WOMEN'S INTERAGENCY HIV STUDY
SECTION 40: WIHS/MACS GEOCODING PROTOCOL

A. BACKGROUND & STUDY PURPOSE

Suppression of plasma HIV RNA with combination ART (cART) has critical implications for the personal health of HIV+ individuals and their ability to transmit infection. Despite these clear associations, a substantial proportion (23%) of HIV-infected people who are prescribed cART in the US do not achieve virologic suppression. Black men and women are less likely than Whites to be prescribed cART and achieve virologic suppression when cART is prescribed.[1] The reasons for these findings are complex, but poverty, income inequality, and access to health care undoubtedly play a significant role.

The link between individuals' socioeconomic status and health outcomes, including HIV infection, has been well established.[2] [3], [4] In addition to individual-level poverty, the socioeconomic context (social and economic features of the individual's environment) is also a determinant of health.[5] Several studies have documented the association of neighborhood-level poverty with morbidity and mortality due to HIV and other diseases, but the pathways between contextual poverty and these outcomes are unclear.[6],[7], [8], [9], [10], [11] In addition to poverty, income *inequality* ("relative disparity with which income is distributed in a population"[2]) is itself associated with increased mortality in the general population.[12], [2], [13] Substantial income inequality, such as the level observed in the US, has also been associated with increased rates of sexually transmitted infections (STIs)I, including HIV infection.[14], [15] Despite the mounting evidence that these contextual and structural inequities affect health, their effects on critical markers of HIV control (e.g., virologic suppression) and effective primary care (e.g., blood pressure control) has received relatively little attention.

Health care availability and quality are also important social determinants of health.[16] About 15% of the US population lacks health insurance; Blacks and Hispanics, racial/ethnic groups that are disproportionately both poor and affected by HIV, are most likely to be uninsured.[17] For example, at the UNC clinic as of 2010, 41% of patients enter care without health insurance, and 31% have public (Medicaid/Medicare) health insurance.[18] A number of studies have documented the association between lack of health insurance and decreased health care utilization. For example, lack of health insurance and changes in insurance result in decreased use of cART.[19], [20] Lack of insurance is also associated with lower rates of blood pressure (BP) control among patients with treated hypertension.[21] as well as uncontrolled glucose among individuals with diabetes.

Eligibility rules for Medicaid, the nation's insurer for poor and low-income individuals and primary insurer for HIV+ individuals, vary considerably in their stringency from state to state.[22] Moreover, Medicaid coverage is unstable, since patients fall in and out of eligibility categories.[23] One analysis demonstrated that almost two-thirds of new Medicaid recipients lost coverage within one year.[24] Not surprisingly, the relationship between Medicaid and HIV outcomes is complex, and publicly insured patients have increased mortality compared to those with private insurance.[25]

Although the association between insurance status and mortality among HIV+ patients has been documented,[26] no published studies have evaluated the effect of insurance on HIV virologic suppression – an outcome that will be critically important if the US is to realize the goals of the National HIV/AIDS Strategy.[27] Changes in health care policy, such as The Affordable Care Act (ACA) and other changes in insurance, are structural interventions that could conceivably alter the effects of socioeconomic context by substantially decreasing the number of (predominantly poor and low-income) people who lack health insurance, thereby dramatically affecting health outcomes for women and men with and at risk for HIV infection.

In our conceptual model, neighborhood-level poverty worsens people's health through deprivation of important resources, such as health care, education, and affordable food and housing. Income inequality at county and state levels, where policies are made and implemented, reflects differences in power (economic, political, and legal) that influence health outcomes through policies and actions that control access to resources. Health insurance permits access to care, with improved control of HIV infection and other disease outcomes. Structural interventions, such as the ACA, could improve health care access and potentially mitigate the effects of poverty and income inequality.

By documenting the role of social determinants, such as poverty, income inequality, and health insurance, on disease outcomes, and evaluating whether these factors provide additional explanatory information beyond individual level demographic and clinical data, this study will improve understanding of pathways and generate evidence to support policies that can decrease inequalities in health outcomes.

This project has potential to add considerable value to the WIHS and MACS. While this concept focuses on poverty and income inequality, spatial mapping of each participant will also allow evaluation of the impacts of many other neighborhood-level factors of importance on a broad range of outcomes, once participants' addresses are linked to census-linked datasets. This concept also provides the infrastructure to support joint analyses in the MACS and WIHS leveraging two decades of longitudinal data collection across the cohorts. ***Finally, this concept represents a unique opportunity for a natural experiment using the premier cohorts of men and women with and at risk for HIV infection in the US to evaluate the effects of the most significant changes in federal health insurance coverage since Medicare and Medicaid were created more than 40 years ago.***

B. SCIENTIFIC AIMS

The overall aim of this study is to evaluate the relationship between virologic suppression and serious AIDS-defining events (both among HIV+ women/men), blood pressure control (among all women/men diagnosed with hypertension), diabetes (among all women/men diagnosed with diabetes) and mortality (among all women/men) and:

- Socioeconomic context (e.g., proportion of population below the federal poverty line in each participant's census block group, income inequality in each participant's county, etc.)
- Local and national policies that affect health
- Participant's health insurance status

Outcome variables: Adverse outcomes

- Among HIV+ and HIV- individuals with a diagnosis of hypertension: Adverse outcomes will be defined as either: 1) Documentation of BP > 140/90 during a study visit after a diagnosis of hypertension or 2) Death.
- Among HIV+ and HIV- individuals with a diagnosis of diabetes: Adverse outcomes will be defined as either: 1) Documentation of HgbA1c >7.0 percent during a study visit after a diagnosis of diabetes or 2) Death.
- Among HIV+ individuals only: Adverse outcomes will be defined as either: 1) HIV RNA > 200 copies/mL more than 6 months after initiation of cART; 2) serious AIDS-defining events (defined as WHO Clinical Stage 4 or diagnosis of tuberculosis) or (3) Death.

Exposure variables: Geographic measures of socioeconomic inequality, including: Median income of population in each participant's census block group; Gini coefficient (a measure of income inequality) among residents in each participant's county; Medicaid income eligibility requirement in each participant's state; expansion of Medicaid eligibility criteria; subsidization of lower income families for purchase of health insurance; and timing of adoption of essential benefits; current insurance status: private insurance with or without prescription drug coverage; Medicare without private insurance (but including Medicaid) i.e., Medicare alone; Medicaid alone; and no health insurance.We will explore how to categorize insurance over time, allowing for availability of coverage.

Covariates: Gender, age, race/ethnicity, body mass index, annual household income, substance use, census block group characteristics (e.g., population density, race distribution). HIV+ individuals: CD4 history, HIV RNA history, previous AIDS-defining condition and cART use, time updated cART exposure.

C. STUDY DESIGN

We propose a multilevel analysis of data collected by the WIHS and MACS. All study participants will be eligible for this study, although some hypotheses will involve subsets of the cohort (e.g., individuals with diagnosis of hypertension, diabetes, or HIV infection).

Most of the information needed to complete this study is routinely obtained during the course of study visits. Standard social and medical history, physical examination, and laboratory examination during baseline and follow-up study visits routinely capture: the participant's address; medical diagnoses, including hypertension and diabetes; interval medical complications since the last visit; medication use, blood pressure, HgbA1c, and determination of plasma HIV RNA levels. Staff also routinely ask each participant about health insurance and estimated annual household income as part of their biannual interview.

In order to collect contextual characteristics, WIHS/MACS staff at each site will collect addresses from consented participants, geocode them (using ArcGIS) and assign a census block group (see **Figure 1**: Data Management; **Appendix A:** Geocoding Protocol; and **Appendix B:** Consent Tools). ArcGIS matches each participant's geocoded location to a Federal Information Processing Standard (FIPS) code that identifies geographic locations in the US. We will use a twelve-digit version of the FIPS code that identifies locations at the census block group level.

Each site creates a limited dataset that contains only the participants' FIPS codes and study IDs (e.g., WIHSID or MACSID) and securely transfers the dataset to the UNC WIHS site, which manages the geocoding data. This limited dataset only includes the study IDs and the corresponding FIPS for each participant, and will only be available to a very limited group of WIHS staff who work directly under the supervision of Dr. Adimora to create analytic datasets for investigators with EC-approved concept sheets.

The UNC WIHS site links the FIPS codes to census-linked datasets, such as the American Community Survey, Decennial Census, American Housing Survey, and Annual Economic Survey, to create group-level variables that describe the locations where participants live. The UNC WIHS creates individualized datasets for investigators, which only contain the study IDs and the values of specific group-level variables requested by the investigator (e.g., percent of individuals under the federal poverty line, median household income, percentage of individuals with high school education). **The FIPS codes that are received by the UNC WIHS site will never be provided as part of any analytic dataset; they will be kept in a password-protected folder on the UNC server, which contains extensive security protections.**

Figure 1. Geocoding Protocol Data Management

D. DATA ANALYSIS

We will measure poverty by the proportion of the census tract that is below the federal poverty line and will also explore other income measures (e.g. median income) of socioeconomic inequality. We anticipate that we will primarily use the Gini measure for income inequality because it is available for all locales, and has been used successfully to measure income inequality at the county level in NC.[32] In contrast, measures like P90/10 are sometimes unavailable because some areas may have 10^{th} percentile incomes of $0.[33] We will measure inequality at the county level because NC MSA's are based on county units[32] and counties administer distribution of meaningful resources (e.g. health departments, social services).

We will use Cox proportional hazards regression models for time-to-event outcomes (e.g., mortality, incident serious AIDS condition, virologic failure) and linear-logistic regression models for binary outcomes (e.g., loss of BP control), as shown in Table 1. We will use a robust variance estimate to account for possible clustering of individuals into census tract groups.[34] We will explore adjustment for the variables footnoted in Table 1, as well as other variables that arise during study refinement with the WIHS/MACS cohort investigators. We will fit continuous exposures and covariates using restricted quadratic splines,[35] or categorize as needed. We will explore the use of propensity scores as an alternative to regression adjustment.[36] For Cox and logistic models, we will explore model fit using a combination of graphical techniques and residual diagnostics. For Cox models, we will explore the proportional hazards assumption using a plot of the log cumulative hazard by time, as well as fitting the product term between key variables and time.

For policy-based analyses, we will treat the date of the policy-related legislative rollout as a time-dependent instrumental variable.[37] In all analyses, all statistical estimates will be tabulated along with their 95% CIs. For each analysis, after obtaining initial results as described above and in Table 1, we will explore secondary analyses to assess the robustness of findings and generate new hypotheses. For example, if we support our hypothesis of group-level associations, we will then explore to what extent

these group-level associations are mediated by individual-level factors (e.g. cART exposure, viral load, CD4) using marginal structural models.[38]

Table 1. Overview of Analysis

Study Population	Exposures	Outcome	Analysis *
HIV+ and HIV-	• Proportion of CTG** population below poverty line • Gini income-inequality coefficient, county level • Health insurance categories (private, public, none) • Calendar date of ACA implementations	All-cause mortality	Cox regression, with clusters accounted via GEE robust variance
HIV+ and HIV- with diagnosis of HTN	• Proportion of CTG population below poverty line • Gini income-inequality coefficient, county level • Health insurance categories (private, public, none) • Calendar date of ACA implementations	18-month BP control (BP<140/90)	Logistic regression, with clusters accounted for via GEE robust variance
HIV+ and HIV- with diagnosis of Diabetes	• Proportion of CTG population below poverty line • Gini income-inequality coefficient, county level • Health insurance categories (private, public, none) Calendar date of ACA implementations	18-month glucose control (A1C)	18-month glucose control (A1C
HIV+	• Proportion of CTG population below poverty line • Gini income-inequality coefficient, county level • Health insurance categories (private, public, none) • Calendar date of ACA implementations	Time to next serious AIDS condition	Cox regression, with clusters accounted via GEE robust variance
HIV+, on ART	• Proportion of CTG population below poverty line • Gini income-inequality coefficient, county level • Health insurance categories (private, public, none) • Calendar date of ACA implementations	Time to virologic failure (VL>200 c/ml)	Cox regression, with clusters accounted via GEE robust variance

***** Covariates will vary by analysis, but include: gender, age, race/ethnicity, BMI, and where applicable HIV serostatus, CD4 cell count, HIV viral load, AIDS defining illnesses, BP, health insurance, individual income, and census block group characteristics (e.g., population density and race distribution).
** Census tract group

Power Calculations: We provide estimates of statistical power for 18-month BP and glucose control as well as virologic failure, which apply to continuous exposures such as (a) poverty and (b) income inequality. Power is estimated using large-sample approximations assuming 2-sided 95% CIs (a type 1 error of 5%).

The power calculations below are based on enrollment in the WIHS. We anticipate that the larger sample size of the MACS will afford greater power for the proposed analyses.

For 18-month BP control, we expect 1140 (30%) of 3800 entire WIHS cohort women to be diagnosed with hypertension. We assume 171 (15%) of these 1140 women will experience lack of BP control within 18-months of diagnosis. We further assume that the clustering at the CBG or county level will cause a doubling of our variance. Comparing the upper to lower tertile of poverty or income inequality, we will have 90%, 80%, and 70% statistical power to detect ORs of 1.48, 1.38, and 1.32, respectively.
For 18-month glucose control, we expect 570 (15%) of 3800 entire WIHS cohort women to be diagnosed with diabetes. We assume 114 (20%) of these 570 women will experience lack of glucose control within 18- months of diagnosis. We further assume that the clustering at the CBG or county level will cause a doubling of our variance. Comparing the upper to lower tertile of poverty or income inequality, we will have 90%, 80%, and 70% statistical power to detect ORs of 1.53, 1.38, and 1.24, respectively.
For virologic failure, we expect at least 1680 (60%) of 2800 entire WIHS cohort HIV+ women to initiate ART. We assume 336 (20%) of these 1680 women will experience virologic failure during a 3-year follow up period. We again further assume that the clustering at the CBG or county level will cause a doubling of our variance. Comparing the upper to lower tertile of poverty or income inequality, we will have 90%, 80% and 70% statistical power to detect hazard ratios of 1.30, 1.24, and 1.20, respectively.

CONTACT FOR QUESTIONS ABOUT THIS PROTOCOL

Catalina Ramirez
Project Director, UNC WIHS
130 Mason Farm Rd. CB 7030
Chapel Hill, NC 27599
catalina_ramirez@med.unc.edu

References

1. Centers for Disease Control and Prevention, *Vital signs: HIV prevention through care and treatment--United States.* MMWR. Morbidity and mortality weekly report, 2011. **60**: p. 1618-23.
2. Lynch, J.W., et al., *Income inequality and mortality in metropolitan areas of the United States.* American journal of public health, 1998. **88**(7): p. 1074-1080.
3. Cunningham, W.E., et al., *The effect of socioeconomic status on the survival of people receiving care for HIV infection in the United States.* Journal of health care for the poor and underserved, 2005. **16**(4): p. 655-76.
4. McMahon, J., et al., *Poverty, hunger, education, and residential status impact survival in HIV.* AIDS and Behavior, 2011. **15**(7): p. 1503-11.
5. World Health Organization. *Social determinants of health*. 2010 3/22/10]; Available from: http://www.who.int/social_determinants/en/.
6. Ellerbrock, T.V., et al., *Heterosexually transmitted human immunodeficiency virus infection among pregnant women in a rural Florida community.* N Engl J Med, 1992. **327**(24): p. 1704-9.
7. Cubbin, C. and M.A. Winkleby, *Protective and harmful effects of neighborhood-level deprivation on individual-level health knowledge, behavior changes, and risk of coronary heart disease.* American journal of epidemiology, 2005. **162**(6): p. 559-68.
8. Zierler, S., et al., *Economic deprivation and AIDS incidence in Massachusetts.* American journal of public health, 2000. **90**(7): p. 1064-73.
9. St. Louis, M.E., et al., *Human immunodeficiency virus infection in disadvantaged adolescents. Findings from the US Job Corps.* JAMA, 1991. **266**(17): p. 2387-91.
10. Joy, R., et al., *Impact of neighborhood-level socioeconomic status on HIV disease progression in a universal health care setting.* Journal of acquired immune deficiency syndromes, 2008. **47**(4): p. 500-5.
11. McDavid Harrison, K., et al., *County-level socioeconomic status and survival after HIV diagnosis, United States.* Annals of Epidemiology, 2008. **18**(12): p. 919-27.

12. Lochner, K., et al., *State-level income inequality and individual mortality risk: a prospective, multilevel study.* American journal of public health, 2001. **91**(3): p. 385-91.
13. Kaplan, G.A., et al., *Inequality in income and mortality in the United States: analysis of mortality and potential pathways.* BMJ, 1996. **312**(7037): p. 999-1003.
14. Aral, S.O., *The social context of syphilis persistence in the southeastern United States.* Sexually transmitted diseases, 1996. **23**(1): p. 9-15.
15. Holtgrave, D.R. and R.A. Crosby, *Social capital, poverty, and income inequality as predictors of gonorrhoea, syphilis, chlamydia and AIDS case rates in the United States.* Sex Transm Infect, 2003. **79**(1): p. 62-4.
16. CSDH, *Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of health*, 2008, World Health Organization: Geneva.
17. DeNavas-Walt, C., B.D. Proctor, and J.C. Smith, *Income, poverty, and health insurance coverage in the United States: 2008*, in *Current Population Reports, P60-236*2009, U.S. Census Bureau: Washington, D.C.
18. Howe, C.J., et al., *Enrollment, retention, and visit attendance in the University of North Carolina Center for AIDS Research HIV clinical cohort, 2001-2007.* AIDS Res Hum Retroviruses, 2010. **26**(8): p. 875-81.
19. Smith, S.R. and D.M. Kirking, *The effect of insurance coverage changes on drug utilization in HIV disease.* Journal of acquired immune deficiency syndromes, 2001. **28**(2): p. 140-9.
20. Smith, S.R. and D.M. Kirking, *Access and use of medications in HIV disease.* Health services research, 1999. **34**(1 Pt 1): p. 123-44.
21. Duru, O.K., et al., *Health insurance status and hypertension monitoring and control in the United States.* American journal of hypertension, 2007. **20**(4): p. 348-53.
22. Kaiser Family Foundation. *Medicaid/CHIP Eligibility*. 2009 January 15, 2012]; Available from: http://facts.kff.org/chart.aspx?cb=56&sctn=151&ch=988.
23. Rosenbaum, S., *Medicaid.* The New England journal of medicine, 2002. **346**(8): p. 635-40.
24. Carrasquillo, O., et al., *Can Medicaid managed care provide continuity of care to new Medicaid enrollees? An analysis of tenure on Medicaid.* American journal of public health, 1998. **88**(3): p. 464-6.
25. Palella, F.J., Jr., et al., *Increased mortality among publicly insured participants in the HIV Outpatient Study despite HAART treatment.* Aids, 2011. **25**(15): p. 1865-76.
26. Goldman, D., et al., *The effect of insurance on mortality in an HIV + population in care.* Journal of the American Statistical Association, 2001. **96**(405): p. 883-894.
27. White House Office of National AIDS Policy. *National HIV/AIDS strategy for the United States*. 2010 August 2, 2010]; Available from: http://www.whitehouse.gov/sites/default/files/uploads/NHAS.pdf.
28. Schootman, M., et al., *Positional accuracy and geographic bias of four methods of geocoding in epidemiologic research.* Annals of Epidemiology, 2007. **17**(6): p. 464-70.
29. Krieger, N., et al., *On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research.* American Journal of Public Health, 2001. **91**(7): p. 1114-6.
30. U.S. Bureau of the Census, G.D. *United States Census 2010: Geographic products and information*. 2009 October 28, 2011 [cited 2012 February 5, 2012]; Available from: http://www.census.gov/geo/www/2010census/.
31. Krieger, N., et al., *Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project.* American journal of epidemiology, 2002. **156**(5): p. 471-82.
32. Brodish, P.H., M. Massing, and H.A. Tyroler, *Income inequality and all-cause mortality in the 100 counties of North Carolina.* Southern medical journal, 2000. **93**(4): p. 386-91.
33. Weinberg, D.H., *U.S. neighborhood income inequality in the 2005-2009 period*, in *American Community Survey Reports*2011, US Census Bureau.

34. Wei, L.J., D.Y. Lin, and L. Weissfeld, *Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions.* Journal of the American Statistical Association, 1989. **84**(408): p. 1065-1073.
35. Howe, C.J., et al., *Splines for trend analysis and continuous confounder control.* Epidemiology, 2011. **22**(6): p. 874-5.
36. Joffe, M.M. and P.R. Rosenbaum, *Invited commentary: propensity scores.* American journal of epidemiology, 1999. **150**(4): p. 327-33.
37. Cain, L.E., et al., *Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable.* American journal of epidemiology, 2009. **169**(9): p. 1124-32.
38. VanderWeele, T.J., *Marginal structural models for the estimation of direct and indirect effects.* Epidemiology, 2009. **20**(1): p. 18-26.

APPENDIX A. GEOCODING PROTOCOL

1. Documenting Changes in Consent Status (WIHS ONLY)

Each site is responsible for reviewing the geocoding protocol with each participant and documenting the participant's consent status (see **Appendix B** for draft consent language and a visual tool to aid with consent).

Any changes in consent status (e.g., participant agreement to participate, or withdrawal of consent) should be documented at each visit using the Consent Tracking Form (CONS, see Figure 2) which is data-entered into Apollo.

NOTE: Sites DO NOT need to fill out a CONS form unless there is a change in consent status.

Figure 2. WIHS Consent Tracking Form (CONS)

WOMENS INTERAGENCY HIV STUDY
CONSENT TRACKING (CONS)

A1. PARTICIPANT ID: |___|-|___|___|___|___|-|___|___|___|___|-|___|

A2. WIHS STUDY VISIT #: |___|___|

A3. FORM COMPLETED BY: |___|___|___|

A4. DATE FORM COMPLETED: ___ ___ / ___ ___ / ___ ___ (M / D / Y)

A5. FORM VERSION: 04/01/16

NOTE: COMPLETE FORM ONLY IF RESPONSE TO A6a, A7a, A8a, A9a, A10a, or A11a IS YES.

CONSENT TYPE	a. CONSENT AVAILABLE FOR THIS VISIT		b. CONSENT STATUS AT THIS VISIT	
	YES	NO	CONSENTED	DID NOT CONSENT
A6. Genetics	1	2 (A7)	1	2
A7. B-cell	1	2 (A8)	1	2
A8. Commercial use of B-cells	1	2 (A9)	1	2
A9. GWAS	1	2 (A10)	1	2
A10. dbGaP	1	2 (A11)	1	2
A11. Geocoding	1	2 (END)	1	2

2. Collecting Participant Addresses

At the time of their visit (or by phone, if an abbreviated visit is conducted) site staff should ask each participant for the address of their primary residence (e.g., where they have spent most nights in the past six months). If a participant reports multiple addresses, ask for the address where he resides while getting medical care. **Site staff should confirm the address with the participant to ensure that it is correct.**

Staff should document this information in a site-specific log (see **Figure 3.** Example Address Tracking Log). We recommend Excel which can be easily imported into ArcGIS – but sites may choose to use another file-type such as SAS or Access, or other existing system to collect this information and then export it into a *csv*, *xls*, or *xlsx.* file.
The site-specific log should include the following six fields:

- WIHSID
- Street Number & Street Name
- Apartment Number (if applicable)
- City

- State

There may be cases that require staff to ask for additional information:

If a participant will only provide a PO Box→ Ask the participant whether it would be possible for them to provide a physical address that will only be used for geocoding purposes. If they are unwilling to provide a physical address, please make a note on your address log to document that.

If a participant is currently incarcerated or in rehab→ Individuals who are incarcerated or otherwise institutionalized often have limited interaction with their immediate environment. For that reason, they are routinely excluded from geocoding-related analyses. If an individual is incarcerated or in rehab, a site should not include the address of the facility. The UNC CRS staff review the residence information provided by participants on their sociodemographic questionnaire and will not be expecting to receive any data from individuals who list their primary residence as "*jail/prison*" or "*rehab.*"

If a participant is currently homeless→ If the participant has been homeless for <3 months or usually sleeps at a residential location (e.g., friend's home), site staff should document their prior residence or the location where they sleep most nights. However, if individuals are "*living on the street,*" they should be excluded from this protocol.

If a participant is currently living outside of the United States→ Ask the participant if he has a second residence in the US where he spends part of the year. If not, the participant will not be able to be geocoded and will be excluded from this protocol.

Figure 3. Example of Address Tracking Log

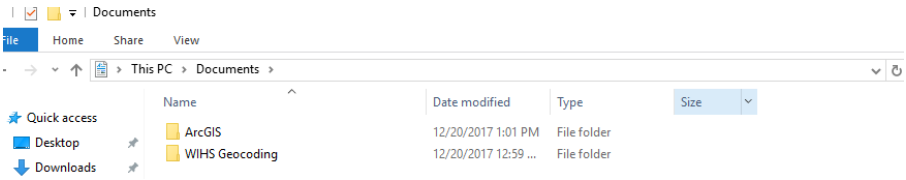
WIHSID	Street	Street2	City	State	Zip	Notes
51001	123 E Franklin St		Chapel Hill	NC	27514	
51002	654 Mason Farm Rd		Chapel Hill	NC	27514	
51003	104 NC 54	Apt 102	Carrboro	NC	27510	
51004	PO Box 1054		Chapel Hill	NC	27516	Declined to provide physical address

3. Preparing Your Geocoding File

After the end of each even visit, site staff should prepare the address tracking log for geocoding by completing the following steps:

- **Ensure that all consented individuals who attended the current visit are included in your log.** This includes individuals who completed a full visit and those who completed an abbreviated visit.
- **Ensure that there is an address, or a reason why address is missing, for all individuals.** Each study ID should have an associated address or a note indicating the reason(s) that an address was not provided (e.g., PO Box only – declined to provide address).
- **Make sure that your file does not include individuals who were institutionalized or homeless for >3 months or "living on the street"**.

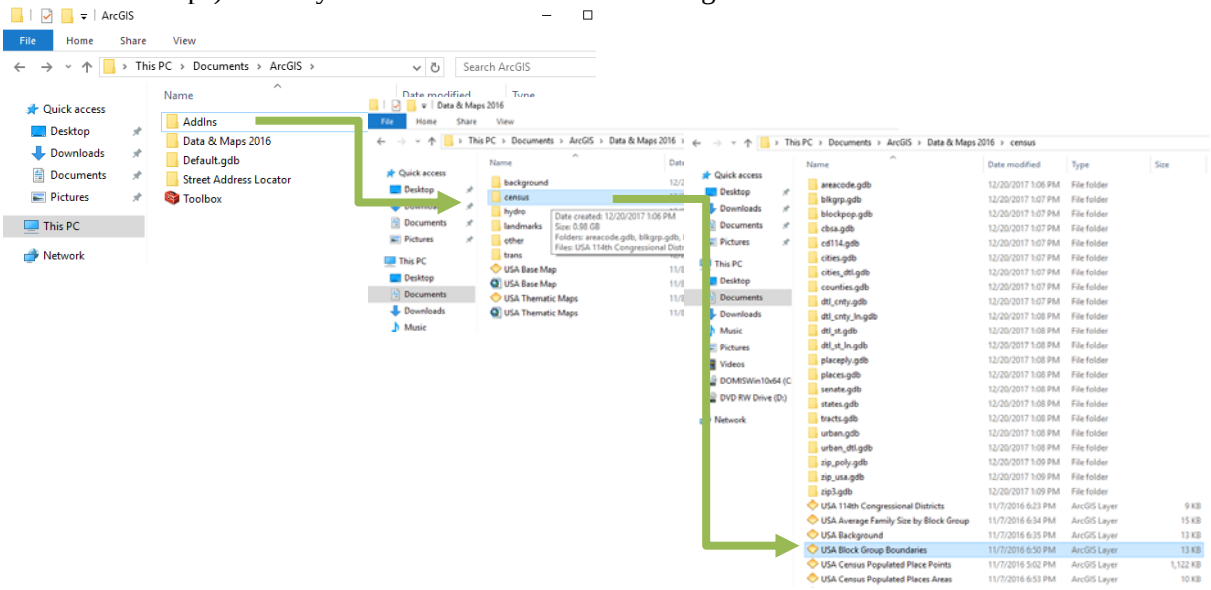
Once you have reviewed the file for completeness and accuracy, save the tracking log file in the geocoding folder called: Documents/WIHS Geocoding/VISIT# so that you can easily access the files during geocoding. A new folder should be created for even visit. The tracking log should be saved in Excel format (e.g., *.xls* or *.xlsx*). In order to easily identify the file, we suggest the following naming convention: VISIT#_Address Tracking_SITE NAME_ DD-MM-YY. We also recommend saving all files in the same location at every visit – see example "Geocoding" folder below.



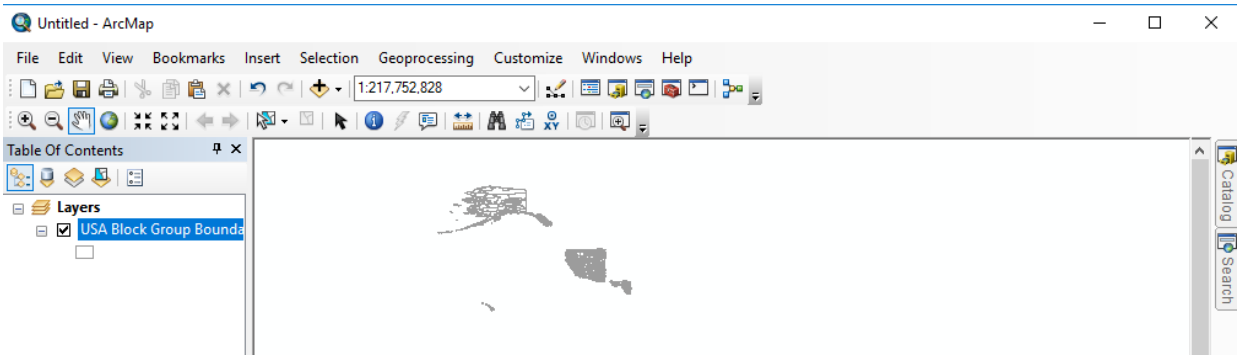
4. Geocoding Participant Addresses

Step 1: To start the geocoding process, you will need to navigate to the USA census block group map layer file for ArcGIS. All UNC-provided computers have been loaded with the necessary map layer files, which can be found in the following folder: **Documents/ArcGIS/Data and Maps 2016/census**.

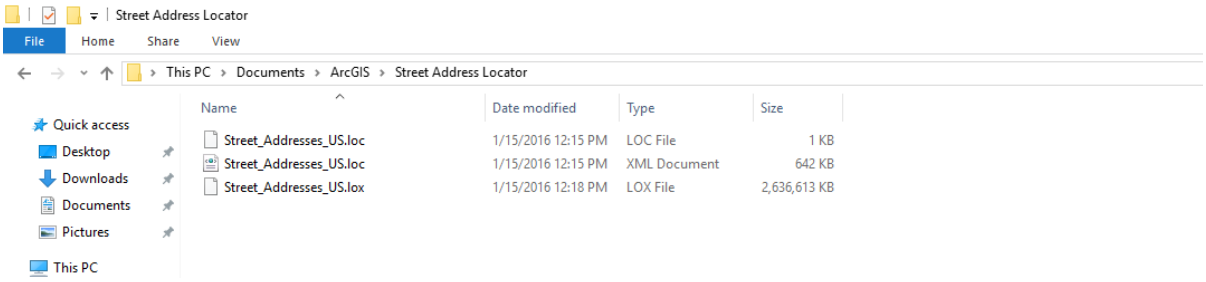
> **NOTE:** If you cannot find the USA census block group map layer file, or are using a non-UNC laptop, then you may download the file from http://www.arcgis.com/home/. If you download the file, make sure to save it in your ArcGIS folder (create a folder called: Documents/ArcGIS/Data and Maps) so that you do not have to download it again in the future.



Step 2: Double-click on the census block group layer file to launch ArcGIS. The file can take up to a minute to open Once the files start to load, ArcGIS should open and you should begin to see a US map appear.

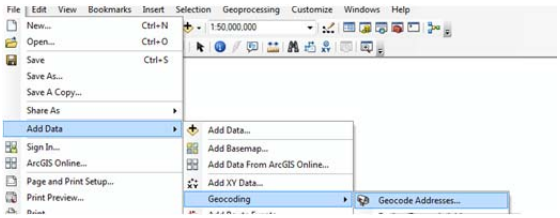


Step 3: Before you can start geocoding, you need to ensure that the Street Address locator file is available. This file contains information that allows the program to locate the addresses in your tracking log and place them on the census map that we loaded in Step 2. All UNC-provided computers were loaded with the necessary locator files, which can be found in the following folder: **Documents/ArcGIS/Street Address Locator**.

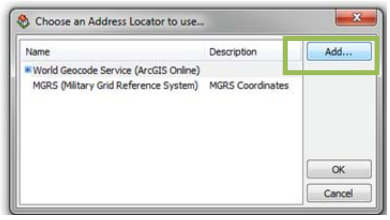


> **NOTE:** If you cannot find the locator files, or are using a non-UNC laptop, then you may request a copy of the locator files from the UNC WIHS site. If you download the files, make sure to save them in your ArcGIS folder (create a folder called: Documents/ArcGIS/Street Address Locator) so that you do not have to download them again in the future.

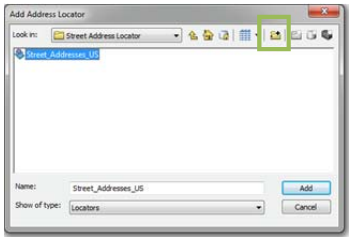
Step 4: Now that ArcGIS is open and has the correct shapefile loaded, you can begin the Geocoding process by going to: File→Add Data→ Geocoding→Geocode Addresses



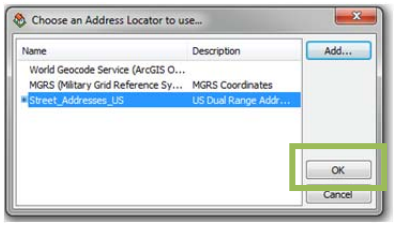
A pop-up box will appear asking you to select the correct "Locator." ArcGIS comes with two pre-loaded locator files, but in order to ascertain census block groups, we will need to add the new Street Addresses_US file that you just identified in Step 3. Start this process by selecting the button that says "Add".



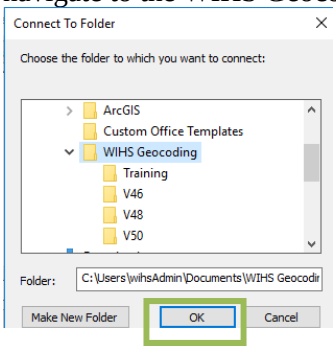
A pop-up screen will appear asking you to direct ArcGIS to the location of the locator file. Use the  button to open a file explorer window and navigate to the locator file. Highlight the locator file (as shown below) and select the option to "Add" the file.



Once you have selected the Street Address file, it should be added to the list of available "Address Locators." Select the Street_Addresses_US file and click "Ok."

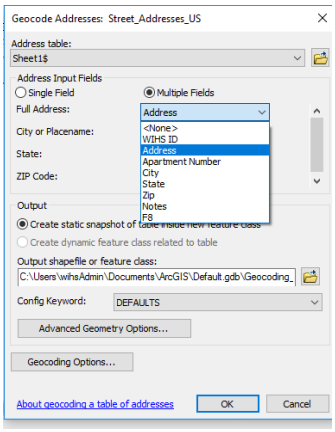


Step 5. Once you have selected the locator file, a pop-up box will appear asking you to identify the address table (the list of addresses for geocoding). Use the button to open a file explorer window and navigate to the WIHS Geocoding folder. Click the "OK" button.



Once you have selected the folder as a source location, navigate to the correct visit folder and select the address tracking log.

Once your address tracking log has been linked, ArcGIS will require information about what input fields need to be used for geocoding. Select the option for "Multiple Fields." The Full Address, City, State and Zip Code fields will each have a corresponding entry in the drop-down menu matching the column header from the address tracking log. Select the appropriate column header for each category (e.g., "street" for full address, "city" for city or placename)

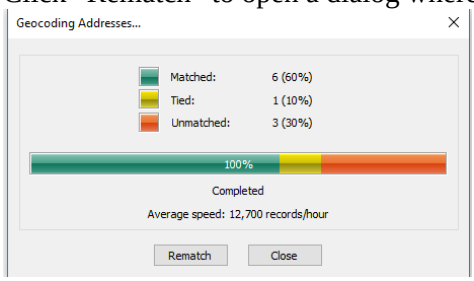


Finally, the bottom half of this pane is the dialog box that controls the geocoding output. You do not need to change the default settings. **Select OK to start the geocoding process.**

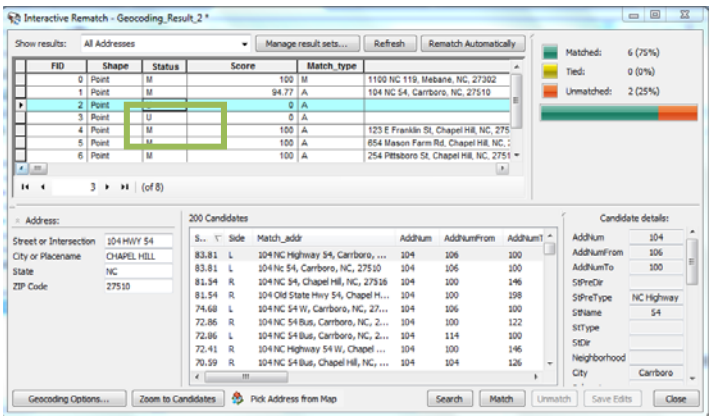
Step 6. A window will automatically open that tracks the geocoding progress. The window will provide an overview of the three possible results:

- **Matched:** Address was successfully matched with an existing US location.
- **Tied:** Address matched with more than one location and will require staff review.
- **Unmatched:** Address did not match any locations listed in the US street maps and will require staff review.

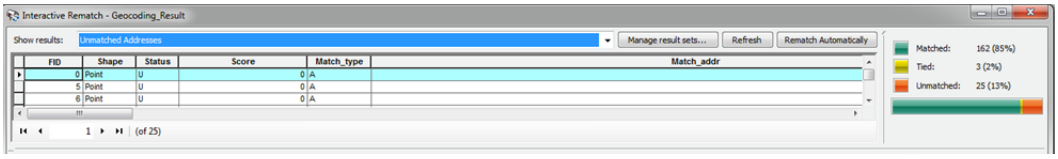
Click "Rematch" to open a dialog where you will be able to review the results.



The rematch dialog screen will show you the address that you entered (bottom left side of the pane) and the candidate match information will be on the right side (under "Candidate details").



Step 7. In order to review the results that were "Unmatched" use menu bar at the top of the review screen (highlighted in the screenshot below) to select "Unmatched Addresses." The results screen will change to only display addresses that were not properly matched.

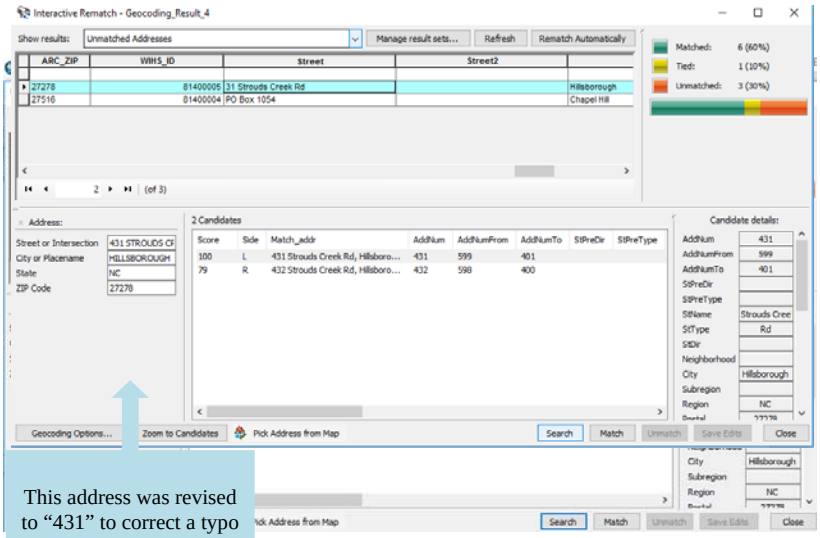


Highlight the first address on the list and review the address information, which is listed on the bottom left corner of the screen and identify revisions that need to be made. Note that if the individual did not provide an address (e.g., PO Box), then no changes are required.

Addresses can be revised directly in the rematch screen that you reviewed. You may edit the address information and revise as needed, for example:

- Revising misspellings in the street name
- Adding in street prefixes (N Main St. vs. Main St.)
- Revising abbreviations (Highway or Hway vs. Hwy)

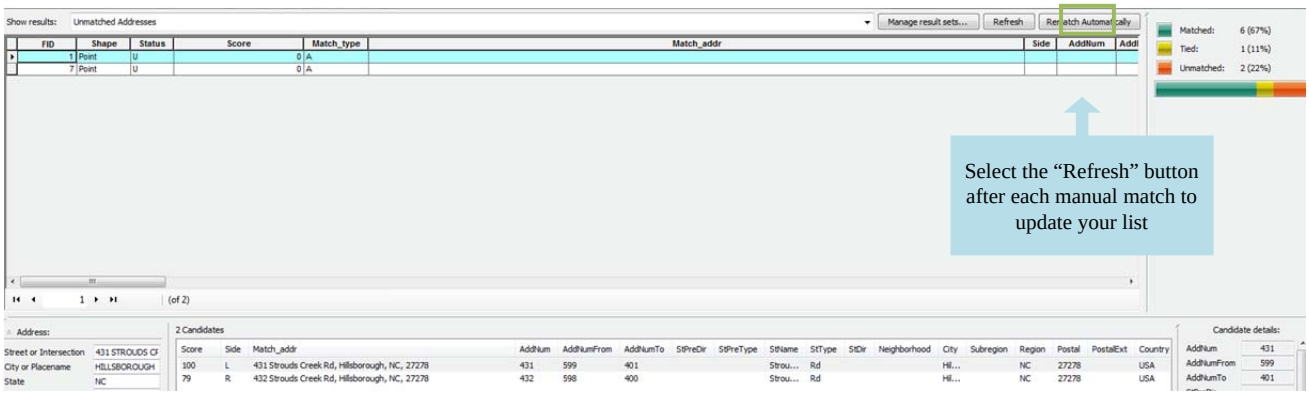
Once you have revised the address and are ready to re-submit it for geocoding, select the button that says "Search."



This address was revised to "431" to correct a typo on the address log. Search results yielded a perfect match for the revised address.

ArcGIS will attempt to geocode the specific address you revised and will list any matches on the right side of the screen. Review your results on the right side of the screen:

- **If there is a correct match** – highlight it and select the button at the bottom of the screen that says "Match." Then, select the option for "Refresh" on the top ribbon and you will see that the results total in the top left corner will change (i.e., the number of matches will increase by one and the number of unmatched will decrease by one, and the address will also disappear from your list of unmatched results). **Repeat the process for all remaining unmatched results.**



- **If there is not a correct match** – if no errors exist, then replace the address with the nearest major intersection that was provided by the participant using the following format: Street1 Name and Street2 Name. Follow the process outlined above to search for intersection matches. If the intersection is not able to be geocoded, further exploration of the address is recommended. Another option is to call the participant to see if the address you collected at the visit is correct. If any addresses remain unmatched, leave them in the rematch pane (do not delete any entries). In the final Excel document, these participants will remain classified as "U" indicating that ArcGIS could not find a match. This may occur, for example, if a participant is living in a new development that cannot be found with older data and maps CDs. **Repeat the process for all remaining unmatched results.**

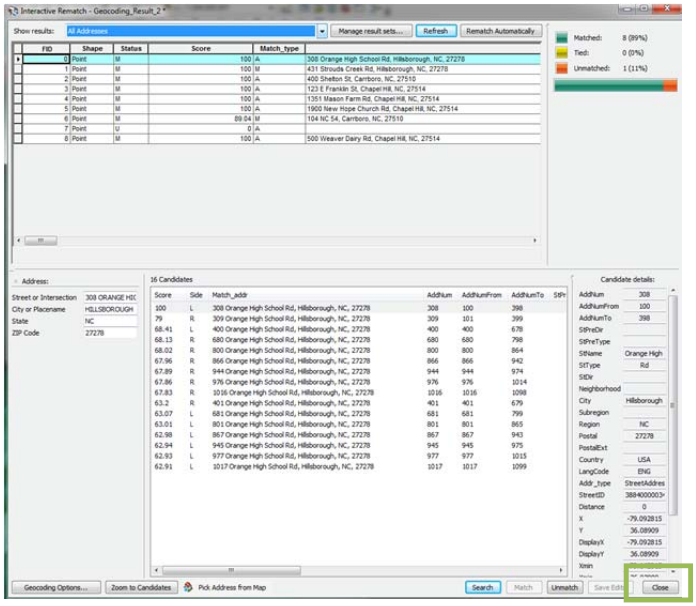
Step 8: Use the same process outlined above to review any matches that were "Tied." Use the menu bar at the top of the review screen to select "Tied Addresses." The results screen will change to only display addresses that matched with one or more possible locations.

Highlight the first address on the list and review the address information, which is listed on the bottom corner of the screen. Review the information and identify revisions that need to be made. In a number of cases, revising the address and searching for new matches may yield a more successful match (e.g., >80% match).

- **If there is a correct match** – highlight it and select the button at the bottom of the screen that says "Match." You will see that the results total in the top left corner will change (i.e., the number of matches will increase by one and the number of ties results will decrease by one, and the address will also disappear from your list of tied results). **Repeat the process for all remaining tied results.**

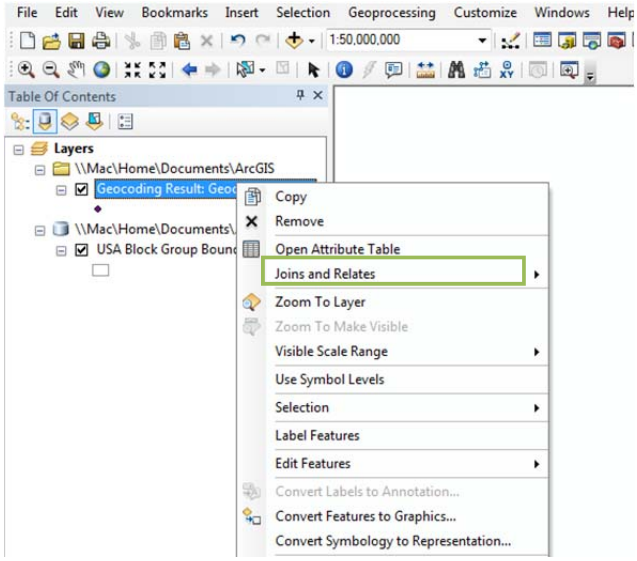
- **If there is not a correct match** – if no errors exist, then review the tied results options and select the most appropriate result, usually the one with the highest score. In most cases, a thorough review of the street address will identify errors or omissions (e.g., missing street prefix, incorrect suffix, misspelling, etc.) that are impeding a complete match. **Repeat the process for all remaining tied results.**

Once you have finished reviewing your results and have followed the steps outlined above to remediate unsuccessful matches, click on the "Close" button to finalize your geocoding results.



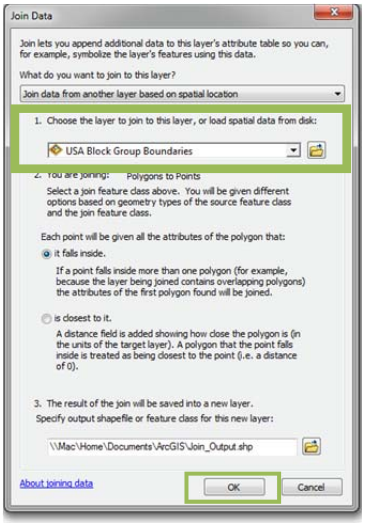
Step 9. Now, join the geospatial data from addresses with the census block group boundary file. When these two datasets are joined, ArcGIS overlays the exact geospatial locations of the addresses over the map of the census block group boundaries and determines the exact census block group that corresponds to each unique address.

First, right click on "Geocoding Result" on the "Table of Contents" pane, which is on the left side of the screen (see below). On the options menu, select "Join and Relates," and then "Join."

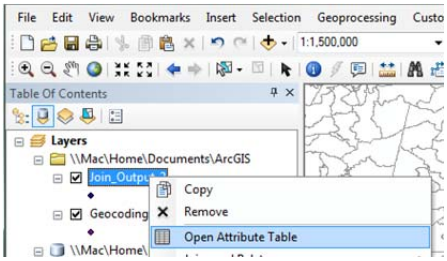


Step 10. A pop-up window will open in which you need to specify how to join the two datasets (see below). In the first pull-down menu, select to "Join data from another layer based on spatial location." Then, select the specific layer that you want to join to your address data; you should be able to see "USA Block Group Boundaries" (this was the file we first opened in Step 1). Indicate that points should be

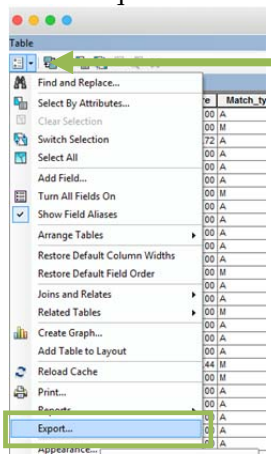
assigned if they "fall inside" and leave the default output folder options unless you specifically want to save the output elsewhere. Click "Ok."



Step 11. Now, export the joined results in a format that can be opened using Excel. First, right-click on "Join_Output" on the "Table of Contents" pane which is on the left side of the screen (see below). On the options menu, select "Open Attribute Table."



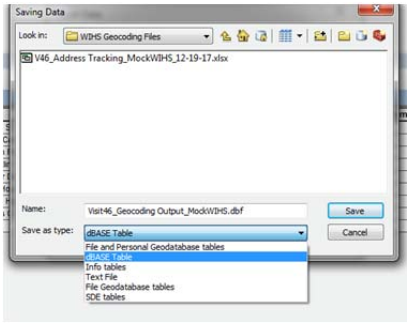
Step 12. A large table will open which contains the joined geocoding and block group data. Select the "Table Options" icon in the upper left , and select the "Export" option.



Step 13. A "Saving Data" window will immediately open. Choose "Export All Records," and then select the browse icon to choose the export file type. Navigate to where you would like the file to be stored. We recommend saving all of your output files in the same location; in our example there is a folder

entitled "WIHS Geocoding Files" that is organized by visit and contains both our raw address input files and the joined output.

Name the file: VISIT#Geocoding Output_SITE NAME. From the drop-down "Save as Type" menu, choose "dBASE Table." . **Note that you file will not open in excel unless you selected the dBASE format.** Click "Save" and "OK." If the program asks if you want to add the new table to the map, select "No."

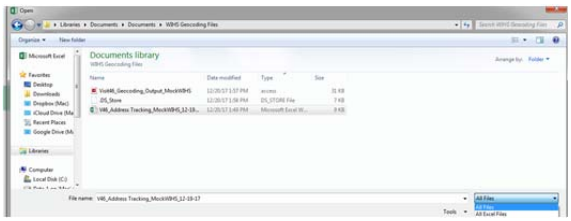


Exit the Attribute Table and ArcMap. **There is no need to save the project.**

5. Formatting Your Geocoding Output File

Before sending the output file to UNC, you will need to format it.

First, open Excel and choose "Open File" option from the File menu. Navigate to the folder with the geocoding output file that you just saved. Excel defaults to opening files that are in its native formats (e.g., .xls or .xlsx); in order to see the saved output filed which we saved as a dBASE file (.dbf), use the drop-down menu at the bottom of the selection window (choose "All Files" as shown below). Select your "VISIT#Geocoding Ouput_SITE NAME.dbf" file. Select "Open."



To increase privacy in the file you will be sharing, delete all variables except WIHSID, Status, Score, Match_type, and FIPS. However, since some addresses may have been unmatched, we suggest deleting the address column last so that you can review the unmatched results and add in notes for addresses that are unmatched.

In our example below, we have removed all of the extra columns and have kept our column entitled "Notes", where we have indicated cases in which only a PO Box was provided or where an address was unmatchable. This is not a required step but will reduce the number of queries that are received by your site.

WIHS_ID	Status	Score	Match_type	Street	FIPS	Notes
81400008	M	100.0000	A	400 Shelton St	371350107051	
81400003	M	89.0400	M	104 NC 54 Bypass	371350107031	
81400002	M	100.0000	A	1351 Mason Farm Rd	371350116023	
81400001	M	100.0000	A	123 E Franklin St	371350114002	
81400009	M	100.0000	A	500 Weaver Dairy Road	371350112052	
81400007	M	100.0000	A	1900 New Hope Church Road	371350109022	
81400006	M	100.0000	A	308 Orange High School Rd	371350110001	
81400005	M	100.0000	M	31 Strouds Creek Rd┬á	371350109011	
81400004	U	0.0000	A	PO Box 1054		Participant consented to geocoding but declined to provide physical address

Once you have reviewed the data, remove the address column. Finally, add in a column to the right of the WIHSID column called "Visit_Number." In the first cell (B2) add in the geocoding visit and then copy that number all the way down to the last row that has participant information.

WIHS_ID	Visit	Status	Score	Match_type	FIPS	Notes
81400008	46	M	100.0000	A	371350107051	
81400003		M	89.0400	M	371350107031	
81400002		M	100.0000	A	371350116023	
81400001		M	100.0000	A	371350114002	
81400009		M	100.0000	A	371350112052	
81400007		M	100.0000	A	371350109022	
81400006		M	100.0000	A	371350110001	
81400005		M	100.0000	M	371350109011	
81400004		U	0.0000	A		Participant consented to geocoding but declined to provide physical address

You should now have a file that only contains the following seven variables: WIHSID, Visit, Status, Score, Match_type, FIPS and your original Notes.

Choose "Save As" from the File menu and save this file as VISIT#_Block Group Only_SITE NAME. **Save the file as an Excel file (.xls or.xlsx).**

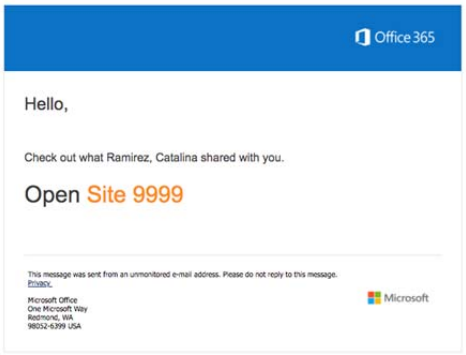
NOTE: If you do not save the file as an Excel file, **all of the changes you have made to the file will be lost** and you will have to re-start the formatting process.

6. Transferring Data to UNC via Secure File Transfer

Files may be transferred to UNC using a variety of IRB-approved secure transfer mechanisms. Instructions for the secure transfer of geocoding files using One Drive, a Microsoft 365 application which is HIPAA-compliant and has been approved for the transfer on sensitive data files are provided below.

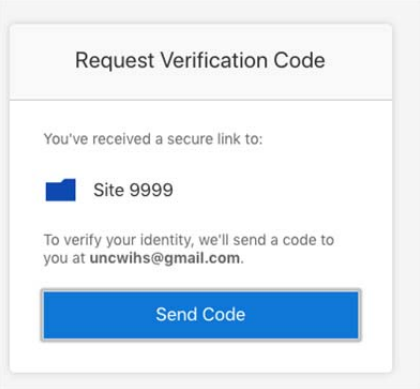
If your institution would like to use another IRB-approved secure transfer mechanism, please contact Catalina Ramirez (catalina_ramirez@med.unc.edu) to make the necessary accommodations.

Step 1. The UNC WIHS Geocoding Staff will set up a secure OneDrive folder which will be used to transfer the final geocoding output file.



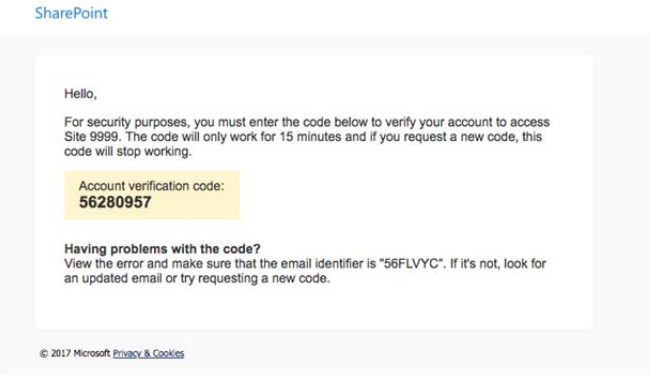
Note: If you do not receive the e-mail message please check your Junk/Spam folders, as external Office365 messages can sometimes be filtered by institutional firewalls.

Step 2. In order to verify the identity of the individual accessing the folder, Microsoft will require you to request a verification code for real-time authentication. The code will be sent to the e-mail address you provided.

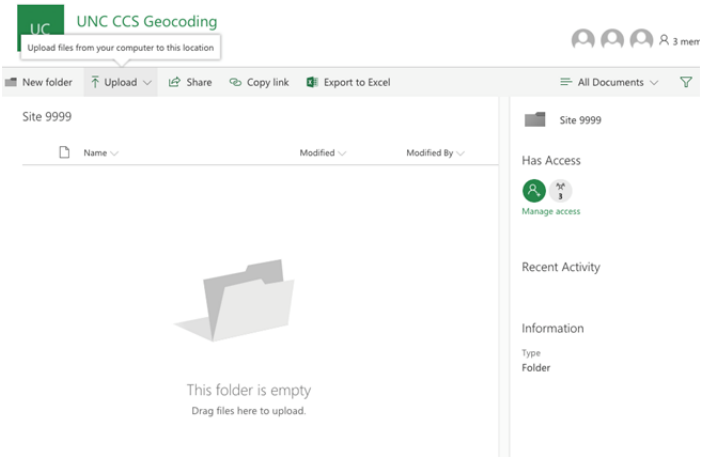


> **Note:** A new authentication code will be required each time you attempt to open the shared folder.

Step 3. Once you receive the authentication code, copy it to the authentication screen and select the option to "Verify." Please DO NOT select the option for "keep me signed in."



Step 4. Once your account has been verified, you will be able to review the files in your folder. You can then use the upload feature to select the final geocoding output file and add it to your folder.



Step 6. Send an e-mail to Catalina Ramirez (catalina_ramirez@med.unc.edu) to notify UNC that you have uploaded data.

APPENDIX B. CONSENT TOOLS

Consent to Participate in a Research Study
Addendum to provide additional information to subject after original consent

---

Consent Form Version Date:
IRB Study #

Title of Study:

Principal Investigator:
Principal Investigator Department:
Principal Investigator Phone number:
Principal Investigator Email Address:

Study Contact:
Study Contact telephone number:
Study Contact email:

Funding Source and/or Sponsor: National Institutes of Health (NIH)

The following information should be read as an addition to the original Consent Form that you read and signed at the beginning of the study. Unless specifically stated otherwise in the following paragraphs, all information contained in that original Consent Form is still true and remains in effect. Your participation continues to be voluntary. You may refuse to participate, or may withdraw your consent to participate at any time, and for any reason, without jeopardizing your future care at this institution or your relationship with your study doctor.

Changes to Study Visit Protocol
We are also asking you to give [YOUR INSTITUTION] authorization to use your address so that we can examine how characteristics of the geographic area in which you live influence your health and wellness.

Study staff will use a computer program called ArcGIS to match your address to a larger geographic area called a census block group (CBG)- which is large enough so that your exact address cannot be determined. Research staff use data from large data sources – like the census-that describes your block group to create datasets that investigators can use. Your block group is never provided to anyone, only the characteristics that describe where you live.

The data that describes where you live – and not your address- will be used by researchers to help them better understand how neighborhood and community factors relate to health and wellness. Your actual address will not be reported, published or shared with anyone else.

You will not receive any additional compensation. Your participation will only take a few minutes.

Subject's Agreement:
I have read the information provided above. I have asked all the questions I have at this time. I voluntarily agree to continue to participate in this research study.

_____ _____

Signature of Research Subject Date

_____

Printed Name of Research Subject

_____ _____

Signature of Research Team Member Obtaining Consent Date

_____

Printed Name of Research Team Member Obtaining Consent