WEB APPENDIX 1: Direct maximum likelihood to account for outcome misclassification

We compare the proposed multiple imputation approach to account for outcome misclassification to a direct maximum likelihood approach outlined in Carroll et al (31) and detailed by Lyles et al (3). The direct maximum likelihood approach for a main study with a validation subgroup specifies the likelihood for the logistic regression model relating exposure to outcome as the product of the likelihood for the main study and the likelihood for the validation subgroup. In both likelihood terms, sensitivity and specificity are based on associations between observed outcome, gold standard outcome, and exposure defined using a logistic model

$$\eta_d = \text{logit}[\Pr(W = 1|D = d, X = x)] = \theta_0 + \theta_1 d + \theta_2 X + \boldsymbol{\theta_3 Z}, \text{ for } d = 0,1.$$

Sensitivity and specificity are calculated as

$$\text{SE}_i = \Pr(W = 1|D = 1, X = x_i, \boldsymbol{Z} = \boldsymbol{z_i}) = \frac{\exp(\eta_{i1})}{1 + \exp(\eta_{i1})},$$

and

$$\text{SP}_i = \Pr(W = 0|D = 0, X = x_i, \boldsymbol{Z} = \boldsymbol{z_i}) = \frac{\exp(\eta_{i0})}{1 + \exp(\eta_{i0})}.$$

The likelihood for the main study is

$$\begin{aligned} L_m = \prod_{i=1}^{n_m} \{ [(1 - \text{SP}_{\text{x}_i}) \times \Pr\big(D = 0 \big| X = x_{i,} \boldsymbol{Z} = \boldsymbol{z_i}\big) \\ + \text{SE}_{\text{x}_\text{i}} \times \Pr(D = 1|X = x_{i,}, \boldsymbol{Z} = \boldsymbol{z_i})]^{w_i} \times [\text{SP}_{\text{x}_i} \times \Pr(D = 0|X = x_i, \boldsymbol{Z} = \boldsymbol{z_i}\ ) \\ + \big(1 - \text{SE}_{\text{x}_\text{i}}\big) \times \Pr(D = 1|X = x_i, \boldsymbol{Z} = \boldsymbol{z_i})]^{(1-w_i)} \}, \end{aligned}$$

and the likelihood for the validation study is

$$\begin{aligned} L_v = \prod_{j=1}^{n_v} \{[\mathrm{SE}_{\mathrm{x}_j} \times \Pr\big(D = 1\big|X = x_j, \boldsymbol{Z} = \boldsymbol{z_j}\big)]^{w_j d_j} \times [(1 - \mathrm{SP}_{\mathrm{x}_j}) \\ \times \Pr\big(D = 0\big|X = x_j, \boldsymbol{Z} = \boldsymbol{z_j}\big)]^{w_j(1-d_j)} \times (1 - \mathrm{SE}_{\mathrm{x}_j}) \\ \times \Pr\big(D = 1\big|X = x_j, \boldsymbol{Z} = \boldsymbol{z_j}\big)]^{(1-w_j)d_j} \times [\mathrm{SP}_{\mathrm{x}_j} \\ \times \Pr\big(D = 0\big|X = x_j, \boldsymbol{Z} = \boldsymbol{z_j}\big)]^{(1-w_j)(1-d_j)} \}, \end{aligned}$$

where *i* indexes the n_m participants in the main study (but not the validation subgroup) (i=1, … , n_m) and *j* indexes the n_v participants in the validation subgroup (j=1, … , n_v). As in the main body of the paper, *D* is an indicator of the gold standard outcome status, *W* is the observed outcome status, *X* is the treatment group, and ***Z*** is the vector of covariates.

The direct maximum likelihood approach differed between the logistic model and log binomial model only in the choice of link function. In the logistic model,

$$\Pr(D = 1|X = x, \boldsymbol{Z} = \boldsymbol{z})] = \frac{\exp(\beta_0 + \beta_1 X + \boldsymbol{\beta_2 Z})}{1 + \exp(\beta_0 + \beta_1 X + \boldsymbol{\beta_2 Z})}\ ,$$

And in the log binomial model,

$$\Pr(D = 1|X = x, \boldsymbol{Z} = \boldsymbol{z})] = \exp(\beta_0 + \beta_1 X + \boldsymbol{\beta_2 Z})\ .$$

WEB APPENDIX 2: SAS code for multiple imputation to account for outcome misclassification

We adapted SAS code from Cole (4) to perform multiple imputation to account for outcome misclassification. Here, we present SAS code to account for outcome misclassification that is differential with respect to treatment group. We will illustrate use of Firth’s correction to prevent separation of data points when modeling the relationship between gold standard and observed exposure in the validation subgroup in the imputation model. The SAS code below could be adapted to account for nondifferential outcome misclassification by removing the interaction between treatment group and observed outcome in the imputation model.

```
*step 1: Fit logistic regression model relating gold standard outcome to observed outcome in
validation subgroup;
data m; col1=.; col2=.; col3=.; col4=.; col5=.; col6=.;
data a; set a; wx=w*x;
proc logistic data=a descending covout outest=b(keep=_name_ intercept w x wx z1 z2) noprint;
        where r=1;
        model d=w x wx z1 z2/firth;
data bb; set b; if  _name_="d"; bh0=intercept; bh1=w; bh2=x; bh3=wx; bh4=z1; bh5=z2;
        keep bh0-bh5;
data cov; set b; if _name_^="d"; keep intercept w x wx z1 z2 ;

*step 2: Sample coefficients for each imputation;
proc iml;
  use cov; read all into cov;        *variance-covariance matrix;
  use bb; read all into mu; mu=mu`;  *means;
  v=nrow(cov);                       *number of variables;
  n=40;                              *number of imputations;
  seed=222;
  l=t(root(cov));                    *cholesky root of cov matrix;
  z=normal(j(v,n,seed));             *generate nvars*samplesize normals;
  d=l*z;                             *premultiply by cholesky root;
  d=repeat(mu,1,n)+d;                *add in the means;
  td=t(d);
  create m from td;                  *write out sample data to sas dataset;
  append from td;
quit;
data m; set m;  retain _imputation_ 0; _imputation_=_imputation_+1;
   b0=col1; b1=col2; b2=col3; b3=col4; b4=col5; b5=col6;
```

```
    keep _imputation_ b0-b5 ;
data aa; merge d bb;  do _imputation_=1 to 40; output; end;
proc sort data=m; by  _imputation_;
proc sort data=aa; by  _imputation_;

*step 3: impute outcome for records not in the validation subgroup;
data c; merge aa m; by  _imputation_; call streaminit(9);
    if r=1 then d_imp=d;
    else  d_imp=rand("bernoulli",1/(1+exp(-(b0+b1*w+b2*x+b3*wx+b4*z1+b5*z2))));

*step 4a: Fit Logistic analysis model in each imputation;
proc logistic data=c outest=e covout noprint desc; by _imputation_;model d_imp=x z1 z2;

*step 4b: Fit Binomial analysis model in each imputation;

proc genmod data=c desc;
model d_imp=x z1 z2 / link=log dist=bin wald type3;
by _imputation_;
ods output parameterestimates=f;
run;

*step 4c: Summarize Logistic results over all imputations;
proc mianalyze data=e;  modeleffects x;
    title " multiple imputation to account for outcome misclassification";
run;

*step 4d: Summarize Binomial results over all imputations;
data f;
            set f;
            if parameter="x";
proc mianalyze data=f;
            modeleffects estimate;
            stderr stderr;
            ods output parameterestimates=mi3(keep= parm estimate stderr);
run;
```

WEB APPENDIX 3. Monte Carlo Simulation Methods

For each of 15 simulated scenarios, records were generated with values for treatment group (*x*), true outcome status (*d*), observed outcome status (*w*), and an indicator of inclusion in the validation subgroup (*r*). Half of the simulated records were assigned to each treatment group. True outcomes were simulated based on a Bernoulli distribution with the probability of being a case generated from a logistic regression model with the β coefficient for acyclovir equal to -0.693 (a true odds ratio of 0.5). Error-prone outcomes were generated based on hypothetical values for the accuracy of the outcome measure; one set of simulations generated observed outcome data with nondifferential misclassification with sensitivity set to be 0.90, 0.60, or 0.30, and the other assumed that outcome classification was more sensitive in exposed participants (sensitivity=0.95 and 0.70 for the two scenarios, respectively) than unexposed participants (sensitivity=0.85 and 0.50). Specificity was 0.90 in all scenarios.

For each scenario, a designated proportion was chosen to be included in the validation subgroup. For each record, *r* was sampled from a Bernoulli distribution with probability equal to the proportion included in the validation subgroup. As in the real data example, the true outcome was assumed to be known only for records where $r = 1$.

Each scenario was simulated 10,000 times. Bias was defined as 100 times the difference between the average estimated log odds ratio and the true log odds ratio. Confidence interval coverage was calculated as the percentage of simulations in which the estimated Wald-type confidence limits included the true value. Bias and precision were considered together using mean squared error, which was calculated as the sum of the square of the bias and the variance. Statistical power was calculated as the percentage of simulations in which the Wald-type confidence interval excluded the null value.