# Supplemental Appendix A. R Code Used to Generate Figure 3.

```r
# This File Runs the Models to Create Figure 3

# The User only Needs to change the working directory where files created by this script will be saved
setwd("C:/...")

# Clear the Environment
rm(list = ls())

# Set up the General Parameters of the Model (Values from Table 1)
parm <- c('gamma'=1/3, 'pi'=1, 'delta.0'=395, 'delta.1'=395, 'phi.0'=78, 'phi.1'=78, 'rho.0'=1e-5, 'rho.1'=1e-5, 'alpha.0'=1e-6, 'alpha.1'=0.73*1e-6, 'mu.HH.0'=1/10, 'mu.HH.1'=1/10, 'mu.E'=1/10)

# Gamma=Duration of Infection
# pi=scaling parameter
# delta=Pathogen shedding into Household Environment
# phi=Pathogen shedding into Community Environment
# rho=Pathogen Pickup from Household Environment
# alpha=Pathogen Pickup from Community Environment
# mu=Rate of Pathogen die-off in the environment

# Setup the initial Conditions of the Model
set.seed(1) # Set the seed for the random number generator, so results can be replicated
N.HH <- 100 # Number of Household in the Community
N.Pop <- 500 # Size of Population (Number of Individuals)
HH.Size <- round(rnorm(N.HH, N.Pop/N.HH, 1.5)) # Create Vector of HH Sizes
N.Pop.Actual <- sum(HH.Size) # Actual Population Size, should be close to N.Pop (500)
N.Pop.Actual
N0 <- 1 # Number of Initial Infecteds

# Create a vector the different coverage levels to be used
coverage.levels <- c(0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)

# How many times to run the model at each level of coverage?
N.runs <- 100

# Create Empty Data Frame to be filled
data.total <- data.frame(coverageNA, runNA, risk.overallNA, risk.0NA, risk.1NA, Inf.HHNA, Inf.E-NA, Inf.TotalNA, prop.HHNA, prop.E-NA, rr.directNA, rr.indirectNA, rr.totalNA, rr.overallNA)

coverage.n <- 0 # A counter for the level of coverage

for (coverage in coverage.levels) { # Loop across Levels of coverage (0-100%, by 10%)
  # Running counts
  coverage.n <- coverage.n+1

  for (run in 1:N.runs) {

    # Report overall progress of the simulations
    percent <- (((coverage.n-1)*N.runs)+run)/(length(coverage.levels)*N.runs))*100
cat("Coverage =",coverage,"\nRun =",run,"\nOverall Progress =",round(percent,1),"\n")

    # Create a vector indicating if HH has improved (1) or unimproved (0) sanitation
    sanitation <- vector(mode='numeric', length=N.HH)
    for (i in 1:N.HH) {
      if (i <= round(coverage*N.HH)) sanitation[i] <- 1
    }

    # Summarize Population according to sanitation
    # (Number of individuals using improved/unimproved sanitation)
  }
}
```

```
N.San.0 <- 0
N.San.1 <- 0
for (i in 1:N.HH) {
  if (sanitation[i]==0) N.San.0 <- N.San.0 + HH.Size[i]
  if (sanitation[i]==1) N.San.1 <- N.San.1 + HH.Size[i]
}

# Create Matrix of Susceptibles, Infected, Recovered, and HH Environments
# 1st Column is the current event, 2nd column is for the next event
HHS <- matrix(nrow=N.HH, ncol=2)
HHI <- matrix(nrow=N.HH, ncol=2)
HHR <- matrix(nrow=N.HH, ncol=2)
HHE <- matrix(nrow=N.HH, ncol=2)
for (i in 1:N.HH) {
  HHS[i,1] <- HH.Size[i]  # Begin with everyone susceptible
  HHI[i,1] <- 0            # Begin with no infected
  HHR[i,1] <- 0            # Begin with no immune
  HHE[i,1] <- 0            # Begin with a clean HH environment
}

# Create vector of the Community Environment
E <- c(0, NA)  # Begin with a clean Community environment

# Create Vectors of Parameters for Each Household
gamma <- vector()
delta <- vector()
phi <- vector()
rho <- vector()
alpha <- vector()
mu.HH <- vector()
mu.E <- parm['mu.E']
for (i in 1:N.HH) {
  gamma[i] <- parm['gamma']
  if (sanitation[i]==0) {
    delta[i] <- parm['delta.0']
    phi[i] <- parm['phi.0']
    rho[i] <- parm['rho.0']
    alpha[i] <- parm['alpha.0']
    mu.HH[i] <- parm['mu.HH.0']
  }
  if (sanitation[i]==1) {
    delta[i] <- parm['delta.1']
    phi[i] <- parm['phi.1']
    rho[i] <- parm['rho.1']
    alpha[i] <- parm['alpha.1']
    mu.HH[i] <- parm['mu.HH.1']
  }
}

# Randomly place initial infected in the population
inHouseholds <- sample(1:N.HH, Io, replace=TRUE, prob=HH.Size)
for (i in inHouseholds) {
  HHI[i,1] <- HHI[i,1]+1
  HHS[i,1] <- HHS[i,1]-1
}

# Create Vectors to keep track of cumulative S, I, R
time <- 0
S.0 <- 0
I.0 <- 0
R.0 <- 0
S.1 <- 0
I.1 <- 0
R.1 <- 0
for (i in 1:N.HH) {
  if (sanitation[i]==0) {
    S.0 <- S.0 + HH.Size[i]
    I.0 <- I.0 + HHI[i]
    R.0 <- R.0 + HHR[i]
  }
  if (sanitation[i]==1) {
    S.1 <- S.1 + HH.Size[i]
    I.1 <- I.1 + HHI[i]
    R.1 <- R.1 + HHR[i]
  }
}
Cum.Inf.HH <- 0 # For tracking cumulative infections from the HH Environment
Cum.Inf.E <- 0 # For tracking cumulative infections from the shared Environment

time.i <- 0 # Start at time=0
iter <- 0
max.time <- 500 # Maximum Number of Days to Run the Simulation

while (time.i < max.time & sum(HHS[,1])>0 & ( sum(HHI[,1])>0 | sum(HHE[,1])>0 | E[1]>0 )) {
    iter <- iter + 1

    # Create Vector of Rates
    recovery <- vector()
    newInfectionHH <- vector()
    newInfectionE <- vector()
    for (i in 1:N.HH) {
        recovery[i] <- gamma[i] * HHI[i]
        newInfectionHH[i] <- HHS[i]*parm['pi']^rho[i]*HHE[i,1]
        newInfectionE[i] <- HHS[i]*parm['pi']^alpha[i]*E[i]
    }

    # Create Vector of the rates of all events
    all.events <- c(recovery, newInfectionHH, newInfectionE)

    # When does the next event happen?
    tau <- rexp(1, rate=sum(all.events))

    # Update Time
    time.i <- time.i + tau

    # What Event Occurred?
    event <- sample(length(all.events),1,prob=all.events)

    # Update States of People
    # Create a vector of State Transitions based on events
    # Corresponds with the vector 'all.events'
    HHS[,2] <- HHS[,1]
    HHI[,2] <- HHI[,1]
    HHE[,2] <- HHE[,1]
    if (event >= 1 & event <= N.HH) { # Recovery
        HHI[event,2] <- HHI[event,1] - 1
        HHE[event,2] <- HHE[event,1] + 1
    }
    if (event >= N.HH+1 & event <= N.HH+2) { # New Infection from HH Environment
        HHS[event-N.HH,2] <- HHS[event-N.HH,1] - 1
        HHI[event-N.HH,2] <- HHI[event-N.HH,1] + 1
        Cum.Inf.HH <- Cum.Inf.HH + 1
    }
    if (event >= N.HH+2 & event <= N.HH+3) { # New Infections from HH Environment
        HHS[event-2*N.HH,2] <- HHS[event-2*N.HH,1] - 1
        HHI[event-2*N.HH,2] <- HHI[event-2*N.HH,1] + 1
        Cum.Inf.HH <- Cum.Inf.HH + 1
    }

    # Update Values for Environmental Compartments
    for (i in 1:N.HH) {
        HHE[i,2] <- HHE[i,1]*exp(-mu.HHI[i]*tau) + (delta[i]*HHI[i]/mu.HHI[i])*(1-exp(-mu.HHI[i]*tau))
    }
    E[2] <- E[1]*exp(-mu.E*tau) + (sum(phi*HHE[,1])/mu.E)*(1-exp(-mu.E*tau))

    # Update Cumulative S, I, R Tracker
    time[i+1] <- time.i
    S.0[i+1] <- 0
    I.0[i+1] <- 0
    R.0[i+1] <- 0
    S.1[i+1] <- 0
    I.1[i+1] <- 0
    R.1[i+1] <- 0

    for (i in 1:N.HH) {
        if (sanitation[i]==0) {
            S.0[i+1] <- S.0[i+1] + HHS[i,2]
            I.0[i+1] <- I.0[i+1] + HHI[i,2]
            R.0[i+1] <- R.0[i+1] + HHE[i,2]
        }
        if (sanitation[i]==1) {
            S.1[i+1] <- S.1[i+1] + HHS[i,1]
        }
    }
}

# Input the Initial Values for Each
if (sanitation[1]==0) {
    S.0[iter+1] <- S.0[iter+1] + HHS[1,2]
    I.0[iter+1] <- I.0[iter+1] + HHI[1,2]
    R.0[iter+1] <- R.0[iter+1] + HHE[1,2]
}
if (sanitation[1]==1) {
    S.1[iter+1] <- S.1[iter+1] + HHS[1]
I.[iter+1] <- I.[iter+1] + HHI[i]
R.[iter+1] <- R.[iter+1] + HHR[i]

}

HHH[,1] <- HHH[,2]
HII[,1] <- HII[,2]
HH[,1] <- HH[,2]
HRE[,1] <- HRE[,2]

# Create data frame of cumulative incidence for each run
data.row = c(coverage=coverage, run=run,
             risk=run=(S.0[1]+S.1[1]-tail(S.0+1, n=1))/(S.0[1]+S.1[1]),
             risk.0=(S.0[1]+tail(S.0, n=1))/S.0[1],
             risk.1=(S.1[1]-tail(S.1, n=1))/S.1[1],
             Inf.HH=Cum.Inf.HH,
             Inf.E=Cum.Inf.E,
             Inf.Total=Cum.Inf.HH+Cum.Inf.E,
             prop.HH=Cum.Inf.HH/Cum.Inf.E,
             rr.direct=NA,
             rr.indirect=NA,
             rr.total=NA,
             rr.overall=NA)
data.total <- rbind(data.total, data.row)

) # Close 'Number of Runs' Loop

data.total <- data.total[,-1]

# Calculate Relative Risks
data.total$rr.direct <- data.total$risk.0 / data.total$risk.1
for (c in 1:n.runs) {
  for (r in 1:n.runs) {
    data.total$rr.indirect[c, r] <- data.total$risk.0[c, r] / data.total$risk.1[c, r]
  }
}

write.csv(data.total, file="Figure 3 - Water (Total).csv")

# Calculate Medians
data.medians <- data.frame(coverage=NA, run=NA,
                           risk=NA, risk.0=NA, risk.1=NA,
                           Inf.HH=NA, Inf.E=NA, Inf.Total=NA, prop.HH=NA, prop.E=NA,
                           rr.direct=NA, rr.indirect=NA, rr.total=NA, rr.overall=NA)

obs <- 0
for (c in 1:length(coverage.levels)) {
  obs <- obs+1
  data.medians[obs,] <- NA
  data.medians$coverage[obs] <- coverage.levels[c]
  data.medians$risk[obs] <- median(data.total$risk.overall[which(data.total$coverage==coverage.levels[c])], na.rm=T)
  data.medians$rr.0[obs] <- median(data.total$rr.0[which(data.total$coverage==coverage.levels[c])], na.rm=T)
  data.medians$rr.1[obs] <- median(data.total$rr.1[which(data.total$coverage==coverage.levels[c])], na.rm=T)
  data.medians$rr.direct[obs] <- median(data.total$rr.direct[which(data.total$coverage==coverage.levels[c])], na.rm=T)
  data.medians$rr.indirect[obs] <- median(data.total$rr.indirect[which(data.total$coverage==coverage.levels[c])], na.rm=T)
  data.medians$rr.total[obs] <- median(data.total$rr.total[which(data.total$coverage==coverage.levels[c])], na.rm=T)
  data.medians$rr.overall[obs] <- median(data.total$rr.overall[which(data.total$coverage==coverage.levels[c])], na.rm=T)
}

for (c in coverage.levels) {
  data.medians$rr.0[c] <- data.medians$rr.0[data.medians$coverage==c] / data.medians$risk.0[c]
  data.medians$rr.1[c] <- data.medians$rr.1[data.medians$coverage==c] / data.medians$risk.0[c]
}

# Close Coverage Level Loop

data.total <- data.total[,-1]

# Calculate Medians

write.csv(data.total, file="Figure 3 - Water (Total).csv")
data.medians$r.r.overall[data.medians$coverage==c] <- data.medians$r.r.overall[data.medians$coverage==c] / data.medians$r.r.overall[1]
}

# Save output data as local files
write.csv(data.medians, file="Figure 3 - Water (Medians).csv")

save(data.total, data.medians, Io, max.time, N.MM, N.Pop, N.PopActual, N.runs, 
parm, sanitization, 
file="Figure 3 - Water.RData")

# Make plot window
rm(list = ls())
load("Figure 3 - Water.RData")
dev.new(width=7.5, height=3)
par(mfrow=c(1,2))
par(mar=c(3,3,1,1))
lw <- 1.5

# Plot of the Median Risk
plot(0,0,pch=NA, ylim=c(0,1), xlim=c(0,100), xlab="", ylab="", main="")
lines(data.medians$coverage*100, data.medians$r.risk, col='black', lty=2, lwd=lw)
lines(data.medians$coverage*100, data.medians$r.risk.1, col='black', lty=1, lwd=lw)
points(data.medians$coverage[1]*100, data.medians$r.risk.0[1], pch=17, cex=1.5)
mtext(side=1, text="Intervention Coverage (%) in the Community", line=2, cex=0.75)
mtext(side=2, text="Proportion Infected", line=2, cex=0.75)
text(100,1,"x", adj=c(1,1), cex=2)

legend("bottomleft", expression(paste("D" ["0"])), expression(paste("Control Group (D" ["0"]", "))), expression(paste("Intervention Group (D" ["1"]", "]))), col='black', lty=c(NA,2,1), lwd=c(NA,1.2,1.2), pch=c(17,NA,NA), pt.cex=c(1.5,NA,NA), cex=0.8, bty="n")

# Plot the Protective Efficacy
plot(data.medians$coverage*100, (1-data.medians$r.r.total)*100, col='black', type='l', lty=2, lwd=lw, ylab=NA, xlab=NA, ylim=c(0,100), xlim=c(0,100))
lines(data.medians$coverage*100, (1-data.medians$r.r.direct)*100, col='black', lty=1, lwd=lw)
lines(data.medians$coverage*100, (1-data.medians$r.r.indirect)*100, col='gray60', lty=1, lwd=lw)
lines(data.medians$coverage*100, (1-data.medians$r.r.overall)*100, col='gray60', lty=2, lwd=lw)
mtext(side=1, text="Intervention Coverage (%) in the Community", line=2, cex=0.75)
mtext(side=2, text="Protective Efficacy (%)", line=2, cex=0.75)
text(100,50,"x", adj=c(1,1), cex=2)

legend("topleft", c(expression(paste("Overall Effect (1-D" ["0"]","D" ["0"]","4W" ["1"]","D" ["1"]","D" ["0"]","D" ["0"]", "))), expression(paste("Total Effect (1-D" ["1"]","D" ["0"]","D" ["0"]", "))), expression(paste("Direct Effect (1-D" ["1"]","D" ["0"]","D" ["0"]", "))), expression(paste("Indirect Effect (1-D" ["0"]","D" ["0"]","D" ["0"]", ")))), col=c('gray60','black','black','gray60'), lty=c(2,4,4,4), lwd=1.2, cex=0.8, bty="n")
### SUPPLEMENTAL APPENDIX B

**SUPPLEMENTAL TABLES FROM THE LITERATURE REVIEW**

**Supplemental Table 1**

Details of studies assessing herd protection from drinking water

<table>
<thead>
<tr>
<th>Setting (reference)</th>
<th>Study design</th>
<th>Outcome</th>
<th>Range of drinking water coverage included in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mortality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Sub-Saharan</td>
<td>Retrospective cohort</td>
<td>Infant mortality (&lt; 1 year of age)</td>
<td>0–100%</td>
</tr>
<tr>
<td>African countries:</td>
<td>form 10-year birth history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>urban</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Sub-Saharan</td>
<td>Retrospective cohort</td>
<td>Infant mortality (&lt; 1 year of age)</td>
<td>0–100%</td>
</tr>
<tr>
<td>African countries:</td>
<td>form 10-year birth history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rural</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nigeria: national²</td>
<td>Retrospective cohort</td>
<td>Infant mortality (0–11 months of age)</td>
<td>Not provided</td>
</tr>
<tr>
<td>from 5-year birth history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nigeria: national²</td>
<td>Retrospective cohort</td>
<td>Child mortality (12–59 months of age)</td>
<td>Not provided</td>
</tr>
<tr>
<td>from 5-year birth history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil: northeast³</td>
<td>Retrospective cohort</td>
<td>Child mortality (&lt; 5 years of age)</td>
<td>Not provided, but mean is 30.1% and standard deviation is 24.5</td>
</tr>
<tr>
<td>from 5-year birth history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil: south and southeast³</td>
<td>Retrospective cohort</td>
<td>Child mortality (&lt; 5 years of age)</td>
<td>Not provided, but mean is 54.5% and standard deviation is 26.7</td>
</tr>
<tr>
<td>from 5-year birth history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nutrition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nigeria: national⁴</td>
<td>Cross-sectional</td>
<td>Childhood stunting (height-for-age z-score &lt; -2)</td>
<td>Not provided</td>
</tr>
<tr>
<td>Malawi: national⁵</td>
<td>Cross-sectional</td>
<td>Childhood stunting (height-for-age z-score &lt; -2)</td>
<td>Not provided</td>
</tr>
<tr>
<td>Ghana: national⁶</td>
<td>Cross-sectional</td>
<td>Low birth weight (smaller than average)</td>
<td>Not provided</td>
</tr>
</tbody>
</table>

Studies are grouped by outcome and presented in the same order as Table 2.
SUPPLEMENTAL TABLE 2  
Details of studies assessing herd protection from sanitation

<table>
<thead>
<tr>
<th>Setting (reference)</th>
<th>Study design</th>
<th>Outcome</th>
<th>Range of sanitation coverage included in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diarrhea</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil: urban⁷</td>
<td>Before and after intervention (2 distinct cohorts)</td>
<td>Diarrhea prevalence in the previous 3–4 days among children &lt; 3 years of age</td>
<td>26–80%</td>
</tr>
<tr>
<td>India: rural⁸</td>
<td>Cross-sectional</td>
<td>Diarrhea prevalence in the previous 2 weeks among children &lt; 4 years of age</td>
<td>0–100%</td>
</tr>
<tr>
<td>Zimbabwe: rural⁹</td>
<td>Cohort</td>
<td>No. of episodes of diarrhea in 45 weeks among children &lt; 5 years of age at baseline</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Brazil: urban¹⁰</td>
<td>Cohort</td>
<td>Duration in days of diarrheal episodes</td>
<td>Not reported, though at least 34.1–55.1%</td>
</tr>
<tr>
<td><strong>Parasitic infection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil: urban¹¹</td>
<td>Before and after interventions (2 distinct cross sections)</td>
<td>Prevalence of <em>Ascaris lumbricoides</em></td>
<td>26–80%</td>
</tr>
<tr>
<td>Brazil: urban¹¹</td>
<td>Before and after interventions (2 distinct cross sections)</td>
<td>Prevalence of <em>Trichuris trichiura</em></td>
<td>26–80%</td>
</tr>
<tr>
<td>Brazil: urban¹¹</td>
<td>Before and after interventions (2 distinct cross sections)</td>
<td>Prevalence of <em>Giardia duodenalis</em></td>
<td>26–80%</td>
</tr>
<tr>
<td><strong>Mortality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Sub-Saharan</td>
<td>Retrospective cohort form 10-year birth history</td>
<td>Infant mortality (&lt; 1 year of age)</td>
<td>0–100%</td>
</tr>
<tr>
<td>African countries:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>urban¹</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Sub-Saharan</td>
<td>Retrospective cohort form 10-year birth history</td>
<td>Infant mortality (&lt; 1 year of age)</td>
<td>0–100%</td>
</tr>
<tr>
<td>African countries:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rural¹</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India: national¹²</td>
<td>Retrospective cohort from 10-year birth history</td>
<td>Infant mortality (&lt; 1 year of age)</td>
<td>0–100%</td>
</tr>
<tr>
<td>Brazil: northeast³</td>
<td>Retrospective cohort from 5-year birth history</td>
<td>Child mortality (&lt; 5 years of age)</td>
<td>Not provided, but mean is 4.23 with standard deviation of 10.4</td>
</tr>
<tr>
<td>Brazil: south and</td>
<td>Retrospective cohort from 5-year birth history</td>
<td>Child mortality (&lt; 5 years of age)</td>
<td>Not provided, but mean is 29.3% with standard deviation of 28.2</td>
</tr>
<tr>
<td>southeast³</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nutrition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ecuador: rural¹³</td>
<td>Cohort study</td>
<td>Stunting (height-for-age z-score &lt; −2) among children &lt; 5 years of age</td>
<td>0–100%</td>
</tr>
<tr>
<td>Nigeria: national⁵</td>
<td>Cross-sectional</td>
<td>Stunting (height-for-age z-score &lt; −2) among children &lt; 5 years of age</td>
<td>Unknown</td>
</tr>
<tr>
<td>Guatemala: urban¹⁴</td>
<td>Cross-sectional</td>
<td>Stunting (length-for-age z-score &lt; −2) among children 6–36 months of age</td>
<td>0–100%</td>
</tr>
<tr>
<td>Guatemala: rural¹⁴</td>
<td>Cross-sectional</td>
<td>Stunting (length-for-age z-score &lt; −2) among children 6–36 months of age</td>
<td>0–100%</td>
</tr>
<tr>
<td>Bangladesh: urban¹⁵</td>
<td>Cohort</td>
<td>Wasting (weight-for-height z-score) in children 0–35 months of age at baseline</td>
<td>About 18% to about 93%</td>
</tr>
</tbody>
</table>

Studies are grouped by outcome and presented in the same order as Table 3.

SUPPLEMENTAL REFERENCES

11. Barreto ML, Genser B, Strina A, Teixeira MG, Assis A, Rego RF, Teles CA, Prado MS, Matos S, Alcantara-Neves NM,


