**P-2 Modeling and Analytics (M&A) goals for P-2.**

**Updated 6/18/15**

**P2-wide M&A Team:**

* Peter Daszak (EHA) Lead
* Kevin Olival (EHA) M&A Coordinator
* Christine Kreuder-Johnson (CKJ) UC Davis PoC
* Damien Joly – Metabiota PoC

**EHA M&A Team:**

* Peter Daszak (EHA) Lead
* Kevin Olival (EHA) M&A Coordinator
* Carlos Zambrana-Torellio
* Andrew Huff
* Toph Allen
* Lizzie Loh
* Emily Hagan
* Allison White
* Erica Johnson
* Oct 1st onwards: Noam Ross

**General Management M&A P-2**

* Weekly EHA M&A meetings
* Liaison to rest of P2 will be at the PREDICT Exec Board meetings.
* Should aim for Monthly PREDICT-wide M&A meetings
* Kevin to sit in on Surveillance calls to represent M&A team.
* Aim to deliver high impact 2-pagers (short, important issues) to USAID every 3-4 weeks. Especially important on topics that Dennis, Andrew, Alisa are thinking through for P2 directions.

**Primary Questions:**

1. What are the BIG goals/projects we want to accomplish by the end of P-2 (like Deep Forest and Hotspots II for P-1)?
2. What data and types of analyses do we need to reach those goals?
3. How should these data be organized to be most useful for analyses (linking with IM team and EIDITH)?
4. How will we best integrate analyses of human behavioral data with wildlife and livestock data?
5. Best ways for M&A to inform surveillance? E.g. What data to collect, number of sites and samples necessary, and selection of pathways in order to test global hypothesis on zoonotic disease emergence.

**M&A projects**

* **Future projections of hotspot maps** – using projected changes in demography, climate, livestock, and trade. CZT; EEJ, AMW; TA
* Better response variable for hotspot models AH/TA
  + all outbreaks of an emerging disease if they’re separate events (not just first emergence event, and not just spreading events)
  + This will involve the use of time-series driver data too, e.g. temporal landuse change datasets when looking at different outbreaks over time.
  + Linking in with DoD Biosurveillance Ecosystem (BSVE) and other global datasets and data mining/data scraping activities
  + CKJ – have grant call to expand the BSVE, she’s planning to submit tomorrow for some of her work – would be good to link in with our work whether funded or not.
  + Mantle will be a way to interface with other disease surveillance partners
* **Analyzing PREDICT data to inform Surveillance KJO; TA**
  + Use P-1 data and targeted P-2 hypotheses to help inform site selection and pathway selection (TA and KJO did this and informed surveillance)
  + Heat map of where viruses are found in PREDICT
  + MK: Do analysis of risk of emergence for Ebola with data from 3 years ago, could we have predicted it?
  + # viruses for each host then mapping host distribution
  + P1 DF data to refine DF continued sampling (CZT will do a 2 pager)
  + Within-country hotspot maps (EEJ/CZT); need within country data sets (check with Woutrina – KJO); need to vary this according to main drivers?
* **DEEP FOREST** CZT/LL (also KM/KJO)
  + Behavioral questionnaire data analysis (LL) waiting for Tom’s translation, and UCD data in August
  + analysis of big datasets on host and viral communities across the landuse gradients
  + Kris Murray will stay involved for DF
  + Meeting to catch up v. soon then a bigger one with UCD
  + Current plan to not continue DF in Uganda to full extent, but keep up some work in Bwindi but expand to other sites around country.
* **What-if scenarios to model disease interventions and human behaviors** (NR/CZT/AH)
  + Scenario models with regional data to test disease intervention strategies
  + Replacement scenarios; e.g. changing of protein sources and patterns of livestock consumption
  + Predicting likelihood of spread of new high priority PREDICT viruses
  + Spatial models of human behavior, can we map globally using proxy datasets?
  + Justin Barshers - from Berkely – looks at individual behaviors and bushmeat consumption and individual economics.
  + Glowurm app (AH, CZT)
* **Analyzing behavioral risk data from P2BR work** (MM/EGH)
  + Hospital cohorts
  + Large pathway groundtruthing cohorts
* **Evolutionary models** -- starting with existing P-1 sequence data (KJO/post-doc TBD)
  + Defining ‘epizones’ (PREDICT M&A-wide paper that includes some specific epizones (see below)
  + Recombination hotspots
  + Phylogeography of disease and hosts
  + Modeling of viral evolution with longitudinal datasets
  + Testing nearest neighbor and other phylogenetic hypotheses on determinants of spillover, including molecular mechanisms (e.g. receptors/binding domains) for certain pathogen groups – leverage EHA’s NIH CoV study in China.
  + MERS analysis: add % GDP per capita, degree of conflict etc., and glowurm analysis – will tell you where MERS will spillover to camels, then where it will cause local chains of tranmisssion (CZT/EEJ)
    - Include political instability and refugee data? If these exist to help predict subsequent spread.
    - Temporal analysis
  + Repeat for SL-CoVs and SL-Covs with ACE2.
  + Ebola epizone (EEJ,CZT)
* **Viral diversity and viral discovery curves across all countries (KJO/CZT w Simon and Tracey)**
  + CKJ looked at country and sample type influence on detection. Using P1 data to see if we’re anywhere near viral discovery saturation.
  + Can do as a group paper.
  + See Simon’s paper on Macaques
* **Mapping our three EID ‘pathways’ globally (all):**
  + We’re analyzing how these pathways affect EID risk using Behavioral Risk and surveillance data. We could extrapolate this globally if we knew where the pathways are, so let’s map them…
  + Map of wildlife consumption and local trade globally: Start in China: XX has the poultry markets for China, need to build maps of some typical chains in few diff Provinces to understand epizone risk. (e.g. Marianne’s paper) Also see PREVENT data, and see PRH’s model results which he NR will follow up on.
  + Analysis of PREVENT data from wildlife market surveys under P1 – in Indonesia and DRC and RC (MM to check on data at DC meeting)
  + Agricultural intensification pathway
  + Land use change pathway (e.g. how recent the change was; cf. EEJ’s paper on malaria) (see Future Earth fragmentation project also)
* **Livestock and Antibiotic-resistant EID hotspots** (AMW)
  + Do antimicrobial resistant (AMR) EIDs correlate with antibiotic use?
  + Is this driven by veterinary or human use of antibiotics? Look at spatial patterns and evolutionary drivers with veterinary and human use and livestock movement, using global data on antibiotics that we now have.
  + PD will email DC group, assess if worth buying dataset.
* **Conflict and political instability as drivers of EIDs** (AH)
  + Draft plan being circulated
* **Economic modeling**
  + Ebola (KCB)
  + Quarantine (KCB)
  + Couple of papers to finish off
  + PD needs to add to this list – incentives for mitigation strategies., eg. Human movement
  + AMR (TBD)
* **Follow up projects from FE Horizon Scan**
* **Tech team**
  + Strategy to manage datasets internally (Mantle)
* **Extra Industries Working Group (EIWG) CZT/AMW/PD +others**
* **Africa Livestock Futures** (CZT)
  + Plan drafted

**Maureen – Behavioral risk slides (MM and EH with M&A team)**

Identifying behaviors and conditional factors for spillover first, maybe look at factors for spread later

**3 Outcome variables for spillover to link with behavioral data and context**

Cross-sectional study (not cohort – except maybe China CoV study)

* 1. Viral testing data: PCR viral data
  2. Serology on people who participated in behavioral survey
  3. Self report symptom surveillance. “verbal autopsy”

**Spread – what questions do we ask to get at spread**

Marm: look at Bangladesh surveys for airport behavior and travel

Discuss data to collect for M&A at outset of qualitative and quantitative BRG plan. Especially those that can inform policy. Look at policies that have been already proposed and what behaviors go around that – eg. How much do you sell animal for, how much does it cost to live in village;

For spread: sense of movement and exchange of people between locations. “Human territoriality” – qualitative interviews: how far do you travel, how does yoru day work; seasonality of movement (festivals, etc.).

~50-60 people in China across 2 pathways.

\*\*KJO to send PREVENT fliers to Maureen

“Prevalence of specific behaviors” – e.g. prevalence of people getting scratched when hunting bats may vary by country and culture based on behaviors.

Problem with existing surveys we used, too comprehensive – interviewer fatigue

In all of P-1 questionnaires, no one asked about illness, access to medical care, or death (included with verbal autopsy –symptoms post-death).

PH: How do you access medical care, and how you move after you’re sick? How do you make those decisions, economic considerations of access to health care … how this affects movement.

Most of this is captured in recent surveys from Bangladesh. Collected data on what symptoms passengers have, and can map back with the pathogens we are interested in.

MK: where you do the survey matters – survey was done at Dhaka airport (travel node), vs. doing survey in village. E.g. like studying predator-prey relationships if you look at prey, not much will happen, but if you study predators there will be kills all the time.

MK: do case study approach, take top 20 EIDs and identify what the risk factors were and work backwards.

**\*PD: need to pick what specific viruses** (or types of viruses, e.g. SARS-like) we are going to test and model for spillover and spread. MERS, Nipah-like, AI, … others?

MK: IATA challenge is linkages between different port – Bangladesh surveys allowed to get to some of those connections.

KJO: can’t we crawl web (e.g. Kayak) to get at connections for flights, only set number

W. J. Edmonds – started simple survey to look at contact between people, that group has done surveys in 6 countries (PLOS Medicine). \*\*Look up and distribute paper.

**More Details – specific ideas from last EHA brainstorming**

**Future demographic, livestock, and climate changes and disease**

Projecting Hotspots forward 5, 10, and 20 years into the future using modeled changes in pop growth, climate, livestock production, and travel and trade expansion.

* We have lots of data on this, GUMP, CIESIN, FAO, etc. already, but should start building other dataset, e.g some specific trade changes. And focus more on scenerios.
* Can we link scenerios with real world projections? Bring in some hard data where available. Look at recent trends in last 10 years to make realistic projections.
* Big thing missing – how will viral distribution change over time. Complicated process that deals with what various wildlife and domestic animal species will do with climate change. Link in with livestock trade.
* Even projecting 5 years in the future would be useful, and can test by end of project.
* Future projections should focus on 3 pathways, but especially on land conversion.

Peter on advisory board for US Global Change Research Program – opportunity to collaborate with others doing other climate change.

Carlos – group in Italy is projecting 5000+ mammals into future range. Potential collaboration with Carlo Rondenini, affiliated with FAO.

Need to combine climate data with livestock futures. Scenario based projections – because models aren’t necessarily giving us the picture 5, 10, 20 years into the future. Can predict climate, but harder to predict the human response and behaviors with climate change – that’s where scenario models come in.

Can we do Asian livestock futures? Split into SE Asia, S Asia, Australasia… etc.

* How will livestock production change in Asia and what will this do for zoonotic risk?
* USAID state in Email that they do not want to do ‘Asian Livestock Futures’. Therefore any work that involves this approach should be framed as:   
  “analysis of how future trends in a wide range of environmental, socioeconomic and agricultural factors will affect EID risk in our countries, including what-if scenarios based approach to generate projections of how specific policies/changes will alter these outcomes.”
* Check with USDA ERS (Economic Research Service) to see if they’ve done that.

What-if modeling of replace protein source, e.g. if you stop eating bats and replace with a livestock protein source, what does this do to risk? Andrew H. has experience with this.

Travel and Trade – What-if: e.g. Future Africa and China trade based on different projections

UN Comtrade data on Livestock trade. Need to be careful because of repackaging and ‘value’ added relabeling of country of origin.

Hotspots with livestock production

Trade movement. Poultry movement between China and Africa over time. Relate this with biosecurity practices by country for risk assessment.

Publicly available troop movement data by country as proxy for military movement. Possible to FOIA additional data.

Look at plant/grain travel also, e.g. did PEDV come in from China in contaminated grain shipments?

**Antibiotic resistance**

* Analysis not big focus of P-2, but maybe could be used to cross check that antibiotic resistance will not be the next big one.
* Use bacteria and E. coli (e.g. like Goldberg) to use as proxy for detecting spillover, loop in diagnostic team.

Recent paper with data on global antibiotic use.

Can use as proxy, country-level predictor variable – what are the laws on antibiotic use in different countries for humans and veterinary.

Global travel and trade and spread of antibiotic resistance, need to include livestock trade. UN COMTRADE data.

**Sicki GRID** – **better response variable for hotspot models**

Best response will be all outbreaks (not just first emergence event as in Jones et al. 2008), maybe do for subset of top (top 100?) pathogens, or those like Chikungunya, Ebola, Nipah, etc. where outbreaks can be tracked decently.

Leverage EHA’s DTRA work and BSVE to help populate response dataset.

Discuss future API for EIDITH to better link in analyses.

* Damien hoping to have in person IM meeting in Jan, may be over phone – develop better way to live-link the EIDITH database to the M&A team, and to include some GIS links with the database.

**Modeling to inform Surveillance**

* Make sure that planned surveillance will end up as a publishable (therefore scientifically defensible) study in each country, e.g. number of samples and sites, etc.
* CKJ to put together list of plans in each country.

Need to be more clear on what are our units of response, number of replicates, etc. for each gradient.

Should only have one pathway per country. Or one pathway at a given site. Maybe find a few double pathway sites in later years to tease out, focus on single pathways initially. Will allow us to tease out processes of disease emergence better.

What we ideally need is an experiment – e.g. one of Malaysia DF sites was chopped down.

RISK assessment

Does exposure to wildlife increase **risk**? Use human serology to test as a response variable (e.g. if spillover has happened). Predictors in addition to land use change and other variables = how frequent and how diverse the viruses are.

**What-if scenarios** for risk mitigation testing. Maureen role in testing risk mitigation at local scale. What are longterm impacts of different mitigations? E.g. draining swamps in south to reduce malaria risk – has negative ecological impacts.

What’s the consequence we’re trying to calculate risk for. Spillover only, DALYs, pandemic size, economic impact? For reemerging viruses may be easier to calculate; For emerging viruses can do different scenario testing – e.g. for a new Ebola what is the likelihood it will get into an urban area like W. Africa outbreak.

* What if scenario around AI interventions – e.g. how many ‘clean days’ you have in live animal markets etc.
* PRH will talk with FAO about their scenarios around clean days etc.

Andrew H and PRH, to collaborate on modeling interventions and policies we can include in what-if scenerios. What effect will different ‘policy cocktails’ have.

**Phylogenetic analysis in P-2**

Predicting viral pathogenicity is exceptionally hard and there are many preconceived notions about it

Test molecular evolution using time-sampled data (longitiudinal data) where available. Can automate and mine GenBank to further populate dataset and add to P-1 molecular data. Can get cell receptor data from Genbank etc.

Test nearest neighbor hypothesis

Can we look at seq we have now (P-1) and where they were collected in pathways – are ones from the same pathway more similar than those from different areas. Does diversity of virus genetics change across the landscape. Target surveillance where viral diversity is predicted to be the highest. From wild to markets, e.g. does diversity decrease across pathway bottleneck – and how does this related to host genetic diversity.

Predict recombination potential across the landscape. P-1 data by site and see how much viral genetic diversity there is.

* Important for flu, but other viruses recombine too.

Phylogenetic relatedness and receptor/host relatedness, 3 hypotheses to test: H0 – do they match; H1 – general decline with phylogenetic distance; H2 – islands of susceptibility scattered across phylogeny.

Test pathogeneicity, is it better than random between viruses of a given clade. E.g. CoVs and pathogenicity – can we predict this? Test a mixed model

**Defining “Epizones”**

* Do quick little paper, white paper for USAID, or maybe short paper – to define epizones more broadly. Need soon, to help inform diagnostics – e.g. do we need more than a short viral sequence for diagnostics and characterizing?

Tentative Definition: “A region where a viral group that has or is likely to emerge (including an understanding of viral strain risk) is present in its reservoir, ***and*** where the human risk behaviors and drivers exist for the viral group to spillover and emerge.”

* Operational definition of ‘epizone’ from M&A
  + Specific geographic area where X number of people and X number of animals overlap –
  + e.g SARS – market from where first cases, plus pool where bats came from and wildlife range of those. Need surveillance data on how wide these ranges are – from surveillance team.
  + Are epizones across diseases similar? Do they correlate, or is each disease so unique that there is no strong overlap of epizones. Start with core e.g. areas where Nipah is emerging, and move out from there to see what extent of epizone.
  + Hazard vs Risk. May have wildlife reservoir but no human behavioral risk = hazard, w no risk.
  + Viral evolution plays in to identify likelihood of emergence – some clades may be more likely to emerge/jump than others. E.g. SARS-like viruses are found across Europe. Need to integrate with viral phylogenetic analyses.