Article DOI: https://doi.org/10.3201/eid2412.180979

Prevalence of Avian Influenza A(H5) and A(H9) Viruses in Live Bird Markets, Bangladesh

Technical Appendix 1

Sample Size Calculations

The selection of 20 large and 20 small live bird markets (LBMs) was necessary to detect a statistically significant difference between 2 groups of LBMs with assumed LBM-level avian influenza virus prevalences of 50% and 10%, respectively (95% significance, 80% power). Assuming a prevalence of infection of 7% in chickens in a contaminated LBM, detecting the infection in chickens traded in a contaminated LBM with a confidence of 95% required the sampling of 40 chickens.

Assuming a prevalence of infection of 7% in chickens and 14% in waterfowls in a contaminated LBM, sampling 40 chickens and 20 waterfowls from each LBM was needed to detect at least 1 infected chicken and 1 infected duck with a 95% significance. For chicken breeds, we aimed to sample 15 broilers, 15 Desi (local in Bengali, indigenous chicken breeds raised in backyard farms) and 10 Sonali (cross-breed of the Rhode Island Red cocks and Fayoumi hens) in LBMs in which more broilers were sold than Sonali, and 10 broilers, 15 Desi and 15 Sonali in in LBMs in which more Sonali were sold than broilers, hypothesizing that their numbers were proportional to their relative numbers sold in Dhaka and Chittagong LBMs. For waterfowl, we aimed to sample 15 ducks and 5 geese/LBM, or 20 ducks were sampled if geese were not present. Assuming that the probability of a pool of 5 environmental samples being contaminated in a contaminated LBM was 0.3 (and the probabilities of environmental sites being contaminated were independent), it was necessary to collect 10 pools to detect at least 1 contaminated pool with a probability of 0.97.

Bayesian Hierarchical Logistic Regression Model

For a given viral subtype (i.e., H5 or H9), the contamination status of a given LBM *m*, ω_m , was assumed to follow a Bernoulli distribution with parameter γ_{α_m} , the probability of a

LBM of type α_m being contaminated. In other words, γ_{α_m} could be interpreted as the LBM-level prevalence for LBMs of type α_m (Equation.2.1):

$\omega_m \sim Bernoulli(\gamma_{\alpha_m})$ (Equation 2.1)

In each iteration, the contamination status of each LBM was first simulated, such that all pools from the same LBM originated from either a contaminated (i.e., some pools could be positive) or noncontaminated LBM (i.e., all these pools were necessarily negative). Because models were simulated separately for poultry and environmental samples, the interpretation of LBM-level prevalence differed accordingly: the LBM-level prevalence estimated from a model based on poultry (or environmental) samples referred to the proportion of LBMs with at least 1 infected poultry (or contaminated environmental site).

The real-time reverse transcription PCR result of pool *i* in LBM *m*, $y_{i,m}$, was assumed to follow a Bernoulli distribution with parameter $\theta_{i,m}$ (Equation 2.2), the probability of pool *i* being contaminated (Equation 2.3):

$y_{i,m} \sim Bernoulli(\theta_{i,m})$ (Equation 2.2)

$\theta_{i,m} = \omega_m \times \lambda_{i,m}$ (Equation 2.3)

$\lambda_{i,m}$ was the probability of pool *i* in LBM *m* being contaminated if this LBM was contaminated. If LBM *m* was not contaminated ($\omega_m = 0$), all pools from this LBM were negative by real-time reverse transcription PCR ($y_{i,m} = 0$). If LBM *m* was contaminated ($\omega_m = 1$), $y_{i,m}$ was then simulated by a Bernoulli trial with parameter $\lambda_{i,m}$ (as $\theta_{i,m} = \lambda_{i,m}$). A pool *i* was positive if at least 1 of its swabs was infected. Therefore, $\lambda_{i,m}$ was expressed as a function of 1) the underlying bird– (or environmental swab)–level prevalence, $\pi_{i,m}$ and 2) the number of birds (or environmental swabs) comprising pool *i* in LBM *m*, $n_{i,m}$ (Equation 2.4):

$\lambda_{i,m} = 1-(1-\pi_{i,m})^{n_{i,m}}$ (Equation 2.4)

$\pi_{i,m}$ was assessed through a Bayesian hierarchical logistic regression to account for the hierarchical data structure. $\pi_{i,m}$ only depended on the type of sample (Equation 2.5):

$$logit(\pi_{i,m}) = \delta_m + \sum_j \beta_j \times \psi_{j,i,m} \text{ (Equation 2.5)}$$

β_j was a regression coefficient for sample of type *j*, and $\psi_{j,i,m}$ an indicator variable, equal to 1 if the pool *i* in market *m* was of type *j*, and null otherwise. δ_m was the LBM-specific intercept.

At the second level, δ_m was assumed to follow the LBM-specific normal distribution (Equation 2.6):

$$\delta_m \sim Normal(\mu_m, \sigma_m{}^2) \text{ (Equation 2.6)}$$

The variance $\sigma_m{}^2$ assumed that prevalence varied between LBMs after adjusting for a LBM-level predictor. The mean μ_m was modeled as a linear function of a LBM-level intercept, ϕ, and 1 of the LBM-level predictors, B_α (Equation 2.7), which was used to differentiate the LBM-level prevalence:

$$\mu_m = \phi + B_\alpha \times \tau_{\alpha_m} \text{ (Equation 2.7)}$$

τ_{α_m} was an indicator variable, equal to 1 if the market *m* was of type α; it was otherwise null. All unknown parameters were specified by weakly informed priors to enable the observed data to be the main contributor to the estimation of the posterior distributions (online Technical Appendix 1 Table).

The models were run by using a Markov chain Monte Carlo simulation in JAGS (*1*) and R.3.4.2 (*2*). After a burn-in period of 5,000 iterations, each model was iterated up to the point where convergence was achieved in all parameters on the basis of the Gelman and Rubin statistic (*3,4*) and the effective sample size (*5*). Although LBM-level prevalence was estimated directly from each model, the underlying bird- and environmental site-level prevalence was estimated by taking the inverse logit transformation of the corresponding regression coefficients. Median and 95% highest density interval are reported. All possible combinations of lower- and higher-level regression predictors were tested. Models were compared with each other on the basis of the deviance information criterion (DIC) (*6*). Half of the variance of the posterior mean deviance was used as an estimate of the effective number of parameters (*7*). Models with lower DIC were considered to better support the data than those with higher DIC if the DIC difference was >5. Finally, a posterior predictive check was performed to assess model adequacy. In each iteration, model parameter values were sampled from their joint posterior distribution. The contamination status of each market, and in contaminated markets, the contamination status of each pool were then simulated. The pool-level prevalence was computed and formed the posterior predictive distribution along with those computed from other iterations. This prevalence was compared with the observed pool-level prevalence by using the Bayesian p value, which represents the probability that the former could be equal to or more extreme than the latter (*7*).

References

1. Plummer M. JAGS version 3.4.0 user manual, 2013 [cited 2018 Aug 23]. http://www.stats.ox.ac.uk/~nicholls/MScMCMC15/jags_user_manual.pdf

2. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016 [cited 2018 Aug 23]. https://www.r-project.org/

3. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat. 1998;7:434–55.

4. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci. 1992;7:457–72. http://dx.doi.org/10.1214/ss/1177011136

5. Kruschke JK. Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. 2nd ed. New York: Academic Press; 2014.

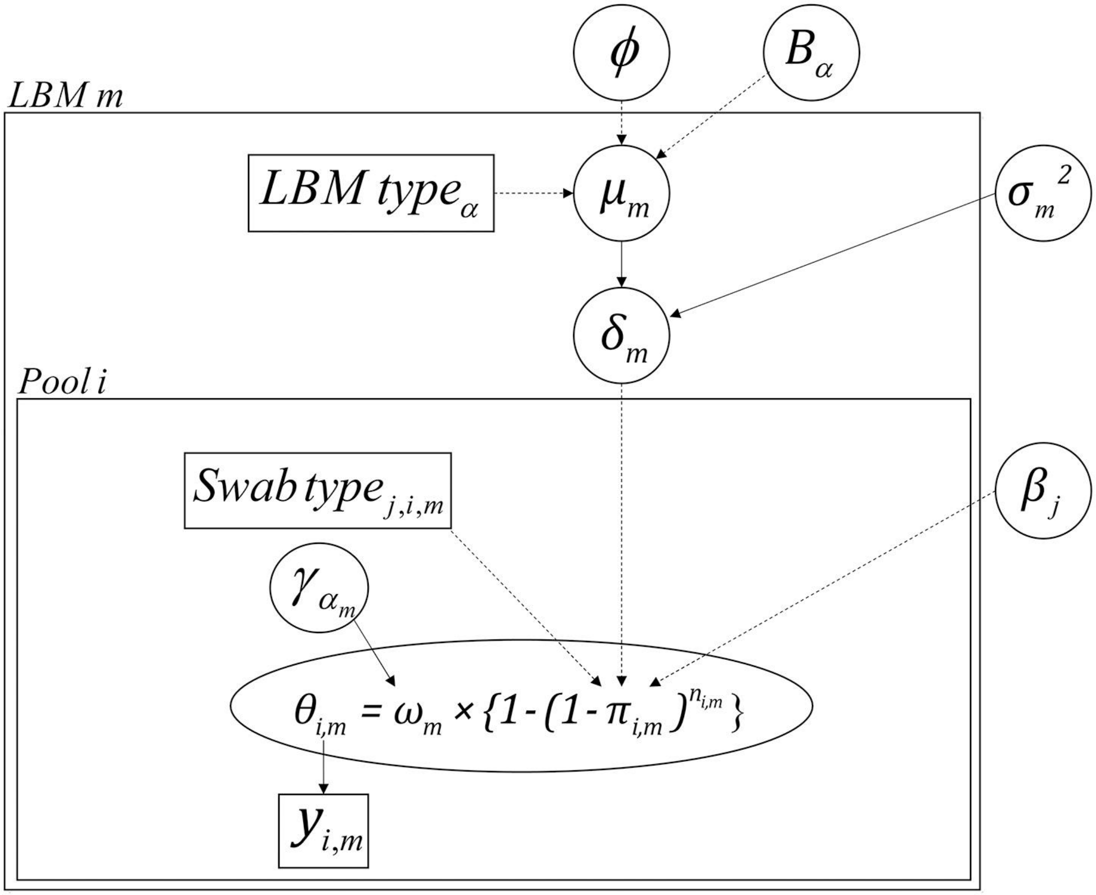
6. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol. 2002;64:583–639. http://dx.doi.org/10.1111/1467-9868.00353

7. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis, 2nd ed. Milton Park (UK): Taylor & Francis; 2003.

Technical Appendix 1 Table. Weakly informed priors used in models for analyzing prevalence of avian influenza A(H5) and A(H9) viruses in live bird markets, Bangladesh

Equation	Notation	Prior distribution*
2.1	γ	Beta (1,1)
2.5	β	Normal (0,1,000)
2.6	σ	Uniform (0, 100)
2.7	B	Normal (0, 1,000)
2.7	ϕ	Normal (0, 1,000)

*Posterior distributions are presented in online Technical Appendix 2 Figures 2, 3 (https://wwwnc.cdc.gov/EID/article/24/12/18-0879-Techapp2.pdf).



Technical Appendix 1 Figure. Model for analyzing prevalence of avian influenza A(H5) and A(H9) viruses in live bird markets, Bangladesh. The 2-level hierarchical relationship between data and model parameters is presented. Rectangles indicate constants, and circles indicate variables. Solid arrows indicate stochastic dependency, and dashed arrows indicate deterministic dependency. Subscript letters correspond to those in the model description.