A. BENEFIT OF KNOWLEDGE – ANALYTIC APPROACHES:: The analyses conducted as part of this project will utilize unprecedented 13-year MLSFH panel data from 2006–18 for 1,400+ mature adults on ***SDPs***, health and life-cycle behaviors, including multiple waves of pre- and post-intervention data. The key analyses of these data to be conducted as part of this project include (Aim 3): **(*a*)** evaluate the ***causal*** impact of health information on **(*i*) *SDPs***, **(*ii*)** mental health and health behaviors, and **(*iii*)** labor supplies, savings, intergenerational transfers and other life-cycle behaviors, and **(*b*)** investigate the ***pathways*** through which ***SDPs*** affect these behaviors and outcomes.

A.0.a. Evaluation the impact of health-information on SDPs and life-cycle behavior: The randomized health-information intervention in 2016 provides exogenous variation among respondents in their knowledge about current mortality and health conditions, with the aim of improving the accuracy of individuals' ***SDPs***. Whether this is the case, and whether updated ***SDPs*** affect subsequent health, health behaviors and other life-cycle behaviors, is evaluated using established econometric methods for program evaluation.[1–5] In essence, we will estimate the difference-in-difference equation $Y_{it} = \alpha + \beta\, Intervention_{vt} + \gamma\, X_{it} + \delta\, Time + \mu_i + \varepsilon_{it}$, where, depending on the model and outcome of interest, Y_{it} is a measure of either ***SDPs***, mental health, health- or other life-cycle-behaviors. Individual fixed-effects μ_i account for all time-invariant differences between individuals (including potential pre-treatment differences across villages), *Time* reflects secular time trends, and X_{it} includes time-varying individual or village characteristics. The intervention variable—$Intervention_{vt} = Post_t \times Treatment_v$—is defined as the interaction between the treatment dummy ($Treatment_v = 1$ for all treatment villages) and an indicator $Post_t$ that equals 1 for all post-intervention time periods. In this diff-in-diff estimation, β captures the *causal impact* of the health-information intervention on the outcome of interest Y_{it}, after controlling intra-individual heterogeneity through the fixed-effects μ_i. We will also examine the impact of the health-information intervention not only on health and behaviors, but also on ***SDPs*** and the determinants of ***SDPs*** such as household shocks or health inputs. Interactions with gender, schooling and HIV status will be considered, as well as analyses whether the effect of the intervention is modified by cognitive function (which may be a more relevant for heterogeneous treatment effects than schooling given the fairly low schooling levels among mature adults). Our identification of the causal effects of the health intervention in this estimation strategy relies on three specific assumptions: **(*a*)** non-random attrition across treatments, **(*b*)** random assignment into treatment, and **(*c*)** parallel time trends across treatment groups. There are several standard ways to check for assumptions (*a*) and (*b*), though not necessarily (*c*), as it requires having multiple pre-intervention data points. One important advantage of our data is exactly this availability of multiple pre-intervention data points, which will allow us to explicitly test the last assumption (*c*) of common time trends across treatment groups (e.g., see prior MLSFH studies[6,7]).

A.0.b. Identifying pathways between SDPs and behaviors: While the above analyses identify the effect of health-information on ***SDPs***, health, and behaviors, they do not provide detailed insights into the *pathways* of how updated ***SDPs*** affect outcomes. To answer questions such as "Did individuals benefit by updating their ***SDPs*** by receiving new health-information, and *how* did this information affect their subsequent behaviors?," we will build on the analytic framework in Delavande & Kohler[8] (forthcoming, *Rev. of Econ. Studies*) that is derived from an economic intertemporal choice model with uncertainty. For example, the probability of individual i choosing a certain health behavior $a_i = 0$ or 1, such as engaging in risky sex, is given as $\text{Prob}(a_i = 1) = \text{Prob}\{V_i(1) + f_i S_i^+ U_i^+ + (1 - f_i)[p_i(1) S_i^+ U_i^+ + (1 - p_i(1)) S_i^- U_i^-] + \varepsilon_{i1} \geq V_i(0) + f_i S_i^+ U_i^+ + (1 - f_i)[p_i(0) S_i^+ U_i^+ + (1 - p_i(0)) S_i^- U_i^-] + \varepsilon_{i0}\}$, where $V_i(a)$ is the immediate utility from behavior a_i, U_i^+ and U_i^- are the status-dependent period 2 utility if i is HIV+ or HIV– respectively, f_i is the subjective probability of being HIV+ at the beginning of period 1, S_i^+ and S_i^- are the HIV-status dependent probabilities of surviving from period 1 to period 2, and ε_i represents unobserved heterogeneity in preferences, etc. $p_i(a)$ is the subjective probability of becoming HIV+ as a function of behavior a. Modification of this model for other health or life-cycle behaviors (smoking, work efforts, savings) is straight forward. Using our extensive data on ***SDPs*** and related expectations, the parameters of this model can be obtained using Roodman's[9] MLE-estimator for a recursive set of equations, where equations reflect **(*i*)** the disease/mortality perceptions *prior* to i's access to the health-information provided as part of our intervention, **(*ii*)** probabilities of attrition, **(*iii*)** updates of the perceptions f_i, S_i^+ and S_i^- based on health-information, behavioral changes and contextual events such as local mortality, and **(*iv*)** individuals' choices about their health behavior a_i as a function of ***SDPs*** (f_i, S_i^+ and S_i^-) and other characteristics. Identification is obtained from exogenous variation in access to health-information resulting from our randomized intervention. Additional exclusion restrictions using exogenous variation in the MLSFH identify other equations of the recursive system, as has been illustrated in our prior work.[8,10] The results from this structural model then reveal the determinants of the initial expectations prior to the intervention, how individual behaviors and access to new information cause individuals to update their perceptions f_i, S_i^+ and S_i^-, and how these updated perceptions affect health behaviors a_i and later life behaviors/outcomes (e.g., health, marriage, work efforts, intergenerational transfers). Importantly, estimated parameters can be used for health-policy simulations, as we have illustrated previously.[8]

A.0.c. Sample size, attrition and power considerations: One indicator for the adequate power of the proposed analyses is our extensive mature-adult research that has informed the proposed research design.[11–16] This conclusion is supported by more detailed power calculations. In 2012, the MLSFH mature adult survey completed 1,266 surveys, 90% of the 1,402 eligible respondents selected based on 2010 enrollment criteria; the 2013 MLSFH mature adult survey reinterviewed 1,203 (95%) of the 2012 respondents (plus additional who were absent in 2012). A $N \approx 1,500$ is expected for the 2016 MLSFH survey, ≈90% (≈1,350) of whom can be expected to be retained until 2018. Power calculations using *Optimal Design Software*[17] for (village-level) clustered randomized experiment with individual outcomes ($N_{2016-18} = 1,350$, power = 80%, $\alpha = .05$, R^2_{L2}=.2) suggest that our study design is able to identify minimum detectable effect sizes (MDES[18,19]) of 19–22% for expected levels of within-village intra-class correlation in health-related outcomes in the range of .07–.15. Changes in ***SDPs***, health and life-cycle behaviors as a result of the health-information intervention of ≈20% or higher will be detectable within our study design. Based on the existing literature, we expect that the health-information intervention will result in substantially larger changes in our measures of ***SDPs***. Power of the study design will be substantially stronger for analyses that utilize the multiple longitudinal measures collected as part of this study.[17]

B. Literature Cited:

[1] Heckman JJ, Vytlacil EJ. Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation. In *Handbook of Econometrics*, editors Heckman JJ, Leamer EE, Amsterdam: Elsevier, vol. Volume 6B, chap. 70, pp. 4779–4874. 2007.

[2] Heckman JJ, Vytlacil EJ. Econometric evaluation of social programs, Part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In *Handbook of Econometrics*, editors Heckman JJ, Leamer EE, Amsterdam: Elsevier, vol. Volume 6B, chap. 70, pp. 4875–5143. 2007.

[3] Abbring JH, Heckman JJ. Econometric evaluation of social programs, part iii: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In *Handbook of Econometrics*, editors Heckman JJ, Leamer EE, Amsterdam: Elsevier, vol. Volume 6B, chap. 72, pp. 5145–5303. 2007.

[4] Moffitt RA. Issues in the estimation of causal effects in population research, with an application to the effects of teenage childbearing. In *Causal Analysis in Population Studies*, editors Engelhardt H, Kohler HP, Fürnkranz-Prskawetz A, Berlin: Springer Verlag, The Springer Series on Demographic Methods and Population Analysis, chap. 2, pp. 9–29. 2009. 5.

[5] Moffitt R. Remarks on the analysis of causal relationships in population research. *Demography* 2005; **42(1)**:91–108.

[6] Baranov V, Bennett D, Kohler HP. The indirect impact of antiretroviral therapy: Mortality risk, mental health, and HIV-negative labor supply. *Journal of Health Economics* 2015; **44**:195–211.

[7] Baranov V, Kohler HP. The impact of AIDS treatment on savings and human capital investment in Malawi, 2014. Population Studies Center, University of Pennsylvania, Working Paper #14–3. http://repository.upenn.edu/psc_working_papers/55.

[8] Delavande A, Kohler HP. HIV/AIDS-related expectations and risky behavior in Malawi. *Review of Economic Studies* 2016; **83(1)**:118–164.

[9] Roodman D. Fitting fully observed recursive mixed-process models with cmp. *Stata Journal* 2011; **11(2)**:159–206.

[10] Delavande A, Kohler HP. The impact of HIV testing on subjective expectations and risky behavior in Malawi. *Demography* 2012; **49(3)**:1011–1036.

[11] Payne C, Mkandawire J, Kohler HP. Disability transitions and health expectancies among adults 45 years and older mature in Malawi: A cohort modeling approach. *PLOS Medicine* 2013; **10(5)**:e1001435.

[12] Kohler IV, Payne CF, Bandawe C, Kohler HP. The demography of mental health among mature adults in a low-income high HIV-prevalence context, 2015. Popuation Studies Center, Univeristy of Pennsylvania, PSC Working Paper Series, WPS 15-1. Revise and resubmit at *Demography*. http://repository.upenn.edu/psc_working_papers/59/.

[13] Payne CF, Kohler HP. The population-level impact of ART rollout on adult health and mortality in rural Malawi, 2015. Unpublished manuscript, under review at *PLoS Medicine*.

[14] Kohler HP, Behrman JR, Kohler IV, Romero J. Survival expectations, work efforts, savings and intergenerational transfers among mature adults in Malawi, 2015. Unpublished Manuscript, Population Studies Center, Universtiy of Pennsylvania, Philadelpia, PA.

[15] Kohler I, Payne C, Kohler HP. The link between functional limitations and mental health among mature and elderly adults in rural Malawi, 2013. Paper presented at the Annual Meeting of the Population Association of America, New Orleans, LA, April 11–13, 2013. http://paa2013.princeton.edu/abstracts/131894.

[16] Myroniuk TW, Kohler HP, Kohler I. Are marital dissolutions actually bad for health? Longitudinal evidence from older Malawians who survived the HIV epidemic, 2015. Manuscript submitted to 2016 Annual Meeting of the Population Association of America.

[17] Raudenbusch SW, *et al.* Optimal design software for multi-level and longitudinal research (version 3.01) (power tests), 2011. http://sitemaker.umich.edu/group-based/.

[18] Dupont WD, Plummer WD. Power and sample size calculations for studies involving linear regression. *Control Clin Trials* 1998; **19(6)**:589–601. http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize.

[19] Dupont WD, Plummer WD. Power and sample size calculations: A review and computer program. *Controlled Clinical Trials* 1990; **11(2)**:116–128. http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize.