Web Appendix 1

We use a simple, hypothetical situation to illustrate the bias in estimating IRD using first events, based on the commonly used approach as described in the Methods section. Suppose the sample is equally divided into a vaccinated and an unvaccinated group, and there are two strata of equal size in each group. The disease incidence rate is constant over time and, without vaccination, equals to 4 and 6 events per person-year in the low and high risk stratum, respectively. On average, time to first event in the low and high risk groups are therefore 1/4 and 1/6 year, respectively, in the unvaccinated group. Suppose an intervention causes an absolute reduction of 2 events per person-year, so that the incidence rate is 2 and 4 events per person-year among vaccinated people in the low and high risk stratum, respectively. On average, time to first event in the low and high incidence risk groups are therefore 1/2 and 1/4 year, respectively, in the vaccinated group. Using all events observed in one year, the estimate for incidence rate in the control group is $(4 + 6)/2 = 5$ and that in the intervention group is $(2 + 4)/2 = 3$. Therefore, IRD$= 5 - 3 = 2$, and that's the truth in this case. But suppose only data on time to first event is used, the estimate for incidence rate in the control group is $2/(1/4 + 1/6) = 4.8$ and that in the vaccine group is $2/(1/2+1/4) = 2.67$. Therefore, IRD$= 4.8-2.67 = 2.13$, which is biased upward. It can be shown in the same way that if the intervention causes a proportional reduction in disease incidence (as opposed to absolute reduction), the use of all events to estimate the IRD will remain unbiased whereas the use of first events will be biased downward. In short, in the presence of heterogeneity, the use of all events for estimating IRD is preferred.

Web Appendix 2

Robust variance estimator in the case of $\mathbf{X}_i = (1, X_{i1})$ and $X_{i1} = 0, 1$

In matrix form,

$$\widehat{\mathrm{Var}}(\beta^*) = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\left(\sum_{i=1}^{n} Z_i^2\hat{e}_i^2\mathbf{X}_i^T\mathbf{X}_i\right)(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$$

Since $\sum_i X_{i1}^2 Z_i = \sum_i X_{i1} Z_i$,

$$\mathbf{X}^T\mathbf{W}\mathbf{X} = \begin{pmatrix} \sum_i Z_i & \sum_i X_{i1}Z_i \\ \sum_i X_{i1}Z_i & \sum_i X_{i1}^2 Z_i \end{pmatrix} = \begin{pmatrix} T_0 + T_1 & T_1 \\ T_1 & T_1 \end{pmatrix},$$

$$(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1} = \frac{1}{T_0T_1}\begin{pmatrix} T_1 & -T_1 \\ -T_1 & T_0 + T_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{T_0} & -\frac{1}{T_0} \\ -\frac{1}{T_0} & \frac{1}{T_0} + \frac{1}{T_1} \end{pmatrix}$$

and

$$\sum_i Z_i^2\hat{e}_i^2\mathbf{X}_i^T\mathbf{X}_i = \begin{pmatrix} \sum_i Z_i^2\hat{e}_i^2 & \sum_i X_{i1}Z_i^2\hat{e}_i^2 \\ \sum_i X_{i1}Z_i^2\hat{e}_i^2 & \sum_i X_{i1}Z_i^2\hat{e}_i^2 \end{pmatrix},$$

therefore, the first and second diagonal elements of $\widehat{\mathrm{Var}}(\beta^*)$ are the variance estimates for β_0^* and β_1, respectively, i.e.,

$$\widehat{\mathrm{Var}}(\beta_0^*) = \frac{1}{T_0^2}\sum_{\{i:X_{i1}=0\}} Z_i^2\hat{e}_i^2$$

and

$$\begin{aligned}\widehat{\mathrm{Var}}(\beta_1) &= \frac{1}{T_0^2}\sum_i(1 - X_{i1})Z_i\hat{e}_{\mathrm{new},i}^2 + \frac{1}{T_1^2}\sum_i X_{i1}Z_i\hat{e}_{\mathrm{new},i}^2 \\ &= \frac{1}{T_0^2}\sum_{\{i:X_{i1}=0\}} Z_i^2\hat{e}_i^2 + \frac{1}{T_1^2}\sum_{\{i:X_{i1}=1\}} Z_i^2\hat{e}_i^2\end{aligned}$$

where $T_g = \sum_{\{i: X_{i1}=g\}} Z_i$ for $g = 0$ or 1. Note that the first line in $\widehat{\text{Var}}(\beta_1)$ is the robust variance estimate for the OLS model that $Y_{\text{new},i} = Y_i/\sqrt{Z_i}$ and $X_{\text{new},i1} = \sqrt{Z_i}X_{i1}$ and $Y_{\text{new},i} = \beta_0^* \sqrt{Z_i} + \beta_1 X_{\text{new},i1} + e_{\text{new},i}$.

Take X_{i1} as a group factor to partition the subjects into two groups according to $X_{i1} = 0$ and $X_{i1} = 1$, and assume that $e_{\text{new},i}$'s are independently and identically distributed (i.i.d.) within each group, i.e, $\text{Var}(e_{\text{new},i} \mid X_{i1} = g, Z_i) = \psi_g$ $(g = 0, 1)$. Moreover, assume $\text{E}(v_i) = \nu = 0$ such that $\hat{\beta}_0^* = \hat{\beta}_0$ is the unbiased estimate for the incidence rate when $X_{i1} = 0$, then it follows that the unbiased estimate for ψ_g is the empirical sample variance, i.e.,

$$\widehat{\psi}_g = \frac{1}{N_g - 1} \sum_{i=1}^{N_g} \hat{e}^2_{\text{new},i}$$

where $N_g = \#\{i : X_{i1} = g\}$ such that the variance estimate for the rate difference proposed by Stukel et al. (11) is

$$\widetilde{\text{Var}}(\beta_1) = \frac{1}{T_0}\widehat{\psi}_0 + \frac{1}{T_1}\widehat{\psi}_1.$$

Therefore, under the i.i.d. assumptions for the $e_{\text{new},i}$'s within each level of X_{i1}, the robust variance estimate $\widehat{\text{Var}}(\beta_1)$ is consistent with $\widetilde{\text{Var}}(\beta_1)$, meaning that $\text{E}\left(\widehat{\text{Var}}(\beta_1) - \widetilde{\text{Var}}(\beta_1)\right) = 0$, since conditional on $\mathbf{X}$ and $\mathbf{Z}$,

$$\begin{aligned}
\text{E}\left(\widehat{\text{Var}}(\beta_1)\right) &= \text{E}\left(\frac{1}{T_0^2}\sum_i (1 - X_{i1}) Z_i \hat{e}^2_{\text{new},i} + \frac{1}{T_1^2}\sum_i X_{i1} Z_i \hat{e}^2_{\text{new},i}\right) \\
&= \frac{1}{T_0^2}\sum_i (1 - X_{i1}) Z_i \psi_0 + \frac{1}{T_1^2}\sum_i X_{i1} Z_i \psi_1 \\
&= \frac{1}{T_0}\psi_0 + \frac{1}{T_1}\psi_1 \\
&= \text{E}\left(\widetilde{\text{Var}}(\beta_1)\right).
\end{aligned}$$

Web Appendix 3

Simulation parameter configurations and data generation processes

Suppose there are n study subjects. Since Y_i $(i = 1, \cdots, n)$ is count data, we generate $Y_i \sim \text{Poisson}(Z_i(\mathbf{X}_i\beta + v_i))$ where v_i is the heterogeneity term. Equivalently, the inter-arrival time between two consecutive events for subject i follows an Exponential distribution with parameter $\mathbf{X}_i\beta + v_i$. The data generation process is as follows:

1. Generate the total length of follow-up time Z_i for subject i from a Uniform distribution: Uniform$(2, 3)$.

2. Generate $\mathbf{X}_i = (1, X_{i1}, \cdots, X_{ik})$ in two configurations:

   (a) No confounding, i.e., $\mathbf{X}_i = (1, X_{i1})$: $X_{i1} \sim \text{Bernoulli}(p)$ where $p = 0.5$ or 0.7, representing balanced and unbalanced study designs, respectively. Set $\beta = (1, -0.5)$ and $(0.05, -0.025)$. These values mimic the incidence per person-year of common events such as AOM and rare events such as RCP in young children, respectively.

   (b) One quantitative confounder i.e., $\mathbf{X}_i = (1, X_{i1}, X_{i2})$: set $\beta = (1, -0.5, 0.05)$ and (0.05, -0.025, 0.005). $X_{i1} \sim \text{Bernoulli}(p)$ where $p = 0.5$ or 0.7. $X_{i2} = 5\beta_0 \times \text{Beta}(\alpha_1 + \alpha_2 X_{i1}, \alpha_3)$. The parameters, α_j $(j = 1, 2, 3)$, of the Beta distribution are chosen to reflect different degrees of collinearity between X_{i1} and X_{i2} and different degrees of skewness. $(\alpha_1, \alpha_2, \alpha_3)$= (5, 2, 5) and (5, 5, 5) make X_{i2} fairly symmetric (skewness are approximately -0.2 and -0.5

for $\alpha_2 = 2$ and $\alpha_2 = 5$, respectively), though not normal. The correlations between X_{i1} and X_{i2} for the two settings are about 0.3 (moderate) and 0.5 (strong), respectively. Another two settings $(\alpha_1, \alpha_2, \alpha_3)$= (2, 2, 1) and (2, 5, 1) represent highly skewed distributions (skewness are approximately -1.0 and -1.6 for $\alpha_2 = 2$ and $\alpha_2 = 5$, respectively) where the correlations between X_{i1} and X_{i2} are about 0.3 (moderate) and 0.5 (strong), respectively.

3. Generate the heterogeneity term v_i according to three scenarios:

   (a) homogeneous: $v_i = 0$.

   (b) discrete: $\Pr(v_i = \frac{\beta_0+\beta_1}{2}) = p$ and $\Pr(v_i = -\frac{\beta_0+\beta_1}{2}) = 1 - p$ where p is same as the Bernoulli probability for $X_{i1} = 1$.

   (c) positive and continuous: $v_i \sim \text{Gamma}(p(1-p), \beta_0 + \beta_1)$.

   Note that with same p, the v_i in scenarios (b) and (c) have the same variance $p(1-p)(\beta_0 + \beta_1)^2$.

4. For the i-th subject, the j-th event time is generated using the inversion method $T_{i,j} = -\ln U/(\mathbf{X}_i\beta + v_i)$ where $U \sim \text{Uniform}(0, 1)$. Trivially, $T_{i,0} = 0$. Iterate until $\sum_{\ell \leq J} T_{i,\ell} > Z_i$ and let $Y_i = m_i = J - 1$. Estimation using all events is based on $\{i : \mathbf{X}_i, Z_i, Y_i\}$. Let $Y_{i,1} = 1$ and $Z_{i1} = T_{i1}$ if $T_{i1} \leq Z_i$. Otherwise $Y_{i,1} = 0$ and $Z_{i1} = Z_i$. The estimation using first events only is based on $\{i : \mathbf{X}_i, Z_{i1}, Y_{i1}\}$.

5. In a main series of simulations the performance of the proposed method using data on all events was evaluated and compared with the estimation using data on

first events and Stukel's method (if without confounder). For the high incidence scenarios with $\beta_1 = -0.5$, sample size was set as $n = 200$ and 1,000. For the low incidence scenarios with $\beta_1 = -0.025$, only $n = 1,000$ was used as no study would use $n = 200$ for such a rare outcome. The performance of the asymptotic and small-sample versions of the robust standard errors was further studied in more details in a supplementary series of simulations. The total sample size varied from $n = 100$ to 1,000 at interval of 100 for the high incidence scenarios and $n = 500$ to 2,000 at interval of 100 for the low incidence scenarios.

Web Appendix 4

STATA codes applied to the data in the case study

The proposed estimation approach requires transformations on the following variables: y: number of events. z: follow-up time. vaccine: 1 denotes receiving vaccine and 0 placebo. gender: 1 denotes male and 0 female. age: age at enrollment. district: 1 denotes Bansang, 2 Basse and 3 Others. ethnic: 1 denotes Mandinka, 2 Fula, 3 Serahule, 4 Wolof and 5 Others.

STATA codes:

```
generate y_wls=y/sqrt(z)
generate x0_wls=sqrt(z)
generate vaccine_wls=x*sqrt(z)
generate gender_wls=gender*sqrt(z)
generate age_wls=age*sqrt(z)
tabulate district,generate(ind_dist)
generate ind_dist1_wls=ind_dist1*sqrt(z)
generate ind_dist2_wls=ind_dist2*sqrt(z)
```

tabulate ethnic,generate(ind_eth)

enerate ind_eth2_wls=ind_eth2*sqrt(z)

generate ind_eth3_wls=ind_eth3*sqrt(z)

generate ind_eth4_wls=ind_eth4*sqrt(z)

generate ind_eth5_wls=ind_eth5*sqrt(z)

regress y_wls ind_eth2_wls−ind_eth5_wls vaccine_wls ind_dist1_wls ind_dist2_wls age_wls gender_wls x0_wls, nocons vce(robust)

Web Table 1. Simulation Results on the Estimates for β_1 With High Incidence in the Absence of Confounding ($n = 200$).

			First Events				All Events					
			Proposed Method						Proposed Method		Stukel	
p	β_1	v_i	Avg_est*	ESD†	Avg_se$_r$‡	CP$_r$§	Avg_est*	ESD†	Avg_se$_r$‡	CP$_r$§	Avg_se¶	CP♯
0.5	-0.5	0	-0.5083	0.1243	0.1220	94.9	-0.5002	0.0784	0.0775	94.8	0.0775	94.8
		±0.25	-0.5183	0.1202	0.1185	94.9	-0.5001	0.0864	0.0852	94.6	0.0852	94.5
		Gamma	-0.5093	0.1367	0.1353	94.7	-0.5001	0.0907	0.0909	94.9	0.0908	94.9
0.7	-0.5	0	-0.5140	0.1492	0.1465	94.8	-0.5010	0.0904	0.0899	94.5	0.0901	94.7
		±0.25	-0.5260	0.1611	0.1581	94.8	-0.4987	0.1018	0.1013	94.4	0.1015	94.4
		Gamma	-0.5193	0.1638	0.1589	94.5	-0.4996	0.1021	0.1014	94.4	0.1016	94.4

* Average of the parameter estimates.

† Empirical standard deviation.

‡ Average of the robust standard error estimates.

§ 95% coverage proportion based on the robust standard error estimates.

¶ Average of the standard error estimates using the method proposed by Stukel et al. (11).

♯ 95% coverage proportion based on the standard error estimates using the method proposed by Stukel et al. (11).

Web Table 2. Simulation Results on the Estimates for (β_1, β_2) With High Incidence, $n = 200$ and X_2 Slightly Skewed.

			Moderate Collinearity						Strong Collinearity					
			First Events			All Events			First Events			All Events		
p	Coefficient	v_i	Avg_est*	Avg_se$_r$†	CP$_r$‡	Avg_est*	Avg_se$_r$†	CP$_r$‡	Avg_est*	Avg_se$_r$†	CP$_r$‡	Avg_est*	Avg_se$_r$†	CP$_r$‡
0.5	$\beta_1 = -0.5$	0	-0.5067	0.1440	94.8	-0.4991	0.0880	94.8	-0.5094	0.1644	94.4	-0.4985	0.1006	94.6
		±0.25	-0.5182	0.1412	95.0	-0.4997	0.0955	94.5	-0.5199	0.1607	94.6	-0.5002	0.1091	94.8
		Gamma	-0.5148	0.1606	95.1	-0.5003	0.1010	94.9	-0.5150	0.1844	94.6	-0.4996	0.1151	94.5
	$\beta_2 = 0.05$	0	0.0506	0.0912	93.9	0.0497	0.0591	94.8	0.0492	0.1015	94.2	0.0499	0.0642	94.6
		±0.25	0.0522	0.0889	94.5	0.0513	0.0641	94.3	0.0516	0.0990	94.2	0.0508	0.0695	94.6
		Gamma	0.0533	0.1011	94.1	0.0508	0.0676	94.6	0.0520	0.1119	94.3	0.0498	0.0732	94.3
0.7	$\beta_1 = -0.5$	0	-0.5128	0.1682	94.6	-0.4998	0.0997	94.5	-0.5181	0.1844	94.5	-0.5005	0.1107	94.3
		±0.25	-0.5266	0.1805	94.8	-0.4990	0.1107	94.5	-0.5314	0.1986	94.4	-0.4999	0.1237	94.5
		Gamma	-0.5199	0.1874	94.7	-0.4995	0.1112	94.4	-0.5224	0.2046	94.5	-0.5000	0.1236	94.1
	$\beta_2 = 0.05$	0	0.0490	0.0845	94.3	0.0495	0.0569	94.5	0.0491	0.0967	94.2	0.0495	0.0639	94.5
		±0.25	0.0525	0.0918	94.8	0.0505	0.0645	94.2	0.0515	0.1051	94.5	0.0500	0.0723	94.8
		Gamma	0.0513	0.0926	94.1	0.0506	0.0646	94.2	0.0500	0.1060	93.9	0.0503	0.0724	94.5

* Average of the parameter estimates.

† Average of the robust standard error estimates.

‡ 95% coverage proportion based on the robust standard error estimates.

Web Table 3. Simulation Results on the Estimates for $(\beta_1,\ \beta_2)$ With High Incidence, $n = 200$ and X_2 Highly Skewed.

			Moderate Collinearity						Strong Collinearity					
p	Coefficient	v_i	First Events			All Events			First Events			All Events		
			Avg_est*	Avg_se$_r$†	CP$_r$‡	Avg_est*	Avg_se$_r$†	CP$_r$‡	Avg_est*	Avg_se$_r$†	CP$_r$‡	Avg_est*	Avg_se$_r$†	CP$_r$‡
0.5	$\beta_1 = -0.5$	0	-0.5066	0.1537	95.1	-0.4996	0.0919	94.7	-0.5100	0.1737	94.6	-0.5002	0.1028	94.4
		±0.25	-0.5205	0.1511	95.0	-0.4988	0.0993	94.7	-0.5210	0.1710	94.8	-0.5016	0.1106	94.9
		Gamma	-0.5120	0.1685	94.5	-0.4999	0.1045	94.6	-0.5145	0.1902	95.1	-0.5011	0.1161	94.8
	$\beta_2 = 0.05$	0	0.0475	0.0700	94.5	0.0499	0.0436	94.6	0.0470	0.0835	94.1	0.0500	0.0498	94.3
		±0.25	0.0492	0.0687	94.7	0.0501	0.0470	94.8	0.0481	0.0820	93.8	0.0505	0.0534	94.4
		Gamma	0.0484	0.0772	94.2	0.0505	0.0494	94.5	0.0480	0.0913	94.3	0.0510	0.0559	94.0
0.7	$\beta_1 = -0.5$	0	-0.5180	0.1776	94.6	-0.4998	0.1037	94.4	-0.5173	0.1986	94.2	-0.4999	0.1160	94.2
		±0.25	-0.5272	0.1907	94.4	-0.4980	0.1150	94.0	-0.5370	0.2140	94.1	-0.4997	0.1283	94.1
		Gamma	-0.5261	0.1926	94.8	-0.5017	0.1149	94.2	-0.5257	0.2145	93.9	-0.5003	0.1285	93.9
	$\beta_2 = 0.05$	0	0.0482	0.0684	94.3	0.0499	0.0446	94.4	0.0468	0.0897	93.6	0.0505	0.0555	94.1
		±0.25	0.0485	0.0742	94.3	0.0501	0.0500	94.5	0.0480	0.0967	93.7	0.0493	0.0617	94.1
		Gamma	0.0476	0.0749	94.8	0.0497	0.0500	94.3	0.0453	0.0973	93.9	0.0491	0.0617	94.4

* Average of the parameter estimates.

† Average of the robust standard error estimates.

‡ 95% coverage proportion based on the robust standard error estimates.