Core learning objectives in epidemiology and biostatistics
(Master’s degree, epidemiology stream, thesis and non-thesis programs)

Department of Epidemiology, Biostatistics, and Occupational Health

McGill University

Document prepared by the Curriculum Committee and the Graduate Studies Committee

18 February 2008
Introduction

This document presents objectives of the core program in epidemiology and biostatistics for MSc students registered in the epidemiology stream. The document was designed to facilitate the planning of courses and other learning activities, and to help students identify learning objectives. The objectives presented here may be met through course work, the conduct of the thesis research or, for the non-thesis MSc degree, the research project, and other academic activities such as departmental seminars. This document focuses on learning objectives as opposed to learning modalities. The stated objectives may be reached by a variety of learning methods proposed by our Department.

This is a public document posted on the website of the Department of Epidemiology, Biostatistics and Occupational Health. The document may be disseminated and used without restriction.
General learning objectives

The general objective of the core teaching program in epidemiology is to introduce students to the principles and methods of epidemiologic research in order to enable them to design, conduct, analyze, and interpret epidemiologic research. Upon the completion of their MSc degree, students should have acquired an understanding of the following broad topics:

a) the contribution of epidemiology and biostatistics to health research
b) design, conduct, and analysis of epidemiologic studies
c) critical appraisal of epidemiologic studies, synthesis and integration of epidemiologic research, and causal inference in epidemiologic research
d) communication of scientific results

In addition, students should have acquired a basic knowledge of some substantive epidemiology, including a general appreciation of broad public health problems in Canada and internationally.
Section I

Basic concepts: populations, health, causality, probability, inference
1. **General introduction to epidemiology and biostatistics**

*(Proficiency level: Koepsell and Weiss 2003, Chapters 1 and 2; general introductory epidemiology and biostatistics textbooks; British Medical Journal papers on qualitative research)*

Students will be able to:

a) Describe the contribution of epidemiology and biostatistics to the scientific study of health and disease.

b) Describe case studies in the history of epidemiology and biostatistics, including, for example, John Snow’s research on cholera, Joseph Goldberger’s work on pellagra, the Salk vaccine trial, or the Framingham heart study.

c) Contrast epidemiologic and biostatistical research methods with other scientific research methodologies, including quantitative and qualitative methods.

2. **Concepts of well-being, health, and disease**

*(Proficiency level: World Health Organization 2001)*

Students will be able to:

a) Define and distinguish the concepts of health, quality of life, impairment, activity limitation, and participation restriction.

b) Describe the concept of *life course* in population health.

3. **Human populations**

*(Proficiency level: World Health Organization, [www.who.int](http://www.who.int))*

Students will be able to:

a) Define the concepts of *population structure* and *population dynamics* and identify factors that influence on them.

b) Define various population models used in epidemiologic research, including closed (or fixed) populations, open (or dynamic) populations, and stable dynamic populations.
c) Describe and compute measures that characterize population dynamics, including birth and mortality rates, fertility rates, sex ratio, and dependency ratios.

d) Define the concepts of demographic and health transition.

4. **Causation and causal inference**

   *(Proficiency level: Rothman 2002, Chapter 2)*

   Students will be able to:

   a) Demonstrate some familiarity with common philosophical perspectives on the concept of cause and discuss these in the context of epidemiologic and biostatistical research.

   b) Relate the epidemiologic concepts of exposure, risk factor, and determinant to causal concepts.

   c) Define counterfactual and discuss relationships between the concept of counterfactual and that of confounding.

   d) Discuss the problems of generalization in scientific research and in epidemiology, and describe relationships between the concept of generalization and that of effect modification.

   e) Distinguish between generalizable and non-generalizable results in epidemiologic research.

5. **Probability**

   *(Proficiency level: Moore and McCabe 2006, Chapter 4)*

   Students will be able to:

   a) Define randomness and probability.

   b) Describe and apply the basic laws of probability.

   c) Define conditional probability.

   d) Use Bayes' theorem to obtain “reverse” probabilities.
e) Define *variable* and *random variable*, and distinguish discrete and continuous random variables.

f) Define *expectation* and *variance* of a random variable.

g) Describe the role of probability in statistical inference, particularly through frequentist and Bayesian interval estimation.

6. **Distributions**

*(Proficiency level: Moore and McCabe 2006, Chapter 5)*

Students will be able to define the properties and describe the main applications of the normal ("Gaussian"), Bernouilli, binomial, hypergeometric, and Poisson distributions, and of normal approximations to these distributions.

7. **Inference: general concepts**

*(Proficiency level: Rothman 2002, Chapter 6; Moore and McCabe 2006, Chapter 6)*

Students will be able to:

a) Define and distinguish *parameter* and *statistic*.

b) Define and distinguish *frequentist* and *Bayesian* inference.

c) Describe the key concepts and statistical principles involved in point and interval estimation, including exact frequentist confidence intervals.

d) Describe the key concepts and statistical principles involved in hypothesis testing, P-values, and power.
Section II

Epidemiologic studies: parameters, strategies, validity
8. **Sources of data in epidemiology**

*(Proficiency level: Kelsey 1996, Chapter 3; MacMahon and Trichopoulos 1996, Chapter 3; Koepsell and Weiss 2003, Chapter 6)*

Students will be able to:

a) Identify various sources of data related to the health of populations, including administrative databases, vital statistics, and survey data.

b) Explain and critically assess the processes that document health care related events leading to the production of miscellaneous data bases, including: paper and electronic health records; administrative databases created for documenting beneficiaries, hospitalizations, pharmaceutical reimbursements, mandatory and optional disease registrations, and mandatory and optional adverse events reports; and vital statistics.

c) Explain the processes that document causes of death and transform the information into mortality data bases and statistics; describe limitations of these processes.

9. **Measurement: health status, exposures, covariates**

*(Proficiency level: Koepsell and Weiss 2003, Chapters 3 and 4)*

Students will be able to:

a) Define and estimate different measures of incidence, including risk (cumulative incidence) and incidence rate.

b) Define and estimate *point prevalence* and *period prevalence*.

c) Describe basic mathematical relationships between *cumulative incidence* and *incidence rate*, and between *incidence* and *prevalence*, and describe conditions under which these relationships hold.

d) Define and estimate: attack rate, case-fatality, neonatal mortality, infant mortality, maternal mortality, life expectancy, person-years of life lost, life expectancy in good health, quality-adjusted life years, and disability-adjusted life years.
e) Address validity, intra-rater reliability, and inter-rater reliability in the development of questionnaires, scales, physical measurements, biologic tests, and other data collection instruments in epidemiologic research.

f) Identify the circumstances when a validation sub-study, comparing convenient exposure assessment methods to better but less feasible methods, can be carried out.

g) Operationalize study variables and discuss basic principles regarding the modeling of exposure and covariate data, including time-dependent variables.

h) Define *kappa coefficient*, *coefficient of variation*, *standard error of measurement*, and *intraclass correlation coefficient*.

i) Define and distinguish the concepts of sensitivity, specificity, predictive values, likelihood ratio, and receiver operating characteristic curves; describe the application of these concepts in clinical epidemiology as well as in non-clinical situations.

j) Describe sources of missing data; define and contrast the main methods used to handle missing data.

10. **Comparative parameters**

    *(Proficiency level: Koepsell and Weiss 2003, Chapter 9)*

    Students will be able to:

    a) Define and distinguish risk difference, risk ratio, rate difference, rate ratio, prevalence difference, and prevalence ratio.

    b) Interpret the parameters listed above and use them in the appropriate contexts.

    c) Describe mathematical relationships between parameters such as, for example, relative risk and incidence rate ratio.
11. **Sampling**  
*(Proficiency level: Armitage et al. 2002)*

Students will be able to:

a) Describe sources of variability in epidemiologic research, including sample size, characteristics of study subjects, data collection instruments, and study design.

b) Explain the need for sampling and the advantages of various methods of sampling in epidemiologic studies.

c) Distinguish between probability and non-probability sampling.

d) Describe the most common methods of probability sampling used in epidemiologic studies: simple random sampling, stratified sampling, systematic sampling, cluster sampling, and multistage sampling.

e) Describe relationships and distinctions between the concepts precision, bias, and generalizability in the estimation of population parameters.

f) Describe methods of non-probability sampling used in epidemiologic studies: convenience sampling and purposive sampling.

12. **Describing the distribution of diseases**  
*(Proficiency level: Koepsell and Weiss 2003, Chapter 7)*

Students will be able to:

a) Describe examples of person, place, and time variables and discuss the contribution of person/place/time studies in epidemiologic research.

b) Discuss the distinctions and the relationships between analyses of rates by age, period (calendar time), and birth cohort.

c) Define *prevalence survey* and describe strengths and weaknesses of such surveys.
13. **Ecologic studies**

*(Proficiency level: Koepsell and Weiss 2003, Chapter 12)*

Students will be able to:

a) Distinguish between individual-level measures and population-level measures in epidemiologic research.

b) Define *ecologic study*.

c) Define *ecologic fallacy* and describe circumstances where this bias is likely to occur.

14. **Experimental studies**

*(Proficiency level: Koepsell and Weiss 2003, Chapter 13; Moher et al. JAMA 2001)*

Students will be able to:

a) Define *experiment* and establish distinctions between experimental and quasi-experimental studies.

b) Define and establish distinctions between broad types of experimental and quasi-experimental studies such as randomized clinical trials and community trials; distinguish applications of experimental and quasi-experimental studies for clinical and non-clinical health research.

c) Define and describe the purpose of randomization, placebos, blinding, and intention-to-treat analyses; distinguish cluster versus individual randomization.

d) Discuss issues in the choice of comparison interventions, including placebo, usual medical care, and others.

e) Define and distinguish the concept of *efficacy* and that of *effectiveness*.

f) Apply the CONSORT criteria (JAMA 2001) in the design and critical assessment of a trial.
15. **Non-experimental studies: cohort studies, case-control studies, and related designs**

*(Proficiency level: Koepsell and Weiss 2003, Chapters 5, 14, and 15)*

Students will be able to:

a) Identify and characterize basic designs in non-experimental epidemiologic research, including cohort and case-control studies, and variants of these designs such as nested case-control studies, case-cohort studies, case-crossover studies, and proportional mortality studies.

b) Describe conceptual relationships between these various study designs.

c) Describe and contrast advantages and disadvantages of each study design; identify applications in clinical and in non-clinical settings.

d) Describe basic issues in the definition of study subjects in non-experimental epidemiologic studies, including the definition of exposure, case series, and denominator series; describe the concept of *person-time*.

e) Identify common sources of bias for each design.

16. **Selection bias**

*(Proficiency level: Koepsell and Weiss 2003, Chapter 15; Rothman 2002, Chapter 5)*

Students will be able to define *selection bias* and identify common situations in which this type of bias may occur, including referral bias and detection bias in case-control studies, selection bias in the recruitment of cohort members, and selection bias associated with losses to follow-up in cohort and in experimental studies.
17. **Information bias**  
*Proficiency level: Koepsell and Weiss 2003, Chapter 10*

Students will be able to:

a) Define *information bias* and describe various sources of information bias such as recall bias, unblinded ascertainment of study variables, etc.

b) Define and distinguish *differential* and *non-differential misclassification*, and describe the impact of measurement errors on epidemiologic estimates.

c) Define the requirements of exposure assessment to maximize validity and precision, taking into account timing of measurement (retrospective or prospective) and size of study group.

18. **Confounding**  
*Proficiency level: Koepsell and Weiss 2003, Chapter 11*

Students will be able to:

a) Define *confounding* and *confounding variable*, and identify situations in which confounding may occur.

b) Identify and use the various methods available to control for confounding, including restriction, matching, randomization, stratification, standardization, and multivariable modeling.

c) Discuss the concept of residual confounding and identify circumstances where the control of confounding may be incomplete.

19. **Interaction (synonym: effect modification)**  
*Proficiency level: Rothman 2002, Chapter 9*

Students will be able to:

a) Define and distinguish the concepts of *interaction*, *synergy*, and *antagonism*.
b) Discuss and contrast the concepts of *additive* and of *multiplicative* interaction, and relate these concepts to the notion of *biologic* interaction.

c) Estimate and interpret additive and multiplicative interaction parameters taking into account the type of data available from different study designs.
Section III

Statistical analysis
20. **General goals and concepts; descriptive and summary statistics**

*(Proficiency level: van Belle et al. 2004; Moore and McCabe 2006; Rosner 2006)*

Students will be able to:

a) Describe and distinguish various types of data, including continuous, categorical, ordinal, and nominal data.

b) Identify the general goals of statistical analysis, including data description, data summarization, parameter estimation, parameter comparison, and inference.

c) Identify and use statistical methods and models appropriate for specific types of study designs, data, and research objectives.

d) Distinguish and use various kinds of tables and graphs to describe a single variable; calculate the mean of a series of observations, the median, percentiles, and identify the range; calculate a proportion (count) and describe it as a mean (sum) of zeros and ones.

e) Distinguish and use various kinds of tables and graphs to show relationships between variables; define Pearson’s coefficient of correlation, describe its properties, and estimate it; describe the method of least squares for the estimation of linear regression parameters; discuss the meaning of a regression line; estimate parameters for simple regression; and describe how simple linear regression with a binary group indicator can be used to compare two means.

21. **Inference for a single parameter: mean, proportion, rate, and survival probability**

*(Proficiency level: Moore and McCabe 2006, Chapters 1, 7, and 8)*

Students will be able to:

a) Calculate exact and large-sample confidence (frequentist) and credible (Bayesian) intervals for a single parameter: mean, proportion, and rate (person-time denominator).

b) Perform significance tests concerning these parameters.
c) Calculate a Kaplan-Meier survival curve, and obtain Greenwood confidence intervals for
the fraction surviving at specified times (using Greenwood’s method).

d) Calculate sample sizes required for a given precision in confidence interval estimates.

22. **Comparing two parameters**

*(Proficiency level: Moore and McCabe 2006, Chapters 2, 7, 10, 11, 12, and 13)*

Students will be able to:

a) Describe and select the appropriate statistical method for the crude and for the adjusted
comparison of two means, proportions, rates, or survival probabilities.

b) Use the following non-regression adjustment methods:

- weighted averages of stratum-specific differences in means or risks
- weighted averages of stratum-specific risk differences or risk ratios
- Mantel-Haenszel procedure for odds, risks, and rate ratios
- Woolf’s method for odds ratios

23. **Nonparametric methods**

*(Proficiency level: Moore and McCabe 2006, Chapter 15)*

Students will be able to:

a) Describe the principles underlying the use of nonparametric tests, their advantages, and
their disadvantages.

b) Use the sign test, the Wilcoxon rank sum test, the Wilcoxon signed rank test, and the
Kruskal-Wallis test.

24. **Generalized linear models for regression analyses**

*(Proficiency level: Rothman and Greenland 1998, Chapters 20 and 21; Moore and McCabe 2006,
Chapters 11 and 16)*

Students will be able to:

a) Describe and apply the general concepts underlying regression, including: linear
predictor, regression coefficient, link, and distribution.
b) Describe and distinguish uses of regression for prediction, adjustment (confounding),
and assessment of interaction.

c) Describe and distinguish uses of regression for etiologic, diagnostic (prevalence
functions), and prognostic research.

d) Describe the principles of model selection and evaluation.

e) Explain the meaning of:

- the intercept in a null model
- the intercept and the slope in models with one binary x variable
- differences of means, proportions, or rates in models with one binary x variable
- regression coefficients in models with two x variables

f) Conduct multiple regression analyses, taking into account:

- collinearity
- the transformation of variables
- the type of response variable, including: quantitative variables (ordinary linear
  regression), binary variables (logistic regression), counts (Poisson regression),
  and survival times (Cox regression)
- time-varying covariates

g) Use logistic regression for the calculation of propensity scores.

25. **Computing and data management**

*(Proficiency level: Delwiche and Slaughter 2003, UCLA Academic Technology Services, Statistical
Computing Resources, [www.ats.ucla.edu/stat]*)

Students will be able to:

a) Computerize and manage simple and more complex research data.

b) Use spreadsheets for small-scale statistical calculations and simulations, managing
simple lists, and other similar tasks.

c) Import and export data using different software packages.

d) Program introductory level tasks such as loops and other control structures, arrays,
iterations, and other similar items.
Section IV

Critical appraisal, protocol development, ethics
26. **Appraisal, synthesis, and use of epidemiologic data**

*(Proficiency level: Koepsell and Weiss 2003, Chapter 8; Moher et al. Lancet 1999)*

Students will be able to:

a) Conduct critical appraisals of published epidemiologic studies and of their own research.
b) Define and distinguish various methods of synthesis and integration of epidemiologic studies, including narrative reviews, systematic reviews, and meta-analysis.
c) Identify basic elements and issues in the conduct of meta-analysis, including the selection of studies, the types of data included, quality scales, basic methods of statistical analysis, and main sources of bias.
d) Apply the QUOROM (Lancet 1999) criteria in the design and critical assessment of a systematic review or meta-analysis.
e) Critically use data from published studies to inform researchers and decision-makers regarding a health problem.

27. **Protocol development**

*(Proficiency level: Koepsell and Weiss 2003)*

Students will learn to develop elements of a research protocol at an introductory level of proficiency. In particular, they will be able to:

a) Formulate, at an introductory level, research objectives based on the identification of gaps in the scientific literature.
b) Demonstrate, at an introductory level, intellectual creativity and originality in the formulation of research questions.
c) Identify common field work problems in the conduct of epidemiologic research.
28. **Research ethics**

*(Proficiency level: Introductory tutorial for the Tri-council policy statement: Ethical conduct for research involving humans)*

Students will be able to:

a) Identify fundamental ethical principles that guide the conduct of research involving human participants.

b) Describe published guidelines regarding the protection of human participants.

c) Recognize when a study requires human participants' protections.

d) Identify ethical issues to consider when selecting participants for a study, including those that apply to special groups.

e) Define *privacy* and *confidentiality* and describe how these can be maintained throughout the research process.

f) Define *informed consent* and describe the elements that should be included in an informed consent document or procedure.

g) Describe conditions that may affect a person’s capacity to consent and the responsibilities of the researcher in seeking consent.

h) Define *institutional review board*. 
Section V

Specialized concepts, methods, and applications

Note: This section focuses on methods and applications as opposed to substantive knowledge.

A core program in epidemiology and biostatistics should include basic elements in public health, infectious disease epidemiology, and genetic epidemiology. Readers will recognize that a myriad of other domains are not included here, including, for example pharmacoepidemiology, psychiatric epidemiology, cancer epidemiology, social epidemiology, cardiovascular epidemiology, etc. Please note, however, that substantive knowledge in these various domains (as opposed to methodologic knowledge) has been included in our core curriculum (see the Substantive epidemiology section).
29. **Public health**

Students will demonstrate a general proficiency in public health research, with a specific focus on the contribution of epidemiology and biostatistics to public health. In particular, students will be able to:

a) **Screening**

*(Proficiency level: Rothman 2002, Chapter 11)*

- Distinguish between screening and case-finding.
- Distinguish between a screening test and a screening program.
- Apply recognized criteria to determine the appropriateness of screening programs.
- Describe appropriate study designs to evaluate screening programs.
- Identify the presence of length-biased sampling and lead time bias in studies evaluating screening programs.

b) **Cluster investigation**


- Define and identify a cluster.
- Describe methodologic challenges in cluster investigation, including the various sources of bias.
- Describe the advantages and disadvantages of different designs for cluster investigation.
- Identify the strategies of analysis for cluster investigation.

c) **Surveillance**

*(Proficiency level: Rothman and Greenland 1998, Chapter 22)*

- Describe the objectives, uses and principles of disease surveillance in public health.
- Describe the major methodologic and analytic challenges of surveillance systems.
- Describe criteria for deciding whether to include a disease in a surveillance system.
- Distinguish active and passive surveillance.
- Describe the major components of a surveillance system for infectious and for chronic diseases.
- Describe capture-recapture methods of disease surveillance.
- Describe sentinel systems of surveillance.
- Describe the data sources for major surveillance systems in Canada and the United states and describe their limitations.
d) Health promotion

(Proficiency level: Tones, Chapter 7.3 in Detels et al, Oxford Textbook of Public Health 2002)

- Describe models of health and health promotion.
- Describe basic theories and ideologies central to the notion of health
- Appreciate the social influences and context of health and health-related practices.
- Discuss values, ethics, and the practice of public health.

e) Environmental health


- Demonstrate a basic understanding of the influence of environmental factors (physical, chemical, and biologic) on human disease in communities.

f) Occupational health


- Describe basic concepts in occupational health and occupational epidemiology, with an emphasis on the study of disease occurrence in worker populations.

30. Infectious disease epidemiology

(Proficiency level: Nelson and Masters Williams 2007; Thomas and Weber 2001, Chapter 2; Koepsell and Weiss 2003, Chapter 19)

Students will be able to:

a) Identify the major differences between infectious and non-infectious disease epidemiology.

b) Identify different types of infectious agents, e.g., viruses, bacteria, and parasites.

c) Identify the major mandatory reportable diseases in Canada and the United States.

d) Describe recent trends in infectious disease morbidity and mortality.

- Demonstrate a basic understanding of infectious disease epidemiologic methods (outbreak investigation, transmission dynamics, and the concept of dependent happenings).
31. **Genetic epidemiology**

*(Proficiency level: Rothman and Greenland 1998, Chapter 30)*

Students will be able to:

a) Describe the justification for considering genetic variants (or other parameters from the genome) as important determinants of complex diseases.

b) Recognize that genetic variants (or other parameters from the genome) have similarities, but also differences with more conventional risk factors.

c) Describe the advantages and limitations of traditional epidemiologic study designs (with unrelated study subjects) to evaluate genetic determinants of disease.

d) Describe the advantages and limitations of association study designs that use families (related study subjects) or a combination of related and unrelated study subjects to evaluate genetic determinants of disease.
Section VI

32. **Substantive epidemiology**


Students will demonstrate a general knowledge of important risk factors and of the broad causes of morbidity and mortality affecting human populations. In particular, students will be able to:

a) Describe the contemporary burden of disease in various population groups in Canada and the rest of the world; population groups may be considered in terms of socio-demographic, occupational, clinical, and other characteristics.

b) Describe the historical evolution of the burden of disease in Canada and the rest of the world.

c) Describe how various proximal and distal determinants of population health (socio-economic, lifestyle, environmental, biologic, and health services) influence the levels and distribution of health in populations.

d) Describe how these risk factors influence the occurrence of major diseases of public health importance in Canada and the rest of the world today.

e) Describe how knowledge of determinants influences the prevention and control of disease.

f) Describe major current disease prevention and control efforts in Canada and the world today, including screening programs.
Section VII

33. Communication of scientific results

(Proficiency level: Szklo and Nieto 2007, Chapter 9)

Students will be able to:

a) Report results of an epidemiologic study in a written format.

b) Report results of an epidemiologic study in the form of a brief oral communication.
References

In several sections of this document, references are made to selected references illustrating the level of proficiency expected of students for a given topic. For example, on the topic of interaction, the expected level of proficiency corresponds to Chapter 9 of Rothman’s textbook *Epidemiology, An introduction* (2002). These references are presented for illustrative purposes only and do not represent recommendations or prescriptions for material to be used for teaching.


British Medical Journal papers on qualitative research:


Websites

Interagency Advisory Panel on research Ethics. www.pre.ethics.qc.ca


UCLA Academic Technology Services, Statistical Computing Resources: www.ats.ucla.edu/stat

World Health Organization: www.who.int
Appendix. Reviewer comments to be reconsidered

This appendix includes reviewer comments that were not incorporated in the current version of this document. However, we felt that these comments should be revisited when the document is updated after the academic year 2008-2009.

a) Item 28 a (Ethics)

Specify the fundamental ethical principles referred to in this section.

b) Section V (Specialized concepts, methods, and applications)

Some objectives are too vague to be evaluated. There remain some important objectives not addressed elsewhere. These might be organized under headings such as “Basic concepts of health and disease” and “Basic concepts of public health and clinical interventions (or services).” Some of the material under Section VI (Substantive epidemiology) might be integrated under the new headings suggested above.

c) Consider an introduction to some concepts in health economics and decision modeling.