Predicting primate sources of Zika virus spillover

Barbara A. Han¹, Subho Majumdar²,³, A. Mojsilović³, Kush Varshney³

¹Cary Institute of Ecosystem Studies, ²Univ. Florida Informatics Institute, ³IBM Watson Research Lab

BACKGROUND
• Non-human primates maintain mosquito-borne flaviviruses in the wild (reservoirs).
• Successful control of Zika virus relies on spillover prevention from wild reservoirs.
• Primate diversity & abundance is exceptionally high in the neotropics, especially Brazil.
• Identifying potential primate reservoirs will require highly targeted surveillance.

OBJECTIVES
• Which primate species are most likely to carry Zika virus in the Neotropics?
• Where are these species distributed?
• Where should human and non-human primate surveillance be focused?

METHODS
• Collect primate trait data to identify primate species with the highest probability of carrying Zika virus.
  • Why trait data? Organismal trait data are useful because:
    • They are immediately available (unlike Zika virus surveillance data)
    • They reliably distinguish one host species from another
    • They underpin species’ intrinsic capacity to harbor zoonoses
• Apply a machine learning method on imputed trait data to recognize primate species resembling known flavivirus hosts to target Zika virus surveillance.

DATA
RESPONSE VARIABLE: Host status for Zika & other mosquito-borne flaviviruses (YFV, JEV, SLEV, DENV; flavivirus) for all primates (N=376); 20 positive species in 376 x 6 matrix; only 2 Zika+ species

METHODS ROLE: Identify species at risk

PREDICTOR VARIABLES: 50 organismal traits (e.g., body mass, longevity, metabolic rate, geographic range area, population density; compiled from 1+)

MODEL PERFORMANCE
• Model classified flavivirus+ primate with 82% accuracy
• Model identified particular primate species as highly likely Zika virus hosts
• Model identified primates as likely hosts for other flaviviruses

MODEL VALIDATION
Out-of-sample validation: Reassigned virus-positive primate to negative, and re-ran the model to obtain risk scores.
• Model assigned high risk scores to majority of known primate hosts.
• Model assigned low risk of flavivirus positivity to some known hosts.
• Data scarcity, even for very common species, may preclude our capacity to predict flavivirus reservoirs.

RESULTS & CONCLUSIONS
A map depicting overlapping species ranges of global primates and their probability of being Zika virus positive.

6 primate species in Central and South America in the 90th percentile probability of being Zika+.

Acknowledgments

PEOPLE: IBM WATSON DATA SCIENCE GROUP, CHRIS HOUGH (IBM graphic artist)
FUNDING: IBM SOCIAL GOOD FELLOWSHIP PROGRAM (S. MAJUMDAR), NSF EID GRANT # ... (B. HAN)