Prepared by:
Gryphon Scientific, LLC
Corey Meyer, Benefit Assessment Task Lead
December 22, 2015

Risk & Benefit Analysis of Gain of Function Research

Supplemental Material – State of Surveillance for Influenza Viruses and Coronaviruses

TABLE OF CONTENTS

1 The State of Influenza and Coronavirus Surveillance

Global surveillance for influenza viruses and coronaviruses in humans and animals plays a key role in several aspects of the benefit assessment. Surveillance data are utilized for several alternative experimental approaches, and circulating influenza viruses detected through surveillance are subjected to pandemic risk assessments, which are informed by GoF research. To provide a foundation for the assessment of those topics in the main report, this supplemental report reviews the current state of influenza and coronavirus surveillance.

2 The State of Coronavirus Surveillance

2.1 Introduction

Comparative analysis of coronavirus sequences from different species or of varying levels of virulence can provide insight into the mechanistic basis of cross-species adaptation and virulence, respectively. As such, comparative sequence analysis represents an alternative approach to the use of GoF approaches that lead to enhanced transmission and virulence in mammals, which also provide insight into coronavirus virulence and mammalian adaptation mechanisms. The success of comparative sequence analysis methods depends on the quality and availability of genetic surveillance data. This report reviews the state of genetic surveillance for SARS, MERS, and SARS/MERS-like bat-CoVs to identify gaps in existing surveillance systems and shortcomings in the quality of available genetic data. The impact of these limitations on the feasibility of conducting comparative sequence analysis approaches as well as the quality of the information that can be generated using these approaches is also evaluated.

2.2 SARS Surveillance Data

SARS-CoV is no longer circulating in human or animal populations, but genetic surveillance data from the outbreaks are available through NCBI's SARS Coronavirus Genome Tool. 1,137 SARS-CoV sequences are available, including 172 complete sequences and 965 partial sequences.[1] Unprecedented international cooperation and rapid sequencing led to the publication of multiple SARS-CoV genomes throughout the 2003 pandemic.[2] Human sequences were isolated largely from hospitalized cases, though limited active surveillance of patient contacts and people within the community was also conducted. In addition, during the outbreak, genetic surveillance in markets associated with human infections produced SARS-CoVs sequences from palm civets and raccoon dogs. Following the outbreak, active surveillance in farm and wild animal populations discovered several other species capable of harboring the virus and identified bats as reservoirs of SARS-like coronaviruses.[3] The WHO advises continued vigilance for the possible reemergence of SARS-CoV, and continued genetic surveillance of bats is vital to determine if and how SARS-CoV or a related coronavirus could potentially cause another pandemic.[4]
These genetic data have been used to study the mechanisms of cross-species adaptation and viral virulence. For example, animal SARS-CoV sequences were compared to human sequences from the early, middle, and late phases of the outbreak to better understand how the virus mutated and spread over the course of the epidemic from introduction to elimination.[5]

2.3 MERS Surveillance Data

2.3.1 Animal Surveillance

A high priority for MERS-CoV research is understanding viral diversity and reservoir dynamics in nature, in particular within bats, the virus reservoir, and camels, which may also function as a reservoir and are

1 NCBI. SARS Coronavirus Genome Alignment Tool. http://www.ncbi.nlm.nih.gov/genomes/SARS/SARS.html. Last Update Accessed December 18, 2015.
2 Donnelly CA *et al* (2004) Epidemiological and genetic analysis of severe acute respiratory syndrome. *The Lancet Infectious diseases* 4**:** 672-683.
3 Ibid
4 WHO (2004) WHO guidelines for the global surveillance of severe acute respiratory syndrome (SARS).
5 Shi Z, Hu Z (2008) A review of studies on animal reservoirs of the SARS coronavirus. *Virus research* 133**:** 74-87.

thought to play a key role in continued zoonotic spillover of the virus to humans.[6] Understanding viral persistence and evolution in natural hosts will enable the identification of major routes of animal to human transmission, which provides a basis for the design of intervention strategies to interrupt zoonotic transmission.[7] Additionally, surveillance data can provide important insight into the mechanisms of cross-species adaptation, i.e. through comparison of bat, camel, and human sequences. Ideally, genetic surveillance for coronaviruses should be carried out in domestic and wild animals that interact with humans to identify possible routes of infection, how the virus mutates during transmission, and how transmission routes can be blocked.[8] Genetic surveillance currently falls short of this goal, but efforts are being made to improve MERS-CoV surveillance networks.

Animal surveillance is currently conducted in some camel slaughterhouses and markets in the Middle East, and some countries have programs testing imported camels.[9] For example, in United Arab Emirates, samples are collected from imported camels at screening centers on the UAE border.[10] Efforts have also been undertaken to determine the prevalence of MERS-CoV in camels in Saudi Arabia through serological and molecular methods.[11] However, little genetic surveillance data for camels are available outside the Middle East (Table S1). Only five countries within the Middle East, United Arab Emirates, Saudi Arabia, Qatar, Oman, and Egypt, have reported MERS-CoV sequences from camels, with the majority of sequences originating from United Arab Emirates.[12] Nigeria is the only country outside of the Middle East within which MERS-CoV surveillance has been conducted on camels and MERS-CoV sequences have been produced.[13]

6 Cai Y *et al* (2014) CD26/DPP4 cell-surface expression in bat cells correlates with bat cell susceptibility to Middle East respiratory syndrome coronavirus (MERS-CoV) infection and evolution of persistent infection. *PloS one* 9**:** e112060.

7 Nishiura, H., et al. (2014). "Missing information in animal surveillance of MERS-CoV." Lancet Infect Dis **14**(2): 100.

8 (2015b) Interviews with coronavirus researchers.

9 Ibid

10 Zulaikha MAH *et al* (2015) Asymptomatic MERS-CoV Infection in Humans Possibly Linked to Infected Camels Imported from Oman to United Arab Emirates, May 2015. *Emerging Infectious Disease journal* 21.

11 Alagaili AN *et al* (2014) Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *mBio* 5**:** e00884-00814.

12 Benson DA *et al* (2005) GenBank. *Nucleic Acids Research* 33**:** D34-D38.

13 Chu DK *et al* (2015) Middle East respiratory syndrome coronavirus (MERS-CoV) in dromedary camels in Nigeria, 2015. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 20.

Table S1. Full and partial MERS-CoV sequences available on GenBank*		
Host Species	**Country**	**Number of Sequences**
Camelus sp.		**130**
	United Arab Emirates	73
	Saudi Arabia	33
	Qatar	11
	Oman	5
	Egypt	2
	Nigeria	6
Homo sapiens		**284**
	Saudi Arabia	203
	South Korea	26
	United Arab Emirates	11
	France	6
	Netherlands	6
	United Kingdom	5
	China	4
	Iran	3
	Qatar	3
	USA	3
	Greece	2
	Jordan	2
	Oman	2
	Philippines	2
	Tunisia	2
	Malaysia	1
	Thailand	1
	Turkey	1
	Not provided	1
Not provided		**6**
	Saudi Arabia	6
Grand Total		**420**
**All sequences from NCBI MERS Coronavirus Resource/ NCBI GenBank.*		

Several shortcomings in the body of available MERS-CoV genetic surveillance data from camels constrain the utility of the comparative sequence analysis approach for studying cross-species adaptation. First, the fact that MERS-CoV viruses circulating in animals are not routinely sampled in the Middle East creates longitudinal gaps that hinder the accurate interpretation of relationships among different virus strains. Continuous, longer-term surveillance is necessary to determine the dynamics of virus circulation in dromedary camel populations. Second, the lack of genetic surveillance data outside the Middle East precludes the comparison of viruses from countries where there has been spillover to humans to viruses from other parts of the world where spillover has not occurred, which could provide insight into the viral features that promote human infection.[14] A large number of geographically diverse camel sequences are needed to support such studies.[15] Third, few paired camel and human sequences are available, i.e. sequences from an infected human and the camel that was likely the source of the infection. Such paired isolates are valuable because they may permit the identification of novel genetic traits associated with cross-species adaptation, which is extremely difficult when comparing genetically diverse human and camel sequences due to the high genetic diversity among CoVs. Finally, the WHO International Health Regulations (IHR) Emergency Committee stated that in general, virological surveillance data has not been shared in a timely manner, hampering research efforts.[16] Overall, the Committee was dissatisfied with the lack of information from the animal sector regarding MERS-CoV.

Analysis of surveillance samples is complicated by USDA regulations that make importing camel samples to the United States difficult. All camel samples are required be routed through the Plum Island Animal Disease Center or other laboratory capable of handling select agents to ensure that the samples are free of select agents.[17] This regulatory requirement has limited the ability of researchers in the United States to obtain natural isolates and sequences from camels, leaving sequencing to be done largely in Saudi Arabia and Europe.[18]

2.3.2 Human Surveillance

Many passive surveillance systems are in place around the world to identify possible cases of MERS-CoV in people with respiratory illness that have recently traveled to the Middle East. Some active surveillance of Haj pilgrims has also been carried out, but there is little continuous active surveillance on other populations to determine the prevalence of MERS-CoV infections in humans or to identify nascent outbreaks.[19] As a result, most human samples are from hospitalized cases rather than active surveillance initiatives. This means that sequences from virus strains that cause mild or asymptomatic infections are not represented.

Overall, genetic information about human MERS-CoV has improved over the last few years but remains limited. There are a total of 284 human MERS-CoV sequences currently available in GenBank through NCBI's MERS Coronavirus Resource, including 76 full-length sequences and 208 partial sequences.[20] In 2013 only 65 human MERS-CoV sequences were available, demonstrating that the availability of genetic information has increased dramatically.[21] Approximately 70% of sequences originate from Saudi Arabia, though there are numerous sequences from South Korea, United Arab Emirates, and several European

14 (2015b) Interviews with coronavirus researchers.
15 Ibid
16 WHO (2015) WHO statement on the tenth meeting of the IHR Emergency Committee regarding MERS.
17 (2015b) Interviews with coronavirus researchers.
18 Ibid
19 Al-Abaidani, I. S., et al. (2014). "Overview of preparedness and response for Middle East respiratory syndrome coronavirus (MERS-CoV) in Oman." Int J Infect Dis **29**: 309-310.
20 Benson DA *et al* (2005) GenBank. *Nucleic Acids Research* 33**:** D34-D38.
21 Cotten M *et al* (2014) Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *mBio* 5.

countries that have experienced travel-associated cases. That said, the WHO currently reports 1,611 laboratory confirmed cases of MERS, meaning that the viral genome sequences available represent only 17% of all human cases.[22] A large amount of genetic information, including possible human adaptations, may be missed by the lack of routine sequencing. In addition, surveillance data at the human-animal interface and in areas where transmission potential is high due to large crowds and increased human-animal contact is lacking.[23]

2.3.3 Initiatives to Improve MERS-CoV Surveillance

The characteristics of existing MERS-CoV surveillance data are summarized in Table S1. International public health and animal health organizations recognize the shortcomings in existing MERS-CoV surveillance networks and have taken steps to improve the quantity and quality of MERS-CoV genetic surveillance data.

The World Organization for Animal health (OIE) has prioritized the gathering of more information on MERS-CoV in camels and is working to establish a laboratory network in camel-rearing countries to better facilitate genetic surveillance.[24]

In addition, two technical meetings on MERS-CoV sponsored by the Food and Agriculture Organization of the United Nations (FAO) were held in Doha, Qatar and Cairo, Egypt. Participants discussed the enhancement of collaboration between human and animal health sectors in field investigations, surveillance, and research. From the meeting in Doha came the Doha Declaration calling for more joint animal-human investigation of cases and surveillance at the animal-human interface.[25] The Doha Declaration recommends establishing targeted, laboratory-based surveillance in camels. Surveillance should focus on farms, slaughterhouses, and other locations of camel gatherings and where large numbers of camels come into contact with humans. To the extent possible, countries should agree on a common MERS-CoV surveillance system for testing imported camels. Additionally, the Doha Declaration recommends that researchers attempt to isolate live virus as often as possible and that all samples should be submitted for sequencing. Researchers should continue characterization of the identified viruses, including comparing genome sequences from human cases to genome sequences from camels. Functional studies of newly obtained viruses should also be performed. Further works should include testing of bats for MERS-CoV and comparison of bat sequences for a better understanding of possible infection dynamics.[26]

Finally, the World Health Organization (WHO) is working with affected countries and international partners to coordinate the global health response to the ongoing MERS-CoV outbreak in the Middle East, including providing technical guidance on surveillance and laboratory testing. The WHO recommends sequencing MERS-CoV nucleic acid from as many positive specimens as possible. Full genome sequencing and sequencing of the spike protein gene directly from clinical samples is particularly encouraged.[27]

---

22 World Health Organization (2015).

23 (2015b) Interviews with coronavirus researchers.

24 WHO (2015) Laboratory testing for Middle East respiratory syndrome coronavirus (MERS-CoV): Interim Guidance.

25 World Health Organization. (2015) "Middle East respiratory syndrome coronavirus (MERS-CoV): Summary of Current Situation, Literature Update and Risk Assessment." WHO/MERS/RA/15.1

26 (2015). Doha Declaration. Regional Workshop on MERS-CoV and One Health. Doha, Qatar, Food and Agriculture Organization of the United Nations.

27 WHO (2015) Laboratory testing for Middle East respiratory syndrome coronavirus (MERS-CoV): Interim Guidance.

2.4 Bat CoV Surveillance

Before the emergence of SARS-CoV in 2002, the research and public health communities had little interest in coronaviruses; as a result, bats had not been tested for the presence of CoVs. Since then, over 1,000 discrete bat CoVs have been discovered, which represents an exponential increase in surveillance coverage. The reservoir of bat coronaviruses is being sequenced at an incredibly fast rate, and large databases are emerging. However, only ten of the greater than 17,000 species of bats have been examined for coronaviruses; in this respect, surveillance coverage is still poor.[28] Bat surveillance has revealed that in the extensive family of bat coronaviruses, some resemble SARS-CoV and MERS-CoV and thus have potential to spill over into human populations.[29] However, due to technical limitations, namely an inability to isolate and grow most bat CoVs in the laboratory, few bat CoVs have been closely examined.[30] For that reason, comparative analysis of coronavirus sequences from bats, humans, and camels could provide valuable information about cross-species adaptation as well as how virulence can increase after a species jump.[31] In addition, the publication of full-genome sequences will allow researchers to reconstruct genetic backbones so that viruses may be studied in the laboratory.[32] Additional bat CoV surveillance data are needed to support these research goals.

2.5 Conclusion

Comparative sequence analyses to generate hypotheses about genetic determinants of cross-species adaptation and virulence require a large number of human and animal sequences. Shortcomings in currently available genetic data sets significantly constrain the utility of comparative sequence analysis studies. The number of SARS-CoV sequences from animal isolates is limited and cannot be augmented due to the fact that SARS-CoV is no longer circulating in nature. Additionally, the number of available sequences for MERS-CoV is insufficient to determine how the virus changes from camels to humans; the number of paired human and camel sequences collected at the animal-human interface in particular is lacking. The number of bat CoV sequences has increased exponentially since the emergence of SARS, but available sequences are unlikely to represent the full spectrum of CoV diversity in bats due to the fact that a small fraction of bat species have been examined for CoVs. Together, these gaps restrict the utility of sequence-based approaches for studying cross-species adaptation. Efforts to use genetic surveillance data for SARS-CoV and MERS-CoV to study virulence are hampered by the fact that most human sequences represent severe, hospitalized cases, as little active community surveillance is currently being conducted in the Middle East nor was conducted during the SARS-CoV outbreak. Furthermore, the limited availability of whole virus genome sequences restricts studies that can be performed using existing genetic data. Of the 420 human and camel MERS-CoV sequences on GenBank, only 94 are complete, limiting the breadth of information that can be gleaned from comparative sequence analyses.[33] Finally, the fact that few isolates of MERS-CoV are available for research use also hinders research progress in these areas.

Long-term systematic surveillance in humans and animals is needed to build libraries of genetic sequences and viral isolates that can be used to investigate viral evolution, pathogenesis, and transmission. Surveillance coverage is currently poor, but is improving as more institutions recognize the necessity of MERS-CoV genetic surveillance.

---

28 (2015b) Interviews with coronavirus researchers.
29 Ibid
30 Ge XY *et al* (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503**:** 535-538.
31 Ibid
32 Ibid
33 Benson DA *et al* (2005) GenBank. *Nucleic Acids Research* 33**:** D34-D38.

3 The State of Influenza Surveillance

3.1 Introduction

Influenza surveillance is foundational to two aspects of the benefit assessment. First, comparative analysis of the genetic sequences of wild type viruses collected through surveillance is an alt-GoF approach that may enable that identification of novel genetic traits associated with a particular GoF phenotypic change (e.g. pathogenicity). Analysis of wild type sequences can also be used to "validate" known molecular markers – that is, to demonstrate that a marker for a particular phenotypic change is selectively enriched in sequences of viruses with that phenotypic trait. For example, certain molecular markers for mammalian adaptation have been shown to be enriched in highly pathogenic avian influenza isolates recovered from human infections relative to isolates recovered from poultry.[34] The success and utility of these alt-GoF approaches depends on the quality and availability of genetic surveillance data. Second, pandemic risk assessments, which may be informed by molecular marker data generated through GoF experimental approaches as well as other data sources, involve evaluating the risk posed by circulating influenza viruses detected through surveillance. As these assessments guide investments in pandemic preparedness activities, the degree to which those activities truly improve preparedness partly depends on the scope of influenza surveillance coverage. Both alt-GoF approaches that involve comparative sequence analysis and the process of pandemic risk assessments are described in detail in the main report. The following report reviews the current state of influenza surveillance, which provides critical context for evaluating the utility and limitations of both.

3.2 Availability of Animal and Human Surveillance Data

3.2.1 Animal Surveillance

The quality of influenza surveillance in animals suffers from gaps in species coverage and geographical coverage. These gaps are highlighted by discrepancies between the distribution of available influenza genetic data versus global poultry and swine populations – the countries with the largest swine or poultry populations are not necessarily the largest producers of genetic surveillance.[35] The scope of genetic surveillance with respect to species and location is reviewed in the following section.

3.2.1.1 Species Gaps

Surveillance quality and coverage varies between species. Avian surveillance varies internationally and can include both active and passive surveillance of live and dead birds. The United States government does not conduct broad scale active avian surveillance on wild birds, but some government offices conduct passive opportunistic sampling of sick or dead birds.[36] Passive surveillance for avian influenza in

[34] Davis CT *et al* (2014) Use of highly pathogenic avian influenza A(H5N1) gain-of-function studies for molecular-based surveillance and pandemic preparedness. *mBio* 5.

[35] Russell, C. A., et al. (2014). "Improving pandemic influenza risk assessment." Elife **3**: e03883.

[36] https://www.aphis.usda.gov/wps/portal/aphis/ourfocus/animalhealth/sa_animal_disease_information/sa_avian_health/sa_biosecurity_for_birds/!ut/p/a1/lZJNc4IwEIZ_Sw8eMRGQj978Fiv9cmyFCxNCgEwhYZKg479vxB7sTGttbrv7bubZdxfEYAdihva0QIpyhqpTHDvJ6mlpDsbQDBYbfwaDx7d56K1d62lpa0GkBZPFaGm7awih7ZkwmI6XU9cPIQyc2_rhL28E_p_BzGIMVONKkGEmpLKBHOmCFNJRVOBxLEHJUp4K5Kc41Z2EWK0RlVSElSp8jKTUUmQJAllORd1Z 8K5vKeIXe pTyiXBraDqqP8VOhaZPKE0mGYgcjOYp76HjRRbyLBt7BibxHDzBzTGrjDgZk6 YHWDN6YIJ2Ghv0WqNE5UYHeV9ly-oO0SP9O-T0Gcb8IXpzi7-A1mMRtrmPn62X1YmfBh-CW4tuhOcGWTet6i4ml3VdGIpZanBxMkJ4KIfit0ulSqkfc92IOHw6FfcF5UpI953YM_tZRcKrD7rgSRvif3V1N9G2zuaWm3m5rzxpWhWd8vHpHq9rX9eju7hMxRplk/?1dmy&urile=wcm%3apath%3a%2Faphis_content_library%2Fsa_our_focus%2Fsa_animal_health%2Fsa_animal_disease_information%2Fsa_avian_health%2Fsa_biosecurity_for_birds%2Fct_bfb_wild_birds

dead wild birds only provides little insight into the diversity of avian influenza virus genotypes circulating globally or risk for future outbreaks in poultry or humans.[37] To address this shortcoming, a number of academic institutions, including those funded by the NIAID Centers of Excellence in Influenza Research and Surveillance (CEIRS) program, conduct active surveillance on live birds and mammals throughout the United States and the world.[38] Globally, surveillance efforts focus largely on bird species that are mass produced for consumption, such as chickens, or sold in live bird markets, such as ducks. Species that are not economically significant may be overlooked; however, surveillance projects targeting exclusively wild non-market bird populations such as gulls and other shorebirds have increased.[39,40] Non-domesticated species that live alongside domesticated species such as sparrows, starlings, wild turkeys, and other game birds can be left out of surveillance efforts that exclusively target other wild species or domesticated birds.[41] For avian influenza, limiting surveillance efforts to only mass produced poultry or only waterfowl neglects a large part of influenza ecology.

Relatively little is known regarding influenza circulating in swine around the world compared with knowledge about avian and human influenza viruses, including information about influenza subtype prevalence, antigenic characterization, and genetic profiles.[42] For swine influenza surveillance, there are many swine producing parts of the world that may have data but do not share it. For example, China has a significant swine population, but very little surveillance data comes from China.[43] Feral pigs are also a neglected population. Feral pig populations are increasing in the southern United States and are well established in other countries. Feral pigs often interact with both domestic pigs and wild birds, but sampling data for this population is lacking. Within the United States, the USDA Animal and Plant Health Inspection Service- Wildlife Services and their State and local partners conduct feral pig control programs that involve surveillance but have not historically tested for influenza.[44] Sampling of predators and other mammals is neglected as well. Raccoons and canines can transmit influenza and prey on or come into contact with avian populations, but are rarely sampled.[45]

3.2.1.2 Geographical Gaps

Global animal surveillance is limited, leaving large geographical gaps in data. The majority of sequences come from a small number of countries, with most countries having little or no genetic surveillance in place.[46] Surveillance efforts in the United States and Europe are more comprehensive than many other regions.[47] Outside of the U.S. and Europe, avian influenza surveillance has increased dramatically in Asia, but is still generally poor outside of Asia.[48] The southern hemisphere has been largely neglected and there are few swine or avian surveillance efforts in South America, Latin America, and Africa.[49,50] Available data skews heavily towards where it is convenient and permissive to do surveillance.[51] Factors such as infrastructure and ease of access to animals play a large role in how surveillance efforts are designed.

---

37 Machalaba, C. C., et al. (2015). "Global avian influenza surveillance in wild birds: a strategy to capture viral diversity." Emerg Infect Dis **21**(4): e1-7.
38 http://www.niaidceirs.org/
39 NIAID CEIRS St. Jude Center of Excellence for Influenza Research and Surveillance
40 NIAID CEIRS Center for Research on Influenza Pathogenesis
41 (2015b) Interviews with influenza researchers.
42 Vincent, A., et al. (2014). "Review of influenza A virus in swine worldwide: a call for increased surveillance and research." Zoonoses Public Health 61(1): 4-17.
43 (2015b) Interviews with Influenza Researchers.
44 www.aphis.usda.gov/wps/portal/aphis/ourfocus/wildlifedamage/sa_programs/sa_nwrc/sa_nwdp/ct_feral_swine/
45 (2015b) Interviews with Influenza Researchers.
46 Butler, D. (2012). "Flu surveillance lacking." Nature **483**(7391): 520-522.
47 (2015b) Interviews with Influenza Researchers.
48 Ibid
49 (2015a) Interviews with CDC representatives.
50 (2015b) Interviews with Influenza Researchers.
51 Ibid

Areas with large swine or poultry industries and locations where wild bird, poultry, and swine populations interact are vital to surveillance, but may be excluded because they are not accessible. Geographical gaps may be due in part to political forces. Countries may be less motivated to share information if they are not guaranteed to share in the benefits, such as influenza vaccines or outbreak support, to which their information contributes. Countries undergoing political or social unrest may lack the infrastructure, personnel, funding, or political will to prioritize surveillance efforts. Political instability may create security issues for researchers, or governments may be openly hostile to foreigners.[52] Surveillance systems may be under resourced, creating difficulties in sharing data or samples in a timely fashion.

3.2.1.3 Commercial Limitations

Powerful commercial interests may underpin species and geographical gaps in surveillance. Surveillance of agricultural animals is principally ad hoc and done on the basis of voluntary cooperation from agricultural companies.[53] Agricultural producers fear business losses if information about influenza detections reaches consumers, competitors, or the government. This fear creates reluctance to submit samples or allow outside researchers to access animals. Certain agricultural industries have powerful political lobbies, such as the swine industry in Brazil or the poultry industry in China, that can resist surveillance efforts and interfere with international transparency.[54] USDA surveillance of mass produced chickens and swine relies heavily on cooperation from industry partners. The swine industry in the United States is dominated by large swine producing companies that own the farms and the pigs. Some use private diagnostic labs, so results do not enter USDA surveillance streams without deliberate reporting of the data.[55] There are limited reporting incentives for samples collected by private companies, so results may or may not be shared.[56] This creates large gaps in swine surveillance. However, since some U.S. diagnostic laboratories have allowed samples to be submitted anonymously, to allay industry fears, the number of samples and locations represented has increased.[57]

3.2.2 Human Surveillance

Human surveillance for influenza is generally more advanced than animal surveillance, but is still in need of improvements. Human surveillance is conducted more extensively in nations with well-developed healthcare systems, such as the United States, Japan, New Zealand, and China, but the quality and continuity of surveillance decreases for many other countries, especially in the Southern Hemisphere.[58] Across all regions, an inherent limitation of passive human surveillance is that it depends on patient interactions with the healthcare system, thus novel infections are not detected when individuals do not seek medical attention.[59] This limitation is exacerbated in areas with few health services; in fact, in some poorly resourced and/or remote areas, it is possible that a novel virus could cause an epidemic that goes unnoticed, particularly if the virus is not highly lethal.[60] Furthermore, additional surveillance is needed on individuals at the interface of human-animal transmission, such as farm and animal market workers, which is critical for early detection of human infections with novel influenza viruses as well as for

52 (2015a) Interviews with CDC representatives.
53 (2015b) Interviews with Influenza Researchers.
54 (2015a) Interviews with CDC representatives.
55 (2015b) Interviews with Influenza Researchers.
56 Machalaba, C. C., et al. (2015). "Global avian influenza surveillance in wild birds: a strategy to capture viral diversity." Emerg Infect Dis 21(4): e1-7.
57 (2015b) Interviews with Influenza Researchers.
58 (2015b) Interviews with Influenza Researchers.
59 Ibid
60 Ibid

gaining insight into the epidemiological factors that promote the transmission of newly introduced viruses.[61]

3.2.3 Summary of Animal and Human Surveillance Gaps

There are many regions of the world and large animal populations for which little or no animal or human surveillance data is available, but where significant amounts of influenza almost certainly exist.[62] Political and economic factors significantly impact where and how much genetic surveillance can be done, resulting in limited numbers of sequences from certain species and regions, such as Africa and South America. In particular, wild birds, agricultural swine, and feral swine are not surveilled at levels that reflect the risk these populations face. Not all regions have comprehensive coverage, as countries with fewer economic resources and poor public health infrastructure are underrepresented in surveillance data. Comprehensive global surveillance is not yet a reality.

3.3 Quality of Surveillance Data

In addition to gaps in the availability of surveillance data from certain species and regions, existing genetic surveillance data suffers several shortcomings that compromise its utility for scientific studies.

3.3.1 Continuity Challenges

Surveillance tends to be carried out sporadically, either in response to disease outbreaks, or as temporary projects as funding allows. Even then, crisis-driven surveillance is incorporated into already overextended systems only when deemed absolutely imperative.[63] Programs implemented during these intermittent periods of concern provide useful data on influenza virus diversity and are essential for responding to outbreaks but do not capture how diversity changes over time.[64] Due to the ad hoc nature of current surveillance, individual countries and organizations often decide that different viruses are most relevant and threatening, and furthermore may alter their focus from year to year. This variability leads to a lack of sustained surveillance covering any one area, population, or viral species, resulting in large gaps in data because many consecutive years of surveillance are needed to see trends. Continuous surveillance is needed to gain a better understanding of how influenza fluctuates and spreads through various populations and how phenotypic traits arise in nature over time.

3.3.2 Methodology and Analysis Gaps

Influenza surveillance is currently scattered among universities, industry, and government agencies and includes passive surveillance carried out at the state and Federal levels as well as surveillance related to research projects from a variety of funding sources. Wide variations in sampling methodology, sample testing procedures, and protocols for subtype characterization between groups engaged in surveillance limit the comparability of existing animal and human surveillance data. [65] Specifically, the use of unstandardized methods complicates comparison of surveillance information between years and locations, thus undermining the usefulness of data to other researchers or government decision makers.[66] Many of

---

61 Russell, C. A., et al. (2014). "Improving pandemic influenza risk assessment." Elife **3**: e03883.

62 Ibid

63 (2012). "Under surveillance." Nature **483**(7391): 509-510.

64 Machalaba, C. C., et al. (2015). "Global avian influenza surveillance in wild birds: a strategy to capture viral diversity." Emerg Infect Dis **21**(4): e1-7.

65 (2015b) Interviews with Influenza Researchers.

66 Machalaba, C. C., et al. (2015). "Global avian influenza surveillance in wild birds: a strategy to capture viral diversity." Emerg Infect Dis **21**(4): e1-7.

the research surveillance streams are targeted surveillance designed to address specific questions, meaning their scope is limited. When data from these efforts is deposited in the public domain, it is open to misunderstandings about why particular samples were collected or particular methods used. Drawing broader inferences from surveillance data that was designed to address narrow research objectives may be difficult.[67] Data produced through targeted surveillance is not necessarily applicable to other locations or situations and may not provide an adequate basis for pandemic risk assessments. Alternatively, sampling methods may be ad hoc. For example, selection of swine samples for sequencing, among those submitted by veterinary diagnostic labs to the USDA's passive surveillance stream for swine influenza, is based largely on subjective expert intuition.[68] Appropriately synthesizing information that has not been collected systematically is also difficult. Notably, for human surveillance, the WHO's databases of human influenza and guidelines for conducting human surveillance alleviate some of the standardization issues that plague animal surveillance.[69]

Improving human and animal surveillance requires expanding the emphasis beyond highly pathogenic avian influenza viruses to include a wider variety of subtypes and sequences. Virus characterization protocols vary as widely as collection methods. Some programs screen only for H5 or H7 virus subtypes, and some screen only for HA or NA subtypes but not for subtype combinations.[70] Surveillance may target the most recent outbreak strain and neglect other subtypes, meaning that even if zero cases are reported, additional subtypes not tested for may be circulating. There are no standardized and comprehensive reporting requirements for animals beyond highly pathogenic avian influenza and H5 and H7 subtypes of low pathogenicity avian influenza viruses.[71] There is also a lack of additional measures to track influenza virus diversity beyond subtype. Capturing information on all eight virus segments enhances understanding of virus dynamics and evolution, but is not routinely done.[72]

Critical metadata are deficient in many animal surveillance reporting systems.[73] Metadata concerning the health and living conditions of animals, type of sample taken, and type of collection site allows influenza results to be contextualized and is essential for larger epidemiological studies. In the United States, many swine producers have samples tested anonymously. While anonymity may increase industry participation, it limits the utility of the sample because little to nothing is known about the farm or animal from which the sample was collected.[74] Epidemiologic metadata such as patient demographics, antiviral treatment, vaccination history etc. should accompany human samples, but is nonstandard and often lacking.[75] In addition, human data in many countries must be de-identified to protect patient privacy, which may require the removal of useful metadata. Beyond collecting data and metadata in a systematic way, data analysis programs and databases also need to be developed to generate meaningful conclusions.[76] Large scale statistical analysis and geographic mapping can be employed to track, compare, and visualize influenza on a regional or global scale, but require technological and computational effort. Broader collaboration is needed between those generating data and those analyzing it to guarantee that data addresses the research question at hand and remains useful for additional studies.

---

67 (2015b) Interviews with Influenza Researchers.
68 Ibid
69 WHO Global Epidemiological Surveillance Standards For Influenza
70 Machalaba, C. C., et al. (2015). "Global avian influenza surveillance in wild birds: a strategy to capture viral diversity." Emerg Infect Dis **21**(4): e1-7.
71 Ibid
72 Ibid
73 Ibid
74 (2015b) Interviews with Influenza Researchers.
75 (2015a) Interviews with CDC representatives.
76 (2015b) Interviews with Influenza Researchers.

While the WHO maintains large databases for human samples, there is no comparable centralized database for animal samples, which inhibits information sharing.[77] Researchers or government stakeholders who are not part of localized research networks or do not have access to smaller databases may not be aware of surveillance results. Within the United States, surveillance data for swine and birds is largely recorded only at the state level and may or may not be shared with the federal government or research institutions.[78] Interoperability and the capacity for information to be linked and shared across multiple databases is also increasingly important but fundamentally lacking.[79]

Another challenge for understanding surveillance is the lack of negative data. Scientific journals and online databases facilitate information sharing from research projects and government surveillance, but often underreport negative findings.[80] Researchers may only publish or share positive data and do not see the value of publishing negative results.[81] This leads to asymmetrical surveillance results which obfuscate the true prevalence of influenza. The lack of interest in negative results also means that locations where influenza is not regularly found are deprioritized in surveillance efforts, creating additional coverage gaps.[82] Promoting greater sharing of negative findings would increase the overall utility of surveillance results and increase tracking potential for influenza viruses.

3.3.3 Sequencing Challenges

Sequencing of surveillance samples can be delayed, limiting their utility for assessing an ongoing outbreak or recent evolution. Months or years can pass between collection of virus samples, sequencing, and release of sequences into public databases.[83] Limited funding for sequencing can contribute to the delay, as can the fact that many samples are sequenced in retrospective research studies. Researchers also contribute to the lag because many will not publicly share sequences until after the associated data have been published.[84] Sequencing efforts related to surveillance should ideally include both consensus and deep sequencing. However, almost all published surveillance sequences are consensus sequences, which do not include minor genetic variations present in viral quasi-species that may represent emerging mutations or antigenic drifts.[85] Interestingly, one expert felt that depth of sequencing partially compensates for incomplete coverage of human and animal populations.[86] Providing a wider landscape of genetic diversity from one sample helps reveal shifting genotypes without sampling a whole population.

3.3.4 Summary of Gaps

Surveillance methodology is compromised by a variety of issues that undermine the utility of surveillance data for research projects. Samples and accompanying metadata may not be collected in uniform ways, which complicates the synthesis of data collected by multiple groups. Surveillance is rarely continuous and is subject to shifting interests, which further reduces the comparability of surveillance data across studies, regions, and years. Sequencing of samples, essential for understanding existing and emerging mutations, can be delayed and the lack of centralized, accessible databases interferes with sharing sequences and their associated metadata. Negative data is underreported, which misrepresents the true

---

77 Ibid
78 Ibid
79 Machalaba, C. C., et al. (2015). "Global avian influenza surveillance in wild birds: a strategy to capture viral diversity." Emerg Infect Dis **21**(4): e1-7.
80 Ibid
81 (2015b) Interviews with Influenza Researchers.
82 Ibid
83 (2012). "Under surveillance." Nature **483**(7391): 509-510.
84 Butler, D. (2012). "Flu surveillance lacking." Nature **483**(7391): 520-522.
85 (2015a) Interviews with CDC representatives.
86 Ibid

prevalence of influenza. The quality of surveillance data would be increased by greater standardization across all steps of the surveillance process from collecting samples to sharing results.

3.4 Conclusion

Researchers have called the current state of influenza surveillance sad and crippled.[87] Gaps in geographic coverage and species diversity and a lack of data standardization significantly limit the current utility of genetic surveillance data for scientific studies. Gaps in coverage also compromise the value of pandemic risk assessments, as knowing whether as-yet-undetected sequences pose high risks is impossible. Future surveillance programs offer a variety of chances to remedy these issues and bring surveillance up to the standards necessary to provide a comprehensive picture of the global burden of influenza.

[87] (2015b) Interviews with Influenza Researchers.