | 0.208 | -0.101 | 0.516 | 0.188 |
| Sex (girl)- between | -0.076 | -0.217 | 0.064 | 0.286 |
| HIV factors |  |  |  |  |
| Mode of infection (horizontal) - between | -0.192 | -0.400 | 0.016 | 0.071 |
| Recent ART initiation (under 2 years prior to baseline) - between | -0.121 | -0.285 | 0.044 | 0.150 |

*95% CI - confidence interval*

**Box S1. Summary of patient file data collection**

**What are Patient Files in the South African health system?**

In South Africa, patient files are a paper-based form available across health facilities for collection of clinical record of participants who have registered in the clinic for antiretroviral treatment (ART). The information on this form summarises the participants health condition and is updated during each visit at the facility where the file is stored. At each healthcare facility in South Africa, some information from patient files may be available in an electronic database called Tier.net (for more information on Tier.Net visit <http://www.publichealth.uct.ac.za/sites/default/files/image_tool/images/108/TIER.Net%20%5B03%20Nov%202010%5D.pdf>). This database is not linked across facilities and is generally on a single computer in each facility, containing information about patients and their health records in that facility only. Tier.net includes information about viral load and CD4 test results/dates and results are reported centrally on a regular basis.

**How were facilities and patient files used to recruit participants into Mzantsi Wakho study.**

At baseline, clinic rosters and registers were used to determine eligible participants for Mzantsi Wakho (MW). Based on this, all living ART-initiated patients who were 10-19 years old at baseline were approached for recruitment into MW, including those who were considered loss to follow-up or defaulters by the healthcare system. For full details of the study sampling method see page 9 of the MW Research Protocol. At the time of recruitment, not all facilities had Tier.Net but, where available, electronic patient rosters were consulted.

**What was purpose of collecting additional patient file data?**

The adolescent questionnaires administered during MW were based on self-report with the assistance of a highly-trained research assistant. In particular, the questionnaires included detailed ART access and adherence questions. The purpose of collecting additional patient file data was to complement this self-reported data with administrative health data collected by the South African health system. Initially, the principle objective was to use viral load data obtained from patient files to validate the self-reported adherence measures. As the project has evolved, patient file data and clinic-based data collection has become a full arm/sub-study, and this data has been used as an outcome for several papers/ analyses.

**Which facilities were searched for additional patient file data and when?**

The original number of health facilities included in the study was 52 clinics and sub-units in large tertiary hospitals. However, at follow-up, when participants were asked about their primary clinic of care in the self-report questionnaire, they listed new facilities (beyond the original 52). Accordingly, patient file data collection had to consider these 15 additional facilities reported by MW participants at follow-up. It was decided that patient files would be collected at all 67 facilities.

**When did data collection take place?**

Patient file data was collected in two rounds. In the first round, retrospective patient file data were collected from the original 52 facilities between January 2014 through December 2015. In the second round, data were collected from the original 52 facilities plus the 15 additional facilities between January 2016 to December 2017. At the 15 additional facilities, both rounds of data collection forms were used, to ensure that the dataset was complete across all facilities.

**How was patient file data collected?**

Data was extracted using a standardised questionnaire with slight differences between the two rounds, according to a protocol previously applied in other studies of adolescents on ART and adapted to the patient file system in the included healthcare facilities. To identify available clinical records at each facility, the research staff searched individually for every MW participant on the clinic’s roster/register of ART patients, independent of their self-reported HIV status. This was necessary since nearly a third of the adolescents living with HIV had not been disclosed their status at baseline – potentially due to their young age. Research staff also searched for all MW participants on each facility’s Tier.net to identify any electronic records that were not in the patient’s paper files yet (and to identify any files that were missing in physical form but available electronically). They made sure to check for alternate spellings of participants’ names, with the supervision of the facility data capturer. All this information was saved in a password protected clinic roster, which is informing the matching of these participants with National Health Laboratory Service data.

**How does the patient file data relate to the MW self-report questionnaire?**

Patient file data were collected for MW participants that were ART-initiated. Extraction of patient file data during the first round (Jan 2014-Dec 2015), including viral load (VL), WHO staging and CD4 count information, were intended to match information gathered from participants’ self-reported information from the MW baseline interviews, and extraction of data during the second round (Jan 2015-Dec 2017), were intended to align with participants self-reported information at MW follow-up.

In the final merged dataset, viral load and CD4 count data extracted from patient files are available at the T1 and T2 timepoints for some participants. A time window of 12 months preceding or following the self-report interview date was used to match patient file data with self-report information at these two timepoints. Depending on the research question, researchers may want to adjust this time window. This matching protocol has resulted in different levels of missing data at each MW self-reported data point.

**Which patient file data were collected?**

The list of items available on patient file data at baseline and follow-up are similar. The data items collected by the search team includes CD4 count, viral load, current ART medication, TB symptoms and treatment including dates. For full details see below.

*List of variables collected during the baseline and second round of patient file data extraction.*

|  |
| --- |
| **ROUND 1 data extraction**Patient file data extraction form at baseline.1. Current Clinic:
2. Person extracting data from files:
3. Serial Number
4. Was a patient file found for this patient
5. Patient file number
6. When was this patient opened
7. HIV DIAGNOSIS and ART INITIATION
8. Year HIV diagnosed
9. ART start date
10. Pre-ART experience – Antiretroviral (ART, HAART, PMTCT, sdNVP, dual RX pMTCT, PEP) use before the above date?
11. FIRST VISIT RECORDS
12. Date of First Visit
13. WHO stage at first visit
14. CD4 count at first visit
15. Viral load at first visit- Write 888 if the viral load was non-detectable (LDL)
16. MOST RECENT INFORMATION
17. Date of the most recent CD4 count in the patient file
18. Most recent CD4 count
19. Date of the most recent Viral Load in the patient file
20. Most recent Viral Load in the patient file
21. Current ART medication – medication names
22. Date of the last time the participant picked up their medication (to be inputted into roster):
23. Ever had TB before?
24. Has preventive TB treatment been reported?
25. Has TB treatment been reported in THIS clinic?
26. Has there been a TB test reported?

 If yes, please indicate the type of TB test: If yes, please indicate the results of the test:1. Has the patient ever had monitoring for TB symptoms reported?
2. LAB REPORTS OUTSIDE OF PATIENT FILE
3. Did you find another CD4 count lab report outside of the patient file for this patient?

What was the CD4 count? What was the date the CD4 count was taken on?1. Did you find another Viral Load lab report outside of the patient file for this patient?

 What was the Viral Load? What was the date the Viral Load was taken on?**ROUND 2 data extraction**The list of the variables collected in the second round of data extract was similar to baseline information:Patient file data extraction form1. Current Clinic:
2. Person extracting data from files:
3. Serial Number
4. Was a patient file found for this patient
5. Patient file number
6. Patient File Number
7. When was this patient file opened?
8. Was this patient transferred in from a different facility?

 If yes, what was the patient’s previous facility for care:1. ART start date at this facility
2. FIRST VISIT RECORDS: complete ONLY if the patient was transferred into this facility from a previous site since last round of data extraction.
3. Date of First Visit
4. WHO stage at first visit
5. CD4 count at first visit
6. Viral load at first visit- Write 888 if the viral load was non-detectable (LDL)
7. MOST RECENT INFORMATION
8. Has this patient made any new visits to THIS clinic since the last round of data extraction?

If NO, what was the reason for stopping treatment at THIS clinic? Please answer the following question but do not complete the remaining sections. If YES, please complete the remaining sections.1. Date of the most recent CD4 count in the patient file
2. Most recent CD4 count
3. Date of the most recent Viral Load in the patient file
4. Most recent Viral Load in the patient file
5. Most recent WHO stage in the patient file
6. Current ART medication – medication names
7. Has TB treatment been reported in THIS clinic?
8. Has there been a TB test reported in THIS facility?

 If yes, please indicate the type of TB test:If yes, please indicate the results of the test:1. Has the patient ever had monitoring for TB symptoms reported in the last year?
2. Were any new diagnoses reported for the patient since the last round of data extraction?

If YES, please indicate the new diagnoses:1. Was the patient treated for an STI?
2. Was the patient ever pregnant?
3. LAB REPORTS OUTSIDE OF PATIENT FILE
4. Did you find another CD4 count lab report outside of the patient file for this patient?

What was the CD4 count? What was the date the CD4 count was taken on?1. Did you find another Viral Load lab report outside of the patient file for this patient?

 What was the Viral Load? What was the date the Viral Load was taken on?Note:TB records available is too sparse to be helpful.  |