

Supplementary material. True Prevalence R Code

```
library("ggplot2")
library("ggh4x")

#-- Confidence interval
CI = 0.95

#----- Function -----
binom_CI <- function(n, N, CI = 0.95){
  p = n / N
  var <- p * (1 - p) / N
  a <- qt((1 - CI)/2, N-1, lower.tail=FALSE)
  return(list(p = p, var = var, LCI = p - a * sqrt(var), UCI = p + a * sqrt(var)))
}

#----- Validity Study -----

TP <- 70
FP <- 15
FN <- 5
TN <- 135

a <- binom_CI(TP, TP + FN, CI)
Se <- a$p
Var_Se <- a$var
CI_Se <- c(a$LCI, a$UCI)

a <- binom_CI(TN, TN + FP, CI)
Sp <- a$p
Var_Sp <- a$var
CI_Sp <- c(a$LCI, a$UCI)

N <- 1000000 #-- Population size
n <- 300 #-- Sample size
Pr <- 0.2 #-- True prevalence in the population
Population <- data.frame(dis = c(rep(1, round(Pr*N)), rep(0, N-round(Pr*N))), test = rep(NA, N))

rep <- 200000
set.seed(123)
a1 <- qt((1 - CI)/2, n-1, lower.tail=FALSE)
tpr = pr = spr = rep(NA, rep)

for (i in 1:rep){
  se <- rnorm(1, mean = Se, sd = sqrt(Var_Se))
  sp <- rnorm(1, mean = Sp, sd = sqrt(Var_Sp))
  Sample <- Population[sample(N, n),]
  Sample[Sample$dis == 1,]$test <- ifelse(runif(nrow(Sample[Sample$dis == 1,]), 0, 1) < se, 1, 0)
  Sample[Sample$dis == 0,]$test <- ifelse(runif(nrow(Sample[Sample$dis == 0,]), 0, 1) < sp, 0, 1)

  pr[i] <- as.numeric(table(Sample$dis)[2])/n #-- Sample Prevalence
  spr[i] <- as.numeric(table(Sample$test)[2])/n #-- Sample Seroprevalence
  tpr[i] <- (spr[i] + sp - 1)/(se + sp - 1)
}

write.csv(data.frame(pr = pr, spr = spr, tpr = tpr), "pr.csv", row.names = FALSE)

#-- Another sample
Sample <- Population[sample(N, n),]
Sample[Sample$dis == 1,]$test <- ifelse(runif(nrow(Sample[Sample$dis == 1,]), 0, 1) < se, 1, 0)
Sample[Sample$dis == 0,]$test <- ifelse(runif(nrow(Sample[Sample$dis == 0,]), 0, 1) < sp, 0, 1)

a <- binom_CI(as.numeric(table(Sample$test)[2]), n, CI)
pr_hat <- a$p
Var_pr_hat <- a$var

#--- Equations 4 and 5 -----
true_pr <- (pr_hat + Sp - 1)/(Se + Sp - 1)
Var_true_pr <- Var_pr_hat/(Se + Sp - 1)^2 +
  ((pr_hat + Sp - 1)/(Se + Sp - 1)^2)^2 * Var_Se +
  ((Se - pr_hat)/(Se + Sp - 1)^2)^2 * Var_Sp
CI_true_pr <- c(true_pr - a1 * sqrt(Var_true_pr), true_pr + a1 * sqrt(Var_true_pr))

Pr
mean(pr)
quantile(pr, probs = c((1-CI)/2, 1-(1-CI)/2), na.rm = TRUE)

mean(spr)
quantile(spr, probs = c((1-CI)/2, 1-(1-CI)/2), na.rm = TRUE)

true_pr; CI_true_pr
```

```

#-- Distributions -----
ggplot(data = dat) +
  geom_density(aes(x=pr), color = "gray50", fill="gray95", size=1) +
  geom_density(aes(x=spr), color = "red", fill="red", alpha=0.1, size=1, linetype="dashed") +
  geom_density(aes(x=tpr), color = "forestgreen", fill="forestgreen", alpha=0.1, size=1, linetype="dashed")
+
  geom_vline(xintercept=Pr, color="magenta", size=0.5) +
  xlab("Prevalence") + ylab("Density") +
  theme_classic() +
  guides(x = "axis_truncated", y = "axis_truncated") +
  theme(aspect.ratio = 6/5,
  axis.text = element_text(size = rel(1.3), color = "black"),
  axis.title = element_text(size = rel(1.5)))

ggsave("Fig 2.png")

#-- Scatter plot -----
fit <- lm(pr ~ tpr + 0, data = dat) #-- No intercept linear regression
temp <- predict(fit, interval="prediction")

ggplot(data = cbind(dat, temp), aes(x = tpr, y = pr)) +
  geom_point(color = "NA", shape=21, fill="dodgerblue", alpha = 0.01, size=1.5) +
  geom_line(aes(y=lwr), color = "red", linetype = "dashed") +
  geom_line(aes(y=upr), color = "red", linetype = "dashed") +
  geom_line(aes(y=fit), color = "red") +
  theme_classic() +
  guides(x = "axis_truncated", y = "axis_truncated") +
  theme(aspect.ratio = 5/5,
  axis.text = element_text(size = rel(1.3), color = "black"),
  axis.title = element_text(size = rel(1.5))) +
  scale_x_continuous(seq(0.05, 0.35, by = 0.05),
  labels = format(seq(0.05, 0.35, by = 0.05), didits=2)) +
  scale_y_continuous(seq(0.1, 0.3, by = 0.05),
  labels = format(seq(0.1, 0.3, by = 0.05), didits=2)) +
  xlab("Seroprevalence") + ylab("True Prevalence")

ggsave("Fig 3.png")

#-- Linear relationship between the True and Apparent Prevalence for various Se and Sp
Se <- c(0.7, 0.8, 0.9)
Sp <- c(0.8, 0.9, 1.0)
dat <- data.frame(pr = NA, tpr = NA, Se = NA, Sp = NA)

for (se in Se){
  for (sp in Sp){
    pr <- seq(0, 1, by = 0.05)
    tpr <- (pr + sp - 1)/(se + sp - 1)

    dat <- rbind(dat, data.frame(pr = pr, tpr = tpr, Se = se, Sp = sp))
  }
}

dat <- dat[-1,]
dat <- dat[dat$tpr>=0 & dat$tpr<=1,]

ggplot(data = dat) +
  geom_line(aes(x=pr, y=tpr, color = as.factor(Se), linetype = as.factor(Sp)), size = 1) +
  xlab("Apparent Prevalence") + ylab("True Prevalence") +
  scale_x_continuous(breaks = c(0, 0.25, 0.5, 0.75, 1)) +
  scale_y_continuous(breaks = c(0, 0.25, 0.5, 0.75, 1)) +
  scale_linetype_manual(values=c("dotted", "longdash", "solid")) +
  theme_classic() +
  theme(aspect.ratio = 6/5,
  legend.key.width=unit(1.7, "cm"),
  legend.justification = c("left", "top"),
  legend.position = c(0.05, 0.97),
  legend.text = element_text(size = rel(1.1)),
  legend.title = element_text(size = rel(1.3)),
  axis.text = element_text(size = rel(1.3), color = "black"),
  axis.title = element_text(size = rel(1.5))) +
  guides(x = "axis_truncated", y = "axis_truncated") +
  guides(color = guide_legend(title = "Sensitivity", order = 1),
  linetype = guide_legend(title = "Specificity", order = 2))

ggsave("Fig 1.png")

```