**C.5.2.3. Analytical Methods:**

A. Defining the Outcomes of Interest: Individual variables will be assessed for their contribution to the outcomes of interest. Three major outcomes will be evaluated as surrogates for VF and HIV DR: ARV (1) **levels** in RBC (and hair sampling secondarily), (2) **adherence** as measured by pill count (and participant self-report secondarily) and (3) **access** as measured by pharmacy refill rates (and clinic visits secondarily). Categories for these outcomes will be defined based upon the distribution of these variables within the study population and using published categories/definitions. To understand the relationship between these outcomes and VF, univariate logistic regression models will be constructed. These analyses will be conducted overall (unadjusted) as well as after stratifying by site and adjusting for factors known to affect VF and DR such as age, gender, duration of ART, baseline viral load and CD4 cell count. A separate model for cases only will evaluate the relationship between categories of ARV levels (and separately for categories of adherence and access) with VF having at least one major HIV DR mutation and VF without any major HIV DR mutation. A receiver operator characteristic (ROC) curve will be utilized in order to optimize the relationship between the category assignments and the chosen endpoints. A separate comparison of stratified and non-stratified analyses will permit a validation of the measures used for ARV levels, access and adherence at each site. Correlation between these three outcome measures will also be assessed. For cases, separate correlation coefficients will be determined for these variables and the number of major resistance mutations.

Steps for A.

1. Univariate unadjusted with endpoints: virologic suppression (VS) and virologic failure (VF) using cases and controls; amongst those with VF, patients with at least one major HIV DR mutation (VFDR) and patients without one major HIV drug resistance mutation (VFDS)

a. Pill count (PC) – using tertiles of distribution OR using a dichotomized cutpoint

b. Therapeutic Drug Monitoring RBC and Hair Samples (TDM) – as above

c. Self Report (SR) – as above

d. Pharmacy Refill (PR) – as above using either MPR or MRA

e. Clinic Visit (CV) – as above using percent or number of visits in period

2. As above but with the following

a. Separately OR stratify by site (compare stratified and separate unstratified to internally validate the adherence measures used at each site)

b. Adjust for age, gender, duration of ART, regimen, baseline viral load and CD4 count

c. Optimize model by identifying best categories of each adherence measure using ROC curve; consider varying definitions of suppressed/failure (different VL cutoffs)

d. Correlation between each level, adherence and access measures and across these measures to internally validate measures with each other and to further optimize ROC curves above

3. Correlation coefficient between each adherence measure (PC, TDM, SR, PR, CV) for those with VF and the following variables

a. Overall number of mutations

b. Number of Drug Classes

c. TAMs

d. Other major mutations

B. Determining the Key Covariates: After establishing these outcome relationships, it will be necessary to resolve each outcome into its factor components independent of and then including ATM, the design variable of interest. This process will involve three key steps: (1) defining the factor domains, (2) determining which of the domains are associated with the outcomes (univariate), (3) building the multivariable model (MV). Each outcome has areas of resolution, and within each area there are several domains. Descriptive statistics will be used to compare risk factors between sites. Then, these domains will be explored individually to determine their relationship to the outcome using unadjusted univariate (UV) analyses. Several forced MV models will be built using factors found to be significant in the UV analyses (P < 0.5) and those shown to be significant in other studies but not found significant here. These models will be performed with and without adjustment for potential confounders such as age, gender, duration of ART, baseline viral load and CD4 cell count. Regression analyses will be conducted separately by site in order to compare significant covariates between sites as well as stratified by site in order to determine covariates overall.

Steps for B.

1. Define factor domains (optimize the responses and scales)

a. Symptoms (Sx)

(1) Scale – use either presence or absence; use severity index

(2) Karnovsky score – categorical variable

(3) Correlations between

(a) Concomitant meds known to increase levels and symptoms

(b) Opportunistic infections (OI) and symptom scale OR Karnovsky score

(c) Food insecurity and symptom scale OR Karnovsky score

b. Psychosocial context (PS)

(1) Cultural factors (race/ethnicity, ethnic group, religion, religious involvement)

(2) Depression – Kessler 10 scale

(3) Stigma – categories

(4) Recall – Neurocognitive testing and plan for prompting adherence/visits

(5) Treatment Support System – needs defining

(6) Intimate Partner Violence – needs defining

(7) Alcohol/Substance Abuse – needs defining

(8) Satisfaction with clinical services – needs defining

c. Socioeconomic Factors (SEF)

(1) Wealth index, sources of income/employment status/occupation

(2) Number of household members and supported members

(3) Transportation mode and distance

(4) Food insecurity

(5) Childcare

d. African Traditional Medicine (ATM) – Categorize as concurrent use, interrupted use, no use based upon responses (needs defining)

2. Determine which domains are associated with the outcomes using univariate analysis

a. Sx with PC, TDM and SR

b. PS with PC, TDM, SR, PR and CV

c. SEF with PR and CV

d. Meds with TDM

e. FI with TDM

f. Among participants with optimal PR and CV, determine if 2a and 2b above changes

3. Build multivariable model with outcomes

1. Levels (TDM RBC and Hair)
2. Using all factors with UV P < 0.5
3. Using all factors shown predictive in the literature
4. Both
5. Above after adjusting for age, gender, duration of ART, regimen, baseline viral load and CD4 count
6. Separately OR stratify by site (compare stratified and separate unstratified to internally validate the adherence measures used at each site)
7. Adherence (PC and SR)

(1) Using all factors with UV P < 0.5

(2) Using all factors shown predictive in the literature

(3) Both

(4) Above after adjusting for age, gender, duration of ART, regimen, baseline viral load and CD4 count

(5) Separately OR stratify by site (compare stratified and separate unstratified to internally validate the adherence measures used at each site)

c. Access (PR and CV)

(1) Using all factors with UV P < 0.5

(2) Using all factors shown predictive in the literature

(3) Both

(4) Above after adjusting for age, gender, duration of ART, regimen, baseline viral load and CD4 count

(5) Separately OR stratify by site

1. ATM will be examined separately with each outcome measure in UV analysis.
2. ATM will then be added to the above forced models for each step under 3a, 3b, 3c
3. ATM will then be forced into a model examining the three endpoints with the most significant factors identified in the above steps OR among those patients with the optimal levels, adherence AND access to determine an independent effect of ATM

***ARV Levels*** – Concomitant medications known to affect ARV drug levels, psychosocial factors, food insecurity and symptoms associated with poor absorption (vomiting or diarrhea) will each be assessed individually using univariate analyses. If any of these variables are found to be significant (P < 0.5), then these factors will be examined in the MV analyses.

***ARV Adherence*** – Based upon preliminary data from MCH, two major areas emerged as the key issues affecting pill-taking by patients: (1) symptoms and (2) the psychosocial context. Since adherence is also affected by access, separate sensitivity analyses will be conducted among participants with perfect access to determine if the key covariates associated with adherence are independent of access.

***ARV Access*** – Similar to findings related to ARV adherence, two primary areas were most commonly described as impacting access to ARVs: (1) the socioeconomic and (2) the psychosocial context.

Symptoms: Using a standardized symptom severity scale, symptoms and functional status (Karnovsky score) will be associated with the ARV Levels and Adherence outcomes of interest. In further exploratory analyses, concomitant medications known to increase ARV drug levels, food insecurity and recent opportunistic infections (OI) will be correlated with symptom severity to determine their contribution to the symptoms reported.

Psychosocial: Each of the following domains will be explored for their contribution to the ARV Levels, Adherence and Access outcomes of interest: cultural factors (race/ethnicity, ethnic group, religious preference), depression (Kessler 10), stigma, recall (as measured by a validated neurocognitive scale for this setting and the existence of a plan for prompting adherence/refills), treatment support system, intimate partner violence, alcohol/substance abuse, and satisfaction with clinical services.

Socioeconomic: Each of the following domains will be examined for contribution to the ARV Access outcome of interest: wealth index, number of household members, transportation/distance to clinic, food insecurity (using a modification of the HFIAS index), childcare, and employment/financial support.

*African Traditional Medicines* – ATM will serve as the design variable of interest. Using mixed quantitative and qualitative questions, participants will be asked to describe their use of ATM and beliefs regarding traditional and alternative medicine. Specific key questions will inform the categorization of participants into (a) ATM concurrent with western medicine/ARVs and (b) ATM interrupting western medicine/ARVs. Thereafter, a three-step approach will be used to determine the association of ATM and the outcomes of interest. First, ATM use will be assessed for its association with the outcomes separately in UV analyses. Second, ATM use will be forced into the MV models described previously. Finally, ATM will be examined in a MV model of VF with the three primary outcomes to determine if ATM has an effect on VF independent of ARV Levels, Adherence or Access to care. Further exploratory analyses will attempt to isolate the most frequent herbals reported in the questionnaire in order to determine if there are specific herbals most highly associated with the outcomes.

C. Model Testing – For randomly selected multiple sets of participants, key covariates will be identified from responses to separate quantitative questions targeted specifically at the outcomes of interest. Using an embedded mixed methods approach98, qualitative responses will inform the strength and ontology of the covariates selected. In a hypothesis-generating fashion, any covariate found at high frequency (>10%) will be forced into a MV model with the respective outcome. Then using forwards, backwards and stepwise approaches beginning with the covariates at lowest frequency, factors will be removed sequentially to determine their impact on the odds ratios of the most significant covariates. These models will be performed with and without adjustment for potential confounders such as age, gender, duration of ART, baseline viral load and CD4 count. Regression analyses will be conducted separately by site in order to compare significant covariates between sites as well as stratified by site in order to determine covariates overall. This final multivariable model will be compared to the final models constructed from step 3.

Steps for C.

1. Create random sets of participants from cases and controls

2. Identify key covariates from access and adherence rank questions within each set

3. Determine strength and hierarchy of these covariates as informed by qualitative responses overall and within an individual

4. Force covariates at 10% frequency into MV with access or adherence outcomes

5. Use forward, backward and stepwise approaches to determine impact on OR

6. Also adjust for age, gender, duration of ART, regimen, baseline viral load and CD4 count

7. Also perform separately OR stratified by site

8. Compare this model to the MV model constructed above for B.

D. Model Validation – All factors found to be significant in this study and those shown to be significant in other studies will be used to perform cross-validation to internally validate our risk factor model. To the extent possible, we will cross-validate the risk factor model externally as well against other data sets. All MV models from steps A-D will then be assessed using the Akaike Information Criterion. The best model will be designated the EWI model. This model will then be applied to new validation sets in order to determine if the prediction is maintained in different settings.

**C.5.2.4. Statistical Methods.** All statistical analyses will be performed using SAS version 9.2 (SAS Institute, Cary, NC). Aim 2 will be an analysis of risk factors associated with virologic failure in peri-urban and rural sites in KZN. We have designed a nested case-control study for this Specific Aim and will use conditional logistic regression to model failure status as a function of risk factors, such as age, gender, duration of ART, baseline CD4 cell count and baseline viral load. The conventional statistical approach to conditional logistic regression is to perform a stratified Cox proportional hazards regression model (e.g. PHREG procedure in SAS) with time equal to a constant, failure status equal to case status, and each matched case-control cluster as separate stratum. Covariates are modeled the same as they are in proportional hazards regression. We will consider models that are both stratified and unstratified on site. This modeling strategy will be used to investigate the adjusted and unadjusted adherence and access to ART effects as well as first steps in exploring the protective effects of ATM on virologic failure. Modeling building and testing will be accomplished using a variety of traditional statistical methods (backward and stepwise deletion, forward selection) as well as modern methods such as L1-regularized conditional logistic regression.

1. Neurocognitive Data

1. Digit Span Forward - Record the highest item with at least one trial completed accurately (1-8)
2. Digit Span Backward - Record the highest item with at least one trial completed accurately (1-7)
3. TMT A - record total seconds
4. TMT B - record total seconds
5. Scoring is determined based upon age and gender (refer to tables)
6. Each raw score is then referenced according to table as Normal, 1 SD, 2 SD
7. If >= 2 referenced scores are >= 1 SD (but <2 SD), then classify as MND/ANI; if >= 2 referenced scores are >= 2 SD, then classify as HAD

2. Depression – unclear if higher rates in VF are related to being told they are a case (VF)

3. Adherence/Access

a. Measures (exclude refill date on same day of enrollment)

1. Pill Count - calculated EITHER by dividing the number of pills taken (not returned) by the number prescribed OR expected pill count = [days per refill - (current date – refill date)] • (doses per day) • (pills per dose)
2. Pharmacy Refill - calculated EITHER as the number of months or days’ supply obtained divided by the total number of months or days in the period (MRA) OR the number of refills obtained divided by the expected number of refills in a given time period (SKT gives 30 days supply but refill every 28 days). Use 90 days for 3-month interval and 180 days for 6-month interval.
3. The expected count is currently defined as the fill quantity less expected dosage since the fill date. If the expected count is positive, *1-(actual-expected)/fill\_quantity* gives adherence since the fill date. And if the expected count is negative, because the refill date has passed without a pickup, *1-(actual-expected)/(fill\_quantity-expected)* will give the adherence since the fill date--adherence goes to 0 as the expected dosage becomes more negative.
4. Delays - the total delay in pickup is defined as the sum of delays over the past 24 weeks. Delays are just the actual fill date less the expected fill date, when this number is positive.
5. High counts can be determined by number of refill dates where given less than 30 days.

b. Determine relationship of ARV duration and VL with adherence/failure/resistance

c. Exclude first 20 VF patients or by date October 12, 2010 to December 18, 2010 (see below)

d. May need to stratify those with positive and those with negative expected pill count

MORE DETAIL FOR ADHERENCE:

IMPORTANT NOTE: MANY PATIENTS HAVE A REFILL DATE ON THE DATE OF ENROLLMENT, PLEASE DO NOT INCLUDE THIS DATE FOR CALCULATING THE TOTAL NUMBER OF PILL DAYS DISPENSED!!!

1. Pill Count

If the Expected Pill Count (see Expected Pill Count #3 below) is

a. Positive: [1 minus (Actual Pill Count minus Expected Pill Count)] divided by Total Number of Pill Days Dispensed (see Pharmacy Refill #2 below)

b. Negative: 1 minus [(Actual Pill Count minus Expected Pill Count) divided by (Total Number of Pill Days Dispensed minus Expected Pill Count)] -----------NOTE, I NEED TO CLARIFY WHETHER BRACKETS [] SHOULD BE PLACED IN THIS EQUATION - i.e. DO WE INCLUDE "1" IN THE NUMERATOR OR NOT?---------NO

c. Actual Pill Count is found in the BN, BT, BZ for controls and CH, CN, CT for cases. I would only use one of the pills to count here. Please choose one ARV in this order of preference depending upon what the patient is taking: EFV, TDF, FTC, 3TC, D4T, DDI, LPV/r

2. Pharmacy Refill - this is pretty straightforward

a. Total Number of Pill Days Dispensed (found by adding up all individual Pill Days Dispensed) divided by 180 days.

b. Individual Pill Days Dispensed are found in the following columns: Controls are CU, CW, CY, DA, DC, DE, DG, and DI. Cases are DL, DN, DP, DR, DT, DV, DX, DZ.

3. Expected Pill Count - this is used for Pill count above

a. Final Equation: Total Number of Pill Days Remaining times Doses per Day times Pills per Dose

b. Total Number of Pill Days Remaining is equal to Total Number of Pill Days Dispensed (as above in Pharmacy Refill #2) minus Total Number of Pill Days Expected Taken.

c. Total Number of Pill Days Expected Taken is equal to the Enrollment Date minus First Refill Date.

d. Doses per day is 1 for EFV, TDF, FTC; 2 for 3TC, NVP, D4T, AZT, DDI, LPV/r

e. Pills per dose is 1 for EFV, TDF, 3TC, FTC, NVP, D4T, AZT; 2 for LPV/r, DDI

4. Delays

a. This is the sum of all days delayed in pharmacy refill over 6 months. It is defined as the actual refill date minus the expected refill date.

b. Each actual refill date can be found in the column prior to the refill dispensed days listed above.

c. Expected refill date is found by adding the number of dispensed days to the refill date. For example if given 30 days on Jan 1, then the expected refill date would be Jan 31.

5. High Counts

Identify any number of Individual Pill Days Dispensed (found in Pharmacy Refill #2 above) that are less than 30 days.

1. Categories

a. Questionnaire

1. Area/Location designation

2. Employment status - 3-4 groups

3. Type of work - Claudia completed

4. Home facilities - 3-4 groups

5. Wall and Floor - 2-3 groups

6. Wealth Index - need scoring from Nachega

7. Clinic Start ARV - 2 groups

8. Transport to Clinic - 2 groups

9. Pay for Clinic - 2-3 groups

10. Remember how to take meds - 2-4 groups

11. Remember how to pickup meds - 2-4 groups

12. Religious Denominations - Sally completed

13. OTC meds - 3-5 groups

14. Recommended to Clinic - 3 groups

15. Safe sex practice - 2-3 groups

16. Safe sex frequency - 2-3 groups

17. Knows living with HIV - 3-5 groups

18. Emotional supporter - 3-5 groups

19. Treatment supporter - 3-5 groups

20. Adherence sessions - 3-5 groups

b. CRF

1. Make categories of concomitant meds

2. Make categories of non-AIDS diagnoses

2. Queries/Errors

a. Questionnaire

1. What wall - MUD appears twice

2. Adherence counseling sessions - 3 appears twice

b. CRF

1. "9" appears in the symptoms list

2. The lab section has two "CD4 3" and two "VL 4" but no "CD4/VL 2"

3. Functional Status does not appear in the report (FUN\_STAT)

4. Combine AIDS diagnoses from 1-3 into one

5. Combine ARVs from 1-3 into one (current)

6. Combine ARVs from 1-3 into one (pre)

7. Combine non-AIDS diagnoses from 1-3 into one

8. Combine concomitant meds from 1-16 into one

Analysis Summary

1. Are Adherence/Access measures associated with VF (cases versus controls) and HIVDR (presence and mutations)?
   1. Adherence Measures
      1. Pill Count Adherence
      2. Dispens per day (refill pickups) Adherence (MRA)
      3. Self-reported Adherence
      4. Delays in getting refills
      5. Number of days with a low refill dispensed amount (which is equivalent to number of days with high pill counts)
   2. Outcomes
      1. Cases versus Controls
      2. Amongst cases
         1. Major mutation or no major mutation
         2. Number of classes resistant
         3. Number of major mutations
         4. Specific mutations (K65R, M184V, NNRTI, TAMs)
   3. Adjusting/Stratifying/Covariates
      1. Exclude patients during the Oct to Dec VL date
      2. Adjusted and unadjusted for ART duration
      3. Adjust for failing VL for cases in the resistance outcome analysis (#b.2. above)
      4. Stratify by those with positive or negative Expected Pill Count
      5. Stratify by those with high or low refill adherence (similar to #c.3. above)
   4. Methods
      1. Perform ROC and/or AIC to determine best categories
         1. Tertiles based on population or literature (<80%, 80 to <95%, >=95% OR other)
         2. Dichotomized at >= or < 95% or 80% OR median OR other
         3. Continuous
      2. Determine Sensitivity/Specificity analysis to determine best combination of measures
      3. Determine Correlation between adherence measures

1. Are Baseline Factors associated with the following outcomes:
   1. VF (cases versus controls)
   2. Resistance (presence, mutations categories)
   3. Pill Count Categories with and without stratifying by positive or negative Expected Pill Count
   4. Dispens per day (refill pickups) Categories

1. Baseline Factors for above are grouped by:

         a. Symptoms (Sx)

                                (1) Scale – use either presence or absence; use severity index

                                (2) Karnovsky score – categorical variable

         b. Psychosocial context (PS)

                                (1) Cultural factors (race/ethnicity, ethnic group, religion, religious involvement)

                                (2) Depression – Kessler 10 scale

                                (3) Stigma – categories

                                (4) Recall – Neurocognitive testing and plan for prompting adherence/visits

                                (5) Treatment Support System

                                (6) Intimate Partner Violence

                                (7) Alcohol/Substance Abuse

                                (8) Satisfaction with clinical services

     c. Socioeconomic Factors (SEF)

                                (1) Wealth index, sources of income/employment status/occupation

                                (2) Number of household members and supported members

                                (3) Transportation mode and distance

                                (4) Food insecurity

                                (5) Childcare