Essential Field Epidemiology

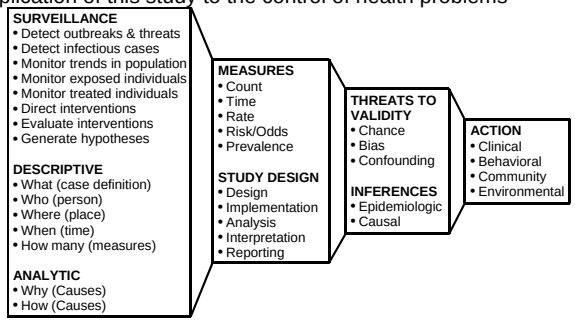
A Quick Reference Guide[1]

Tomás Aragón, Wayne Enanoria, Arthur Reingold
Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
http://www.idready.org
Version 2006-09-06

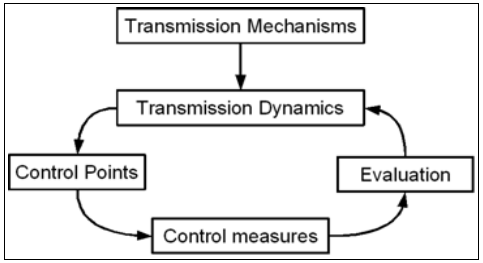
This Quick Reference Guide is based on our course. Column 1 is an overview, followed by a detailed outline. We hope you find this Guide useful. Please send us feedback on how we can improve it.

The epidemiologic approach: Steps to public health action

Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems



Concepts for the control of infectious diseases



Conducting an outbreak investigation in 7 steps

1. Case investigation
2. Cause investigation
3. Control measures
4. Conduct analytic study (if necessary)
5. Conclusions (epidemiologic and causal inferences)
6. Continue surveillance
7. Communicate findings

1 Available for download at http://www.epitools.net

Concepts for the control of infectious diseases

Transmission mechanisms

Chain model of infectious diseases

1. Causative agent
   (a) Transmissibility = P(Transmission|Exposure*)
   (b) Infectivity = P(Infection|Transmission)**
   (c) Pathogenicity = P(Disease|Infection)
   (d) Virulence = P(Complication|Disease)

* Exposure to external source; could be within species or between species (e.g., avian influenza to human)
** Or P(Infection|Colonization); i.e., source could be endogeneous

2. Reservoir (or source)
   (a) Human
      i. Symptomatic illness
      ii. Carriers
         A. Asymptomatic (no illness during infection)
         B. Incubatory (pre-illness)
         C. Convalescent (post-illness recovery)
         D. Chronic (persistent infection)
   (b) Animal (zoonoses)
   (c) Environmental
3. Portal of exit (respiratory, gastrointestinal, urogenital, wound)
4. Mode of transmission
   (a) Contact
      i. Direct contact (touch, kissing, sex)
      ii. Indirect contact (intermediate objects, fomites)
      iii. Vertical transmission (before, during, and after birth)
   (b) Respiratory droplets/secretions (cough, sneeze, runny n.)
   (c) Airborne (droplet nuclei, dust)
   (d) Vehicle-borne (ingestion, instrumentation, injection, infusion)
   (e) Vector-borne (mechanical, biologic)
5. Portal of entry (respiratory, gastrointestinal, urogenital, wound, mucosal, cutaneous, percutaneous, etc.)
6. Susceptible host

Natural history of infection/infectiousness

1. Latent period (*LP*): time from infection until infectiousness
2. Infectious period
3. Incubation period (*IP*): time from infection until symptoms
4. Asymptomatic infectiousness (*LP<IP*)
5. Symptomatic period
6. Removed states (non-infectious and/or non-symptomatic, or dead)

Convergence model of microbe-host interaction

1. Microbial agent
2. Human host
3. Genetic and biological factors
4. Physical environmental factors
5. Ecological factors
6. Social, economic, and cultural factors

Transmission dynamics

Reproductive number (erspective of an infectious case)

1. The basic reproductive number is the average number of secondary infectious cases that are produced by a single index case in a completely susceptible population in the absence of control strategies.
   $R_0 = c\ p\ d$, where *c* is the contact rate to susceptible hosts per unit time, *p* is the transmission probability per contact, and *d* is the duration of infectiousness.
2. The effective reproductive number is the average number of secondary infectious cases over time:
   $R(t) = R_0\, x(t)$, where *x* is the population fraction susceptible, *h* is the vaccine efficacy, and *f* is the fraction vaccinated.
3. The control reproductive number is the effective reproductive number in the presence of control strategies. Here's an example: $R_c(t) = R_0(1 - h\ f)$, where *h* is vaccine efficacy, and *f* is vaccine coverage (the fraction vaccinated).

Infection rate among susceptibles

From the perspective of the susceptible host:
$I(t) = c\ p\ P(t)$, where *P* is the probability the source is infectious (e.g., prevalence of infectious sources).

Generation Time (serial interval)

The time from one transmission of the infection to the next. Longer generation times allow more time for interventions to interrupt transmission.

Transmission containment (control measures)

Remember that interventions can be applied to the susceptible host and/or infectious source (See Table 1, p. 5).

1. Contact rate (*c*)
   (a) Reduce contact rate
      i. Behavior change (host and/or source)
      ii. Case finding for intervention (e.g., isolation)
      iii. Contact tracing for intervention (e.g., quarantine)
      iv. Isolation of cases
      v. Quarantine of exposed (individual, community, geographic boundary [Cordon sanitaire])
      vi. "Reverse" isolation (isolation of non-exposed)
      vii. Reduce number of infectious sources
      viii. Social distancing (school closures, restrict mass gatherings, etc.)
2. Probability source is infectious (*P*)
   (a) Reduce prevalence of infectious sources
      i. Case finding for intervention (e.g., treatment)
      ii. Identify and control infectious sources
      iii. Vaccination (if applicable)
3. Transmission probability (*p*)
   (a) Reduce infectiousness
      i. Pre/post-exposure vaccination
      ii. Treatment of case
   (b) Interrupt transmission
      i. Infection control practices
      ii. Barrier methods (e.g., masks, condoms)
      iii. Environmental disinfection

(c) Reduce susceptibility
   i. Pre-exposure vaccine, immune globulin, or drug
   ii. Post-exposure vaccine, immune globulin, or drug
   iii. Treatment of co-factor (e.g., ulcerative STD)
4. Fraction susceptible (*x*)
   (a) Increase herd immunity (consider naturally-acquired immunity, vaccine coverage, vaccine efficacy)
5. Duration of infectiousness (*d*) [see 3 (a)]

Conducting an outbreak investigation in 7 steps

Introduction

1. How are outbreaks recognized?
   (a) Practitioners (astute clinician, infection control professional, laboratory worker)
   (b) Patient or patient's family
   (c) Public health surveillance data (uncommon)
   (d) Local media (newspaper and television)
2. Reasons for investigating outbreaks
   (a) Prevent additional cases in the current outbreak
   (b) Prevent future outbreaks
   (c) Learn about a new disease
   (d) Learn something new about an old disease
   (e) Reassure the public
   (f) Minimize economic and social disruption
   (g) Teach epidemiology
3. Constraints of field outbreak investigation
   (a) Urgency to find source and prevent cases
   (b) Pressure for rapid conclusion
   (c) Statistical power often limited
   (d) Media reports may bias interviewees
   (e) Pressures because of legal & financial liability
   (f) Delays can limit human/environmental samples for testing

Step 1. Case investigation

1. Confirm outbreak (terms: cluster, outbreak, epidemic)
   (a) Confirm diagnoses
      i. Clinical syndrome
      ii. Epidemiologic risk (person, place, time)
      iii. Laboratory and diagnostic tests
   (b) Case definition
      i. Inclusion criteria
         A. Clinical criteria (symptoms, signs, & onset)
         B. Epidemiologic criteria (person, place, and time)
         C. Laboratory criteria (culture results & dates)
      ii. Case classification (suspect, probable, confirmed)
      iii. Exclusion criteria (for suspect and probable)
      iv. Operating characteristics (sensitivity, specificity, PV+)
   (c) Case line listing started (adapt case report form)
      i. Case definition data (clinical, epidemiologic [person, place, time], laboratory, and exclusion criteria)
      ii. Established or suspected risk factors (causes)
      iii. Demographic information
      iv. Contact information
   (d) Case finding (why)
      i. Determine extent of outbreak
      ii. Determine population at risk
      iii. Identify secondary cases (suggests person-to-person)
      iv. Direct control measures at population at risk
      v. Identify infectious sources
      vi. Identify contacts to infectious cases
   (e) Case finding (how to)
      i. Passive surveillance
         A. Enhance passive
         B. Stimulated passive
      ii. Active surveillance (may involve surveys)
      iii. Media outreach
   (f) Cases interviews
      i. Known and likely exposures/causes (use case report forms, if available)
      ii. Hypothesis generating questions (open ended)
      iii. Case characteristics (sufficient to determine case status [Clinical, Epidemiologic (person, place, time), Laboratory])
      iv. Document medical care and treatments [name & phone number if you need to contact doctor])
      v. Document disposition, complications, death
      vi. Demographic and contact information
   (g) Complete line listing
      i. May include at-risk subjects & non-cases
      ii. May be sufficient to describe cases & test hypotheses
      iii. May require survey
   (h) Case descriptive epidemiology (counts, times, rates, risks)
      i. Orient cases by person, by place, and by time (epidemic curve). For example, an epidemic:
         A. Can suggest agent or incubation period
         B. Can suggest magnitude and time course
         C. Can suggest pattern of spread: Common source (point, intermittent, continuous); Propagated (person-to-person spread); Time limited vs. ongoing outbreak
         D. Can show where we are in course of epidemic
         E. Can be used for evaluation/monitoring
         F. Can provide additional clues (outliers, etc)
   (i) Establish baseline occurrence of cases
   (j) Rule out alternative explanations
      i. Chance: Random error
         A. Confidence interval (precision)
         B. P value (observed vs expected)
      ii. Bias: Systematic error
         A. Selection bias
         B. Measurement bias (information bias)
      iii. Confounding
2. Establish preliminary causal hypotheses
   (a) Clues from clinical syndrome
   (b) Clues from etiologic agent, if known
   (c) Clues from case interviews (have in common?)
   (d) Clues from existing knowledge base (see Step 2: Cause investigation)

Step 2. Cause investigation

1. Systematically review known causal factors (Transmission mechanisms and transmission dynamics)
   How is this organized?
   (a) Epidemiologic/clinical investigation
   (b) Environmental investigation
   (c) Laboratory investigation
   (d) Veterinary/Vectorborne
   (e) Forensics/Law enforcement
2. Prioritize likely causes to guide control measures (Step 3)
3. Generate testable hypotheses to conduct analytic study (Step 4) if cause remains unknown or control measure not working

Step 3. Control measures

Design from the consideration of control points, strategies, and interventions (see p. 1)

Step 4. Conduct analytic study (if necessary)

1. Design (Study protocol)
   (a) Primary question(s)
   (b) Significance (importance, background)
   (c) Study design (address confounding)
      i. Time frame (prospective vs. retrospective)
      ii. Design approach (cohort vs. case-control)
   (d) Subjects (address selection bias)
      i. Selection criteria
         A. Inclusion criteria (person, place, time)
         B. Exclusion criteria (to enhance validity)
      ii. Sampling design (simple random, stratified, clustered)
      iii. Recruitment plan
   (e) Variables (address measurement bias, confounding)
      i. Outcome variables (case definitions)
      ii. Exposure variables (causes: exposures, risk factors)
      iii. Confounders and Effect modifiers
      iv. Demographic and contact information
   (f) Statistical issues (address random error, power, confounding)
      i. Hypotheses
      ii. Sample size
      iii. Analytic approach
2. Implement (Operations manual, training)
3. Analyze and Interpret

Step 5. Conclusions (epidemiologic & causal inferences)

1. Epidemiologic inference (address threats to validity)
   (a) Descriptive epidemiology
      i. For a specific outcome, make comparisons and note differences across one or more dimensions (e.g. time series)
      ii. Seek known explanations to account for observed differences (rule out chance, bias, confounding as explanations)
      iii. Draw conclusions from descriptive study (epidemiologic inference #1)
   (b) Analytic epidemiology
      i. Generate hypotheses from descriptive studies
      ii. Design and conduct studies to test hypotheses (control for chance, bias, confounding)

iii. Draw conclusions from analytic study (epidemiologic inference #2)

2. Causal inference
   (a) Key characteristics of causes
      i. Essential attributes include association, time order, and directionality
      ii. Causes include microbial agent, host, and environmental factors (see convergence model)
      iii. Causes include active and static conditions
      iv. Causes many be either positive (presence induces diseases) or negative (absence induces disease)
   (b) Models of causation
      i. Sufficient-component cause model (Rothman)
         A. Component causes
         B. Necessary component (without it, outcome never occurs)
         C. Sufficient component (with it, outcome always occurs)

Step 6. Continue surveillance

1. Goals of surveillance that triggers a public health action
   (a) Detect outbreaks & public health threats
   (b) Detect infectious cases (case finding)
   (c) Monitor trends in a target population
   (d) Monitor exposed individuals for symptoms
   (e) Monitor treated individuals for complications
   (f) Direct & evaluate public health interventions
   (g) Generate hypotheses for further evaluation
2. Establishing a surveillance system
   (a) Define purpose and goals
   (b) Assign personnel
   (c) Develop case definition
   (d) Review elements of a good system
   (e) Get the system started
   (f) Analyze, interpret, and disseminate findings
   (g) Evaluate system
3. Elements of a good surveillance system
   (a) Simplicity (structure and ease of operation)
   (b) Flexibility (adapt to changing information needs)
   (c) Data quality (completeness and validity)
   (d) Acceptability (willingness of reporting entities)
   (e) Sensitivity and specificity (case definition, detect outbreak)
   (f) Predictive value positive (PV+) or negative (PV-)
   (g) Representativeness (of target population over time)
   (h) Timeliness
   (i) Stability (reliability and availability)

Step 7. Communicate findings

1. Communicate preliminary assessments and recommendations (letter, memo)
2. Prepare interim/final report (here is a template)
   (a) Background
      i. Detection, reporting, and notifications
      ii. Preliminary findings
      iii. Preliminary control measures
      iv. Relevant questions and importance (cite literature)
      v. Primary question(s) addressed by this investigation +/- study
   (b) Methods
      i. Epidemiologic investigation
         A. Case-definition
         B. Case finding
         C. Study design
         D. Sampling and subject recruitment
         E. Measurements
         F. Statistical methods & analytic approach
      ii. Environmental investigation
      iii. Laboratory investigation
   (c) Results
      i. Epidemiologic investigation
         A. Descriptive epidemiology (epidemic curve)
         B. Analytic study results
      ii. Environmental investigation
      iii. Laboratory investigation
   (d) Discussion (summarize inferences and supporting evidence [from literature review] , strengths, limitations)
   (e) Recommendations
      i. Prevention and control
      ii. Further investigations
   (f) References
   (g) Figures and Tables
   (h) Appendices (for quality assurance, educational, and training purposes)
      i. Survey instrument
      ii. Data file description and repository
3. Prepare manuscript (Introduction; Methods; Results; Discussion; References; Figures and Tables, Appendixes)
4. Risk communication strategy (what to say)
5. Media communication strategy (how to say it)

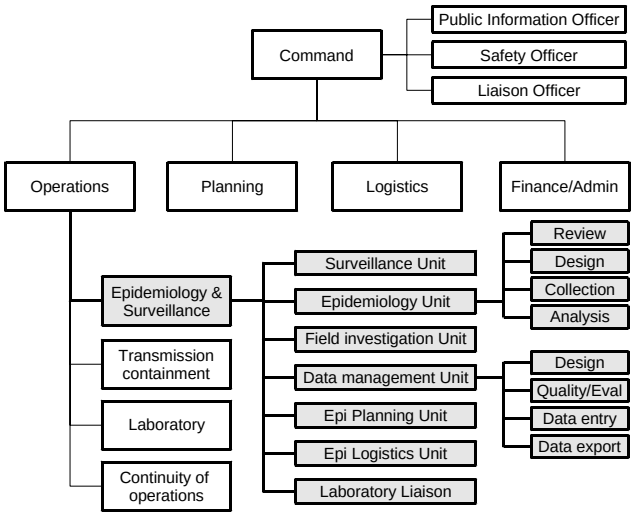


Figure 1: Incident command structure for an infectious disease emergency with details for epidemiology and surveillance functions

Operational aspects of field epidemiology

Figure 1 displays the incident command structure (ICS) for an Infectious Disease Emergency Response (IDER). Under IDER Operations is the Epidemiology and Surveillance (Epi) Group. This structure is a suggested approach. An ICS structure consists of Command, Operations, Planning, Logistics, and Finance/Administration Sections. ICS is based on functions, not agency or organizational hierarchy. Conceptually, these functions apply at all levels and are useful for organizing your thoughts and actions.

Command ("incident commander"") sets and disseminates the operations objectives to achieve the mission. The Operations function carries out these objectives. In this scenario, the county health officer might be the IC, and communicable disease controller might be the Epi Group leader. All other functions support Operations. The Planning function studies, anticipates, and communicates Operations needs and demands. Planning also develops the incident Action Plan (IAP) for the operational period. The IAP is approved by Command. The Logistics Section provides logistical support (human and material resources). Finance/Admin tracks costs and provides administrative support.

The Command Staff functions includes Public Information Officer(s), Liaison Officer(s), and Safety Officer(s). The PIO coordinates internal and external communications. The LO coordinates with assisting agencies, and the SO assures the safety of all response personnel.

The Epi Group can be under Operations or Planning. For example, they may be part of Operations, although their analyses will help inform and guide Planning. Below is a suggested Epi Group emergency operations structure. The structure is designed to scale up to potentially involve many team members implementing discrete response tasks.

Case finding for surveillance purposes belongs in the Epi Group. In contrast, case finding for isolation, contact tracing for quarantine, and case management should be under Containment activities, yet need to be monitored.

In a public health emergency, the health department operations center (DOC) will have an ICS structure similar to Figure 1. Under DOC Operations (not shown), it may have several branches including Infectious Disease Emergency Response, Environmental Health Emergency Response, Mental Health Emergency Response, Mass Casualty Care, etc.

Epidemiology & Surveillance Group operations

A. Surveillance Unit
   I. Passive and Active Surveillance [Review, Design, Collect, Analyze, Preliminary Report]*
B. Epidemiology Unit
   I. Descriptive and Analytic Epidemiology [Review, Design, Collect, Analyze, Preliminary Report]
   II. Outbreak Investigation [Review, Design, Collect, Analyze, Preliminary Report]

C. Field Investigation Unit [Field investigators, Biologic collection, Environmental collection]-Do field activities for EIS Group
D. Data Management Unit [Design, Quality Control/Evaluation, Data Entry, Data Export]
E. Epi Planning Unit [Review, Interim and Final Reports, Documentation]
F. Epi Logistics Unit [Orientation, "Just in Time" Training, Communications, IS/IT Support]
G. Laboratory Liaison Unit (Collection, Transport, Prioritization)

*The Teams are in brackets. These are discrete functional tasks that can run in parallel to save time.

ABCs of ICS: Manage by Objectives for Operational Period

A) **Assess** command, control, communications, and resources
B) **Begin** planning and setting priorities for operational period
C) **Communicate** objectives and delegate tasks
D) **Document and track** all:
   (a) Assignments of goals, objectives, or tasks (in and out)
   (b) Rsource requests (in and out)
E) **Evaluate & summarize**
   (a) Progress on meeting objectives, and
   (b) Recommendations for command and next shift

Job Action Sheets

Job action sheets provide immediate, intermediate, or extended actions to be taken by staff. Some actions apply to all personnel, to subdivisionS (e.g., Epi Group), or to specific Units (e.g., Surveillance Unit). Here's a sample JAS:

Job name: Passive Surveillance leader — Appointed By:
Functional Area:: — Reports To:
Assigned Location: — Direct Reports:

Area of responsibility

The Passive Surveillance Leader supervises, assists, and backs up Passive Surveillance Review, Design, Data Collection, and Data Analysis Team Leaders.

Specific Action Items

Immediate

1. Read this entire Job Action Sheet
2. Highlight or circle your position on the organizational chart
3. Obtain briefing from supervisor and participate in meeting to formulate objectives for operational period
4. Check communications equipment for this position
5. Assess equipment needs for position (copier, computer, etc.)

Intermediate

1. Set up passive surveillance system
2. etc.

Extended

1. Observe all staff for signs of stress or inappropriate behavior
2. Plan for an extended activation and determine staffing needs.
3. Prepare an end of shift report and brief oncoming relief.
4. Ensure that all Documentation Forms and other required documentation are kept and submitted afterwards
5. Keep copies of all logs, reports, messages, and other documents used and received while in the DOC.

Basic formulas

Operating characteristics

		Disease: Yes	Disease: No
Meets case definition	Yes	True positive	False positive
	No	False negative	True negative

$$Sensitivity = P(T+|D+) = \frac{TP}{TP+FN}$$

$$Specificity = P(T-|D-) = \frac{TN}{TN+FP}$$

$$PV+ = \frac{(Sensitiviy)(Prior\ Probability)}{(Sensitiviy)(Prior\ Prob)+(1-Specificity)(1-Prior\ Prob)}$$

Take home points

1. Sensitive ("loose") case definition (low FN) will pick up more cases (at the expense of specificity [more FP]);
2. Use sensitive case definition if benefits of picking up mild cases outweigh the burden of picking up false positives.
3. Sensitive case definition (low FN) better for ruling out a case (because you have more confidence in a negative result); ("SnOut")
4. Specific ("tight") case definition reduces misclassification in analytic studies, improving the validity of findings;
5. Specific case definition (low FP) better for ruling in a case (because you have more confidence in a positive result); ("SpIn")
6. Predictive value positive [P(D+|T+)] is a function of sensitivity, specificity, and prior probability. Prior probability is affected by disease prevalence and other relevant information.

Cohort analysis for binomial data

		Disease: Yes	Disease: No	
Exposure	Yes	a	b	N_1
	No	c	d	N_0

$$Risk\ Ratio = RR = \frac{R_1}{R_0} = \frac{a/N_1}{c/N_0}$$

$$Odds\ Ratio = OR = \frac{R_1/(1-R_1)}{R_0/(1-R_0)} = \frac{a/b}{c/d} = \frac{a\,d}{b\,c}$$

Reporting descriptive results

Variables (e.g. Demographics)	Number	Proportion
Variable 1 (e.g. Gender)		
Male	x	x/(x+y+z)
Female	y	y/(x+y+z)
Transgender	z	z/(x+y+z)

Reporting analytic results in a table

	Exposed				Not exposed					
RF	Ill	Not ill	Total	AP	Ill	Not ill	Total	AP	RR	p
RF_1	a_1	b_1	N_{11}	$\frac{a_1}{N_{11}}$	c_1	d_1	N_{01}	$\frac{c_1}{N_{01}}$	$\frac{a_1/N_{11}}{c_1/N_{01}}$	
RF_2	a_2	b_2	N_{12}	$\frac{a_2}{N_{12}}$	c_2	d_2	N_{02}	$\frac{c_2}{N_{02}}$	$\frac{a_2/N_{12}}{c_2/N_{02}}$	

AP = attack proportion or risk (sometimes referred to as "attack rate" which is only accurate if it is actually a rate)

$RR = Risk\ Ratio$

Cohort analysis for person-time data

		Disease: Yes	Disease: No	
Exposure	Yes	a	b	PT_1
	No	c	d	PT_0

$$Rate\ Ratio = \frac{r_1}{r_0} = \frac{a/PT_1}{c/PT_0}$$

Case-control analysis

		Disease: Yes	Disease: No
Exposure	Yes	a	b
	No	c	d
		M_1	M_0

$$Odds\ Ratio = \frac{O_1}{O_0} = \frac{a/c}{b/d} = \frac{a\,d}{b\,c}$$

Reporting analytic results in a table

	Cases				Controls					
RF	Exp	Not Exp	Total	Frac Exp	Exp	Not Exp	Total	Frac Exp	OR	p
RF_1	a_1	b_1	N_{11}	$\frac{a_1}{N_{11}}$	c_1	d_1	N_{01}	$\frac{c_1}{N_{01}}$	$\frac{a_1 d_1}{b_1 c_1}$	
RF_2	a_2	b_2	N_{12}	$\frac{a_2}{N_{12}}$	c_2	d_2	N_{02}	$\frac{c_2}{N_{02}}$	$\frac{a_2 d_2}{b_2 c_2}$	

Exp = exposed; Frac Exp = Fraction exposed; OR = Odds Ratio

Designing infectious disease control measures

Table 1: Control measures are developed by selecting interventions that address strategies

Control strategies						
Reduce contact rate	***Reduce Pr{source is infectious}***	***Reduce infectiousness***	***Interrupt transmission***	***Reduce susceptibility***	***Increase herd immunity***	These six control strategies act through the five control points: contact rate (c), probability source is infectious (P), transmission probability (p), fraction susceptible (x), and duration of infectiousness (d). The control points come from two equations: • Reproductive number: $R_c(t)=R_0(1-h\,f)=c\,p\,d\,(1-h\,f)$, where R_c = control reproductive number, f = fraction vaccinated, h = vaccine efficacy • Infection rate among susceptibles: $I(t)=c\,p\,P(t)$ The SIR approach: • Surveillance: Detection and monitoring (nonexposed, exposed, cases; target population) • Investigation: Epidemiologic/clinical, Laboratory, Environmental, Veterinary/Vectorborne, Forensics/Law enforcement • Response: Clinical, Behavioral, Community, Environmental, Societal, National, International
1	***2***	***3***	***4***	***5***	***6***	***Control interventions***
X						Change behavior of susceptible host
X						Change behavior of infectious source
A	B					Control (A) and reduce number (B) of infectious sources
A	A	B				**Case finding** for intervention (A=isolation; B=treatment; C=surveillance)
A		B		B		**Contact tracing** for intervention (A=quarantine; B=post-exposure prophylaxis; C=surveillance)
A	A	B				**Case management** (A=isolation and/or restriction/clearance; B=treatment)
X						**Quarantine** of exposed (individual, community, geographic boundary [Cordon sanitaire], work q.)
X						"Reverse" isolation (isolation of non-exposed)
X						"**Social distancing**" • Closing of schools and workplaces • Cancellation of mass gatherings and public transportation • Community-based confinement (within homes) of asymptomatic persons • Border controls • Travel restrictions
	X			X	X	**Targeted vaccination** vs. **mass vaccination**
	X			X	X	**Targeted prophylaxis** vs. **mass prophylaxis**
	x			X	A	**Pre-exposure prophylaxis** (A=vaccine, immune globulin, or drug)
	x			X	A	**Post-exposure prophylaxis** (A=vaccine, immune globulin, or drug)
				X		Treatment of co-factor (e.g., ulcerative STD)
			X			**Infection control** practices (hospital and community) • Source (hand hygiene, cough etiquette, respiratory hygiene, face mask) • Susceptible (hand hygiene, personal protective equipment) • Precautions (standard, contact, droplet, airborne)
			X			Barrier methods (e.g., condoms, face masks, gowns, etc.)
			X			**Environmental** measures, including disinfection

Types of investigative questions

These questions were generated by systematically reviewing the components of each conceptual model and equation. Additionally, for each of these you could pose descriptive questions: what, who, where, when, how many? and analytic questions: why and how? There are no shortage of investigative questions. The idea is to stay focused on those questions that matter most.

These questions are part of the cause investigation. Keep in mind who is most qualified to conduct components of the investigation. The investigation might be organized into epidemiologic/clinical, environmental, laboratory, and criminal investigations.

What important questions are missing? Send feedback to Tomás Aragón at aragon@berkeley.edu.

Transmission mechanisms

Chain model

1. What is the agent?
2. What is the natural reservoir?
3. How long does microbe survive outside of reservoir? (e.g., environment)
4. What are the non-reservoir sources?
5. What are the portals of exit?
6. What are the modes of transmission?
7. What are the portals of entry?
8. What is the role of host susceptibility play? (age, immunocompromised, etc.)

Natural history

1. How long is the incubation period?
2. Is the latent period shorter that the incubation period (asymptomatic infectiousness)? If yes, what is the duration?
3. How long is the infectious period?
4. Questions about clinical syndrome:
   (a) What are the clinical symptoms and stages?
   (b) Is there a well-defined prodrome?
   (c) What is the natural history untreated?
   (d) What is the natural history under treatment?
   (e) What proportion never develop symptoms?
   (f) What proportion develop mild, nonspecific symptoms?
   (g) What proportion develop chronic infection?
   (h) What proportion develop chronic silent disease?
   (i) What proportion develop chronic symptomatic disease?
   (j) What proportion develop complications (acute and long term)?
   (k) What proportion die from the disease (acute and long term)?
   (l) What proportion develop long lasting immunity?
   (m) What proportion become silent carriers?
   (n) What are the case complication rates, including death rates?

Convergence model

1. Any environmental risk factors for transmission?
2. Any ecological factors affecting transmission?
3. Any social factors affecting transmission?
4. Any political factors affecting transmission?
5. Any economic factors affecting transmission?
6. Any cultural factors affecting transmission?
7. Any host biological or genetic risk/preventive factors for transmission?
8. Any microbe biological or genetic risk/preventive factors for transmission?

Transmission dynamics

Reproductive number for infectious cases

$$R_c(t) = R_0(1 - h\,f) = c\,p\,d\,(1 - h\,f)$$

1. What is the basic reproductive number?
2. What constitutes meaningful contact permitting transmission?
3. What is the contact rate between infectious sources and susceptible hosts?
4. What is the variance of the contact rate?
5. Is contact homogeneous or heterogeneous?
6. What factors modify infectious sources contacting susceptible hosts? (modifies contact rate)
7. What is the transmission probability per meaningful contact?
8. How does source infectiousness modify the transmission probability?
9. What non-host, non-case factors modify the transmission probability? (ambient temperature and humidity, etc.)
10. What is the duration of infectiousness?
11. What factors modify the contact rate, transmission probability or duration of infectiousness?
12. What factors interrupt transmission?

Infection rate among susceptibles

$$I(t) = c\,p\,P(t)$$

1. What factors modify susceptible hosts contacting infectious cases? (modifies contact rate)
2. How does host susceptibility modify the transmission probability?
3. What is the probability that a source is infectious?
4. What factors modify that the source is infectious? (age, immunocompromised, failure of therapy, antimicrobial resistance,)
5. What is the prevalence of infectious sources?
6. What is the infection rate?

Generation time

1. What is the generation time?

Evaluating controls measures (strategy + intervention)

For all potential control measures ask the following:

1. Can it work? (efficacy: e.g., randomized control trials)
2. Does it work? (effectiveness)
3. Which is better? (efficiency: e.g., cost-effectiveness)
4. What modifiable factors are associated with better or worse outcomes?

Conducting an outbreak investigation

1. Case investigation
   (a) What is the sensitivity and specificity of the case definition?
2. Cause investigation
   (a) See transmission mechanisms and dynamics?
   (b) What causal pathways are operating?
3. Control measures (see above)
   (a) What is the efficacy?
   (b) What is effectiveness?
   (c) What is the cost-effectiveness?
4. Conduct analytic study (if necessary)
5. Conclusions (epidemiologic and causal inferences)
   (a) What are the potential biases?
   (b) What are the potential confounding factors?
6. Continue surveillance
   (a) What is the completeness of reporting?
   (b) What is the accuracy of reporting?
   (c) What is the timeliness in reporting?
   (d) What are barriers to reporting?
   (e) What strategies enhance reporting?
   (f) What is the sensitivity, specificity, or predictive value of an outbreak or aberration detection system?
7. Communicate findings

Bibliography

1. Aschengrau A, Seage GR. Essential of Epidemiology in Public Health. Jones and Bartlett Publishers, 2003, 1st Edition
2. Dr. Janet Mohle-Boetani, Medical Epidemiologist, Infectious Diseases Branch, Division of Communicable Disease Control, California Department of Health Services (Personal communication)
3. Gregg MB. Field Epidemiology. Oxford University Press, 2002, 2nd Edition
4. Dr. Sandra Huang, Director, Communicable Disease Control Unit, Community Health Epidemiology and Disease Control, San Francisco Department of Public Health (Personal communication)
5. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Designing Clinical Research: An Epidemiologic Approach. Lippincott Williams & Wilkins, 2001, 2nd Edition
6. Institute of Medicine. Microbial Threats to Health: Emergence, Detection, and Response. National Academy Press 2003
7. Reingold A. If Syndromic Surveillance Is the Answer, What Is the Question? Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 2003; 1(2):1
8. Rothman KJ. Epidemiology: An Introduction. Oxford University Press 2002, 1st Edition
9. Thomas JC, Weber DJ. Epidemiologic Methods for the Study of Infectious Diseases. Oxford University Press, 2001, 1st edition

Analysis template in R

We provide an analysis template using R (and the EpiTools package)[2]. Examples involve human West Nile virus surveillance, other data sets (AIDS, measles, hepatitis B, etc.). Analogous templates will be prepared for Epi Info 6, Stata, and SAS.

#Read data

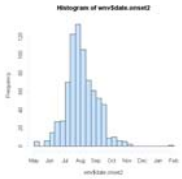
```
# Human West Nile virus disease surveillance, California, 2004
wnv <-read.table("http://www.medepi.net/data/wnv/wnv2004raw.txt",
                 sep = ",", header = TRUE, na.strings = ".")
str(wnv)  #display data set structure
head(wnv) #display first 6 lines
edit(wnv) #browse data frame
fix(wnv)  #browse with ability to edit (be careful!!!)
```

#Convert non-standard dates to Julian dates

```
wnv$date.onset2 <- as.Date(wnv$date.onset, format="%m/%d/%Y")
wnv$date.tested2 <- as.Date(wnv$date.tested, format="%m/%d/%Y")
```

#Display histogram of onset dates (epidemic curve)

```
hist(wnv$date.onset2, breaks= 26, freq=TRUE, col="slategray1")
```



#Describe a continuous variable (e.g., age)

```
summary(wnv$age) # no standard deviation provided
range(wnv$age, na.rm=TRUE); mean(wnv$age, na.rm=TRUE)
median(wnv$age, na.rm=TRUE); sd(wnv$age, na.rm=TRUE)
```

#Describe continuous by categorical variable

```
tapply(wnv$age, wnv$sex,  mean, na.rm = TRUE)
tapply(wnv$age, wnv$county,  mean, na.rm = TRUE)
```

#Display a continuous variable

```
hist(wnv$age, xlab="x", ylab="y", main="title", col="skyblue")
```

#Describe a categorical variable (e.g., sex)

```
sex.tab <- xtabs(~sex, data = wnv)
sex.dist <- prop.table(sex.tab)
cbind(sex.tab, sex.dist)
```

#Display a categorical variable (e.g. sex)

```
barplot(sex.tab, col="pink", ylab="Frequency", main="title")
```

#Re-code continuous variable to categorical (e.g., age)

```
wnv$age3 <- cut(wnv$age, breaks=c(0,45,65,100), right=FALSE)
age3.tab <- xtabs(~age3, data = wnv)
age3.dist <- prop.table(age3.tab)
cbind(age3.tab, age3.dist)
```

#Describe two categorical variables (e.g. sex and age)

```
sexage <- xtabs(~sex + age3, data = wnv)
sexage
prop.table(sexage)    #joint distribution
prop.table(sexage, 1) #row distribution
prop.table(sexage, 2) #column distribution
```

2 R is a freely available, open source program for statistical computing and graphics (http://www.r-project.org). EpiTools is an R package customized for epidemiology (http://www.epitools.net).

#Plot age vs sex distribution

```
barplot(sexage, legend.text=TRUE,
        xlab="Age", ylab="Frequency", main="title")
barplot(sexage, legend.text=TRUE, beside=TRUE,
        xlab="Age", ylab="Frequency", main="title")
barplot(t(sexage), legend.text=TRUE, ylim=c(0, 650),
        xlab="Sex", ylab="Frequency", main="title")
barplot(t(sexage), legend.text=TRUE, beside=TRUE, ylim=c(0, 300),
        xlab="Sex", ylab="Frequency", main="title")
```

#Hypothesis testing using 2-way contingency tables

```
# From the main menu select Packages > Install Package(s)...
# Select CRAN mirror near you. Select 'epitools' package
library(epitools) #load 'epitools'; only needed once per session
tab.age3 <- xtabs(~age3 + death, data = wnv)
epitab(tab.age3)          #default is odds ratio
epitab(tab.age3, method = "riskratio")
prop.table(tab.age3, 1)   #display row distribution (2=column)
prop.test(tab.age3[,2:1]) #remember to reverse columns
chisq.test(tab.age3)      #Chi-square test
fisher.test(tab.age3)     #Fisher exact test
```

#Graphical display of epidemiologic data

#Histogram (continuous numbers or date objects)

```
hist(wnv$age, xlab="x", ylab="y", main="title", col="skyblue")
hist(wnv$date.onset2, breaks= 26, freq=TRUE, col="slategray1")
```

#Bar chart (categorical variable)

```
barplot(table(wnv$sex), col="skyblue", xlab="Sex", ylab="Freq",
        main="title", legend = TRUE, ylim=c(0,600))
```

#Stacked bar chart (2 or more categorical variables)

```
barplot(table(wnv$sex, wnv$age3), col=c("blue","green"),
        xlab="Sex", ylab="Freq", main="WNV Disease, Sex by Age",
        legend = TRUE, ylim=c(0,400))
```

#Group bar chart (2 or more categorical variables)

```
barplot(table(wnv$sex, wnv$age3), beside=TRUE, xlab="Sex",
        ylab="Freq", main="Sex by Age", col=c("blue","green"),
        legend = TRUE, ylim=c(0,250))
```

#Proportion bar chart (2 or more categorical variables)

```
sexage <- xtabs(~sex + age3, data = wnv)
barplot(prop.table(sexage, 2), xlab="Sex", ylab="Proportion",
        main="WNV Disease, Sex by Age", col=c("blue","green"),
        legend = TRUE, ylim=c(0,1.2))
```

#Time series (single *x* values vs. single *y* values)

```
# Use United States measles data
measles <- read.table("http://www.medepi.net/data/measles.txt",
             sep="", header=TRUE)
str(measles); head(measles)
plot(measles$year, measles$cases, type="l", lwd=2, col="navy")
plot(measles$year, measles$cases, type="l", lwd=2, log="y")
```

#Time series (multiple *x* values vs multiple *y* values)

```
# Use U.S. AIDS and hepatitis B surveillance data
aids <- read.table("http://www.medepi.net/data/aids.txt",
        sep="", header=TRUE, na.strings=".")
hepb <- read.table("http://www.medepi.net/data/hepb.txt",
        sep="", header=TRUE)
years <- cbind(aids$year, hepb$year)
cases <- cbind(aids$cases, hepb$cases)
matplot(years, cases, type="l", lwd=2, col=1:2, main="title")
legend(x=1980, y=80000, legend= c("AIDS","Hepatitis B"),
       lty=1:2, col=1:2, lwd=2)
```

#Working with dates and times

```
# Convert non-standard dates to standard Julian dates
dates <- c("11-02-1959","1959Nov02","November 2, 1959")
jdates<-as.Date(dates, format=c("%m-%d-%Y","%Y%b%d","%B %d, %Y"))
jdates; julian(jdates)
# Converting non-standard dates and times to R date-time object
dtim <- c("4/19/1940 12:30 AM", "4/18/1940 9:45 PM")
std.dt <- strptime(dtim, format="%m/%d/%Y %I:%M %p")
std.dt
# Try 'help(strptime)' to see all format options
```

#Manually creating an epidemic curve

Single variable

```
labs <- c("Sun", "Mon","Tue","Wed","Thu","Fri", "Sat")
cases <- c(0, 25, 15, 5, 10, 20, 0)
names(cases) <- labs
barplot(cases, space=0, col="skyblue", xlab="Day", ylab="Cases",
        main="Title")
```

Single variable—Change x-axis labels to perpendicular

```
xv <- barplot(cases, space=0, col="red", xlab="Day",
            ylab="Cases", main="Title", axisnames=FALSE)
axis(side=1, at=xv, labels=labs, las=2)
```

Stratified by second variable

```
male.cases <- c(0, 15, 10, 3, 5, 5, 0)
female.cases <- c(0, 10, 5, 2, 5, 15, 0)
cases2 <- rbind(Male = male.cases, Female = female.cases)
colnames(cases2) <- labs
xv <- barplot(cases2, space=0, col=c("blue", "green"),
             xlab="Day", ylab="Cases", main="Title",
             axisnames=FALSE, legend.text=TRUE, ylim=c(0, 30))
axis(side=1, at=xv, labels=labs, las=2)
```

#Running batch jobs

```
source("c:/myoutbreak/job01.R") #run program file called job01.R
```

#Creating output log files

```
# From within job01.R program file
x <- 1:5; y <- x^2
# sink printed objects to log file
sink("c:/temp/job.log"); print(x); sink()
# capture output w/o requiring print command
capture.output(cbind(x, y),
               file="c:/temp/job.log", append=TRUE)
```

#Multivariable analysis (not part of course)

##Logistic regression (binomial data: cohort, case-control)

```
# WNV data using 'age3' variable created previously
mod1 <- glm(death ~ age3, family=binomial, data=wnv)
summary(mod1) #full results
exp(mod1$coef) #calculate odds ratio
mod2 <- glm(death ~ age3 + sex, family=binomial, data=wnv)
summary(mod2) #full results
exp(mod2$coef) #calculate odds ratio
```

#Conditional logistic regression (matched case-control)

```
# Case-control study of myocardial infarction (Kleinbaum 2002)
# One case matched to 2 controls on age, race, and sex
library(survival) #load survival package
chd <- read.table("http://www.medepi.net/data/chd.txt", sep=",",
header=TRUE)
head(chd)
chd$mi2 <- ifelse(chd$mi=="Yes", 1, 0) #re-code case status
mod1 <- clogit(mi2~smk+strata(match), data=chd)
summary(mod1)
mod2 <- clogit(mi2~smk+sbp+strata(match), data=chd)
summary(mod2)
mod3 <- clogit(mi2~smk+sbp+ecg+strata(match), data=chd)
summary(mod3)
anova(mod1,mod2,mod3, test="Chisq") #compare nested models
```