Epidemiology: who, what, when, where and why?

February 1996: an outbreak of Salmonella infections in Denver, Colorado.

The first twenty patients examined tested positive for S. enteriditis.

This *Salmonella* species is often involved in food poisoning outbreaks, so public health personnel asked routine questions:

Had they all bought food at the same supermarket?

Had they all eaten at the same restaurant?

What had they eaten, and when?

However, no common thread among the patients could be found. Meanwhile, the number of cases continued to rise.

Health officials considered other possibilities.

Further questioning revealed that all patients had recently been to the Denver Zoo and visited the Komodo dragon exhibit.

Like many reptiles, Komodo dragons often harbor *Salmonella* in their intestines.

One study showed that *Salmonella* spp. was detected in 40% of tested pet reptiles (63% of lizards, 53% of snakes and 26% of chelonians).

Bacteria on the reptile's feet contaminated the enclosure, and visitors became infected after touching the enclosure and then touching their mouths with their contaminated hands.

All patients had diarrhea, half of which were considered severe. Six individuals required hospitalization, and health officials speculated that there **might have been several hundred more unreported cases**.

Epidemiology is the study of how disease occurs in populations and can be broken down into three major steps.

- 1. Understand the factors that influence a disease's **frequency** and **distribution**.
- 2. Predict future problems.
- 3. Make recommendations that might prevent disease or limit the number of people affected.

The birth of epidemiology: The first modern epidemiological study identified **cholera** as a *waterborne* disease.

In 1853 a cholera outbreak in London killed 10,675 people. In 1854, another outbreak killed 127 people in three days.

During this time, residents in the affected part of London obtained their drinking water from one of two water sources.

The Southwark and Vauxhall Company or the Lambeth Company.



Both pumped their water from the River Thames.

Broad Street pump, modern replica; Broadwick Street, London. **John Snow**, a local physician, suspected that contaminated water could transmit the disease.

Snow followed the **incidence** of the disease (number of new cases per specified time period).

Snow compared a person's likelihood of contracting cholera to where that person obtained water.

In the first 45 days of the epidemic, **3%** of Southwark and Vauxhall's customers died of cholera, while **0.4%** of Lambeth's customers died of cholera. **Why a seven-fold difference**?

Snow discovered that Southwark and Vauxhall's water came from a pump near Broad Street--an area where **untreated sewage** entered the Thames. The Lambeth Company drew its water from a cleaner section of the river upstream of the city.



Map showing areas within walking distance of the Broad Street pump (circled). Snow gathered statistical evidence that showed a high incidence of cholera in homes and businesses within a short walk of the pump. Black bars indicate cholera deaths. Let's expand on the concept of disease epidemics. Two general types exist.

1. A single, contaminated site can give rise to common source epidemics.

2. Host-to-host epidemics are spread from infected to non-infected individuals.



In common source epidemics, there is a steep rise in the number of cases. In host-to-host epidemics there is a slow increase in cases, as the disease agent begins to spread through the population, and a more gradual decline.

Common source epidemics.

These occur when a large number of individuals become infected from the **same** original source.

Examples:

Event	Source
London cholera epidemic (1853)	Broad Street pump
Denver Salmonella outbreak (1996)	Komodo dragons
<i>E. coli</i> -related GI illness at MSU (2008)	Contaminated lettuce
Pseudomonas skin infections, Oregon (2003)	Body piercing parlor

Common source epidemics are usually characterized by a very **rapid** rise in the number of affected individuals.

Why do the number of cases continue to increase rapidly?

Once the source is identified, cases will continue to be reported for a time period approximately equal to the *duration of one incubation period* (remember, the time between exposure to the organism and the first onset of symptoms). Why one incubation period?

However, when the common source is *not* identified, cases may be reported indefinitely.



Epidemic curve of a Salmonella outbreak in Estonia. Chicken soup served at the school was reported as the source of the contamination. Dontšenko et al., 2008.

Host-to-host epidemics

Host-to-host epidemics are characterized by no single infection source. Rather, <u>many potential sources</u> of infection are available, since every infected individual (human or animal) can serve as a source.

Epidemics of this type can start with one infected person.

Then how does the number of infected individuals get so high?

An epidemic of the flu is a common example.



Epidemic curve for influenza A (H1N1) school outbreak (averaged over weekends), West Sussex, England, 4-26 May 2004. From: Goddard, et al. 2004.

Host-to-host epidemics start out slowly, and then spread as the number of potential sources of infection grows. A peak is reached, and then the epidemic then slowly wanes.



From: http://www.cdc.gov/flu/weekly/

Why do epidemics occur?

1. Weather is one key.

Bacterial meningitis - caused by Neisseria meningitidis.

The world's largest recorded outbreak took place along Africa's **meningitis belt** during the dry season of 1996 - 1997.



250,000 people were ill, resulting in 25,000 deaths.

Meningitis is transmitted from person to person through droplets of respiratory or throat secretions.

Close and prolonged contact (e.g. kissing, sneezing and coughing on someone, living in close quarters or dormitories (military recruits, students), sharing eating or drinking utensils, etc.) facilitate the spread of the disease.

Epidemiologists noted that people stay in their homes more during the dry season, making it easier to catch meningitis from others. The epidemics usually stop when the summer rainy season begins and farmers move out to the fields.

Just the opposite can also be true...the wet season is associated with disease.

Cholera - caused by Vibrio cholerae.

In Bangladesh, a study in 2005 showed that cholera epidemics were related to bacteriophage concentrations in the water.



Figure 2. The monthly number of shigellosis/dysentery, typhoid fever, and cholera cases reported in Vietnam during 1991–2001.

During most of the year, the bacteria-killing viruses keep the choleracausing bacteria in check.

However, during the monsoon season (May-October) in Southeast Asia, the phage becomes diluted in flooded water bodies.

Because of the dilution, fewer *V. cholerae* are killed by the phage and the likelihood of an epidemic increases.

2. Epidemics become more likely when fewer people in a population are resistant.

Immunity to a disease pathogen may last months, years, or even a lifetime. Individuals who have never been exposed to the pathogen in question remain susceptible.

The proportion of the population that is immune to an infectious agent is known as the **herd immunity** for that microorganism.

For example, if **40%** of the population is immune to a strain of the flu, then the herd immunity in that population is assumed to be **40%**.

Herd immunity for the polio virus can approach 100% (because almost everyone has been vaccinated).

Herd immunity is always changing in a population. Why?

a. When people recover from an infection, they are likely to be immune to subsequent infections with the same pathogen – **increased herd immunity**.

Vaccination achieves the same result, without the need for actual disease.

b. Natural immunity can decrease over time - decreased herd immunity.

Herd immunity declines when babies are born, or when susceptible newcomers enter the population.



Typical immunity curves of what occurs after the introduction of one infectious case of measles in a population of susceptible hosts. From; Trottier and Phillipe, 2002.

If herd immunity for a particular pathogen is low, a pathogen entering the population can easily cause an epidemic because there are so many susceptible potential hosts.

As an epidemic begins (green line, day 20 above), herd immunity (red line) starts to rise.

As infected people recover, they are subsequently immune to a second infection.

At some point herd immunity is so high that there are few susceptible hosts (~day 40 above), and transmission cannot occur frequently enough to sustain the epidemic.

Herd immunity will continue to slowly rise, reaching its highest level as the last few people become ill at the **end of the epidemic** (day 60).

Herd immunity explains why so many host-to-host epidemics occur in **cycles**.

Example: Before the widespread use of the chickenpox vaccine, almost all children starting school at age 5 were susceptible to the virus.

At the beginning of each new school year, a new group of potential susceptible hosts came together in crowded classrooms where the chickenpox virus could be spread by respiratory transmission.





(National Epidemiological Surveillance of Infectious Diseases: Data based on the reports received before December 9, 2004)

Intectious Agents Surveillance Report



Dowell, 2001. Centers for Disease Control and Prevention

A question: Can high herd immunity prevent a disease from exceeding the normal occurrence of the disease (epidemic threshold)?

Yes, and it varies for different pathogens.

Herd immunity thresholds for several vaccine-preventable diseases

Disease	Transmission	$\underline{\mathbf{R}}_{0}$	<u>Herd immunity threshold</u>
Diphtheria	Saliva	6-7	85%
Measles	Airborne	12-18	83 - 94%
Mumps	Airborne droplet	4-7	75 - 86%
Pertussis	Airborne droplet	12-17	92 - 94%
Polio	Fecal-oral route	5-7	80 - 86%
Rubella	Airborne droplet	5-7	80 - 85%
Smallpox	Social contact	6-7	83 - 85%
Influenza	Airborne droplet	1-3	75%

 \mathbf{R}_0 is the **basic reproduction number**, the average number of secondary infectious cases that are produced by a single case in <u>completely susceptible population</u>

The different threshold values usually reflect how easily a given pathogen is transmitted to new hosts.

The threshold is commonly a function of virulence, ease of spread, population density, etc.

Vaccination programs are designed to maintain herd immunity above epidemic thresholds and prevent epidemics.



3. Certain populations of people are more susceptible to diseases due to social and economic status...

... or because of previous exposures to a disease. Example: Influenza A H1N1

Only 1% of swine flu cases in the United States are in people over the age of 65.

Why?

Antibodies against some seasonal flu strains from prior years may be active against the new H1N1 swine flu currently circulating the globe.

Blood samples taken from 359 participants in flu vaccine studies conducted from 2005 to 2009 showed that **33% of the samples from people over 60 years old had antibodies that protected against the swine flu virus**, as compared to 6%-9% of the samples from people aged 18–64 years, and none of the samples collected from children.

The flu virus - "success" of the flu virus has been evident throughout history.

An epidemic described by Hippocrates from the 5th century B.C. is thought to have been influenza.

In the last ~900 years, there are more than 300 records of flu-like epidemics.

One epidemic somewhere in the world approximately every 2.4 years.

Since the early 1700's, there have been 22 recorded influenza pandemics.

In 1918 the "Spanish Flu" killed an estimated 20 million people worldwide.

Other outbreaks in the US occurred in the late 50s and then again in the late 60s.



Rows of cots filled with patients stricken in the 1918 influenza epidemic at a naval training station in California. It is believed that the massive movements of troops and refugees during World War I helped to spread the virus. From: U.S. Naval Center.



This May 29, 1919 photograph showed rows of tents that had been set up on a lawn at Emery Hill in Lawrence, Massachusetts where victims of the 1918 influenza pandemic were treated.

Why is the flu such a "successful" disease?

1. The virus appears in many different forms, called strains.

Influenza virus is an \underline{RNA} virus - it encodes its genome in RNA rather than DNA.

The viral RNA genome is composed of 8 segments, each of which carries different genes



Structure of the Influenza A virus. The eight RNA segments carry the genes used by the virus for synthesizing various proteins. Among those proteins are the envelope glycoproteins hemagglutinin (HA) and neuraminidase (NA), which are the primary antigens recognized by the immune system.

Two viral genes, hemagglutinin (H) and neuraminidase (N) are of particular importance to both the virulence and epidemiology of influenza.

Each is located on a different segment of the genome and essential to the completion of viral replication cycle.

Hemagglutinin is the viral envelope glycoprotein - responsible for *attachment* to and *penetration* of host respiratory cells.

Neuraminidase is an envelope enzyme - allows newly formed viral particles to *leave the host cell*.

These proteins are the principle antigens recognized by the host immune system, i.e., when you are infected with influenza, i.e., you produce antibodies against the H and N antigens.

Memory to these antigens is protective against a second infection by the same strain. So, why do humans often suffer from multiple bouts of the flu in their lifetimes?

This is because different stains of influenza differ in their H and N antigens.

13 known variations of hemagglutinin and 9 neuraminidase immunological types are known to exist, all of which cause different immune responses.

Immunologic memory against one type of H or N does not guarantee protection against other types.

Flu epidemic	Causal flu antigenic strain	
1918 Spanish Flu	H1N1	
1957 Asian Flu	H2N2	
1968 Hong Kong Flu	H3N2	
2003 Bird Flu	H5N1	
2009 Swine Flu	H1N1	

2. Epidemic outbreaks of influenza occur as the virus changes genetically.

The segmented RNA genome of the flu virus has two important consequences for influenza epidemiology.

First, like all RNA viruses, influenza virus mutates rapidly, much faster than DNA viruses, bacteria, or eukaryotes.

No repair mechanisms in RNA viruses.

Allows a large number of mutations to be constantly being introduced, therefore changing the viruses genetically.

Some of the mutations might affect proteins (e.g. H and N) that are recognized by a host immune system as antigens – called **antigenic drift**.

Your immune system might not recognize the virus anymore.





Second, sometimes a cell can be infected by **two** influenza viruses at the same time.

Each of these viruses will replicate all 8 of its RNA segments, but when the new viruses are assembled, they might get some of the eight segments from one parent virus and some from the other.

This reshuffling of the genetic material can cause dramatic new viral strains to emerge for which herd immunity in humans is essentially absent.

This process is called **antigenic shift**, and is the principle reason for the emergence of new viral strains.



How antigenic shift, or reassortment, can result in novel and highly pathogenic strains of human influenza. From: Wikipedia.



Because of antigenic drift and shift, influenza is a moving target for epidemiologists.

This is the reason why this year's flu shot will likely not work for next year's flu strain.

Epidemiologist attempt to predict what strains will be in circulation during the coming year in order to develop the flu vaccine.

By giving the vaccine in the fall, protection should be high at the time the influenza season peaks



When will the next influenza pandemic strike...

A sneak preview of the recent swine flu outbreak occurred in 1997, when 18 people in Hong Kong contracted a strain of influenza A, previously known only to infect birds.

The flu spread quickly and six of the patients died because herd immunity was close to zero.

Epidemiologists determined that all of the affected humans had been in contact with infected poultry, and this particular strain of the virus was unable to spread from **human** to **human**.

As a result, health authorities in Hong Kong killed all 1.6 million chickens in the city to eliminate possible reservoirs and prevent future outbreaks of what was termed **avian influenza** or "bird flu" (H5N1 virus).

A second rash of outbreaks began in 2003 in Southeast Asia.

As of 2008, the number of human cases of avian influenza was been low, but carried an estimated mortality rate of **60%**.

So, if the H5N1 virus underwent antigenic drift or shift, then it could possibly use humans as a host and cause a very severe pandemic.



March, 2005: Women carry a basket of chickens through Hanoi, Vietnam. From; AP/Worldwide photos.

Investigating disease outbreaks

Many disease outbreaks are quite perplexing and health authorities are unable to quickly identify a specific cause.

Epidemiologists rely on a defined protocol, in which clues and evidence are gathered in a step-by-step manner, to help them understand what is happening.

1. Define the problem - develop a case definition.

A list of common symptoms, obtained from doctors and the patients themselves.

Important because now the public and other doctors know what to look for in other patients (new cases).

2. Obtain time, place, and personal characteristics of the new disease.

Provides clues to the disease's identity and identifies new patients who suffer from the same conditions.

Time characteristic - when are people are getting sick?

What are people doing at this time of year? Are they all going to the same places, eating the same foods or engaging in the same activities?



Temporal trends in legionellosis cases occurrence in the greater Philadelphia metropolitan area, 1995–2003. Disease occurs with summertime seasonality. From: Fisman et al., 2005.

Place characteristic - *where* are affected individuals living, or where they were when they became ill?

Personal characteristics - who is getting sick?

Are they mainly of one sex, age, or cultural background?

Environmental characteristics – is the outbreak timing impacted by weather, or the life cycles of animal hosts?

Once time, place, and personal characteristics are defined, the next step is to develop a **case-control study**.

This is used to determine what factor or factors link the affected individuals and distinguish them from unaffected individuals.

The affected people are the "cases", and each case is matched with a "control"--an individual that has not become ill.

The matched case and control should be as similar as possible in terms of age, sex, place of residence, and other factors that investigators consider important.

Epidemiologists look for any other difference between the cases and controls that could explain why the cases got sick and the controls did not.

Case-control studies often provide the valuable missing clue that helps epidemiologists crack a difficult mystery.

Examples:

Salmonella infections have been linked to contact with reptiles...like we discussed at the beginning of the chapter.

Potential risk factor ^a	No. of cases/total (%)	Weighted percentage of controls ^b	Univariate OR (95% CI)	Multivariate OR (95% CI) ^c	P	PAF, % (95% CI) ^d
Any reptile or amphibian contact ^e	43/453 (9)	5	1.8 (1.3-2.5)	1.6 (1.1-2.2)	.009	3.4 (1.0-5.1)
Reptile or amphibian in home						
All	33/460 (7)	4	2.1 (1.5-3.0)	1.6 (1.1-2.4)	.01	2.8 (0.8-4.2)
Snake	9/452 (2)	1	2.7 (1.4-5.0)	1.6 (0.8-3.2)	.2	
Turtle	6/454 (1)	1	1.5 (0.7-3.3)	0.8 (0.4-1.8)	.6	
Iguana	3/451 (1)	1	1.1 (0.4-3.4)	0.4 (0.1-1.4)	.1	
Non-iguana lizard	12/453 (3)	1	5.2 (3.1-8.7)	2.7 (1.5-5.0)	.001	
Amphibian	12/454 (3)	1	1.9 (1.1-3.3)	1.6 (0.9-2.8)	.1	
Touched reptile	23/453 (5)	3	1.7 (1.1-2.5)	1.6 (1.0-2.4)	.04	1.8 (0.1-3.0)
Visited place with reptile	43/456 (9)	8	1.1 (0.8-1.6)	0.9 (0.7–1.3) ^d	.7	
Chronic, non-diarrheal illness	85/441 (19)	11	1.9 (1.5-2.3)	2.5 (1.9-3.2)	<.001	11.5 (9.3-13.2)
International travel	48/462 (10)	2	7.3 (5.7-9.4)	8.4 (6.1-11.5)	<.001	9.2 (8.7-9.5)
Ate pink hamburger in restaurant	14/392 (4)	2	1.7 (1.0-2.9)	1.3 (0.8-2.3)	.3	
Ate eggs in restaurant	109/430 (25)	17	1.6 (1.3-2.0)	1.6 (1.2-1.9)	<.001	9.1 (5.0-12.3)

Table 1. Association between serogroup B or D Salmonella infection and potential risk factors.

^a Demographic factors associated with increased risk for *Salmonella* infection on multivariable analysis were female sex (OR, 1.2; 95% CI, 1.0– 1.5), household income <\$15,000 per year (OR, 1.6; 95% CI, 1.2–2.1), age (OR, 1.25; 95% CI, 1.2–1.3, for every 10 year decrease in age), season, and site. Potential risk factors not associated (*P* > .05) with increased risk for *Salmonella* infection on univariate analysis were consuming alfalfa sprouts, tomatoes, cantaloupe, apple cider, meat, poultry, hamburger at home, pink hamburger at home, ground beef, steak, roast beef, sausage, hot dogs, turkey, chicken, eggs, eggs at home, and runny eggs at home, and having a known immunosuppressive illness. PAF, population-attributable fraction.

^b Weighting based on 7618 population-based controls.

^c Controlling for age, sex, site, season, income, and other risk factors included in the final model.

^d Calculated only for major risk factors associated with illness (P<.05).

^e Reptile or amphibian in home or touched reptile.





*Per 10,000,000 population.
[†]Reptile-associated serotypes are isolates from nonhumans reported to CDC and the U.S. Department of Agriculture that are isolated from reptiles ≥50% of the time.

By using this defined protocol, epidemiologists can often determine the cause of a disease outbreak and *sometimes* be able to predict the next one.