Temporal analysis of epidemics: Disease progress over time

- **Previously:**
  - Considered concept of epidemic (MHV, chapter 1)
  - Measurement of disease intensity (chapter 2)
  - Models for relations (response:predictor) (chapter 3)
    - Introduction to model fitting (parameter estimation)

- **Now:**
  - *Dynamics* of disease intensity over time
    - Perhaps the central feature of epidemiology
  - Major emphasis on: disease progress curve
    - *Plot of disease intensity versus time*
      - The depiction of the epidemic
      - A population growth curve

- Must understand disease progress curves to understand epidemics
  - In some ways, quantifying a disease progress curve is equivalent to characterizing the disease development over time in a host population

- Investigators have many, and overlapping, objectives in quantifying disease progress curves, including:
  - **Comparison:** Are two or more curves different (statistically) in some way?
  - **Understand** how the host cultivar, pathogen genotype, and specific environment affect disease progress?
  - **Principles** of epidemics
    - Epidemic "threshold"?
    - Final "size" of the epidemic?
  - **Control** strategies
  - **Prediction**
A key premise of epidemiology is that **models** are very valuable to characterize disease progress curves.

Disease progress curves may be composed of numerous observations of disease intensity:
- e.g., weekly assessments in annual crops,
- Monthly assessments in some perennials,
- Yearly assessments for some long-lived perennials (trees).

A model consists of just a few parameters (generally):
- **Thus, the complexity of reality is reduced to a manageable number of terms**
  - Reduction in dimensionality, leading to a parsimonious description of the epidemic
- If model is appropriate, considerable understanding is possible.
Typically, we consider models for the **dynamics** of disease intensity in a host population

- Focus on **rates of change** (with time)
  - Differential equations, ...(see below & later…)

Under some circumstances, one is less interested in the dynamics

- Then, one can use more empirical statistical modeling approaches for comparing epidemics (but less so for prediction, or understanding mechanisms)
  - **Repeated-measures ANOVA**
    - Disease intensity *profiles* over time (time becomes an experimental factor)
  - ANOVA with **area under the disease progress curve (AUDPC)**
  - Others (e.g., *time to a certain disease intensity is reached*)

**Note**: We are focusing on **temporal** processes here. Thus, we have narrowed our attention of a general epidemic to just **disease over time**.

We further narrow the scope of consideration (for now):

- Disease intensity **increase** in a host population (not just change)

**How does an epidemic occur?**

- 'Contact' between pathogen 'inoculum' and disease-free host
  - Units for host can depend on the situation.
    - Entire plants, leaves, branches, fruits, roots, leaf area (e.g., # cm²)
  - For convenience of expression, we often refer to just **plants** or **individuals**
    - (even for unit areas: sites)
  - Often refer to disease-free as **'healthy'**
'Inoculum':

○ Although we often try to be very precise in definitions, here it is useful to be a little vague, to emphasis general concepts

○ "the pathogen or its parts that can cause infection" (Agrios)
  ▪ Fungal spores (conidia, sporangia, …), fungal mycelia
  ▪ Bacterial cells
  ▪ For plant viruses (and phytoplasmas), a little more difficult
    □ Virus particles (virions)
    □ But vectors (insects, etc.) often are required
      ♦ Thus, the 'inoculum' comprises the number of viruliferous vectors
  ▪ Similar comments for phytoplasmas and spiroplasmas

○ Of course, contact between inoculum and the host does not guarantee infection. This is part of what is studied in epidemiology

---

**Epidemic classification:**

Plant disease epidemics can be classified based on the 'source of the inoculum' that comes in contact with disease-free host plants

○ **First type:** Inoculum that can cause infections is produced during the current epidemic by the pathogen on or in plants (leaves, … [individuals]) that had been previously infected during the current epidemic.
  ▪ Late blight, caused by *Phytophthora infestans*
    □ Spore (sporangium) lands (deposited) on a healthy leaf [contact] ---> infection ---> ultimately a lesion ---> production of sporangia ---> dispersal of spores ---> some land on healthy leaves (or healthy parts of same leaf) [contact] ---> infection ---> … ---> …
Epidemic classification (continued) -- 1st type

- Inoculum that can cause infections is produced during the current epidemic by the pathogen on or in plants (leaves, ...[individuals]) that had been previously infected during the current epidemic.
  - New infections result in new inoculum that can result in new infections, and so on...
  - Known as polycyclic diseases, and the epidemics are known as polycyclic epidemics
    - That is, multiple disease cycles ("infection chains") -- cycles of infection (with new disease) -- during the current epidemic
    - Also known as compound-interest diseases (reason given later)
  - "Spread from plant to plant"
- In many disciplines, this is the only scenario that would be considered an epidemic (as in 'disease increase')

Epidemic classification (continued)

- Second type: inoculum that can cause infections is not produced during the current epidemic by the pathogen on or in plants (leaves, ...[individuals]) that had been previously infected during the current epidemic.
  - The pathogen may still produce inoculum during the current epidemic, but it is not available to cause infections during the epidemic
  - Example: Verticillium wilt of potatoes (& early dying), caused by Verticillium dahliae
    - Spores (microsclerotia) in soil --> root contact with spores --> infection of root, and systemic movement within plant --> production of spores in vascular tissue --> spore release only as plants decay in soil and on surface (after harvesting)
Epidemic classification (continued) -- 2nd type

- Inoculum that can cause infections is **not produced during the current epidemic by the pathogen on or in plants (leaves, ...[individuals])** that had been previously infected during the current epidemic.

  - Other scenarios also can be considered which preclude production of inoculum that can cause infections currently
    - **Insufficient time** for the 'infection cycle' to repeat within the constraints of the growing season
      - Crop matures and is harvested before new inoculum is produced
      - Long latent period (defined later)
    - Required conditions (host susceptibility, environmental conditions) only occur during narrow temporal windows -- *also resulting in insufficient time...*
      - Fusarium head blight of wheat
        - Infection only of flowers: at anthesis, or for a short time afterwards (a short time window)

- A host species may only be susceptible to infection, but new inoculum is produced only on *other* susceptible host species

- Virus vectors are only active (feeding on plants) *once* per growing season
- Vectors may not be mobile (stay at the infected plant), or cannot move very far (fungal vectors, nematode vectors)
- Vector may be only able to acquire the virus from one host plant species and transmit to another species
  - **Tomato spotted wilt of tomato (thrips vector)**
    - Adults transmit the virus, but do not acquire the virus, and thrips do not (may not) reproduce on tomato. Larvae (immature stage) of thrips acquire the virus.
- Also, the inoculum of many pathogens simply cannot move very far (from plant to plant)
  - In container media (pots in greenhouses), soil-borne inoculum cannot reach other plants
Epidemic classification (continued)--2nd type

○ Inoculum that can cause infections is not produced during the current epidemic by the pathogen on or in plants (leaves, ...[individuals]) that had been previously infected during the current epidemic.
  - Known as monocyclic diseases and the epidemics known as monocyclic epidemics
    □ There is only one infection cycle per epidemic
  - Also known as simple-interest diseases (reasons given later)
    □ "No spread from plant to plant"

○ Disease intensity can be just as high with monocyclic diseases as with polycyclic diseases -- but disease progress curves look different

○ Modeling, analysis, and ultimately control strategies differ for the two classes of diseases
  - (monocyclic vs. polycyclic)

Epidemic classification (continued)

Polycyclic <---> Monocyclic

• In reality, sometimes the distinction disappears (or is diminished)
• Depends on host units of interest
  - For entire plants, then many soil-borne pathogens do cause monocyclic diseases (no plant-to-plant spread)
  - However, there can be spread between roots (or along individual roots).
    □ Thus, inoculum produced in current epidemic can result in new root infections in same epidemic
• Time units are important, as well
  - Diseases that are monocyclic during one growing season (a single epidemic), may be polycyclic over multiple seasons
    □ If one considers the epidemic to be a long-term process, then the inoculum produced currently could ultimately cause new infections (some later year...)
• For trees, the time scale is years (decades, century)
  - Within-year process are not the determining factor

Polyetic:
Monocyclic in one season, but polycyclic over several separate seasons
Epidemic classification (continued)
Polycyclic $\leftrightarrow$ Monocyclic

Primary versus secondary infection
- **Primary infections**
  - Result from contact between host plants and inoculum produced elsewhere, or in a different epidemic.
    - Monocyclic epidemics consist only of primary infections
- **Secondary infections**
  - Any infections that ultimately result from the primary infections in the current epidemic
    - Infections resulting from inoculum produced during the current epidemic
    - Secondary infections occur only in polycyclic epidemics
    - But there are also some primary infections in polycyclic epidemics

We can now consider (simple) models of plant disease epidemics
- $y$: disease intensity (e.g., incidence or severity, including lesion counts), on a proportion scale (0-1; 0-100%)
- $Y$: disease intensity, in absolute-unit scale, not a proportion.
  - Number of diseased individuals
  - Area of lesions
- If $M$ is the size of the host population (number of plants, number of leaves, ...; total area [such as LAI]),
  - Then, $y = Y/M$
    - Could use "$N$", but want to be more general (area, not just numbers)
- $Y$ is either a continuous variable or, often, a count variable with a natural denominator (discrete)
  - For lesion counts, consider a "known" number of "infection sites" on plant units
  - If discrete, we consider it well approximated as a continuous variable (large $M$ or $N$) for model fitting (but not necessarily for theoretical developments)

Primary infections initiate polycyclic epidemics (initial infections)
Simple models of plant disease epidemics (continued)

- Time is represented by $t$
  - Can be measured in units of days (common), weeks, months, years
- For modeling purposes, $t$ is considered to be continuous
  - All time values are possible between the beginning (often $t = 0$) and end of the epidemic
- Models we consider are for a **continuous dynamic process** ("*continuous-time models*")
  - That is, disease intensity is changing **continuously** throughout the epidemic
  - In reality, $y$ does not continuously change, because conditions are only favorable for disease development during some periods.
  - However, the range of different rates are **averaged out**, giving a reasonable simplification for modeling.
  - Generally, here, consider **deterministic models**
- "*Continuous-time deterministic models*"
  - *Stochastic error term added for model fitting*

Models, at least in their original construction, involve $\frac{dy}{dt}$ as a function of $y$ and parameters

- "$\frac{dy}{dt}$" is a **rate**:
  - the change in $y$ with an infinitesimal (i.e., very small) change in $t$.
  - Referred to as the "*absolute rate of change in disease*" (or "*absolute rate of disease increase*")
- As stated previously, $\frac{dy}{dt}$ is a continuous function (because of continuous time).
  - That is, there are no **abrupt** changes from $\frac{dy}{dt} = 0$ to a positive number, and then back to 0 again, …
- The parameters in the models generally represent the effects of host, pathogen, and environment on disease development
- Useful models, which make biological sense, can be developed that involve only **one to three parameters**.
  - In principle, parameters characterize the epidemics
Models for epidemics
Models are considered in increasing order of complexity

Exponential model
- For polycyclic epidemics (early in the epidemic)
- Use goes back to at least 1790s for human population growth (a different meaning for y or Y)
- Model states that the absolute rate of disease increase is directly proportional to disease intensity
- In mathematics,
  - \( \frac{dy}{dt} = r_E \cdot y \)
    - Note: \( r_E \) is a parameter (constant)
    - Represents all the factors and variables that affect disease increase other than intensity of disease (which is explicitly in the model)
    - Units of 1/time

Exponential model: \( \frac{dy}{dt} = r_E \cdot y \)
- \( r_E \), known as the "rate parameter", represents all the factors and variables that affect disease increase other than intensity of disease (which is explicitly in the model)
- Magnitude of \( r_E \) is affected by environment, host susceptibility, and pathogen aggressiveness.
  - More directly, \( r_E \) is affected by:
    - Production of 'inoculum' by infected individuals,
      - (relative 'infectiousness' or aggressiveness of diseased individuals)
    - Probability of 'inoculum' reaching a disease-free host from an infected individual ("dispersal"),
    - Probability of 'inoculum' in contact with disease-free host causing infection ("infection efficiency"),
    - Time for newly infected individual to produce 'inoculum'
    - Time that an infected individual produces 'inoculum'
  - \( r_E \) may be known as a transmission rate in theoretical epidemiology, and the "intrinsic rate of increase" in demographics: "secondary infection rate"
Exponential model: \( \frac{dy}{dt} = r_E y \)

- We start with the model in "rate form" (differential equation form) because this is the natural approach to take for a dynamic process.
- The model states that the increase in \( y \) increases with \( y \):
  - More disease (\( \uparrow y \)) means more inoculum which means more disease... (and so on)... [the nature of polycyclic disease]
- However, one does not measure or observe or count rates of change directly:
  - Rather, one measures or observes disease intensity (\( y \)) at specific times:
    - A plot of these values (possibly with lines connecting the observed intensities) is the disease progress curve.
- To relate the exponential model to observed data, one must integrate the differential equation...
Exponential model: \( \frac{dy}{dt} = r_E \cdot y \)

Standard rules of calculus leads to:
\[ y = y_0 \cdot e^{r_E t} \]

- Where \( e \) is base of the natural log system (2.718…)
- Can also be written as: \( y = y_0 \cdot \exp(r_E t) \),
  - Where \( \exp(\cdot) \) is notation for "e raised to a power"
- Here, \( y_0 \) is a parameter (constant)
  - Known as the constant of integration
    - (Whenever one integrates an equation, one constant is added to the model—a rule of calculus)
    - The notation is intentional, because this "new constant" equals the intensity of disease at \( t = 0 \).
      - Could be written also as \( y(0) \), but \( y_0 \) is more common
- Went from one (for \( \frac{dy}{dt} \)) to two parameters (for \( y \) function)

Note that \( y_0 \) is known as initial disease intensity
- Is the result of primary infections that initiated the epidemic (from inoculum in the soil, other fields, seed infection, and so on) -- summarizes all aspects of primary infections
- Together, \( r_E \) and \( y_0 \) fully define the epidemic for the polycyclic disease (if this model is appropriate)
- A plot of \( y \) versus \( t \) shows classic exponential growth, including the growth of money at a compound-interest rate of \( r_E \).
Exponential model: \[ \frac{dy}{dt} = r_e y \rightarrow y = y_0 e^{rt} \]

- Model is **nonlinear in the parameters** (rate parameter appears in exponential expression)
- It is intrinsically linear, by taking natural logs of both sides:
  \[ \ln(y) = \ln(y_0) + rt \quad \text{or} \quad y^* = y_0^* + r_t \]
- Reminder, the "*" (asterisk) symbol" means a transformation that produces a linear model
- Note, \( y_0^* \) is a *transformed* version of the initial disease intensity (a transformation of a constant is a constant)
  - Represents here the **natural log** of initial intensity
- **Caution**: using natural logs, not base-10 logs
  - Note: \( \ln(\cdot) = \log_e(\cdot) \)
  - In SAS, natural log is \( \log(\cdot) \); in Minitab it is \( \text{loge}(\cdot) \)

In case you are not sure what transformation is being done by some computer program, see what you get for these \( Y \) values:

<table>
<thead>
<tr>
<th>( a )</th>
<th>( \log_e(a) )</th>
<th>( \log_{10}(a) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-4.60517</td>
<td>-2</td>
</tr>
<tr>
<td>0.10</td>
<td>-2.30259</td>
<td>-1</td>
</tr>
<tr>
<td>1.00</td>
<td>0.00000</td>
<td>0</td>
</tr>
<tr>
<td>10.00</td>
<td>2.30259</td>
<td>1</td>
</tr>
</tbody>
</table>
With the transformed, linear, scale, one can easily see effects of different rate parameters and initial disease intensities on disease progress.

- With the transformed scale (i.e., a linear model) one can easily calculate (estimate) $r_E$, with as few as two assessment times.
- One can easily project future levels of intensity and determine the implications of changes in initial disease intensity and the rate parameter.
- We will perform these calculations in class.

**Assignment:**
Read pages 63-84 of MHV (i.e., sections 4.1 to 4.5 of Chapter 4).
Exponential model (some calculations):

\[ \frac{dy}{dt} = r_E y \]

\[ \ln(y) = \ln(y_0) + r_E t \quad \text{or} \quad y^* = y_0^* + r_E t \]

One obtains a linear (and, in fact, a straight-line) model for transformed \( y (= y^*) \) as a function of \( t \).

Advantages:
- Easy to extrapolate and interpolate (prediction)
- Easy to "see" changes in initial \( y \) and the increase over time
- Easy to fit model (estimate parameters)

\( r_E \): all the factors that affect disease increase, other than \( y \) (SECONDARY INFECTIONS);
\( y_0 \): initial \( y \) (at \( t = 0 \)) (PRIMARY INFECTIONS)

Late Blight (continued from class #1)
Temporal analysis of epidemics (continued): Disease progress over time

Review:

○ How does an epidemic occur?
  - Contact between 'inoculum' and disease-free host individuals (broad definition of inoculum)

○ Classification:
  - Source of the inoculum
    - Current epidemic versus previous/other epidemic
    - Polycyclic versus monocyclic epidemics
    - Secondary versus primary infections

○ Models
  - Continuous time population-dynamic models ($dy/dt$) for disease intensity ($y$) as a continuous variable

NOW: Continue with basic models for disease progress

Exponential model: ($y = y_0 e^{rE t}$)

- To make it clear that this is a model for transformed $y$ as a function of time, one could write: $y(t)$ instead of just $y$, $\ln(y(t))$, $y^*(t)$:

  - $y(t) = y(0) e^{rE t}$
  - $y^*(t) = y^*(0) + rE t$

- Any point along the straight line can be the starting point for predictions
- Any two times can be identified as $t_1$ and $t_2$:

  $y(t_2) = y(t_1) e^{rE (t_2-t_1)}$

  $\ln(y(t_2)) = \ln(y(t_1)) + rE(t_2-t_1)$

  $y^*(t_2) = y^*(t_1) + rE(t_2-t_1)$
All problems here can be handled with straight-line manipulation

\[
\ln(y(t_2)) = \ln(y(t_1)) + H_E(t_2 - t_1)
\]
**Time to reach \( y = 0.05 \):**

\[ y(0.05) = -4.605 + 0.1 \cdot 0.05 \]

\( -2.996 = -4.605 + 0.1 \cdot t \)

\[ t = 16.09 \text{ days from now (or 10 + 6.09 = 26 days from start)} \]

---

**New Problem:**

**What is \( r_E \)?**

\[ r_E = \frac{\ln(y(t_2)) - \ln(y(t_1))}{t_2 - t_1} \]

\[ r_E = \frac{-2.996 - (-4.605)}{10} = 0.08 \text{ day} \]