# Targeting surveillance in apparently healthy versus diseased wild mammals for zoonotic virus discovery

## Authors

1. Levinson, Jordan “Ms. Jordan Levinson, MSc” (EHA)[[1]](#footnote-1)
2. [Bogich](http://orcid.org/0000-0002-8143-5289), Tiffany (L) “Dr. Tiffany L Bogich, PhD” (EHA; EEB; NIH)1
3. [Daszak](http://orcid.org/0000-0002-2046-5695), Peter “Peter Daszak, PhD” (EHA)[[2]](#footnote-2),[[3]](#footnote-3)

## Contributors

* Olival, Kevin (J) “Kevin J Olival, PhD” (EHA)
* Epstein, Jonathan (H) “Jonathan H Epstein, DVM, MPH” (EHA)
* [Johnson](http://orcid.org/0000-0001-6673-8743), Christine (K) “Prof. Christine Kreuder Johnson, DVM, MPVM, PhD” (UCDavis)
* Karesh, William “Billy Karesh, DVM” (EHA)

## Affiliations

* [EcoHealth Alliance (EHA)](http://www.ecohealthalliance.org/) - New York, NY, USA
* [Fogarty International](http://www.fic.nih.gov/) < [National Institutes of Health (NIH)](http://www.nih.gov/) - Bethesda, Maryland, USA
* Department of Ecology and Evolutionary Biology (EEB) < [Princeton University](http://www.princeton.edu/main/) - Princeton, NJ, USA
* School of Veterinary Medicine < [University of California Davis (UCDavis)](http://ucdavis.edu/) - Davis, CA, USA

## Keywords

* Emerging Infectious Diseases
* wildlife disease
* surveillance
* zoonoses
* epizootic mortality

## License

[CC0-1.0](http://creativecommons.org/publicdomain/zero/1.0/legalcode).

## Copyright

© 2013 Ms. Jordan Levinson, MSc; Dr. Tiffany L Bogich, PhD; Peter Daszak, PhD; Kevin J Olival, PhD; Jonathan H Epstein, DVM, MPH; Prof. Christine Kreuder Johnson, DVM, MPVM, PhD; Billy Karesh, DVM

## Abbreviations

* EID: Emerging Infectious Diseases
* ICTV: International Committee on the Taxonomy of Viruses
* PCR: Polymerase Chain Reaction

## Abstract

We analyzed a database of mammal-virus associations to ask whether zoonotic disease surveillance targeting diseased animals is the best strategy to identify potentially zoonotic pathogens. Though a mixed healthy and diseased surveillance strategy is generally best, surveillance of apparently healthy bats and rodents would likely maximize discovery potential for most zoonotic viruses.

## Introduction

Nearly two-thirds of emerging infectious diseases (EIDs) (Eidson, et al. 2001) that affect humans are zoonotic, and three-quarters of these originate in wildlife, making surveillance of wildlife for novel pathogens part of a logical strategy to prevent future zoonotic EIDs (Taylor, Latham and Woolhouse 2001, Jones, et al. 2008, Woolhouse and Gowtage-Sequeria 2005, USAID 2009). In this study, we adhere to the following definition of an Emerging Infectious Disease:

Infectious diseases whose incidence in humans has increased in the past 2 decades or threatens to increase in the near future have been defined as ‘emerging.’ These diseases, which respect no national boundaries, include: new infections resulting from changes or evolution of existing organisms, known infections spreading to new geographic areas or populations, previously unrecognized infections appearing in areas undergoing ecologic transformation, and old infections reemerging as a result of antimicrobial resistance in known agents or breakdowns in public health measures.

—The Centers for Disease Control and Prevention

With respect to zoonotic virus discovery, Simon Anthony said, “we could feasibly find most of the viruses that exist in mammals in the next 20 years”. Wildlife are thought to harbor a high diversity of unknown pathogens, but global characterization of this diversity would be costly and logistically challenging (Morse 1993). Given limited resources for pandemic prevention, there is public health benefit in focusing pathogen discovery on those species most likely to harbor novel zoonoses (Woolhouse and Gowtage-Sequeria 2005, USAID 2009). One strategy to maximize the likelihood of discovering novel pathogens is surveillance of animal die-offs, outbreaks in wildlife, or diseased wildlife. Here, we analyze a database of known zoonotic viruses in mammal hosts to answer the driving question of whether we should stratify surveillance strategies (i.e. visibly diseased versus apparently healthy animals) by wildlife host groups to best detect novel pathogens with zoonotic potential. In answering this question, we can better determine how host and virus taxonomy might influence our decisions about applying limited surveillance resources to a growing global health problem.

## Methods

We focused our analysis on mammalian hosts and viruses as they are more likely to be associated with human EIDs than any other host-pathogen type (Cleaveland, Laurenson and Taylor 2001, Woolhouse and Gowtage-Sequeria 2005). We constructed a database of all human emerging viruses previously identified as originating in wildlife (Jones, et al. 2008), supplemented with all zoonotic viruses from the International Committee on the Taxonomy of Viruses (ICTV) database ([www.ictvdb.org](http://www.ictvdb.org/)) with non-human, mammalian hosts. For each zoonotic virus, we conducted a literature search for reports of infection in any mammalian host, using virus name and relevant synonyms ([www.ictvdb.org](http://www.ictvdb.org/)) as keywords in ISI Web of Knowledge, Wildlife Disease Association Meeting Abstracts, Google Scholar, and the Global Mammal Parasites Database ([www.mammalparasites.org](http://www.mammalparasites.org/)). The resulting 605 host-pathogen relationships included 56 unique viruses from 17 viral families and 325 unique mammals from 15 orders. [[4]](#footnote-4)

We then conducted a secondary literature search to determine whether viruses in our database cause signs of disease in their wildlife hosts, using an aggregate of all publications available on PubMed, ISI Web of Science, BIOSIS Previews, and Biological and Agricultural Index Plus, with search terms consisting of virus names and ICTV synonyms, host genus and species names and common names (reconciled to the 2005 version of Mammal Species of the World (Wilson and Reeder 2005)). All resulting abstracts and available full text reports were examined until the first ‘robust’ report of visible disease was encountered. A report was considered ‘robust’ only if infections were confirmed by PCR analysis or virus isolation and clinical signs were explicitly recorded to have occurred during active infection. We excluded studies only reporting serology because of potential cross-reactivity among related viruses and poor correlation between serologic status and concurrent infection. For mammal-virus pairs without visible disease the search was exhaustive.

Viruses were identified as causing visible disease in a host if individual or epizootic mortality, or grossly visible or otherwise observable signs of morbidity such as high fever, loss of mobility, or severe reduction in body condition were reported. We considered diseases to be nonpathogenic in their hosts only if actively infected animals were explicitly reported to be free of visible disease. Animals with less clear signs such as nasal discharge or neonatal mortality were not considered ‘asymptomatic’ because of the low detection probability associated with these traits in wild mammal surveillance. We rejected reports of experimentally induced disease because of the risk that dosage and inoculation technique would not be consistent with naturally occurring infections. However, we included experimental studies if actively infected animals remained asymptomatic, with the assumption that clinical signs of infection were most likely to be seen in animals monitored in laboratory settings than in the wild, and that stressful conditions in captivity would heighten the likelihood of a normally benign pathogen leading to clinical signs (Williams and Barker 2008). Furthermore, experimental infections often involve more direct routes of inoculation than naturally occurring infections, and are therefore more likely to induce disease.

We first determined the percentage of reports of host-virus pairs in Supporting Dataset 1 in which observable disease was described, ‘Symptomatic’, no observable disease was described, ‘Asymptomatic’, or no description of disease was included, ‘No data’. We plotted the results in a pie chart using the default pie function and the RColorBrewer library (Code 1).

Code 1. R code for creating Figure 1A. Authors: Dr. Tiffany L Bogich, PhD (EEB)[[5]](#footnote-5)[[6]](#footnote-6). License: [CC0-1.0](http://creativecommons.org/publicdomain/zero/1.0/legalcode). Data from: Supporting Dataset 1.

rm(list=ls())

library(RColorBrewer)

counts<-c(293,88,224)

pc<-c("48%","15%","37%")

labs<-c("No Data","Symptomatic","Asymptomatic")

#piechart of all drivers

postscript("~/12-1042-F1A.eps", width=5, height=4,pointsize=10, onefile=FALSE, horizontal=FALSE, paper="special")

par(mar=c(2,0,1.5,0))

pie(counts, labels=pc, ps=14,edges=400,radius=0.5, col=brewer.pal(9,"Greys")[c(1,3,6)],lty=1,lwd=1, init.angle=-90)

legend(0.6,-0.1, labs,col=brewer.pal(9,"Greys")[c(1,3,6)],fill=brewer.pal(9,"Greys")[c(1,3,6)],bty="n",cex=0.9)

dev.off()

We then conducted a logistic regression analysis of host apparent disease as a function of host taxonomic group and virus taxonomy for the subset of mammal-virus pairs for which the host order or virus family had at least 3 records in the database using Firth’s bias reduction procedure (Firth 1993) in R statistical software package ‘brglm’ (R v2.15-2) and the brglm function for the removal of the leading term from the asymptotic expansion of the bias of the maximum likelihood estimator (Equation 1).

Equation 1. Authors: Dr. Tiffany L Bogich, PhD (EEB)[[7]](#footnote-7).

We then calculated odds ratios for each host order and virus family relative to the reference categories (Flaviviridae and Artiodactyla) and the predicted probability of being symptomatic for all species order-virus family combinations.

## Results

Figure 1. Percentage of reports of host-virus pairs in which observable disease was described, ‘Symptomatic’, no observable disease was described, ‘Asymptomatic’, or no description of disease was included, ‘No data’ (A). The proportion of hosts symptomatic by mammal order (B) and the proportion of virus families for which hosts are reported symptomatic (C) are given, both with standard error bars, calculated assuming binomial error structure. The total number of each host order or virus family included in the database is given above each bar. Note, all host orders and virus families in the database are included here, but analyses are limited to those host orders or virus families with at least three entries in the database. Authors: Ms. Jordan Levinson, MSc[[8]](#footnote-8); Dr. Tiffany L Bogich, PhD[[9]](#footnote-9). License: [CC0-1.0](http://creativecommons.org/publicdomain/zero/1.0/legalcode). Code from: Code 1; Supporting Code 1. Data from: Supporting Dataset 1.

Our search of 605 mammal-virus associations investigated yielded explicit information on host health in 52% of the 312 mammal-virus pairs. Of these, approximately 28% of infected wildlife hosts were reported to present with visible disease (n = 88) and 72% (n = 224) were reported without evidence of visible disease (Figure 1A). The proportion of hosts that were symptomatic differed across host order (Figure 1B) and virus family (Figure 1C).

We found that virus family and host order were significant predictors of disease status (χ2=88.70, p<0.001 and χ2=59.45, p<0.001, respectively). Species infected with paramyxoviruses, poxviruses, and reoviruses were more likely to be visibly diseased whereas species infected with bunyaviruses were less likely to be visibly diseased relative to the reference category. Hosts infected with filoviruses were marginally more likely to be visibly diseased (Table 1).

Table 1. Logistic regression analysis with bias reduction of whether a host presents with disease for 234 mammal–virus pairs from 5 taxonomic orders of mammals and 10 taxonomic families of viruses. The subset of data used was selected by using a cutoff of at least 3 records in the database to avoid making inference about host orders or virus families, for which we had very little information. Authors: Kevin J Olival, PhD[[10]](#footnote-10); Prof. Christine Kreuder Johnson, DVM, MPVM, PhD[[11]](#footnote-11). License: [CC0-1.0](http://creativecommons.org/publicdomain/zero/1.0/legalcode). Data from: Supporting Dataset 1. Code from: Supporting Code 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Predictor[[12]](#footnote-12) | Coefficient | SE | Test statistic (Z) | p-value | OR | 95% CI |
| Constant | -0.33 | 0.58 | -0.56 | 0.58 | 0.72 | (0.23, 2.26) |
| Virus Family (Reference category: Flaviviridae) | | | | | | |
| Bunyaviridae | -1.74 | 0.64 | -2.71 | 0.01 | 0.18 | (0.05, 0.62) |
| Filoviridae | 3.26 | 1.83 | 1.78 | 0.08 | 26.07 | (0.72, 944.49) |
| Herpesviridae | 0.10 | 0.65 | 0.16 | 0.87 | 1.11 | (0.31, 3.94) |
| Paramyxoviridae | 3.43 | 1.42 | 2.41 | 0.02 | 30.95 | (1.90, 503.52) |
| Picornaviridae | 1.12 | 0.76 | 1.48 | 0.14 | 3.08 | (0.69, 13.68) |
| Poxviridae | 2.29 | 0.81 | 2.82 | <0.001 | 9.90 | (2.01, 48.72) |
| Reoviridae | 2.13 | 1.05 | 2.02 | 0.04 | 8.39 | (1.07, 66.12) |
| Rhabdoviridae | 9.20 | 2.39 | 3.85 | <0.001 | NA[[13]](#footnote-13) | NA2 |
| Togaviridae | -0.36 | 0.63 | -0.58 | 0.56 | 0.70 | (0.20, 2.38) |
| Species Order (Reference category: Artiodactyla) | | | | | | |
| Chiroptera | -6.47 | 1.81 | -3.57 | <0.001 | 0.00 | (0.00, 0.05) |
| Perissodactyla | 0.58 | 0.76 | 0.77 | 0.44 | 1.79 | (0.40, 8.03) |
| Primates | -0.16 | 0.68 | -0.24 | 0.81 | 0.85 | (0.22, 3.24) |
| Rodentia | -1.12 | 0.67 | -1.66 | 0.10 | 0.33 | (0.09, 1.22) |

Species in the order Chiroptera (e.g. *Pipistrellus pipistrellus*) were less likely to be visibly diseased and species in the order Rodentia were marginally less likely to be visibly diseased, relative to the reference category (Table 1). Species in the order Chiroptera have a lower probability of visible disease than in other orders (Figure 2), though **all Chiroptera infected with non-rabies rhabdoviruses have a high probability of visible disease**. In the dataset, all host-pairs with rhabdoviruses were found in Chiroptera and were reported with visible disease in that host (Figure 1).

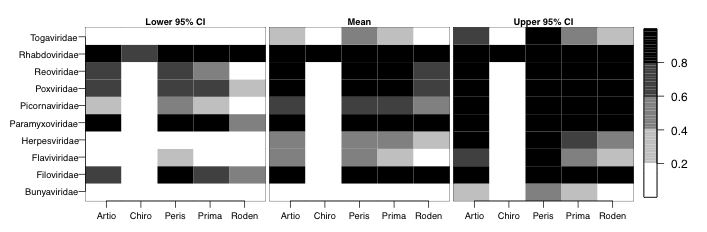
****

Figure 2. The probability of being symptomatic based on a logistic regression analysis with bias reduction of whether or not a host presents with disease for 234 mammal-virus pairs from 5 orders of mammals and 10 families of viruses, including the lower 95% confidence interval (left), mean (center), and upper 95% confidence interval (right). Probabilities are based on the predicted values of the logistic regression and are given on a five-point gray scale from white (0.0 – 0.2) to black (0.8 - 1.0). Confidence values were calculated as the coefficient plus 1.96 times the standard error (from Table 1). Authors: Dr. Tiffany L Bogich, PhD (NIH)[[14]](#footnote-14); Ms. Jordan Levinson, MSc[[15]](#footnote-15); Prof. Christine Kreuder Johnson, DVM, MPVM, PhD[[16]](#footnote-16). Contributors: Kevin J Olival, PhD1. License: [CC0-1.0](http://creativecommons.org/publicdomain/zero/1.0/legalcode). Data from: Table 1, Supporting Dataset 1. Code from: Supporting Code 1.

## Conclusions

Our data suggest that Chiroptera and Rodentia, two of the three main mammalian orders often targeted for zoonotic disease surveillance (the third being non-human primates (Leendertz, et al. 2006, Wolfe, Daszak, et al. 2005, Wolfe, Escalante, et al. 1998) are less likely to present with visible disease than other orders (Figure 1). The mechanism behind this relationship is an important area for additional research. Generally, we found that the probability of presenting with visible disease depends on the host and virus taxonomy, and the only host order for which a single strategy (in this case healthy animal surveillance) can be applied across nearly all virus families (excluding Rhabdoviridae) is for Chiroptera. Therefore, particularly for the case of novel virus detection, our results point to a mixed strategy of targeted syndromic and healthy animal surveillance across host and virus taxonomies. A mixed strategy could combine apparently healthy animal surveillance (particularly in Chiroptera) with syndromic surveillance in other wildlife and domestic animal hosts, as syndromic surveillance has previously proven useful where secondary animal hosts are involved (e.g. surveillance for West Nile virus (Eidson, et al. 2001), henipaviruses (Mohd, Gan and Ong 2000, Selvey, et al. 1995), and Ebola virus (Leroy, et al. 2004)) (Text Box 1).

Text Box 1. Surveillance for West Nile virus with crow deaths. Author: Dr. Tiffany L Bogich, PhD. License: [CC0-1.0](http://creativecommons.org/publicdomain/zero/1.0/legalcode).

Crow deaths were used as a sentinel surveillance system for West Nile virus detection in the northeastern part of the United States during the outbreak in the summer and fall of 1999 (Eidson, et al. 2001). From August to December 1999, 295 dead birds weere laboratory-confirmed with West Nile virus infection 89% of which were American Crows (*Corvus brachyrhynochos*). The complete genome of the West Nile virus isolated from crows is available on GenBank (accession number [KJ501319.1](http://www.ncbi.nlm.nih.gov/nuccore/KJ501319)). Bird deaths were critical in identifying West Nile virus outbreak and provided a sensitive method of detecting West Nile virus ahead of detection in humans.

There are limitations to our study, particularly ascertainment and reporting biases, as acknowledged in previous studies of EIDs (Woolhouse and Gowtage-Sequeria 2005, Jones, et al. 2008). In addition, differences in the number of species belonging to each order, the difficulty of testing inaccessible species and limits to reliable diagnoses of emerging viruses have an impact, especially in resource-poor settings. Further, many disease states are not recognizable in free-ranging mammalian species under field conditions. Lastly, there is a risk that an animal may be co-infected with a number of agents, only one of which causes disease; or that co-infection may have an additive or synergistic effect on clinical signs, and that anthropozoonotic viruses artificially inflate the ‘disease’ count of some mammalian orders over others. However, our findings are based on an aggregation of the best data available to date on host health as it relates to zoonotic viruses and have useful implications for public health.

Our analysis supports a holistic, probability-based approach to zoonotic virus discovery, specifically, continued analysis of passively- and actively-reported mortality events and increased investment in broad surveillance of healthy wildlife. The latter could be targeted geographically to those regions most likely to generate novel EIDs (Jones, et al. 2008) or taxonomically to those groups which are reservoirs for the highest proportion of zoonoses (Woolhouse and Gowtage-Sequeria 2005, Cooper, et al. 2012). These efforts could be envisaged as part of a strategy for ‘smart surveillance’, heightening the opportunity for discovery of novel zoonoses, particularly if wildlife are sampled at key interfaces where contact with people or domestic animals and thus the opportunity for spillover is highest.

## Funding

|  |  |  |  |
| --- | --- | --- | --- |
| Funder | Grant | Targets | Notes |
| Emerging Pandemic Threats PREDICT program (PREDICT) < [United States Agency for International Development (USAID)](https://www.usaid.gov/) – Washington, D.C., USA |  | The work |  |
| [National Institute of Allergy and Infectious Diseases (NIAID)](http://www.niaid.nih.gov/) – Bethesda, MD, USA | R01 AI079231 “Non-biodefense emerging infectious disease research opportunities award” | Peter Daszak, PhD |  |
| Research and Policy for Infectious Disease Dynamics program (RAPIDD) < Science and Technology Directorate < [U.S. Department of Homeland Security (DHS)](http://www.dhs.gov/) |  | Dr. Tiffany L Bogich, PhD |  |
| [Fogarty International Center](http://www.fic.nih.gov/) < [National Institutes of Health (NIH)](http://www.nih.gov/) – Bethesda, MD, USA |  | Dr. Tiffany L Bogich, PhD |  |
| [Fogarty International Center](http://www.fic.nih.gov/) < [National Institutes of Health (NIH)](http://www.nih.gov/) – Bethesda, MD, USA | 3R01TW005869-06S1 “American Recovery and Reinvestment Act award (ARRA)” | EHA |  |

## Disclosures

* Billy Karesh, DVM serves as president of: Working Group on Wildlife Diseases < [World Animal Health Organization (OIE)](http://www.oie.int/) – Paris, France.

## Acknowledgements

* The authors and contributors acknowledge the editorial review work of: [Kilpatrick, A Marm](http://orcid.org/0000-0002-3612-5775) “Dr. A Marm Kilpatrick, PhD” (Department of Ecology & Evolutionary Biology < [University of California Santa Cruz (UCSC)](http://www.ucsc.edu/) – Santa Cruz, CA, USA); “Anonymous reviewer 1”; “Anonymous reviewer 2”.

## Supporting Information

Supporting Dataset 1. Full dataset of host-virus pairs and disease state. Authors: Ms. Jordan Levinson, MSc[[17]](#footnote-17); Dr. Tiffany L Bogich, PhD[[18]](#footnote-18). License: [CC0-1.0](https://spdx.org/licenses/CC0-1.0.html).

<http://wwwnc.cdc.gov/eid/article/19/5/12-1042-techapp1.xlsx>

Supporting Code 1. R code used to generate Figure 1A, B, and C; Figure 2; and for the logistic regression analysis results presented in Table 1. Authors: Dr. Tiffany L Bogich, PhD[[19]](#footnote-19). License: [CC0-1.0](https://spdx.org/licenses/CC0-1.0.html). Data from: Supporting Dataset 1.

##clear workspace

rm(list=ls())

##load relevant libraries (make sure they're installed first)

library(lme4); library(lattice); library(MASS); library(nlme); library(reshape2);library(gdata); library(Hmisc); library(fields); library(multcomp); library(plotrix);library(RColorBrewer); library(brglm)

##change directory to read data file

d<-read.csv("~/EID\_data\_new.csv",header=TRUE)

d$Symp=d$DisSymp

d$Symp[which(d$DisSymp==0)]=2

##prep for Figures 1B and 1C

#check class of each column

str(d)

#checking to see what the data table looks like

tax.mat<-xtabs(DisSymp~VirusFamily+SpeciesOrder, d,sparse=FALSE,drop.unused.levels=TRUE)

tax.mat

checksum.h<-apply(tax.mat,2,sum); checksum.v<-apply(tax.mat,1,sum)

hrem<-which(checksum.h<3); vrem<-which(checksum.v<3)

tax.mat2<-tax.mat[-vrem,-hrem]

vf<-dimnames(tax.mat2)[[1]]

so<-dimnames(tax.mat2)[[2]]

#remove records that did not report disease

dclean<-na.omit(d);dclean<-drop.levels(dclean)

### FIGURE 1B and 1C PLOTS ###

sympVFam<-tapply(dclean$DisSymp,dclean$VirusFamily,FUN=sum)

totVFam<-tapply(dclean$Data,dclean$VirusFamily,FUN=sum)

psympVFam<-sympVFam/totVFam

sympHO<-tapply(dclean$DisSymp,dclean$SpeciesOrder,FUN=sum)

totHO<-tapply(dclean$Data,dclean$SpeciesOrder,FUN=sum)

psympHO<-sympHO/totHO

##VIRUS BARPLOT

prev=psympVFam

tot<-totVFam

se=(prev\*(1-prev)/tot)^.5 ## regular standard error when prevalence isn't 0/1 where gd is calculated prevalence, and N is total number of individuals sampled, and f1 is your data frame

Pmax.all.zeros<-function(N,Pr=0.05){ 1 - exp( log(Pr) / N )} ##create a function for se when prevalence is 0/1

se[which(se==0)]=Pmax.all.zeros(tot[which(se==0)])/2

#deal w/1s and 0s

liw<-c(se[1:5],0,se[7:9],0,se[11:16])

uiw<-c(se[1:7],0,se[9:13],0,0,se[16])

labs<-names(tot[order(psympVFam)])

labs.short<-substr(labs,1,3); labs.short[2]="Picob"; labs.short[11]="Picor"; labs.short[1]="Hepe"; labs.short[5]="Hepa"

##change directory to write figures

postscript(file="~/12-1042-F1C.eps", width=7.5, height=5,pointsize=10, onefile=FALSE, horizontal=FALSE, paper="special")

par(mar=c(7.2,4,4,2))

par(mgp=c(1.75,0.5,0))

plotCI(barplot(psympVFam[order(psympVFam)]\*100,las=1,space=0.2,width=0.5, col="light grey",ylab="% Symptomatic",ylim=c(0,110),xaxt="n"),psympVFam[order(psympVFam)]\*100,uiw=uiw[order(psympVFam)]\*100,liw=liw[order(psympVFam)]\*100,ui=1,li=0,add=TRUE,pch=NA,lwd=1.2)

axis(1, at=seq(0.33,9.35,0.6), tick=FALSE, labels=labs.short, cex.axis=0.8, font=3)

text(x=seq(0.33,9.35,0.6),y=(prev[order(psympVFam)]\*100+uiw[order(psympVFam)]\*100+5),labels=tot[order(psympVFam)], cex.axis=0.8)

dev.off()

##HOST BARPLOT

prev=psympHO

tot<-totHO

se=(prev\*(1-prev)/tot)^.5 ## regular standard error when prevalence isn't 0/1 where gd is calculated prevalence, and N is total number of individuals sampled, and f1 is your data frame

Pmax.all.zeros<-function(N,Pr=0.05){ 1 - exp( log(Pr) / N )} ##create a function for se when prevalence is 0/1

se[which(se==0)]=Pmax.all.zeros(tot[which(se==0)])/2

#deal w/1s and 0s

liw<-c(se[1],0,se[3:4],0,0,se[7:11],0,se[13],0)

uiw<-c(se[1:2],0,se[4:9],0,se[11:14])

hlabs<-names(tot[order(psympHO)])

hlabs.short<-substr(hlabs,1,3)

postscript(file="~/12-1042-F1B.eps", width=7.5, height=5,pointsize=10, onefile=FALSE, horizontal=FALSE, paper="special")

par(mar=c(7.2,4,4,2))

par(mgp=c(1.75,0.5,0))

plotCI(barplot(psympHO[order(psympHO)]\*100,las=3,space=0.2,width=0.5, col="light grey",ylab="% Symptomatic",ylim=c(0,110),xaxt="n"),psympHO[order(psympHO)]\*100,uiw=uiw[order(psympHO)]\*100,liw=liw[order(psympHO)]\*100,add=TRUE,pch=NA,lwd=1.2)

axis(1, at=seq(0.33,8.5,0.6), tick=FALSE, labels=hlabs.short, cex.axis=0.8)

text(x=seq(0.33,8.5,0.6),y=(prev[order(psympHO)]\*100+uiw[order(psympHO)]\*100+5),labels=tot[order(psympHO)], cex.axis=0.8)

dev.off()

###PREP FOR FIGURE 2####

#remove records with hosts or viruses with less than three entries in the db

keep<-intersect(which(as.character(dclean$SpeciesOrder) %in% so),which(as.character(dclean$VirusFamily) %in% vf))

dclean<-dclean[keep,]

dclean<-drop.levels(dclean)

#remove rhabdoviruses - complete separation

#dclean<-dclean[-which(dclean$VirusFamily=="Rhabdoviridae"),];dclean<-drop.levels(dclean)

xtabs(DisSymp~VirusFamily+SpeciesOrder, dclean,sparse=FALSE,drop.unused.levels=TRUE)

#anova for model comparison (walds chisq test)

a1<-brglm(DisSymp~1,data=dclean,family=binomial)

a2<-brglm(DisSymp~SpeciesOrder,data=dclean,family=binomial)

a3<-brglm(DisSymp~VirusFamily,data=dclean,family=binomial)

a4<-brglm(DisSymp~VirusFamily + SpeciesOrder,data=dclean,family=binomial)

anova(a1,a2,a3,a4, test="Chisq")

anova(a4,a3,test="Chisq") #test for species order sig

anova(a4,a2,test="Chisq") #test for virus fam sig

#three way Freq table to have a look

mytable<-xtabs(~DisSymp+VirusFamily+SpeciesOrder,data=d)

ftable(mytable)

summary(mytable)

#par(mar=c(16,4,4,4))

#boxplot(DisSymp~VirusFamily\*SpeciesOrder,data=dclean,las=3)

#log reg model with bias reduction

m1<-brglm(DisSymp~relevel(VirusFamily,ref="Flaviviridae")+relevel(SpeciesOrder,ref="Artiodactyla"),data=dclean,family=binomial)

m12<-brglm(DisSymp~relevel(VirusFamily,ref="Herpesviridae")+relevel(SpeciesOrder,ref="Primates"),data=dclean,family=binomial)

#add confidence intervals to coefficients

sum.coef<-summary(m1)$coefficients

confint.brglm(m1)

#odds ratios

est<-exp(sum.coef[,1])

#95% CI

upper.ci<-exp(sum.coef[,1]+1.96\*sum.coef[,2])

lower.ci<-exp(sum.coef[,1]-1.96\*sum.coef[,2])

cbind(est,upper.ci,lower.ci)

##predicted probabilites using unique set of virus families and species orders

VirusFamily<-unique(dclean$VirusFamily);VirusFamily=drop.levels(VirusFamily) #just virus families)

SpeciesOrder<-unique(dclean$SpeciesOrder);SpeciesOrder=drop.levels(SpeciesOrder) #just species order

dclean$index<-1

nvfam<-tapply(dclean$index,dclean$VirusFamily,sum)

nspp<-tapply(dclean$index,dclean$SpeciesOrder,sum)

vfam.sord<-as.data.frame(expand.grid(VirusFamily,SpeciesOrder)); names(vfam.sord)=c("VirusFamily","SpeciesOrder")

#predicted values from model with both virus and host taxonomy

new.all<-predict.glm(m1,newdata=vfam.sord,type="response",se.fit=TRUE);

vfam.sord$predict<-new.all$fit

upper.res<-new.all$fit+1.96\*new.all$se.fit; upper.res[which(upper.res>=1)]=1

lower.res<-new.all$fit-1.96\*new.all$se.fit; lower.res[which(lower.res<=0)]=0

vfam.sord$upper.res<-upper.res

vfam.sord$lower.res<-lower.res

data.mat<-melt(vfam.sord, id=c("VirusFamily","SpeciesOrder","predict","upper.res","lower.res"))

new.dmat<-acast(data.mat,VirusFamily~SpeciesOrder,value.var="predict")

new.dmat.u<-acast(data.mat,VirusFamily~SpeciesOrder,value.var="upper.res")

new.dmat.l<-acast(data.mat,VirusFamily~SpeciesOrder,value.var="lower.res")

###FIGURE 2

so.short<-substr(so,1,5);

postscript(file="~/12-1042-F2\_LeftPanel.eps", width=5.5, height=4, onefile=FALSE, horizontal=FALSE, paper="special")

par(mgp=c(2.5,0.5,0))

par(mar=c(6,7,1,7))

image(t(new.dmat.l),col=(palette(gray(seq(0,1,len=5))))[seq(5,1,-1)],breaks=c(0,0.2,0.4,0.6,0.8,1),xaxt="n",yaxt="n",main="Lower 95% CI", cex.main=0.8)

axis(1,at=seq(-0.1,1.1,0.24)[1:5]+0.12,labels=so.short,las=1,cex.axis=0.8)

axis(2,at=seq(-0.05,1.05,0.11)[1:10]+0.055,labels=vf,las=1,cex.axis=0.8)

dev.off()

postscript(file="~/12-1042-F2\_CenterPanel.eps", width=5.5, height=4, onefile=FALSE, horizontal=FALSE, paper="special")

par(mgp=c(2.5,0.5,0))

par(mar=c(6,7,1,7))

image(t(new.dmat),col=(palette(gray(seq(0,1,len=5))))[seq(5,1,-1)],breaks=c(0,0.2,0.4,0.6,0.8,1),xaxt="n",main="Mean",yaxt="n", cex.main=0.8)

axis(1,at=seq(-0.1,1.1,0.24)[1:5]+0.12,labels=so.short,las=1,cex.axis=0.8)

#axis(2,at=seq(-0.05,1.05,0.11)[1:10]+0.055,labels=vf,las=1,cex.axis=0.8)

dev.off()

postscript(file="~/12-1042-F2\_RightPanel.eps", width=5.5, height=4, onefile=FALSE, horizontal=FALSE, paper="special")

par(mgp=c(2.5,0.5,0))

par(mar=c(6,7,1,7))

image(t(new.dmat.u),col=(palette(gray(seq(0,1,len=5))))[seq(5,1,-1)],breaks=c(0,0.2,0.4,0.6,0.8,1),xaxt="n",main="Upper 95% CI",yaxt="n", cex.main=0.8)

axis(1,at=seq(-0.1,1.1,0.24)[1:5]+0.12,labels=so.short,las=1,cex.axis=0.8)

#axis(4,at=seq(-0.05,1.05,0.11)[1:10]+0.055,labels=vf,las=1,cex.axis=0.8)

dev.off()

#use this for legend

postscript(file="~/12-1042-F2\_legend.eps", width=1, height=4, onefile=FALSE, horizontal=FALSE, paper="special")

par(mgp=c(2.5,0.5,0))

par(mar=c(6,7,1,7))

plot(NA)

image.plot(t(new.dmat),col=(palette(gray(seq(0,1,len=5))))[seq(5,1,-1)],breaks=c(0,0.2,0.4,0.6,0.8,1),legend.only=TRUE)

dev.off()

# Bibliography

Cleaveland, S., M.K. Laurenson, and L.H. Taylor. 2001. "Diseases of Humans and Their Domestic Mammals: Pathogen Characteristics, Host Range and the Risk of Emergence." *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 356 (1411): 991-999.

Cooper, Natalie, Randi Griffin, Mathias Franz, Moshood Omotayo, and Charles L Nunn. 2012. "Phylogenetic Host Specificity and Understanding parasite sharing in primates." *Ecology Letters* 15 (12): 1370-1377.

Eidson, Millicent, Nicholas Komar, Faye Sorhage, Randall Nelson, Ton Talbot, Farzad Mostashari, and Robert McLean. 2001. "Crow deaths as a sentinel surveillance system for West Nile virus in the northeastern United States, 1999." *Emerging Infectious Diseases* 7 (4): 615-620.

Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80: 27-38.

Jones, Kate E., Nikkita Patel, Marc A. Levy, Adam Storeygard, Deborah Balk, John L. Gittleman, and Peter Daszak. 2008. "Global Trends in Emerging Infectious Diseases." *Nature* 451 (7181): 990-993.

Leendertz, Fabian H., Georg Pauli, Kerstin Maetz-Rensing, Wayne Boardman, Charles Nunn, Heinz Ellerbrok, Siv Aina Jensen, Sandra Junglen, and Boesch Christophe. 2006. "Pathogens as Drivers of Population Declines: The Importance of Systematic Monitoring in Great Apes and Other Threatened Mammals." *Biological Conservation* 131 (2): 325-337.

Leroy, Eric M., Pierre Rouquet, Pierre Formenty, Sandrine Souquière, Annelisa Kilbourne, Jean-Marc Froment, Magdalena Bermejo, et al. 2004. "Multiple Ebola Virus Transmission Events and Rapid Decline of Central African Wildlife." *Science* 303 (5656): 387-390.

Mohd, Nor M.N., C.H. Gan, and B.L. Ong. 2000. "Nipah virus infection of pigs in peninsular Malaysia." *Revue Scientifique Et Technique De L Office International Des Epizooties* 19 (1): 160-165.

Morse, Stephen S., ed. 1993. *Emerging Viruses.* New York: Oxford University Press.

Selvey, Linda A., Rachel M. Wells, Joseph G. McCormack, Anthony J. Ansford, Keith Murray, Russell J. Rogers, Peter S. Lavercombe, Paul Selleck, and John W. Sheridan. 1995. "Infection of humans and horses by a newly described morbillivirus." *The Medical Journal of Australia* 162 (12): 642-645.

Taylor, Louise H., Sophia M. Latham, and Mark E.J. Woolhouse. 2001. "Risk factors for human disease emergence." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 356 (1411): 983-989.

The Center for Food Security & Public Health; Institute for International Cooperation in Animal Biologics; World Organisation for Animal Health. 2009. "Rabies Factsheet." Accessed July 2, 2012. http://www.cfsph.iastate.edu/DiseaseInfo/disease.php?name=rabies.

USAID. 2009. "USAID launches Emerging Pandemic Threats program." Accessed July 2, 2012. http://www1.usaid.gov/press/releases/2009/pr091021\_1.html.

Williams, Elizabeth S., and Ian K. Barker, . 2008. *Infectious Diseases of Wild Mammals.* 3rd Edition. Ames, Iowa: Iowa State Press.

Wilson, Don E., and DeeAnn M. Reeder, . 2005. *Mammal species of the world: a taxonomic and geographic reference.* 3rd Edition. Vol. 12. Baltimore, MD: Johns Hopkins University Press.

Wolfe, Nathan D., Ananias A. Escalante, William B. Karesh, Annelisa Kilbourn, Andrew Spielman, and Altaf A. Lal. 1998. "Wild Primate Populations in Emerging Infectious Disease Research: The Missing Link?" *Emerging infectious diseases* 4 (2): 149-158.

Wolfe, Nathan D., Peter Daszak, A. Marm Kilpatrick, and Donald S. Burke. 2005. "Bushmeat hunting, deforestation and prediction of zoonotic emergence." *Emerging Infectious Diseases* 11 (12): 1822-1827.

Woolhouse, Mark E.J., and Sonya Gowtage-Sequeria. 2005. "Host range and emerging and reemerging pathogens." *Emerging Infectious Diseases* 11 (12): 1842-1847.

1. Authors contributed equally. [↑](#footnote-ref-1)
2. Email: [daszak@ecohealthalliance.org](mailto:daszak@ecohealthalliance.org). Twitter: [@PeterDaszak](https://twitter.com/peterdaszak). Tel: [+1-212-380-4460](tel:+1-212-380-4460). Fax: [+1-212-380-4465](fax:+1-212-380-4465). Address: EcoHealth Alliance, 460 West 34th Street, 17th Floor, New York, NY 10001. [↑](#footnote-ref-2)
3. Senior author. [↑](#footnote-ref-3)
4. We excluded rabies from our analysis because the intense research effort on this virus and its high pathogenicity in almost all of its wide range of hosts (The Center for Food Security & Public Health; Institute for International Cooperation in Animal Biologics; World Organisation for Animal Health 2009) would skew the data disproportionately. [↑](#footnote-ref-4)
5. Email: [tbogich@princeton.edu](mailto:tbogich@princeton.edu). Twitter: [tiffbogich](https://twitter.com/tiffbogich). [↑](#footnote-ref-5)
6. Wrote the code. [↑](#footnote-ref-6)
7. Wrote the equation. [↑](#footnote-ref-7)
8. Prepared the figure. [↑](#footnote-ref-8)
9. Created the study. [↑](#footnote-ref-9)
10. Conducted the analysis. [↑](#footnote-ref-10)
11. Senior author. [↑](#footnote-ref-11)
12. Virus and host reference groups were selected as those for which sample size was sufficiently large and symptomatic infection was moderate (see Figure 1). [↑](#footnote-ref-12)
13. All host–virus pairs were symptomatic. [↑](#footnote-ref-13)
14. Conducted the analysis. [↑](#footnote-ref-14)
15. Collected the data. [↑](#footnote-ref-15)
16. Prepared the figures. [↑](#footnote-ref-16)
17. Collected the data. [↑](#footnote-ref-17)
18. Designed the study. [↑](#footnote-ref-18)
19. Wrote the code. [↑](#footnote-ref-19)