Appendix A. Simple Simulation Design R Code

```
# Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys
# Supplemental code
# Code (1) Simulates data with survey structure
#         (2) Estimates treatment effect via several different methods (naive estimate, with survey weights only,
#         with propensity score methods only, or combining propensity scores and survey weights)

# Required R packages
#install.packages("survey")
#install.packages("MatchIt")
library("survey")
library("MatchIt")


############################################################
## Generate data
############################################################

# Define number of simulations
nsims <- 1000
# Define total population size
N <- 90000
# Stratum indicator (3 equal sized stratum)
stratum <- c(rep(1,30000), rep(2, 30000), rep(3,30000))

# Simulate covariate X as random normal variable
# Mean of X varies by stratum
X <- c(rnorm(N/3, mean=-.25), rnorm(N/3, mean=0), rnorm(N/3, mean=.25))

# Selection model (based on one in Cole & Stuart, 2010)
# Mean P(selection) = 0.1
beta0 <- -3.8
beta1 <- -log(4)
# Define sampling probabilities for each individual
# Sampling probability varies by X
prob.sel <- exp(1+beta0+beta1*X)/(1 + exp(1+beta0+beta1*X))
# Define sampling weights as inverse probability of selection
S.WT <- 1/prob.sel

# Treatment model (has to be done in population to define PATT)
tau0 <- -2
tau1 <- log(4)
# Define probability of treatment for each individual
# Treatment probability varies by X
prob.t <- exp(1+tau0+tau1*X)/(1+exp(1+tau0+tau1*X))
# Binary treatment variable (T) generated by random binomial draws with treatment probability
T <- rbinom(N, 1, prob.t)

# Generate potential outcomes
# Y0 = Potential outcome under control condition (T = 0)
# Y1 = Potential outcome under treatment condition (T=1)
# Potential outcomes vary by X
alpha0 <- 1
alpha1 <- 1
Y0 <- alpha0 + alpha1*X + rnorm(N, mean=0, sd=.5)
```

```
gamma1 <- .2
gamma2 <- .1
Y1 <- Y0 + gamma1 + gamma2*X + rnorm(N, mean=0, sd=.5)
# Define observed outcome Y
Y <- ifelse(T==1, Y1, Y0)

# Final simulated data
data <- as.data.frame(cbind(x=X,stratum=stratum,y0=Y0, y1=Y1, t=T, y=Y, s.wt=S.WT))


############################################################
## Estimate Treatment Effects
############################################################

# Estimate true average effect
# Note: SD of Y is about 1.1 so these are essentially effect sizes
# Average Treatment Effect
ate <- mean(Y1)-mean(Y0)
# Average Treatment Effect on the Treated
att <- mean(Y1[T==1])-mean(Y0[T==1])

# Define empty matrices for estimated tx effects
effects <- coverage.ate <- coverage.att <- matrix(NA, ncol=12, nrow=nsims)
# Column names, indicating various analyses
colnames(effects) <- colnames(coverage.ate) <- colnames(coverage.att) <-
  c("Naive", "Surv", "ATEWt","ATTWt", "ATESub",  "ATTSub","NN", "ATEWtSurv",
"ATTWtSurv","ATESubSurv",
    "ATTSubSurv", "NNSurv")

# Iterate drawing samples and estimating treatment effects
for (i in 1:nsims) {
  # dta1 represents total population in Stratum 1
  dta1 <- data[1:30000,]
  # S1 is sample from Stratum 1 (sampled with stratum-specific probabilities)
  S1 <- dta1[sample(1:nrow(dta1), 4000, replace=FALSE, prob=prob.sel[1:30000]),]
  # Define sampling weights for S1
  S1$s.wt <- S1$s.wt * (30000/4000)
  # dta2 represents total population in Stratum 2
  dta2 <- data[30001:60000,]
  # S2 is sample from Stratum 2 (sampled with stratum-specific probabilities)
  S2 <- dta2[sample(1:nrow(dta2), 3000, replace=FALSE, prob=prob.sel[30001:60000]),]
  # Define sampling weights for S2
  S2$s.wt <- S2$s.wt * (30000/3000)
  # dta3 represents total population in Stratum 3
  dta3 <- data[60001:90000,]
  # S3 is sample from Stratum 3 (sampled with stratum-specific probabilities)
  S3 <- dta3[sample(1:nrow(dta3), 2000, replace=FALSE, prob=prob.sel[60001:90000]),]
  # Define sampling weights for S3
  S3$s.wt <- S3$s.wt * (30000/2000)
  # Final sample
  samp <- rbind(S1, S2, S3)
  Ns <- 9000

#################
  # First naive estimate
  temp <- lm(y ~ t + x, data=samp)
```

```
  effects[i,"Naive"] <- summary(temp)$coef[2,1]
  sd <- summary(temp)$coef[2,2]
  # Calculate 95% CI coverage rates for ATE estimate
  coverage.ate[i,"Naive"] <- ifelse(effects[i,"Naive"]-2*sd < ate & effects[i,"Naive"] +2*sd > ate, 1, 0)
  # Calculate 95% CI coverage rates for ATT estimate
  coverage.att[i,"Naive"] <- ifelse(effects[i,"Naive"]-2*sd < att & effects[i,"Naive"] +2*sd > att, 1, 0)

##################
  # Survey weighted estimate
  # Set up survey design with stratum and sampling weights
  svy.design <- svydesign(ids=~1, strata=samp$stratum, weights=samp$s.wt, data=samp)
  lm.svy <- svyglm(y ~ t+x, design=svy.design)
  effects[i,"Surv"] <- summary(lm.svy)$coef[2,1]
  sd <- summary(lm.svy)$coef[2,2]
  coverage.ate[i,"Surv"] <- ifelse(effects[i,"Surv"]-2*sd < ate & effects[i,"Surv"] +2*sd > ate, 1, 0)
  coverage.att[i,"Surv"] <- ifelse(effects[i,"Surv"]-2*sd < att & effects[i,"Surv"] +2*sd > att, 1, 0)
##################
  # Propensity score weighted estimate
  # Estimate propensity scores to use in propensity score weighted analyses
  t.model <- glm(t ~ x, data=samp, family="binomial")
  pscore <- predict(t.model, data=samp, type="response")
  # Define ATE propensity score weights (IPTW)
  samp$ate.wt <- ifelse(samp$t==1, 1/pscore, 1/(1-pscore))
  # Define ATT propensity score weights (weighting by the odds)
  samp$att.wt <- ifelse(samp$t==1, 1, pscore/(1-pscore))
  # ATE version
  # Set up survey design with strata and propensity score weights
  tate.design <- svydesign(ids=~1, strata=samp$stratum, weights=samp$ate.wt, data=samp)
  lm.t <- svyglm(y ~ t+x, design=tate.design)
  effects[i,"ATEWt"] <- summary(lm.t)$coef[2,1]
  sd <- summary(lm.t)$coef[2,2]
  coverage.ate[i,"ATEWt"] <- ifelse(effects[i,"ATEWt"]-2*sd < ate & effects[i,"ATEWt"] +2*sd > ate, 1, 0)
  # ATT version
  # Set up survey design with strata and propensity score weights
  tatt.design <- svydesign(ids=~1, strata=samp$stratum, weights=samp$att.wt, data=samp)
  lm.t <- svyglm(y ~ t+x, design=tatt.design)
  effects[i,"ATTWt"] <- summary(lm.t)$coef[2,1]
  sd <- summary(lm.t)$coef[2,2]
  coverage.att[i,"ATTWt"] <- ifelse(effects[i,"ATTWt"]-2*sd < att & effects[i,"ATTWt"] +2*sd > att, 1, 0)

##################
  # Propensity score weights + survey weights analysis
  # ATE version
  # Define combined weights (product of propensity score and survey sampling weights)
  samp$comb.ate <- samp$ate.wt*samp$s.wt
  # Set up survey design with strata and combined weights
  comb.ate.design <- svydesign(ids=~1, strata=samp$stratum, weights=samp$comb.ate, data=samp)
  lm.comb <- svyglm(y ~ t+x, design=comb.ate.design)
  effects[i,"ATEWtSurv"] <- summary(lm.comb)$coef[2,1]
  sd <- summary(lm.comb)$coef[2,2]
  coverage.ate[i,"ATEWtSurv"] <- ifelse(effects[i,"ATEWtSurv"]-2*sd < ate & effects[i,"ATEWtSurv"] +2*sd > ate,
1, 0)
  # ATT version
  # Define combined weights (product of propensity score and survey sampling weights)
  samp$comb.att <- samp$att.wt*samp$s.wt
  comb.att.design <- svydesign(ids=~1, strata=samp$stratum, weights=samp$comb.att, data=samp)
```

```
  lm.comb <- svyglm(y ~ t+x, design=comb.att.design)
  effects[i,"ATTWtSurv"] <- summary(lm.comb)$coef[2,1]
  sd <- summary(lm.comb)$coef[2,2]
  coverage.att[i,"ATTWtSurv"] <- ifelse(effects[i,"ATTWtSurv"]-2*sd < att & effects[i,"ATTWtSurv"] +2*sd > att,
1, 0)

##################
  # Nearest neighbor matching approach, without survey design
  # Use MatchIt to estimate propensity scores and create matched sample (1:1 nearest neighbor approach)
  m.out <- matchit(t ~ x, data=samp)
  # Subset to matched sample
  m.data <- match.data(m.out)
  # Estimate ATT effects on matched data
  lm.match <- lm(y ~ t+x, data=m.data)
  effects[i,"NN"] <- summary(lm.match)$coef[2,1]
  sd <- summary(lm.match)$coef[2,2]
  coverage.att[i,"NN"] <- ifelse(effects[i,"NN"]-2*sd < att & effects[i,"NN"] +2*sd > att, 1, 0)

##################
  # Nearest neighbor matching approach, with survey design
  # Set up survey design with stratum and sampling weights
  match.design <- svydesign(ids=~1, strata=m.data$stratum, weights=m.data$s.wt, data=m.data)
  # Estimate ATT effects on matched data
  lm.matchwt <- svyglm(y ~ t+x, design=match.design)
  effects[i,"NNSurv"] <- summary(lm.matchwt)$coef[2,1]
  sd <- summary(lm.matchwt)$coef[2,2]
  coverage.att[i,"NNSurv"] <- ifelse(effects[i,"NNSurv"]-2*sd < att & effects[i,"NNSurv"] +2*sd > att, 1, 0)

##################
  # Subclassification (both with and without survey design)
  # Use MatchIt to estimate propensity scores and create matched sample (subclassification approach)
  m.out.subclass <- matchit(t ~ x, data=samp, method="subclass")
  # Subset to matched sample
  data.subcl <- match.data(m.out.subclass)

  # Define empty vectors for effect estimates and variances
  # For unweighted estimates
  effects.sub <-   vars <- rep(NA, max(data.subcl$subclass))
  # For survey weighted estimates
  effects.sub.wt <- vars.wt <- rep(NA, max(data.subcl$subclass))
  # Empty vectors for subclass specific sample sizes
  N.t.s <- N.s  <- SumWts.s <- SumWts.t.s <- rep(NA, max(data.subcl$subclass))
  SumWts <- sum(samp$s.wt)
  SumWts.t <- sum(samp$s.wt[samp$t==1])
  Nt <- sum(samp$t)

  # Run regression model within each subclass
  # All subclassification effect estimates are generated from the same subclass-specific effect estimates
  for(s in 1:max(data.subcl$subclass))
  {
    # Subclass specific estimates, no survey design
    tmp <- lm(y ~ t + x, data=data.subcl, subset=subclass==s)
    effects.sub[s] <- tmp$coef[2]
    vars[s] <- summary(tmp)$coef[2,2]^2
    # Subclass specific estimates, accounting for survey design
    sub.design <- svydesign(ids=~1,
```

```
                    strata=data.subcl$stratum[data.subcl$subclass==s],
                    weights=data.subcl$s.wt[data.subcl$subclass==s], data=data.subcl[data.subcl$subclass==s,])
   lm.sub <- svyglm(y ~ t+x, design=sub.design)
   effects.sub.wt[s] <- summary(lm.sub)$coef[2,1]
   vars.wt[s] <- summary(lm.sub)$coef[2,2]^2
   # For ATE estimates
   # Number of individuals in each subclass
   N.s[s] <- sum(data.subcl$subclass==s)
   # Sum of sampling weights in each subclass
   SumWts.s[s] <- sum(data.subcl$s.wt[data.subcl$subclass==s])
   # For ATT estimates
   # Number of treated individuals in each subclass
   N.t.s[s] <- sum(data.subcl$subclass==s & data.subcl$t==1)
   # Sum of sampling weights for treated individuals in each subclass
   SumWts.t.s[s] <- sum(data.subcl$s.wt[data.subcl$subclass==s & data.subcl$t==1])
  }

  # ATE, no survey weights: weight subclasses by total # in subclass
  effects[i, "ATESub"] <- sum((N.s/Ns)*effects.sub)
  sd <- sqrt(sum((N.s/N)^2*vars))
  coverage.ate[i,"ATESub"] <- ifelse(effects[i,"ATESub"]-2*sd < ate & effects[i,"ATESub"] +2*sd > ate, 1, 0)
  # ATT, no survey weights: weight subclasses by # treated in subclass
  effects[i, "ATTSub"] <- sum((N.t.s/Nt)*effects.sub)
  sd <- sqrt(sum((N.t.s/Nt)^2*vars))
  coverage.att[i,"ATTSub"] <- ifelse(effects[i,"ATTSub"]-2*sd < att & effects[i,"ATTSub"] +2*sd > att, 1, 0)
  # ATE, with survey weights: weight subclasses by sum of weights in subclass
  effects[i, "ATESubSurv"] <- sum((SumWts.s/SumWts)*effects.sub.wt)
  sd <- sqrt(sum((SumWts.s/SumWts)^2*vars.wt))
  coverage.ate[i,"ATESubSurv"] <- ifelse(effects[i,"ATESubSurv"]-2*sd < ate & effects[i,"ATESubSurv"] +2*sd >
ate, 1, 0)
  # ATT, with survey weights: weight subclasses by sum of weights of treated in subclass
  effects[i, "ATTSubSurv"] <- sum((SumWts.t.s/SumWts.t)*effects.sub.wt)
  sd <- sqrt(sum((SumWts.t.s/SumWts.t)^2*vars.wt))
  coverage.att[i,"ATTSubSurv"] <- ifelse(effects[i,"ATTSubSurv"]-2*sd < att & effects[i,"ATTSubSurv"] +2*sd >
att, 1, 0)
}

###################
# Calculate bias of ATE estimates
bias.ate <- effects - ate
print("ATE bias")
print(round(apply(bias.ate, 2, mean),3))
# Calculate mean squared error of ATE estimates
mse.ate <- (effects - ate)^2
print("ATE MSE")
print(round(apply(mse.ate, 2, mean),4))
# 95% CI coverage for ATE estimates
print("ATE Coverage")
print(round(apply(coverage.ate, 2, mean),2))
# Calculate bias of ATT estimates
bias.att <- effects - att
print("ATT bias")
print(round(apply(bias.att, 2, mean),3))
# Calculate mean squared error of ATT estimates
mse.att <- (effects - att)^2
print("ATT MSE")
```

```
print(round(apply(mse.att, 2, mean),4))
# 95% CI coverage for ATT estimates
print("ATT Coverage")
print(round(apply(coverage.att, 2, mean),2))

# Diagnostic plots for ATE estimates (across all methods)
barplot(abs(apply(bias.ate[,c("Naive", "Surv", "ATEWt", "ATESub", "ATEWtSurv", "ATESubSurv")], 2, mean)),
cex.names=.3, main="ATE Bias")
barplot(apply(mse.ate[,c("Naive", "Surv", "ATEWt", "ATESub", "ATEWtSurv", "ATESubSurv")], 2, mean),
cex.names=.3, main="ATE MSE")
barplot(apply(coverage.ate[,c("Naive", "Surv", "ATEWt", "ATESub", "ATEWtSurv", "ATESubSurv")], 2, mean),
cex.names=.3, main="ATE Coverage")
abline(h=.95)
# Diagnostic plots for ATT estimates (across all methods)
barplot(abs(apply(bias.att[,c("Naive", "Surv", "ATTWt", "ATTSub", "NN", "ATTWtSurv", "ATTSubSurv",
"NNSurv")], 2, mean)), cex.names=.3, main="ATT Bias")
barplot(apply(mse.att[,c("Naive", "Surv", "ATTWt", "ATTSub", "NN", "ATTWtSurv", "ATTSubSurv",
"NNSurv")], 2, mean), cex.names=.3, main="ATT MSE")
barplot(apply(coverage.att[,c("Naive", "Surv", "ATTWt", "ATTSub", "NN", "ATTWtSurv", "ATTSubSurv",
"NNSurv")], 2, mean), cex.names=.3, main="ATT Coverage")
abline(h=.95)
dev.off()
# Results tables
write.table(effects, row.names=FALSE, col.names=TRUE, sep=",")
write.table(coverage.ate, row.names=FALSE, col.names=TRUE, sep=",")
write.table(coverage.att, row.names=FALSE, col.names=TRUE, sep=",")
```

Appendix B. Analysis Models By Estimation Method and Stata Comamnds

Target Population	Estimand	Survey Command	Subclassification Post-Estimation	Stata Command
Survey subjects who choose a non-primary care physician as their usual source of care	Sample ATT (SATT)	Weighting: svyset PSU [pweight = ATT_weight] STRATA Others: svyset PSU [pweight = _n] STRATA	Average across subclasses using the number of treated in the sample in each subclass	svy, subpop(subset): glm Y X_1 X_2 … X_K, family(gamma) link(log)
Adults in the US who choose a non-primary care physician as their usual source of care	Population ATT (PATT)	Weighting: svyset PSU [weight = (ATT_IPTW*SAQWT08F)] STRATA Others: svyset PSU [weight = SAQWT08F] STRATA	Average across subclasses using the number of treated people in the population in each subclass (i.e., the sum of the weights of the treated subjects in each subclass)	svy, subpop(subset): glm Y X_1 X_2 … X_K, family(gamma) link(log)
Survey subjects who report having a usual source of care	Sample ATE (SATE)	Weighting: svyset PSU [pweight = ATE_weight] STRATA Others: svyset PSU [pweight = _n] STRATA	Average across subclasses using the number of survey subjects in each subclass	glm Y X_1 X_2 … X_K [pweight = ate_weight], family(gamma) link(log)
Adults in the US who have a usual source of care	Population ATE (PATE)	Weighting: svyset PSU [pweight = (ATE_IPTW*SAQWT08F)] STRATA Subclass: svyset PSU [pweight =_ SAQWT08F] STRATA	Average across subclasses using the total number in the population in each subclass (i.e., the sum of the weights in each subclass)	svy, subpop(subset): glm Y X_1 X_2 … X_K, family(gamma) link(log)

Appendix C. Supplementary Methods and Discussion on Example 2: The Association Between Usual Source of Care and Health Care Expenditures

It is generally thought that generalist physicians are more effective at managing health care costs than specialist physicians. A study by Phillips and colleagues (2009) using four years of the Medical Expenditure Panel Survey (MEPS) found that individuals who have a family practice or general practitioner as their usual source of care have lower health care costs than individuals who have an internist, specialist, or other usual source of care. However, since a person's usual source of care is not randomly assigned, this comparison may be subject to selection bias unless confounding is carefully addressed. Individuals with a primary care physician as their usual source of care may be substantially different from those seeing a specialist on important confounders, such as health status and willingness to use health care services.

Study Data and Methods

This example used the 2008 Medical Expenditure Panel Survey (MEPS) Household Component, a survey of individuals and families' health and health care services. The Household Component survey is administered to a nationally representative subsample from the National Health Interview Survey (NHIS). The MEPS public-use file contains data on a wide range of items including individual demographics, health conditions, health status, medical service use, health insurance coverage, and employment status.

The MEPS reflects the NIHS sampling design, but also includes unique features. The NHIS uses a multistage area-level probability sampling strategy with stratification, clustering, and oversampling of blacks, Hispanics, and Asians (Machlin et al. 2010). The MEPS also oversamples households based on income (Chowdhury 2011; Machlin et al. 2010). The MEPS

includes several person-level and household-level cross-sectional and longitudinal survey weights. We use SAQWT08F, a person level weight for those responding to the supplemental self-administered questionnaire. When weighted, the MEPS Household Component sample represents the US civilian non-institutionalized population.

We modeled this analysis after one aspect of Phillips et al. (2009), which compared annual medical spending for adults who had identified a generalist or a specialist physician as his or her usual source of health care. An individual's usual source of care was determined from a series of access to care survey questions that are collapsed into a variable indicating 24 types of providers. We classified the control group as those individuals who identified a general practice or family physician; specialists included subspecialties including cardiology, endocrinology, gastroenterologist, nephrologist, surgeons, and "MD – other."

As is common with expenditure data, our outcome of interest, total health care expenditures in 2008, exhibited a right skewed distribution with no negative measures and a nontrivial number of zeros. The traditional approach is to log transform the dependent variable, which sometimes generates a normal distribution. However, this approach is subject to a re-transformation bias (Manning 1998; Manning and Mullahy 2001). The box cox and GLM family test indicated that a gamma distribution with log link fit best in our data. Thus, a generalized linear model with a gamma distribution and log link was used for all outcome analyses and the recycled predications method was used to re-transform estimates back from the log scale (Doshi and Glick 2010).

As in Phillips et al. (2009), covariates included sex, race/ethnicity, marital status, poverty, health insurance status, a binary indicator for age over 65, education, self-reported health, urban/rural indicator, geographical region, and the physical and mental health components of the

Short Form health survey (SF-12). Unlike Phillips et al., we did not include the EuroQol health status questionnaire items (EQ-5D) because this item was not reported in the 2008 MEPS. We also did not include emergency department visits or hospital discharges because we considered these variables to be highly collinear with the outcome of interest. The propensity score model included these covariates as well as the survey weight for the self-administered questionnaire.

Standard propensity score methods require complete data on all individuals. Four survey items had missing values: education (0.70%), the SF-12 physical component score (0.79%), the SF-12 mental component score (0.77%), and self-reported health status (0.85%). In bivariate analyses, indicators for these missing items were not statistically significantly associated with the outcome, total health care expenditures. We performed single imputation of these variables in Stata 12 using the mi ice command based upon all covariates included in the propensity score model.

To judge the success of each propensity score procedure in terms of creating groups that look similar on the observed covariates ("balance"), we use the standardized bias for each covariate: the difference in the means between the treated and control groups divided by the standard deviation in the treatment group (Ho et al. 2011). Generally, standardized difference in means less than 0.25 or 0.20 indicate that the groups are well balanced. Additionally, we assessed propensity score overlap between the treated and control groups.

Propensity scores were calculated in R (version 2.14.2) using the *MatchIt* package (Ho et al., 2011) and exported to Stata for analysis. Weighted outcome models, accounting for the MEPS complex survey design, were conducted in Stata version 12 using the *svyset* command (StataCorp, 2011).

Discussion

The findings from the regression analysis (see Table 2) suggest that the estimated effects for the SATT and PATT and SATE and PATE are similar. This suggests that in this example, the survey sample is similar to that for the target population. The average treatment effect—having a specialist physician as a usual source of care—is associated with higher spending compared to those with a primary care physician.

We find that the average effect of having a specialist as one's usual source of care was associated with nearly a $2,500 increase in health care spending for US adults. These estimates are somewhat higher than Phillips et al.'s (2009) estimated effect of $2,642 in an unadjusted model and $1,430 in a fully adjusted model. These larger estimates may be a result of changes over time as well as the difference in study design.

While these models suggest that there is a cost effect associated with physician specialty, these results have several limitations. While these models sought to address confounding on observed health status, these models could not balance the two groups on unobserved differences in health and preferences for medical care. Similarly, the geographic adjustment is limited in the MEPS public use file to an urban rural indicator and regional variable, and information captured by the MEPS survey weight. Lastly, the sample size of the treatment group is small, which may unduly bias our results.

Appendix C References

Chowdhury, S. R. 2011. "Relative Importance of Poverty and Education in MEPS Weighting Adjustment." *Methodology Report No. 25.* pp. 25. Rockville, MD: Agency for Healthcare Research and Quality.

Doshi, J., H. Glick 2010. "Computer workshop: Analysis of cost." ISPOR 14th Annual International Meeting.

Ho, D. E., K. Imai, G. King, and E. A. Stuart. 2011. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42(8).

Manning, W. G. 1998. "The logged dependent variable, heteroscedasticity, and the retransformation problem." *J Health Econ* 17(3): 283-95.

Machlin, S., S. Chowdhury, T. Ezzati-Rice, R. DiGaetano, H. Goksel, L.-M. Wun, W. Yu, and D. Kashihara. 2010. "Estimation Procedures for the Medical Expenditure Panel Survey Household Component." *Methodology Report #24.* pp. 54. Rockville, MD: Agency for Healthcare Research and Quality.

Manning, W. G. 1998. "The logged dependent variable, heteroscedasticity, and the retransformation problem." *Journal of Health Economics* 17(3): 283-95.

Manning, W. G. and J. Mullahy. 2001. "Estimating log models: to transform or not to transform?" *J Health Econ* 20(4): 461-94.

Phillips, R. L., M. S. Dodoo, L. A. Green, G. E. Fryer, A. W. Bazemore, K. I. McCoy, and S. M. Petterson. 2009. "Usual source of care: an important source of variation in health care spending." *Health Aff (Millwood)* 28(2): 567-77.

StataCorp. 2011. *Stata Statistical Software: Release 12* (College Station, TX: StataCorp LP)