This readme document provides the basic information on the creation of the final dataset. The data are in Stata 9.0 format. The .do files produce the tables and figures in the paper "The Demand for and Impact of Learning HIV Results: Evidence from a Field Experiment" by Rebecca Thornton. If there are any questions, please contact the author at rebeccal@umich.edu.

The final dataset consists of:

1. 2004 Baseline data collected by MDICP4 for men, women, boys, and girls. See http://www.malawi.pop.upenn.edu/Level%203/Malawi/level3_malawi_main.html for more information on the survey instruments and data access to the full data set.
2. Randomly assigned locations of VCT clinic creating the distance to the clinic variables
3. 2005 Follow-up data collected as dissertation field work by the author.

The following variables are included:

- site = region of Malawi where study was conducted (1=Mchinji 2=Balaka 3=Rumphi)

- rumphi = dummy indicator if in Rumphi

- balaka = dummy indicator if in balaka

- villnum = village ID

- m1out = Survey outcome in 1998 (2=completed survey)

- m2out = Survey outcome in 2001 (2=completed survey)

- survey2004 = Completed baseline survey in 2004 (1=completed)

- got = Indicator if obtained HIV results (1= learned HIV results)

- zone = indicator of assigned VCT center/zone (1-16 for each VCT center)

- distvct = Distance from the randomly assigned VCT center (km)

- tinc = Total value of the incentive in Malawi Kwacha

- Ti = Discrete values of the incentive (indicating 0; 10-50; 60-100; 110-200; 210-300 Malawi Kwacha)

- any = Indicator of receiving any positive valued incentive

- under = Indicator of living under 1.5 km from the assigned VCT center

- over = Indicator of living over 1.5 km from the assigned VCT center

- simaverage = Average simulated average of distance from a VCT center by randomly drawing placement of a household within a VCT zone

- age = Reported age of respondent in 2004

- age2 = Age-squared

- male = Indicator if respondent is male

- mar = Indicator if respondent is married at the baseline (From self-reported indicator and marital histories)

- educ2004 = Years of completed education at the baseline

- timeshadsex_s = Times per month reported had sex (From a sub-sample of respondents in Rumphi and Balaka)

- hadsex12 = Indicator if reported having sex in past 12 months from the baseline (taken from variables s5b in the adolescent sample and s7 in the adult sample)

- eversex = Indicator if ever had sex at baseline (Assigned to 1 for the married sample and taken from S4 for the adolescent sample)

- usecondom04 = Indicator if reported using a condom during last year at the baseline (taken from s15a for the adult sample and s20a for the adolescent sample)

- tb = Indicator if reported having an HIV test before the baseline (taken from a25)

- thinktreat = Indicator if thinks there will be ARV treatment in the future at the baseline (taken from T_h32)

- a8 = reported likelihood of HIV current infection (0= none; 1=Low;2=Medium;3=High;88=Don't know; 99=Missing)

- land2004 = Indicator if owned any land at baseline (taken from e5)

- T_consentsti = Indicator if consented to STI test at the baseline

- T_consenthiv = Indicator if consented to HIV test at the baseline

- T_final_tric = Trich result

- T_final_result_ct = Chlamydia result

- T_final_result_gc = Gonorrhea result

- hiv2004 = HIV results (-1=Indeterminent)

- test2004 = Indicator if accepted an HIV test

- followup_tested = Indicator if was tested after conclusion of "incentives enrollment"

- followupsurvey = Indicator if was interviewed at the follow-up survey in 2005

- havesex_fo = Indicator if reported having sex between the time of the baseline and the time of the follow-up survey

- numsex_fo = Number of partners reported having sex with between the time of the baseline and the time of the follow-up survey

- likelihoodhiv_fo = Reported likelihood of current HIV infection at the time of the follow-up (1=none; 2=low-medium; 3=high; 88=Don't know)

- numcond = Number of condoms purchased at the follow-up survey

- anycond = Indicator of any condoms purchased at the follow-up survey

- bought = Indicator if bought any condoms at the time of the follow-up survey

A second dataset, Theoretical Dist. Of Incentives, provides the theoretical distribution of financial incentives, separately for each site (Mchinji, Balaka and Rumphi). The following variables are included:

- tinc = Total value of the incentive in US dollar
- dist = Probability of receiving the incentive amount
- site = region of Malawi (1=Mchinji 2=Balaka 3=Rumphi)
- theory = indicator that these observations are theoretical observations