Influenza A Reassortment – Final Report

Matt Ingham

xxxxxxxx

1. Phenomenon:

The haemagglutinin (HA) and neuraminidase (NA) proteins both exist in many different versions, and form many combinations in various strains, of the influenza virus. Thus, many different subtypes of influenza exist (eg. H1N1, H3N2). These subtypes come about through reassortment, a phenomena in which two differing subtypes of the virus invade a host at the same time and their genomes become mixed, resulting in a new subtype. These subtypes vary greatly in their pathogenicity (the ability of a virus to invade a new host) and virulence (the amount of damage incurred by the host due to the infection) (Capua, Alexander, 2002). Typically, mammals are able to mount a sufficient immune response against any subtype of the influenza virus that remains within mammalian organisms. Additionally, avian subtypes of the virus are usually incapable of properly binding mammalian cells in order to cause an infection. However, when an avian flu subtype is able to infect a mammal, typically a pig (although human infections have been recorded in recent years), immune response is far weaker as no memory cells or antibodies exist to counteract the new subtype. Usually such infections do not spread easily, for reasons mentioned above (Zambon, 2001). Occasionally, reassortment can result in a subtype which contains a combination of HA and NA proteins which has both high pathogenicity and virulence, and results in a serious pandemic (Capua, Alexander, 2002).

Several factors exist that may affect the danger posed by a new subtype of influenza. These include its pathogenicity, virulence, interaction of its HA and NA proteins, amount of avian influenza that initially entered the population to cause the reassortment, and the

likelihood of reassortment itself. These factors are as of yet not well understood nor quantifed for either avian subtypes or new subtypes from reassortment.

The theory that the amino acid sequence of a protein, and to a lesser degree nucleotide sequence which encodes that amino acid sequence, determines function is widely accepted. It is reasonable to believe that a relationship between the sequences of these various subtypes and how the combination of particular HA and NA proteins interact with both each other and with host proteins exists (Wagner *et al*, 2001). These interactions determine the levels of pathogenicity and virulence of various subtypes (Zambon, 2001). The model uses such sequence information to determine how new subtypes will behave, and how often highly pathogenic or virulent strains will occur given current human and avian influenza subtypes. Ideally, this model will help determine why certain subtypes are capable of causing pandemics, while others are not, as well as why certain factors dictate the occurrence of such pandemics and how many deaths result.

2. Model Design:

The model was created in NetLogo in which the various 'turtle' objects represent humans. At any given time some of these objects may be infected with one or more subtypes of influenza. In cases where two or more occur in one object, reassortment may occur with a given probability. If it does, a new subtype will be created and behave based on pre-calculated values for pathogenicity, virulence and protein interaction. Each virus is given a set of viral attributes for these factors, though in some cases such as reassortment rate, the value is constant. For others, such as pathogenicity, it is based on similarity with known viral subtypes.

Two experiments were conducted with the model. The first (Experiment 1) was an attempt to model many various subtypes of influenza, mostly avian, to predict how avian

subtypes may enter the human population and reassort, and the potential of pandemic associated with such activity. Within the model, a pandemic is described as the occurrence of novel virus spreading across large portions of the population, at least more than those observed as infected by the two common subtypes H1N1 and H3N2. While the model was constructed to account for many variables, sufficient data on existing avian influenza subtypes and reassortment rates does not yet exist to incorporate into the model. As such, experiments were conducted based on almost entirely on predicted values from sequence similarity.

The second experiment (Experiment 2) was conducted is an attempt to determine the levels for various attributes, particularly those that would result in wide scale death due to influenza. In this case, only a single avian influenza (H4N6) and the two reassorted subtypes (H4N1 and H4N2) possible were considered. The simulation was then run for a range of values for different attributes to investigate the significance of each on mortality due to influenza reassortment. The following steps were taken to design the model on which these experiments were run.

i. Determine similarity between various subtypes.

In order to determine similarity between subtypes, a custom database was created of the amino acid sequences from all types of HA proteins known to infect humans. This database was used to conduct BLAST (Basic Local Alignment Search Tool) searches in which each HA sequence from various subtypes was used to search through a database of subtypes capable of human infection. This was conducted to determine the level of similarity between versions of the protein. These similarities were used to determine the viral attribute values for new subtypes based on their similarity to the proteins in existing subtypes. Amino acid sequences were used for comparison as similarity found between amino acid sequences

is more significant than between nucleotide sequences. Differences in nucleotide sequences do not necessarily result in a difference in amino acid sequence, and thus difference in function or functional efficiency. It is assumed that the more similar the sequences of two versions of a protein, the more similar their characteristics will be, and that the differences between the subtypes are represented by the differences in sequence.

ii. Quantification of Pathogenicity and Virulence of various subtypes:

In order to determine how fast a subtype will spread, and how many of those infected it will kill, some sort of quantification of pathogenicity and virulence is necessary. These were represented in the model by the odds that two objects will pass an infection when they are in the same space for pathogenicity, and the odds that an object will die once infected with a given subtype for virulence. These odds are based on data from H1N1 and H3N2 subtypes currently circulating, as they are the only well studied subtypes. Odds for new subtypes will be calculated relative to these two subtypes. Mortality rates were based on those recorded clinically (University of Maryland Medical Center, 2005). It is also assumed that inter-species infection rates are far lower than intra-species infection, meaning when an avian flu enters the human population it has a low infection rate, and the initial infection event does not happen as often as that of types H1N1 or H3N2. (Zambon, 2001).

iii. Quantification of HA/NA version combination:

Different versions of HA and NA proteins work together to varying degrees. Any combination that currently is capable of infecting any species is assumed to work equally well together as the manner in which HA and NA proteins interact is still unclear. The degree which new subtypes was predicted based on the most similar HA sequence, and will act to diminish the pathogenicity of the new subtype as new combinations have not evolved

to work together. This factor was taken into account only in Experiment 1, as in Experiment 2 only two reassorted subtypes were being considered and were judged that both combinations would have about equal efficacy.

iv. Simulation of infection in the human population:

NetLogo was used to create a population of human objects. Initially each human is given a random age between 0 and 80. Any human that reaches the age of 80 is assumed to die of natural causes. A birth rate is incorporated to account for this, in which new humans are born at about the rate seen globally today versus the population. This results in a population growth rate (ignoring deaths from influenza) of about around 1%, roughly that of the world currently (US Census Bureau, 2004).

Initially, some humans are infected with either influenza H1N1 or H3N2, as well as others starting with immunity to one or both subtypes. This is meant to simulate the current environment in which these two subtypes circulate throughout the population at all times. Pathogenicity attributes have been empirically determined by finding those which result in about one person per 150 being infected with either subtype. This is based on the amount of predicted cases of influenza per year in the United States (Center for Disease Control, 2005). This method was also used to determine population density and general rates of infection for all subtypes. Humans became infected with various subtypes, and were all capable of infecting each other with any influenza subtype they may themselves be infected with. Infection had a certain probability of occurring based on the values in the similarity tables listed above whenever two humans were in the same vicinity.

If one object was infected with two different subtypes, a reassortment had a certain chance of occurring. All subtypes were assumed to have equal chance of reassortment. In

Experiment 2, the rate of reassortment, infection and mortality was varied in order to compare the affect each variable has on total mortality due to influenza.

Immunity was represented by a human having resistance to infection by a subtype it has previously encountered for 150 steps, or about a year and a half, based on tests with vaccines (Zambon, 2001). This is based on the assumption that after infection an organism is immune only as long as it has memory cells and antibodies for that strain of the subtype. These cells die, and also small mutations result in the inability of the memory cells to recognize new strains, thus resulting in a loss of immunity. As such, immunity is not absolute, and the chances of those immune being infected again increases with time, until immunity level reaches zero.

The model was allowed to run for a period of time representing years, in order to accurately model the reaction of the population to new subtypes due to reassortment and to model multiple occurrences of the rare event of reassortment. In experiment 1, the levels of various subtypes was plotted and used as measurement of virus activity. In experiment 2, the amount of deaths attributed to each virus subtype was recorded at the end of each simulation as well. This was used to compare the relative effect of reassortment, pathogenicity and virulence.

The lack of solid estimates for experiment 1 led to the development of experiment 2, in which the values for one virus were varied and evaluated. Three factors were studied: reassortment rate, pathogenicity and virulence (Appendix – Table 2). Tests were conducted by keeping rates for two variables constant for the potentially pathogenic virus H4N2 and altering one. The simulation was then allowed to run for ten years (represented by 1000 steps), and the number of deaths attributed to H1N1, H3N2, H4N6, H4N1 and H4N2 were tabulated. Values used were hypothetical, but reasonable in comparison to the model

constructed around experiment 1. The initial value was doubled then tripled to evaluate three values for each variable.

The nature of the model is that of a simulation based on analytical equations. Simple equations were developed for phenomena such as infection, and placed together in Netlogo to create a simulation. Each object is given a set of values for different variables that dictate its ability to survive. These variables take the form of values associated with different virus subtypes, which dictate whether or not a certain object is capable of being infected, infecting others once infected, or suffer a premature death due to infection. Other factors include global values and coefficients which dictate phenomena such as reassortment rate and infection rate for all viruses. As a result, observations can be made about the behaviour of different subtypes, and which factors play the largest role in mortality.

3. Assumptions and Simplifications:

In both models, it is assumed that avian influenza subtypes are less pathogenic but more virulent than human subtypes, as they are not well adapted to humans, but humans have developed no immunity to them. This has been seen in H5N1 and H9N2 avian influenza infections in Asia (Capua, 2002).

For experiment 1, it is assumed that a new HA will work proportionate to how similar it is to another HA when combined with an NA. For instance, if H4 has a very high similarity to H1 but low similarity to H3, it is assumed that H4N1 will have a high coefficient of protein interaction, but H4N2 will not if H1N1 and H3N2 were in the database.

Although virulence is typically higher in infants and elderly as they lack the defenses that normal adults do, this factor has been removed by simply taking a virulence level that would represent an average for all infected objects. Combined with the assumption that the

object could be of any age this accounts for the age factor. Additionally, mutations create a range of strains of a given subtype. This will be ignored as the loss of immunity is the main result of such mutations, and is included. The other results of mutation are changes in pathogenicity and virulence that are not accurately predictable as of yet, and thus should not be included. Such mutations would involve recalculating sequence similarity which would be difficult from within NetLogo. Thus, mutation will be ignored for simplicity.

An additional simplification is that of considering only the N1 and N2 types of neuraminidase. Since only two virus subtypes have been found to infect humans that do not include one of these protein types (H7N3 and H7N7), and such infections are very rare, these types of neuraminidase have not been considered as the odds of their reassortment in reality are extremely low in humans.

Since only humans are involved in the model, reassortment can only occur in the human population (as opposed to the avian source which is not present in the model). In reality, reassortments resulting in human infection are thought to most commonly occur in pigs. However, since the purpose of this model is to investigate the possibilities of avian subtypes being incorporated into humans, and the reassortments that may follow, only reassortment in humans is being considered. This is largely due to the fact that relevant data for such phenomena as avian to pig infection and information regarding virus spread within a pig population is very limited, and as such would be extremely difficult to model accurately.

Conforming to the Phenomena:

The initial technique of conforming to influenza behaviour was modeling the activity of H1N1 and H3N2. Since rates of infection are very difficult to predict as so many variables can come into play, this was set initially at a one in ten chance for all non-immune humans in the vicinity. From there, variables such as population density, a global infection rate

coefficient and initial levels of infection and immunity were adjusted in order to arrive at a somewhat steady state of H1N1 and H3N2 levels. In earlier versions, the levels of these types would eventually arrive at zero, meaning the unrealistic case of eradication of these types had occurred. For this reason, rarely (about one human every few steps) a random human is infected with one of these types. Since the model represents a small portion of the global population, this represents someone leaving that section of the world and returning infected, which is realistic and comparable to residents of specific area leaving, becoming infected, and returning.

These levels for H1N1 and H3N2 served as the basis for all other subtypes as they were created with viral attributes relative to those of these two initial virus subtypes. Another demonstration of the phenomenon is the behaviour of pandemic viruses. These viruses are those with high pathogenicity for which there is little or no immunity in the population. In most cases these infections resulted in a sudden surge of infection, followed by a sudden drop as a surge in immunity soon followed. This pattern of infection coincides with past pandemics.

One factor that affected the degree to which novel viruses spread is the density of the population in which they first appeared. A high relative density was required for the virus to spread easily, as it must infect enough people right away since the infection is short-lived. This is also true in reality. Influenza will tend to spread in cities far better, especially places like universities with a high density. This helps add to the stochastic nature of infection seen in reality, as it resulted in the same virus spreading in some cases, and disappearing due to lack of hosts in others, which is also seen in nature.

The appearance of new viruses due to reassortment is not common. Thus, in experiment 1 the reassortment rate was set so that emergence of a new virus was rare. Also, usually only low amounts of the new virus are found, and only spread if it is highly

pathogenic, as was seen as well in experiment 1. In experiment 2, the reassortment rate was varied to evaluate its significance, and was purposefully different from an optimal rate in order to simulate different rates. This is acceptable as the reassortment rate is not well documented in influenza. The value used in experiment 1 is a rough estimate developed empirically. This does represent the idea that reassortment doesn't happen every single time two viruses infect the same host. In order for this to occur, the two viruses must have similar tropism to infect the same cells, they much both be replicating in the cell at the same time, the capsid must enclose around different genome segments and the genome segments must be the right size for the viral capsid to enclose it properly in order to still be functional. If all this occurs, then a new virus subtype can appear and infect humans.

Insights

Insights in experiment 1 were difficult to arrive upon as there was a considerable lack of data regarding subtypes not currently infectious to humans. Sequence similarity was utilized to estimate the viral attributes, but since research on the significance of differences in sequence similarity as they relate to pathogenicity and virulence is still in its early stages, these estimates were rough. Some insights were still possible from this approach, however.

The result of constructing experiment 1 was a model in which most viruses were predicted to behave like the H5N1 or H9N2 subtypes currently observed, which have high virulence, but low pathogenicity. One subtype that did stand out was the H4N6 subtype, which had very high similarity to H3N2 (Appendix - Table 1). This led to the theory that it was close enough to H3N2 that a reassortment resulting in an H4N2 subtype would create a virus with the pathogenicity of H3N2, which spreads easily throughout the human population, but also be different enough that no immunity would exist for it. As the only potentially pandemic subtype in an array of those of relatively little harm to the population,

this allowed for insights into how often such a reassortment will result in a pandemic. It was found that not every occurrence of the H4N2 subtype resulted in a pandemic. In many cases the virus would start to spread, but not infect enough people, and disappear, similar to the manner described in section 4. The random nature of infection would also account for this pattern. Observing such a pattern indicates we should increase surveillance of subtypes such as H5N1 which have had multiple small outbreaks. In experiment 1 it often took several occurrences over several years for the H4N2 pandemic to occur, and this could be the case with H5N1 as well. Other insights reached from experiment 1 were that a reduction of pathogenicity even by half results in an exponential decrease in the spread of the virus. This was observed through the monitoring of a myriad virus subtypes with a range of pathogenicity values and evaluating patterns in their occurrence and spread.

In experiment 2, the greatest difference in deaths attributed to H4N2 came as a result of changing the pathogenicity. The amount of deaths at an infection score of 1500 was over 50 times those seen at 500. Also, while the number of deaths due to H4N1 and H4N2 were about the same at 500 (H4N1 had a value of 300 for all simulations), it was much higher (about a factor of 6) in the latter two simulations. It was also seen that the number of deaths due to H4N1 increased as well, due to the fact that a large increase in H4N2 led to frequent reassortment with H1N1 to create many H4N1 viruses. This is an important insight, as it reveals that if reassortment levels prove to be as high as in this model, two subtypes using the new HA protein could emerge. For instance, a pandemic of H5N1 could result in a large amount of H5N2 as well. Although it may not be as pathogenic or virulent, the lack of human immunity to the second subtype could still result in many deaths.

The second most influential factor was that of reassortment rate. While both virus types (H4N1 and H4N2) were affected by this global variable, the higher pathogenicity of H4N2 still resulted in it having several times more deaths attributed to it. One difference worth

noting is that the amount of deaths rose roughly proportional to the odds or reassortment in H4N2, but increased far more rapidly in H4N1. This is likely attributable to the reassortment phenomenon mentioned before, especially since the rate of reassortment increased. This shows that increases in reassortment do not have the same effect as pathogenicity on a single novel virus, but can result in a large secondary pandemic associated with the primary subtype. This has yet to be seen in any pandemic to date, but is still something to be considered in vaccine production.

Finally, increases in virulence actually resulted in a decrease in deaths. Considering that numbers decreased for H4N1 as well, there is likely something about these trials that resulted in low death rate for reassorted subtypes, and further trials are needed for all three factors. However, one explanation is that the very high death rate resulted in a loss of population density around those still infected, and causes the virus to disappear. Although this seems unlikely considering those who do not die will be immune, the chance that those immune may still get infected may be enough to allow the virus to spread enough to survive for a longer period of time. A higher amount of trials with a longer period of time each would help determine if there is any validity in this explanation.

Overall, this model allowed for several insights into the potential activity of influenza and new subtypes that appear due to reassortment. Due to the stochastic nature of the model, more precise insights could be drawn if it were allowed to run tens or hundreds of times in order to best draw conclusions about different factors. Additionally, as accurate information regarding the pathogenicity of avian influenza and reassortment rates comes to light, it can be incorporated in order to better predict which subtypes in particular pose the greatest threat. However, the findings thus far indicate that pathogenicity is the key factor in potential to cause a pandemic, and should be the main target in vaccine and antiviral drug development.

References:

1. Zambon, MC. The pathogenesis of influenza in humans. Rev Med Virol. 2001 Jul-Aug;11(4):227-41.
2. Capua, I, Alexander, DJ. Avian influenza and human health. Acta Trop. 2002 Jul;83(1):1-6.
3. Wagner, R, *et al.* Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Rev Med Virol. 2002 May-Jun;12(3):159-66.
4. Macken, C., Lu, H., Goodman, J., & Boykin, L., "The value of a database in surveillance and vaccine selection." in *Options for the Control of Influenza IV.* A.D.M.E. Osterhaus, N. Cox & A.W. Hampson (Eds.) Amsterdam: Elsevier Science, 2001, 103-106.
5. Altschul, SF, *et al.* Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10.
6. University of Maryland Medical Center website, April 7, 2005. http://www.umm.edu/non_trauma/flu.htm
7. US Census Bureau, Internation Programs Centre (graph on BBC news site), March 23, 2004. http://news.bbc.co.uk/1/hi/world/americas/3560433.stm
8. Center for Disease Control, Atlanta, USA, flu website, April 7, 2005 http://www.cdc.gov/flu/

Appendix

Query	Subject	%Identical	Expect	Query	Subject	%Identical	Expect	Query	Subject	%Identical	Expect
H1N1				H4N6				H7N3			
											1.00E
H1N1	H1N1	98.0198	0	H4N6	H3N2	63.11031	0	H7N3	H3N2	46.5035	15
			1.00E-				1.00E-				1.00E
H1N1	H1N2	76.7313	156	H4N6	H5N1	42.25621	122	H7N3	H5N1	42.15501	12
			1.00E-				3.00E-				2.00E
H1N1	H5N1	44.28152	83	H4N6	H1N2	39.07104	77	H7N3	H1N2	38.04627	7
			2.00E-				6.00E-				4.00E
H1N1	H9N2	36.875	61	H4N6	H9N2	38.03681	63	H7N3	H1N1	33.14917	5
			5.00E-				5.00E-				7.00E
H1N1	H3N2	33.5277	54	H4N6	H1N1	33.72434	53	H7N3	H9N2	34.375	4
H1N2				H5N1				H7N3	H1N2	41.17647	7
H1N2	H1N2	98.4252	0	H5N1	H5N1	100	0	H7N3	H1N1	41.17647	7
			1.00E-				1.00E-				
H1N2	H1N1	76.7313	156	H5N1	H1N2	56.38889	127	H9N2			
			1.00E-				1.00E-				
H1N2	H5N1	55.55556	125	H5N1	H3N2	41.8	121	H9N2	H9N2	100	
			4.00E-				3.00E-				2.00E
H1N2	H9N2	46.01227	88	H5N1	H1N1	45.16129	85	H9N2	H1N2	46.62577	8
			6.00E-				6.00E-				3.00E
H1N2	H3N2	38.75339	76	H5N1	H9N2	43.39623	79	H9N2	H5N1	43.39623	7
											1.00E
H1N2	H5N1	43.47826	1.8	H5N1	H1N2	43.47826	2.5	H9N2	H1N1	37.5	6
											2.00E
H3N2				H6N1				H9N2	H3N2	36.61538	5
H3N2	H3N2	100	0	H6N1	H5N1	57.97373	0				
			1.00E-				1.00E-				
H3N2	H5N1	41.8	121	H6N1	H3N2	41.49909	127				
			6.00E-				1.00E-				
H3N2	H1N2	39.02439	77	H6N1	H1N2	54.64481	123				
			3.00E-				8.00E-				
H3N2	H9N2	36.61538	58	H6N1	H9N2	48.60681	88				
			6.00E-				1.00E-				
H3N2	H1N1	33.81924	55	H6N1	H1N1	42.07493	72				

Table 1: Results for sequence similarity search of all HA sequences against a database of those known to infect humans.

Reassortment Rate	H1N1	H3N2	H4N6	H4N1	H4N2
1 in 10	28	42	119	15	147
1 in 20	29	26	115	66	367
1 in 30	18	24	86	184	521
Pathogenicity					
5%	29	30	125	15	18
10%	28	46	94	56	376
15%	16	21	112	173	963
Virulence					
25%	29	38	106	120	357
50%	21	33	140	57	327
75%	35	33	177	26	312

Table 2: Results for number of deaths for each subtype in a ten year test simulation in conducted during experiment 2.