eAppendix A

eTable 1. Results from multivariate logistic regression model: Outcome regression (outcome = reported commercial sexual contact).

	OR	**(95% CI)**
Age (years)		
15-19	7.97	(3.10 to 20.49)
20-24	5.59	(2.43 to 12.89)
25-29	2.30	(0.99 to 5.31)
30-34	2.09	(0.92 to 4.74)
35-39	1.57	(0.69 to 3.56)
40-44	1.21	(0.51 to 2.86)
45-49	1.44	(0.58 to 3.54)
50-54[a]	1.00	
Type of residence		
Capital to large city	1.09	(0.71 to 1.66)
Small city	1.71	(1.02 to 2.85)
Town	1.22	(0.78 to 1.91)
Countryside[a]	1.00	
Religion		
Muslim	1.72	(0.73 to 4.03)
Christian	0.60	(0.09 to 4.17)
Other	0.90	(0.24 to 3.32)
Hindu[a]	1.00	
Missing[b]	0.0001	(0.00003 to 0.0003)

Wealth index	1.35	(1.05 to 1.73)
Current marital status (married/not married)	0.07	(0.01 to 0.54)
Husband has the right to have sex with another woman (yes/no)	0.90	(0.40 to 2.04)
Missing	1.25	(0.43 to 3.60)
STD diagnosis past 12 months (yes/no)	47.10	(14.94 to 148.50)
Missing	1.16	(0.23 to 5.79)
Genital discharge past 12 months (yes/no)	2.41	(1.46 to 3.96)
Missing	1.99	(0.73 to 5.41)
Number lifetime sexual partners		
1[a]	1.00	
2	1.47	(0.41 to 5.29)
3-4	13.16	(4.02 to 43.12)
5 or more	7.19	(2.36 to 21.94)
Missing[b]	35.45	(9.13 to 137.61)
Frequency of alcohol consumption		
Almost every day	1.64	(1.17 to 2.31)
About once a week	2.95	(2.01 to 4.32)
Less often than once a week	4.95	(3.06 to 7.99)
Never consumed alcohol[a]	1.00	
Missing[b]	0.00003	(0.00001 to 0.00008)
Education (years)	1.01	(0.91 to 1.12)
Missing[b]	0.00009	(0.00002 to 0.0004)
Education^2 (years)	0.99	(0.99 to 1.00)

Age at first intercourse (years)	1.17	(1.10 to 1.24)
Current marital status (married/not married) and number of lifetime sexual partners		
Married, 2 partners	10.14	(4.02 to 25.57)
Married, 3-4 partners	18.15	(7.53 to 43.75)
Married, 5 or more partners	16.36	(6.95 to 38.54)
Current marital status (married/not married) and age at first intercourse (years)	0.91	(0.84 to 0.99)
Husband has the right to have sex with another woman and current marital status (married/not married)	1.92	(0.91 to 4.06)
Husband has the right to have sex with another woman (yes/no) and religion		
Yes, Muslim	0.16	(0.04 to 0.65)
Yes, Christian	4.79	(1.01 to 22.67)
Yes, Other	0.45	(0.10 to 2.06)
Husband has the right to have sex with another woman (yes/no) and education (years)	1.07	(0.99 to 1.16)
STD diagnosis past 12 months (yes/no) and education (years)	0.67	(0.54 to 0.85)
Number of lifetime sexual partners and religion		
2 partners, Muslim	0.65	(0.16 to 2.67)
2 partners, Christian	2.13	(0.27 to 16.46)
2 partners, Other	0.55	(0.06 to 5.11)
3-4 partners, Muslim	1.43	(0.50 to 4.10)

3-4 partners, Christian	0.04	(0.004 to 0.56)
3-4 partners, Other	4.01	(0.91 to 17.59)
5 or more partners, Muslim	1.38	(0.51 to 3.72)
5 or more partners, Christian	0.38	(0.04 to 3.32)
5 or more partners, Other	2.77	(0.63 to 12.06)
Number of lifetime sexual partners and wealth		
2 partners, Wealth index	1.06	(0.76 to 1.48)
3-4 partners, Wealth index	0.72	(0.52 to 0.98)
5 or more partners, Wealth index	1.01	(0.76 to 1.33)

[a] Reference category.

[b] Reference category is not missing.

eTable 2. Results from multivariate logistic regression model: Inverse probability weighting (outcome = complete information on commercial sexual contact).

	OR	**(95% CI)**
Age (years)		
15-19	9.63	(6.26 to 14.81)
20-24	4.40	(3.11 to 6.21)
25-29	5.03	(3.67 to 6.91)
30-34	4.87	(3.42 to 6.93)
35-39	4.89	(3.38 to 7.08)
40-44	2.65	(1.91 to 3.67)
45-49	1.80	(1.31 to 2.48)
50-54[a]	1.00	
Religion		
Muslim	1.25	(0.85 to 1.83)
Christian	1.44	(0.73 to 2.84)
Other	1.74	(0.94 to 3.19)
Hindu[a]	1.00	
Missing[b]	1971.40	(774.56 to 5017.06)
Current marital status (married/not married)	68.39	(48.21 to 97.01)
Occupation		
Not working	0.50	(0.34 to 0.73)
Professional, technical, manager	0.96	(0.65 to 1.42)
Clerical	1.57	(0.92 to 2.65)
Sales	1.22	(0.91 to 1.63)

Agriculture employee	0.94	(0.74 to 1.21)
Services	0.75	(0.52 to 1.10)
Skilled and unskilled manual[a]	1.00	
Missing[b]	2.63	(0.37 to 6.84)
Number trips away from home to past 12 months	1.00	(0.97 to 1.04)
Missing[b]	1.01	(0.35 to 2.88)
Husband justified in beating wife if: She is unfaithful (yes/no)	0.77	(0.44 to 1.34)
Missing[b]	1.05	(0.58 to 1.93)
STD diagnosis past 12 months (yes/no)	0.89	(0.41 to 1.95)
Missing[b]	0.87	(0.32 to 2.36)
Number lifetime sexual partners		
1[a]	1.00	
2	1.44	(1.07 to 1.95)
3-4	1.77	(1.22 to 2.58)
5 or more	5.77	(3.42 to 9.73)
Missing[b]	0.29	(0.10 to 0.89)
Education (years)	1.03	(0.99 to 1.08)
Missing[b]	0.35	(0.08 to 1.54)
Education^2 (years)	1.00	(1.00 to 1.00)
Genital sore/ulcer in past 12 months	1.61	(0.87 to 3.01)
Missing[b]	0.56	(0.31 to 1.00)
Husband justified in beating wife if: She is unfaithful		

(yes/no) and occupation		
Yes, Not working	1.55	(0.80 to 3.03)
Yes, Professional, technical, manager	0.46	(0.24 to 0.90)
Yes, Clerical	0.60	(0.26 to 1.42)
Yes, Sales	0.67	(0.38 to 1.18)
Yes, Agriculture employee	1.30	(0.82 to 2.06)
Yes, Services	1.42	(0.66 to 3.04)
Current marital status (married/not married) and number of trips away from home to past 12 months	0.94	(0.90 to 0.98)
Current marital status (married/not married) and number lifetime sexual partners		
Married, 2 partners	0.97	(0.63 to 1.51)
Married, 3-4 partners	0.95	(0.52 to 1.74)
Married, 5 or more partners	0.22	(0.10 to 0.49)
Genital sore/ulcer in past 12 months and education (years)	0.94	(0.88 to 1.01)
Husband justified in beating wife if: She is unfaithful (yes/no) and age		
Yes, 15-19	0.68	(0.34 to 1.36)
Yes, 20-24	1.53	(0.78 to 3.02)
Yes, 25-29	0.98	(0.53 to 1.80)
Yes, 30-34	1.23	(0.62 to 2.46)
Yes, 35-39	0.96	(0.49 to 1.89)
Yes, 40-44	0.88	(0.41 to 1.89)
Yes, 45-49	0.99	(0.49 to 2.00)

Number of trips away from home to past 12 months and religion		
Number of trips, Muslim	1.04	(0.98 to 1.10)
Number of trips, Christian	0.99	(0.91 to 1.08)
Number of trips, Other	1.02	(1.02 to 1.12)

[a] Reference category.

[b] Reference category is not missing.

eAppendix B. SAS code to obtain adjusted point estimates

```
data nfhs_3;
set nfhs_3;

/* create indicator for non-response */
if paid = . then miss = 1;
if paid > . then miss = 0; run;

/* complete data model */;
proc surveylogistic data=nfhs_3;
      weight mweight;
      strata mv024;
      cluster mv001;
      class age_cat (ref='8') alcohol (ref='0') plife5 (ref='1') religion
            (ref='11') reside (ref='11')/param=ref;
      model paid (event='1')=
            married plife5 married*plife5 mplife5 religion mreligion std_dx
            mstd_dx reside extrasex mextrasex discharge mdischarge alcohol age_cat
            malcohol agesex extrasex*religion married*agesex edu medu wealth
            edu*edu married*extrasex edu*extrasex std_dx*edu plife5*religion
            wealth*plife5;
      output out=orpred predicted=mhat; run;
proc sort data=or_pred; by mcaseid; run;

/* missingness model */
proc surveylogistic data=nfhs_3;
      weight mweight;
      strata mv024;
      cluster mv001;
      class age_cat (ref='8') plife5 (ref='1') religion (ref='11') occup
            (ref='11')/param=ref;
      model miss (event='0')=
            married age_cat plife5 mplife5 edu medu occup moccup married*plife5
            religion mreligion away_cat maway married*away_cat beat_cheat
            mbeat_cheat beat_cheat*occup std_dx mstd_dx edu*edu beat_cheat*age_cat
            away cat*religion ulcer mulcer edu*ulcer;
      output out=miss_pred predicted=phat; run;
proc sort data=miss_pred; by mcaseid; run;

/* create inverse probability weights and pseudo-outcomes */
data nfhs_3_final;
      merge orpred misspred; by mcaseid;
      * total population weights;
      w = 1/phat;
      w2 = w*mweight;
      * married sub-population weights;
      w_m = married/phat;
      w2_m = w_m*mweight;
      * unmarried sub-population weights;
      unmarried = (married = 0);
      w_um = unmarried/phat;
      w2_um = w_um*mweight;
      * doubly robust psuedo-outcome;
      paid_dr = paid;
      if miss2=1 then paid_dr=0;
      y_dr = (1-miss)*(paid_dr - mhat)/phat + mhat;
run;
```

```
/* complete case */
proc surveylogistic data=nfhs_3_final;
      weight mweight;
      strata mv024;
      cluster mv001;
      model paid (event='1') = ;
run;

/* inverse probability weighting - total population*/
proc surveylogistic data=nfhs_3_final;
      weight w2;
      strata mv024;
      cluster mv001;
      model paid (event='1') = ;
run;

/* inverse probability weighting - married sub-population*/
proc surveylogistic data=nfhs_3_final;
      weight w2_m;
      strata mv024;
      cluster mv001;
      model paid (event='1') = ;
run;

/* inverse probability weighting - unmarried sub-population*/
proc surveylogistic data=nfhs_3_final;
      weight w2_um;
      strata mv024;
      cluster mv001;
      model paid (event='1') = ;
run;

/* outcome regression */
proc surveyreg data=nfhs_3_final;
      weight mweight;
      strata mv024;
      cluster mv001;
      model mhat = ;
      domain married;
run;

/* doubly robust estimation */
proc surveyreg data=nfhs_3_final;
      weight mweight;
      strata mv024;
      cluster mv001;
      model y_dr = ;
      domain married;
run;
```

e**Appendix C. Formulas and MATLAB code to obtain 95% confidence intervals for adjusted point estimates.**

Variance formulas

Inverse probability weighted regression variance:

$$\widehat{Var}(\hat{\mu}_{IPW}) =$$

$$\frac{1}{N^2}\sum_{i=1}^{N} W_{i,S}\left[\frac{C_i}{\pi_i(\tilde{M},\hat{\alpha})}\left(\hat{Y}_i - \hat{\mu}_{IPW}\right) - \right.$$

$$\left.\begin{bmatrix}\sum_{i=1}^{N}\left(W_{i,S}\frac{C_i}{\pi_i(\tilde{M}_i;\hat{\alpha})^2}\pi_i(\tilde{M}_i;\hat{\alpha})\left(1-\pi_i(\tilde{M}_i;\hat{\alpha})\right)(Y_i-\hat{\mu}_{IPW})\tilde{M}_i^T\right)\\ \left(\sum_{i=1}^{N}\left(W_{i,S}\pi_i(\tilde{M}_i;\hat{\alpha})\left(1-\pi_i(\tilde{M}_i;\hat{\alpha})\right)\tilde{M}_i\tilde{M}_i^T\right)\right)^{-1}\\ \tilde{M}_i\left(C_i-\pi_i(\tilde{M}_i;\hat{\alpha})\right)\end{bmatrix}\right]^2 *$$

$$\left[\frac{1}{\left[\frac{1}{N}\sum_{i=1}^{N} W_{i,S}\frac{C_i}{\pi_i(\tilde{M},\hat{\alpha})}\right]^2}\right]$$

Outcome regression variance:

$$\widehat{Var}(\hat{\mu}_{OR}) =$$

$$\frac{1}{N^2}\sum_{i=1}^{N} W_{i,S}\left[\left(b_i(\tilde{L}_i;\hat{\beta}) - \hat{\mu}_{OR}\right) + \left[\begin{array}{c}\sum_{i=1}^{N}\left(W_{i,S}b_i(\tilde{L}_i;\hat{\beta})\left(1-b_i(\tilde{L}_i;\hat{\beta})\right)\tilde{L}_i^T\right)\\ \left(\sum_{i=1}^{N}\left(W_{i,S}b_i(\tilde{L}_i;\hat{\beta})\left(1-b_i(\tilde{L}_i;\hat{\beta})\right)\tilde{L}_i\tilde{L}_i^T\right)\right)^{-1}\\ \tilde{L}_i\left(Y_i-b_i(\tilde{L}_i;\hat{\beta})\right)\end{array}\right]^2\right.$$

Doubly robust estimator variance:

$$\widehat{Var}(\hat{\mu}_{DR}) =$$

$$\frac{1}{N^2}\sum_{i=1}^{N} W_{i,S}\left[\left(\hat{Y}_{i,DR} - \hat{\mu}_{DR}\right) - \left[\sum_{i=1}^{N}\left(W_{i,S}\frac{C_i}{\pi_i\left(\widetilde{M}_i;\hat{\alpha}\right)^2}\pi_i\left(\widetilde{M}_i;\hat{\alpha}\right)\left(1-\pi_i\left(\widetilde{M}_i;\hat{\alpha}\right)\right)\widetilde{M}_i^T\left(Y_i - b_i\left(\tilde{L}_i;\hat{\beta}\right)\right)\right)\left(\sum_{i=1}^{N}\left(W_{i,S}\pi_i\left(\widetilde{M}_i;\hat{\alpha}\right)\left(1-\pi_i\left(\widetilde{M}_i;\hat{\alpha}\right)\right)\widetilde{M}_i\widetilde{M}_i^T\right)\right)^{-1}\widetilde{M}_i\left(C_i - \pi_i\left(\widetilde{M}_i;\hat{\alpha}\right)\right)\right] + \left[\sum_{i=1}^{N}\left(W_{i,S}\left(1-\frac{C_i}{\pi_i\left(\widetilde{M}_i;\hat{\alpha}\right)}\right)\tilde{L}_i^T b_i\left(\tilde{L}_i;\hat{\beta}\right)\left(1-b_i\left(\tilde{L}_i;\hat{\beta}\right)\right)\right)\left(\sum_{i=1}^{N}\left(W_{i,S}b_i\left(\tilde{L}_i;\hat{\beta}\right)\left(1-b_i\left(\tilde{L}_i;\hat{\beta}\right)\right)\tilde{L}_i\tilde{L}_i^T\right)\right)^{-1}\tilde{L}_i\left(Y_i - b_i\left(\tilde{L}_i;\hat{\beta}\right)\right)\right]\right]^2$$

MATLAB code

```
% INVERSE PROBABILITY WEIGHTED REGRESSION VARIANCE
% import data from excel
[ndata to headertext] = xlsread('C:\other.xls');
[m to headertext] = xlsread('C:\covariates.xls');

N = size(ndata,1);      % number of subjects
M = size(m,2);          % number of covariates
weight = ndata(:,1);    % survey weight
miss = ndata(:,2);      % indicator for item non-response
C = 1 - miss;           % C = 1 for observations with complete data
paid = ndata(:,3);      % observed response
phat = ndata(:,4);      % predicated probability of complete data
mu = 0.01110163;        % survey-weighted mean based on IPW

z = zeros(N,1);
w1 = zeros(1,M);
w2 = zeros(M,M);
x = zeros(N,1);
w4 = zeros(1,1);

for i = 1:N;
    tempw1 = zeros(1,M);
    tempw2 = zeros(M,M);
    tempw4 = zeros(1,1);
    tempw1 = weight(i,1)*(C(i,1)/phat(i,1))*phat(i,1)*(1 -
      phat(i,1))*(paid(i,1)*mu)*m(i,:);
    tempw2 = weight(i,1)*phat(i,1)*(1-phat(i,1))*m(i,:)'*m(i,:);
    tempw4 = weight(i,1)*(C(i,1)/phat(i,1));
    w1 = w1 + tempw1;
    w2 = w2 + tempw2;
    w4 = w4 + tempw4;
end;

for j = 1:N;
    w3 = zeros(M,1);
    z(j,1) = (C(j,1)/phat(j,1))*(paid(j,1)-mu);
    w3 = m(j,:)'*(C(j,1)-phat(j,1));
    x(j,1) = weight(j,1)*(z(j,1)-(w1*inv(w2)*w3))^2;
end;

var = (sum(x)/(N^2))*(1/((1/N)*w4)^2)
ci95_low = mu - (1.96*sqrt(var))
ci95_high = mu + (1.96*sqrt(var))
```

```
% OUTCOME REGRESSION VARIANCE
% import data from excel
[l to headertext1] = xlsread('C:\covariates.xls');
[ndata to headertext2] = xlsread('C:\other.xls');

N = size(l,1);          % number of subjects
L = size(l,2);          % number of covariates
weight = ndata(:,1);    % survey weight
mhat = ndata(:,2);      % predicated probability of response
paid = ndata(:,3);      % observed response
mu = 0.01101958;        % survey-weighted mean using OR

z = zeros(N,1);
w1 = zeros(1,L);
w2 = zeros(L,L);
w3 = zeros(L,1);
x = zeros(N,1);

for i = 1:N;
      tempw1 = zeros(1,L);
      tempw2 = zeros(L,L);
      tempw1 = weight(i,1)*(mhat(i,1)*(1-mhat(i,1))*l(i,:));
      tempw2 = weight(i,1)*(mhat(i,1)*(1-mhat(i,1))*l(i,:)'*l(i,:));
      w1 = w1 + tempw1;
      w2 = w2 + tempw2;
end;

for j = 1:N;
      tempw3 = zeros(size(l,2),1);
      z(j,1) = mhat(j,1) - mu;
      tempw3 = l(j,:)'*(paid(j,1)-mhat(j,1));
      x(j,1) = weight(j,1)*((z(j,1) + (w1*inv(w2)*tempw3))^2);
end;

var = sum(x)/(N^2)
95ci_low = mu - (1.96*sqrt(var))
95ci_high = mu + (1.96*sqrt(var))
```

```
% DOUBLY ROBUST ESTIMATION VARIANCE
% import data from excel
[l to headertext] = xlsread('C:\covariatesl.xls');
[m to headertext] = xlsread('C:\covariatesm.xls');
[other to headertext] = xlsread('C:\other.xls');

N = size(l,1);          % number of subjects
L = size(l,2);          % number of covariates to OR model
M = size(m,2);          % number of covariates to IPW model
weight = other(:,1);    % survey weight
miss = other(:,2);      % indicator for item non-response
C = 1 - miss;           % C = 1 for observations with complete data
mhat = other(:,3);      % predicated probability of response
phat = other(:,4);      % predicated probability of complete data
paid = other(:,5);      % observed response
y_dr = other(:,6);      % doubly robust pseudo-outcome
mu = 0.01092076;        % survey-weighted mean using DR

z = zeros(N,1);
w1 = zeros(1,M);
w2 = zeros(M,M);
w3 = zeros(M,1);
w4 = zeros(1,L);
w5 = zeros(L,L);
w6 = zeros(L,1);
x = zeros(N,1);

for i = 1:N;
      tempw1 = zeros(1,M);
      tempw2 = zeros(M,M);
      tempw4 = zeros(1,L);
      tempw5 = zeros(L,L);
      tempw1 = weight(i,1)*C(i,1)*(1/(phat(i,1)^2))*phat(i,1)*(1-
            phat(i,1))*m(i,:)*(paid(i,1)-mhat(i,1));
      tempw2 = weight(i,1)*phat(i,1)*(1-phat(i,1))*m(i,:)'*m(i,:);
      tempw4 = weight(i,1)*(1-(C(i,1)/phat(i,1)))*l(i,:)*mhat(i,1)*(1-mhat(i,1));
      tempw5 = weight(i,1)*mhat(i,1)*(1-mhat(i,1))*l(i,:)*l(i,:)';
      w1 = w1 + tempw1;
      w2 = w2 + tempw2;
      w4 = w4 + tempw4;
      w5 = w5 + tempw5;
end;

for j = 1:N;
      w3 = zeros(size(m,2),1);
      w6 = zeros(size(l,2),1);
      z(j,1) = y_dr(j,1) - mu;
      w3 = m(j,:)'*(C(j,1)-phat(j,1));
      w6 = l(j,:)'*(paid(j,1)-mhat(j,1));
      x(j,1) = weight(j,1)*((z(j,1)-(w1*inv(w2)*w3)+(w4*inv(w5)*w6))^2);
end;

var = sum(x)/(N^2)
95ci_low = mu - (1.96*sqrt(var))
95ci_high = mu + (1.96*sqrt(var))
```