Ph.D. Comprehensive Examination in Epidemiology

2024 Study Guide

This document contains methodologic concepts in epidemiology and biostatistics that students are expected to understand for the comprehensive exam in epidemiology at the Dalla Lana School of Public Health, University of Toronto.

"Hey, this is a marathon, not a sprint."

Epidemiology Topics

Person, place, and time	• Descriptive epidemiology
	  ○ Data summaries
	  ○ Descriptive statistics
	  ▪ Correlation
	  ▪ Graphing (histograms, boxplots, bar graphs, scatterplots)
	  ▪ Measures of distribution (mean, median, standard deviation)
	  ▪ Z-scores
	• Age, period, cohort effects
	• Calculating person-years
	• Life table methods
	• Target populations
	• Types of populations
	  ○ Closed and open populations
	  ○ Steady state
	  ○ Source population
	  ○ Study population
	• Sampling frames
	  ○ Simple random sample
	  ○ Systematic sample
	  ○ Multistage samples
	  ○ Probability samples
	• Social determinants of health

Measures of Disease Occurrence	• Incidence
	  ○ Incidence density (risk)
	  ○ Cumulative incidence
	  ○ Infant mortality rate
	  ○ Case fatality rate
	  ○ Disease specific mortality rate
	  ○ Attack rate, secondary attack rate
	• Prevalence
	• Relationship between incidence and prevalence
	• Standardization
	  ○ Direct standardization
	  ○ Indirect standardization

Measures of association and impact	• Relative measures
	  ○ Risk Ratio
	  ○ Rate Ratio
	  ○ Odds Ratio

	• Absolute measures
	  ○ Risk difference
	  ○ Rate difference
	  ○ Number needed to treat
	• Standardized rate difference, rate ratio
	• Standardized risk difference, risk ratio
	• Standardized mortality rate
	• Attributable fractions
	• Population attributable risk, PAR %
	• Quality adjusted life years
	• Interpretation of each measure
	• Conceptual differences between measures
	• Limitations of each measure

Effect heterogeneity	• Concepts of homogeneity & heterogeneity
	  ○ Pooling
	  ○ P-values
	  ○ Mantel Haenszel methods
	• Assessment of effect measure modification
	• Additive & multplicative scale
	• Interaction
	  ○ Quantitative and qualitative interaction
	  ○ Joint effects framework
	  ○ Additive scale
	  ○ Multiplicative scale
	  ○ Continuum of interaction
	  ○ Statistical modeling, tests for interaction
	  ○ Relative excess risk of interaction
	  ○ Public health relevance
	  ○ Sufficient cause interaction
	  ▪ Synergism
	  ▪ Antagonism
	  ▪ Mechanistic
	  ▪ Monotonicity
	  ○ Biologic interaction
	  ○ Gene-environment interactions

Reliability and validity	• Measures of reliability
	  ○ Inter-rater, intra-rater
	  ○ Internal consistency
	  ○ Test-retest
	  ○ Measures of agreement (statistical, graphical, tabular)
	• Measures of validity

	○ Content ○ Criterion (concurrent, predictive) ○ Construct (discriminant, convergent) ○ ROC curves ○ Sensitivity, specificity, PPV, NPV ○ Likelihood ratio • Questionnaire design ○ Types of questionnaires ○ Principles and approaches for questionnaire design • Validation study designs (to assess validity/reliability of measures) • Sensitivity analysis
Study design	• Asking good questions ○ PICOT ○ Types of research questions (descriptive, predictive, etiologic research questions) • Appropriate justification of using different types of study designs, including strengths and weaknesses of each design • Primary vs secondary data sources • 'Big data' sources ○ Electronic health records ○ Administrative data ○ Record linkage • Health claims databases • Consortium studies • Challenges of studying the health of populations • Measurement of exposure, outcomes, and covariates • Prevention of selection bias, confounding, information (measurement) bias via study design • Matching • Sample size and power calculations • Generalizability ○ Population health impact ○ Knowledge translation frameworks • Transportability

Concepts of causal inference	• Introductory concepts of causation (e.g., historical perspectives, Bradford Hill criteria)
	• Formal causal models
	  ○ Sufficient component cause model ("causal pies")
	  ○ Counterfactual/potential outcomes models
	  ▪ Notation
	  ▪ Causal response types
	• Measures of causal effect
	  ○ Individual causal effects
	  ○ Average causal effects
	  ○ Causal contrasts
	  ○ Marginal and conditional effects
	• Causation vs. association
	• Identifiability assumptions
	  ○ Exchangeability
	  ○ Consistency
	  ○ Positivity
	  ○ Model misspecification
	  ○ SUTVA
	• DAGs
	  ○ Terminology
	  ○ D-separation rules (backdoor paths)
	  ○ Colliders, confounders, mediators
	  ○ Sufficient adjustment sets

Confounding	• Structure of confounding
	• Time-varying confounding
	• Identification of confounders
	  ○ Conceptual principles
	  ○ Via DAGs
	  ○ Analytic approaches
	• Control via study design
	  ○ Randomization
	  ○ Restriction
	  ○ Matching
	• Control via analytic approaches
	  ○ Adjustment
	  ○ Stratification

	• E-value • Residual confounding • Unmeasured confounding • Overadjustment

Selection bias	• Definition
	  ○ Selection without bias
	  ○ Magnitude and direction of bias
	  ○ Difference between confounding and selection
	• Structure of selection bias
	  ○ Definition using DAGs
	  ○ Selection bias vs. collider stratification
	• Selection into study sample:
	  ○ Selective survival (survivor bias)
	  ○ Improper control selection
	  ○ Healthy worker bias, volunteer bias
	• Selection out of study sample:
	  ○ Informative censoring, loss to follow-up
	• Collider stratification bias
	  ○ DAGs M-bias structure

Measurement bias	• Measurement bias
	  ○ Definition using DAGs
	  ○ Structure of measurement bias
	  ○ Differential, non-differential
	  ○ Dependent, independent
	  ○ Strength and direction of measurement bias
	  ○ Mismeasurement of exposure, outcome, confounders, colliders
	
	Measurement error
	Misclassification (PPV, NPV, Sensitivity, Specificity)

Randomized Controlled Trials	• Justification of when RCTs can/should be used
	  ○ Strengths and weaknesses of RCT
	  ○ Clinical equipoise
	  ▪ Superiority, equivalence, non-inferiority
	  ○ Feasibility & ethics
	  ○ Efficacy vs. effectiveness
	• Measurement of exposure, outcomes, covariates

	• Confounding, measurement, selection bias as they relate to RCT
	
	• Randomization
	  ○ Purpose of randomization
	  ○ Best practice approaches for implementing randomization
	  ○ Stratified
	  ○ Blocking
	
	• Structure of RCTs
	  ○ Parallel RCT
	  ○ Cross-over
	  ○ Factorial
	  ○ Group designs
	
	• Allocation concealment
	  ○ Blinding types
	  ○ Placebo control, active control
	  ○ Compliance
	  ○ Detection
	
	• Analysis
	  ○ Intention to treat
	  ○ Per protocol
	  ○ Data analysis considerations for RCT (e.g., regression modeling, measures of change, absolute and proportional change)
	  ○ Interim analyses
	  ○ Ethics & stopping rules, role of DSMB
	
	• Additional RCT designs
	  ○ Cluster trials
	  ○ Community-based interventions
	  ○ Sequential designs
	  ○ Pragmatic designs

Observational Studies
For each study design, should be able to:
○ *Describe key features*
○ *Measurement of key variables*
	Cohort studies
	○ Prospective & retrospective
	○ Target trial framework
	  ○ Asking a causal question
	  ○ Emulating a target trial

○ *Justify when the design should be used* ○ *Strengths and weaknesses*	**Case control studies** ○ Cumulative ○ Case cohort Nested case control with incidence density sampling
	Cross-sectional studies
	Surveys
	Ecologic studies
	Quasi-experimental designs ○ Natural experiments ○ Difference in difference ○ Regression discontinuity ○ Interrupted-time series

Clinical epidemiology	• Diagnostic and prognostic screening studies
	• Clinical and subclinical disease
	• Chronic disease vs. infectious disease
	• Latency, incubation period
	• Natural history of disease
	• Diagnosis and prognosis
	• Diagnostic test accuracy
	  • Sensitivity, specificity, PPV, NPV
	  • Sequential vs. simultaneous testing
	• Screening studies
	  • Disease screening in populations
	  • Screening ethics
	  • Cost/benefit ratio

Field Epidemiology & Surveillance	• Dynamics of disease transmission
	• Endemic, epidemic, pandemic
	• Outbreak investigation
	• Herd immunity
	• Disease surveillance programs
	  ○ Active surveillance
	  ○ Passive surveillance
	  ○ Sentinel health events
	• National and international disease surveillance programs

Systematic reviews and meta-analysis	• Systematic reviews
	  ○ Types of reviews
	  ○ Narrative reviews
	  ○ Qualitative vs. quantitative synthesis
	• Meta analysis
	  ○ Individual patient data meta analysis
	  ○ Network meta analysis

	• Searching for studies to include
	  ○ Assessment of quality
	  ○ Publication bias
	  ▪ Funnel plot
	• Extracting data
	  ○ Statistics for data synthesis
	  ○ Graphical approaches
	  ○ Heterogeneity assessment
	  ▪ I^2
	  ▪ L'abbe Plots
	  ○ Pooling
	  ○ Fixed and random effects
	  ▪ Dersimonian and Laird model
	  ○ Subgroup analyses
	  ○ Meta regression
	• Interpreting results

Regression Modeling & Interpretation	• Basic types of regression models
	  ○ Constant
	  ○ Linear
	  ○ Exponential
	  ○ Logistic
	  ○ Generalized linear models (GLM)
	  ○ Binomial (RR)
	  ○ Poisson (IRR)
	  ○ Polytomous regression models (ordinal, multinomial)
	  ○ Additive models
	• Dose response and trend analyses
	• Interpreting results from regression analysis including crude and multivariable adjusted models, including interaction terms
	• Regression post-estimation
	  ○ Predicted probabilities
	• Approaches for Longitudinal data analysis
	  ○ Analyses of change
	  ○ Repeated measures
	  ○ Clustered (multilevel) data
	  ○ Generalized estimating equations (GEE)

	○ Mixed effects models -Missing data
Time to Event Analyses	• Time scales • Censoring and truncation ○ Interval ○ Left, right • Kaplan Meier curves • Pooled logistic regression for discrete failure time data • Time to event (survival) analysis ○ Survival function ○ Restricted mean survival time • Survival curves • Non-parametric, semi-parametric, parametric survival models ○ Weibull ○ Exponential ○ Gompertz • Cox PH model ○ Hazard ratios ○ Hazards of hazard ratios ○ Time varying exposures ○ Assessment of proportional hazards ○ Time dependent covariates ○ Frailty • Accelerated failure time model • Additive hazard models • Competing risks ○ Fine and Gray models ○ Sub-distribution hazards ○ Semi-competing risks

Biostatistics Topics

Types of data	• Nominal, ordinal, rank, discrete, continuous variables
	• Probability
	  ○ Laws of probability
	  ○ Conditional probability
	  ○ Random variables
	• Probability distributions
	  ○ Types (normal, binomial, Poisson)
	  ○ Range, IQR
	  ○ Variance, standard deviation
	  ○ Mean, median, mode
	  ○ Transformations

Inference for means, proportions, and counts	• Hypothesis testing • T-tests • ANOVA/MANOVA • Type I and II error • P-values • Multiple testing • Confidence intervals (calculation and interpretation) • Paired and unpaired samples • Chi-squared test • Fishers exact test • Mantel-Haenszel test • Frequentist and Bayesian perspectives
Non-parametric statistics	• Sign test • Rank sum test • Wilcoxon signed rank test • Kruskal Wallis test

Regression	• Generalized linear models • Regression diagnostics • Model building and selection techniques • Forward and backward selection • AIC • BIC • R^2 criteria • Goodness of fit assessment • Collinearity • Re-scaling variables • Centering variables • Interpreting results from regression models • Analysis of variance/ covariance • Multinomial and ordinal regression models • Analysis of matched data

Some helpful resources:

Textbooks in Epidemiology:
Epidemiology Beyond the Basics (Szklo and Nieto)
Modern Epidemiology: 4th Edition (Lash et al.)
Causal inference: What If? (Hernan and Robins)
Epidemiology by Design (Westreich)
Epidemiology: An Introduction (Rothman)
Causal Inference- The Mixtape (Cunningham)
Epidemiologic Methods: Studying the Occurrence of Illness (Koepsell & Weiss)
Critical Appraisal of Epidemiological Studies and Clinical Trials (Elwood)
Fundamentals of Clinical Trials 5th Ed (Friedman et al)
Gordis Epidemiology (Celentano)
Essentials of Epidemiology in Public Health (Ashengrau and Seage)

Textbooks in Statistics:
Introduction to the Practice of Statistics (Moore and McCabe)
Biostatistics: The Bare Essentials (Streiner)
Categorical Data Analysis (Agresti) Fundamentals
of Biostatistics (Rosner) Introductory Applied
Biostatistics (D'Agostino)
Regression Methods in Biostatistics (Vittinghoff, Glidden, Shiboski, McCuloch)
Regression Modeling Strategies (Harrell)