

Self-study guide on data

Table of Contents	Page
I. Background	1
II. Guide	2
III. Suggested order	2
IV. Component difficulty	3
V. Component source links and comments	3
VI. More good web resources	5

I. Background

The purpose of this guide is to identify web-based resources for users interested in obtaining information about and increasing their skills in handling data. The need for training in data use and interpretation is identified in the Alaska Diabetes Strategic Plan; this guide is intended to be an interim step toward meeting this need. Technical assistance is available on an ongoing basis (please call Gail Stolz at 269-8034 for more information), and future training events will build on the concepts presented in this guide.

Development of this guide started with finding a good introduction to basic data analysis (NWCPHP). Components were added to supplement its presentation on rates, starting with discussion of confidence intervals and statistical tests for interpreting rates. Data users in Alaska also need to understand issues associated with reports on race and small populations; the papers from the State of Washington address these topics. Then components on data collection, sample selection and conducting surveys were added because many programs need to fill information gaps, and these resources may help.

Each component in this guide is independent of the others, and some topics are presented in more than one component. (A successful strategy for becoming comfortable with using and interpreting data is to return to resources that have been helpful, and to review information about the same topic on several different resources. Repetition helps!) Many resources are available that cover some or all of the topics included in this guide. Components were chosen based on these attributes:

- ✓ Accurate information
- ✓ Straight-forward style
- ✓ Length (prefer relatively short, but complete)
- ✓ Level of detail (prefer moderately complex – e.g., formulas provided, not proven)
- ✓ Estimated internet band width required (prefer .pdf files)
- ✓ Continuing education credit available (unfortunately, it is only available for the components selected from one source in this group).

This guide organizes the components in several ways. Part II lists the components by the topics they cover, generally moving from finding, through analyzing, to presenting data. Part III offers a suggested order that incorporates both the difficulty of the presentations and the topics being discussed. It includes complete URLs, in case the user does not have an electronic version of this guide or there is a problem with a hyperlink in Part II. Part IV very roughly ranks the

difficulty of the components based on the competency that they help users achieve. Part V contains brief descriptions of the component sources and the components themselves.

II. Guide by topic(s) covered

Topic		Component (hyperlinked)	
Finding data			
Alaska data sources		Data sources document	
Collecting data			
Measurement		Columbia University	
Samples & Sampling		PA Tools of the Trade on Surveys	
		PA Tools of the Trade on Probability Sampling	
Questionnaire design		Deakin University	
Analyzing data			
Descriptive statistics – mean, median and mode		NWCPHP	
		FOCUS on variables and distribution	
Odds and risk ratios		Supercourse on measures of association	
Rates	How rates differ from ratios and proportions	Supercourse on rates, ratios and proportions	
	How they are calculated	Crude	
		Category specific	} NWCPHP WA guideline on rates
		Age adjusted	
	Prevalence and incidence		
Mortality and case-fatality	NWCPHP		
Confidence intervals		WA guideline on confidence intervals	
Chi-square and ANOVA; p-values		FOCUS on simple statistics tests	
Data quality considerations			
Race and ethnic groups		WA guideline on race	
Small numbers		WA guideline on small numbers	
Presenting data			
Tables and graphs		NWCPHP	

III. Suggested order – Analyzing data, data quality and using data

The hyperlinks point to webpages that introduce the components. Consequently, users may need to scroll through some introductory or background information before arriving at the link for the presentation itself.

1. NWCPHP <http://www.nwcphp.org/training/courses-exercises/courses/data-interpretation>
2. FOCUS (variables and distribution)
<http://www2.sph.unc.edu/nccphp/focus/vol3/issue5/index.htm>
3. Supercourse (measures of association)
<http://www.pitt.edu/~super1/lecture/lec19091/index.htm>
4. Supercourse (rates, ratios and proportions)
<http://www.pitt.edu/~super1/lecture/lec0441/index.htm>
5. FOCUS (simple statistics tests)
<http://www2.sph.unc.edu/nccphp/focus/vol3/issue6/index.htm>

6. WA Data Analysis Guideline (confidence intervals)
http://www.doh.wa.gov/Data/Guidelines/WordDocs/CI_guidelines.pdf
7. WA Data Analysis Guideline (race)
<http://www.doh.wa.gov/Data/Guidelines/Raceguide1.htm>
8. WA Data Analysis Guideline (rates)
<http://www.doh.wa.gov/Data/Guidelines/WordDocs/Rateguide.doc>
9. WA Data Analysis Guideline (small numbers)
<http://www.doh.wa.gov/Data/Guidelines/WordDocs/SmallNumbers.doc>
10. Washington State (race) <http://www.doh.wa.gov/Data/Guidelines/WordDocs/Raceguide.doc>

If you need to collect data...

- A. Columbia University (measurement) <http://ccnmtl.columbia.edu/projects/qmss/index.html>
- B. PA Tools of the Trade (entire population and sample surveys)
<http://www.health.state.pa.us/hpa/stats/techassist/popsample.htm>
- C. PA Tools of the Trade (probability sampling)
<http://www.health.state.pa.us/hpa/stats/techassist/probability.htm>
- D. Deakin University
http://www.deakin.edu.au/buslaw/bowater/research/pdf/questionnaire_design.pdf

IV. Component difficulty

These topics have complicated features. Directions about how to manipulate data often become quite complex as well. One way to assess the difficulty of these components is based on the apparent skill or competency level that would be achieved by someone who completed them.

- **Awareness Level** competencies - a basic level of knowledge or understanding about a topic.
- **Performance Level** competencies –the level of skill or ability is higher than *awareness*, and builds on it.
- **Technical Level** competencies – the skill or ability level is higher than *performance* and has become a specialized skill.
- **Planning & Management Level** competencies - apply to individuals who perform leadership and policy-making functions

<u>Component source</u>	<u>Competency level</u>
Columbia University	awareness
Deakin University	awareness/performance
FOCUS on Field Epidemiology	performance
Northwest Center for Public Health Practice (NWCPHP)	awareness
PA Dept of Health Health Statistics Technical Assistance - Tools of the Trade (PA Tools of the Trade)	awareness/performance
Supercourse	awareness/performance
WA Dept of Health Data Analysis Guidelines (WA guideline)	performance/technical

Written components (as opposed to slide shows) included in this guide have a fairly advanced reading level.

V. Component comments

[Columbia University Center for New Media Teaching and Learning](#) was created to promote teaching and learning through new media. *Quantitative Methods in Social Sciences e-lessons* is one of six social sciences courses available on its website. The first lesson (Measurement) is a good background discussion about the potential distance between how people use data to answer a question and what the data actually measured.

[Deakin University, Bowater School of Management and Business Research Fundamentals](#) is a series of modules that were created to provide on-line teaching tools for college students. For this guide, *Questionnaire design* is a little long, but has useful details.

[North Carolina Center for Public Health Preparedness](#) (NCCPHP) produces *FOCUS on Field Epidemiology* (a web-based periodical) as one strategy for offering training in epidemiology. Online self-evaluation questions complement each issue, continuing education credit is available (one hour of CME/AMA PRA category 2 per issue), and accompanying PowerPoint presentations are posted on the FOCUS website. The guide includes two FOCUS issues:

- ▶ *Data Analysis Basics: Variables and Distribution* - Offers more detailed information than NWCPHP on the types of data, strategies for improving data accuracy, and simple analysis techniques.
- ▶ *Data Analysis Basics: Simple Statistical Tests* - Builds on the *Variables and Distribution* issue, and includes confidence intervals, chi-square calculation, p-values, and analysis of variance.
- ▶ NCCPHP also offers for Survey Design Part 1: Sampling (URL: http://www2.sph.unc.edu/nccphp/training/HEP_SDP1/certificate.htm), which has continuing education credit. It's longer than the *Tools of the Trade* components included in the guide and contains similar information. Registration is required, but there are no costs attached. It is a good overview of reasons to use samples when conducting surveys and data quality issues. It also has a very good presentation on sample size and factors that might enlarge or reduce it. Part II is expected to be available in 2007.

[Northwest Center for Public Health Practice](#) (NWCPHP) offers *Data Interpretation for Public Health Professionals* as one of several self-paced online training modules. It is an approachable and non-threatening overview in a slide show format. Users can download slides, the presentation plus the slides, or watch the presentation (~ an hour). The presentation uses quizzes to check comprehension.

[PA Dept of Health, Health Statistics – Technical Assistance](#) *Tools of the Trade* is a series of special reports that present and discuss the various statistical measurements, analytical methods, graphics, and data collection and processing methods commonly used with health statistics. Entries vary in length; the two selected for the guide are short and sweet:

- ▶ *Entire population and sample size surveys* – a very nice introduction to the reasons samples might be used in surveys and how a population and a population sample should correspond.
- ▶ *Probability sampling* – a good introduction to the various strategies that can be used to draw a sample from a population.

[Supercourse](#) website contains many slide show presentations on a broad variety of topics related to epidemiology. For some reason, the size of the slides for all of them are limited to about a 2.5 x 3.25 inch rectangle. Some slides have accompanying comments.

- ▶ *Measures of Association* presents risk calculations (absolute, relative and attributable risk) from an epidemiological perspective
- ▶ *Rates, Ratios and Proportions* discusses the differences between rates, ratios and proportions and how they are calculated and used in public health.

[WA Dept of Health, Health Data, Data Analysis Guidelines](#): Each paper contains good information, but all are more technical and demanding than the NWCPHP presentation.

- ▶ *CIs*: Introduces confidence intervals, including when they should be used and how they are calculated in various circumstances.
- ▶ *Race*: Discusses issues associated with self-reported race and ethnicity, including appropriate questions to use and suggestions for presenting these data.
- ▶ *Rates*: Introduces rates, including what they are, the differences between crude and age-adjusted rates, and how they are calculated.
- ▶ *Small numbers*: Discusses data issues associated with small denominators and what to do about them.

VI. More good web-based resources

General

BMJ Coggon D, Rose G and Barker DJP. *Epidemiology for the Uninitiated*:

<http://www.bmj.com/collections/epidem/epid.dtl> (1997)

An on-line book that is somewhat more complex than *Statistics at Square One*: an epidemiology text instead of a statistics one.

BMJ Scrimshaw, TDV. *Statistics at Square One*: <http://www.bmj.com/collections/statsbk/> (1997)

An on-line book that has good basic information, but typographical errors interfere somewhat with comprehension.

Dallal, GE. *The Little Handbook of Statistical Practice*:

<http://www.tufts.edu/~gdallal/LHSP.HTM> (2001)

An on-line college-level book on statistics. It has well organized good information, with reasonably straight-forward explanations.

National Center for Health Statistics *Public Health Data: Our silent partner*

<http://www.cdc.gov/nchs/products/training/phd-osp.htm>

A three-module course that can be taken as a self-study or downloaded and presented in a training event. This series has good information that is well presented. The modules are not included in the guide because they are relatively long and incorporate a fairly detailed introduction to birth and death certificate data. While these are very important data sources, they are not necessarily as relevant as others for community-based program staff.

Statistics Canada *Statistics: Power from Data!*

<http://www.statcan.ca/english/edu/power/toc/contents.htm>

A great resource created by Canada's National Statistics Agency. The website has essays on the full range of data- and statistics-related topics.

Statpages.net: <http://statpages.org/>

A compilation of websites that have information about and/or can assist in calculating statistics.

Schoenbach, V. *Understanding the fundamentals of epidemiology: an evolving text*

<http://www.epidemiolog.net/evolving/>

Log in required to get to the table of contents: a quick review suggests that it includes formulas.

Information about other data collection methods

CDC Division of Diabetes Translation: *Seeing the Faces, Hearing the Stories, Learning from Our Communities: Using Focus Groups to Gain an Understanding of Living with Diabetes in Various Communities Chapter 3: Conducting Effective Focus Groups*

<http://www.cdc.gov/diabetes/pubs/focus/conducting.htm>

A very good web publication on focus groups including developing a plan, logistics, developing a discussion guide and analyzing results

University of Illinois Extension Program Planning and Assessment– *Key Informant Interviews*

<http://ppa.aces.uiuc.edu/KeyInform.htm>

A series of fact sheets about all aspects of conducting key informant interviews

Web-based calculators

Research Randomizer <http://www.randomizer.org/>

Provides randomly selected numbers within a range based on parameters that users set.

The Survey System *Sample Size Calculator* <http://www.surveysystem.com/sscalc.htm>

Gives users the size of the sample needed to generate a result that has a specified likelihood of being within a specified range. (Users decide what confidence interval and confidence level they want to achieve and indicate the size of the population they are studying, and the calculator produces the size of the sample that needs to be taken.)