

Fall  
2011

## Public Health 691F: Practical Data Management and Statistical Computing

Final Exam

**DUE: Friday, December 16 by noon.**  
Late exams will not be accepted. Early ones will be.

---

This exam uses data from a study of functional status change among patients who have undergone **percutaneous coronary interventions** (PCI) at Baystate Medical Center. **Data made available here form only a subset of the complete study data and are for use in this exam only. They are not to be used for any other purpose.**

### Introduction

Diabetics make up a disproportionate number of the over 1 million patients who undergo PCI each year in the United States. The goal of this study is to compare functional status outcomes following PCI among diabetic and non-diabetic patients undergoing these procedures.

Between September 1, 1995 and December 31, 1999 patients undergoing elective PCI at Baystate Medical Center were asked to complete a standardized functional status questionnaire, the SF-12™, on the day of surgery. They were later mailed another copy to complete at ~6-months post-surgery. The SF-12™ measures both mental and physical function, and is a standard scale used in many medical studies to evaluate functional status. During the study period, post-surgical chart audit was used to collect data on patient demographic characteristics, co-morbidities, and surgical complications.

### Data

Data are available in SAS format. There are 3 data files:

1. Data from the clinical chart audit ([clindata.sas7bdat](#))
2. Summary scores from the pre-operative SF-12™ ([preSF12.sas7bdat](#))
3. Summary scores from the post-operative SF-12™ ([pstSF12.sas7bdat](#)).

The data files have not yet been checked for duplicate or problem records.

You are asked to prepare the data for analysis in SAS. **That is**, create a labeled, formatted analysis dataset, including appropriate handling of missing values, and provide a report for the investigator summarizing the data that is available for analysis, and information on cases excluded and reasons for exclusion.

Your final analysis dataset should include only:

1. Patients with complete SF-12™ data for **both** pre- and post- operative evaluations
2. Patients undergoing PTCA (percutaneous angioplasty) only -- no other procedure
3. Patients **without** peri-operative MI or peri-operative cardiac arrest

Data include a unique study ID number for each patient that can be used to link data across files. Many patients elected not to respond to the survey at one or both occasions and did not turn in forms at that occasion. **Only patients who turned in a survey form are included in the SF-12™ dataset at that occasion.** In addition, some patients turned in only partially complete forms, so that summary scores could not be computed using the standard algorithm. In such cases, the mental function (MF) score and/or the physical function (PF) score is missing, though a record for the patient is included in the dataset.

In addition to creating an analysis data file, write a brief report that summarizes the data available for analysis and cases eliminated for various reasons. In particular, your report should address the following questions:

- How many patients underwent PCI (all types) during the study period?
  - Among these, how many patients underwent PTCA only?
  - Among all PTCA only patients, how many had poor outcomes: peri-op MI or cardiac arrest?
- How many patients turned in SF-12™ forms:
  - at pre-op?
  - at post-op?
  - on both occasions?
- How many were complete:
  - at each occasion?
  - at both occasions?
  
- How many patients are included in final analysis file: those with complete scores at both occasions **AND** PTCA only **AND** no peri-op MI/cardiac arrest?
  
- In the analysis file, for each subject compute change in functional status score (SF-12 change) from pre-op: (post minus pre) for both mental and physical function.
  
- Using the analysis file, present summary statistics **OR** graphical displays on the distributions of race, age, gender, pre-operative functional status scores, and change in functional status -- **by diabetic status.**
  
- In addition present a table reporting on the number and percent of diabetic and non-diabetic patients with each of the comorbidities and complications of surgery.
  
- **Note: You are not asked to perform any statistical tests, merely to describe and present appropriate summary data.**

## Guidelines for completing the exam:

**You must do your own work on this exam.** You may discuss strategies with your classmates, but you must do all your own programming, and write your own report. All questions concerning the exam should be directed to the course instructor.

Permissible discussion with classmates includes:

- general strategies for organizing your work
- strategies for dealing with problem data
- strategies for programming -- e.g., type of procedure or data step to use
- helping troubleshoot programs that aren't working -- reading error messages, finding problems that don't lead to error messages, suggesting ways to correct

**Help that is not permissible:**

- copying programs or sections of programs directly from classmates
- copying output, using results produced by another student
- writing programs or reports with other students
- getting help from other faculty, consulting services, other students not in your class, or other professionals

When in doubt, talk to me, send me an email.

Your report, addressing the above questions, should be neat, coherent and focused. The report should be 5-10 pages, inclusive of graphs, tables and charts. I won't read beyond 10 pages.

In addition, **attach an appendix** to your report containing:

- A table documenting your programs:

Program Name	Date run	Purpose	Data read in	Data created	Data Description
--------------	----------	---------	--------------	--------------	------------------

- Copies of your program **logs** (please -- logs only, not programs plus logs). These should be **from final edited runs of the program, not everything you tried and discarded along the way**. If you use put statements that generate long lists of records in the log -- do not include these. In other words, don't give me pages and pages of stuff to lug around and sift through to find the work that you did.

**I do not want any additional output tables or lists that are not used directly in the report. The exception could be inclusion of proc contents output of your final dataset as part of the appendix.**

The chart review data are contained in the file [clindata.sas7bdat](#). The data correspond to the following variables:

Variable Name	Description	Format / Codes
IDNO	study ID number	4-digit numeric
PDATE	Procedure Date	mm/dd/yy
SEX	patient gender	character data: M, F
Race	patient race/ethnicity	1-digit numeric code: 1 Caucasian 2 Black 3 Native American 4 Asian or Pacific Islander 5 Hispanic 6 Other 9 Unavailable
AGE	patient age in years	numeric
INTV	Intervention (PCI type)	character: PTCA CABG Both
<b>Comorbidities:</b>		
diab	Diabetic status	1-digit numeric code: 0 No 1 Yes
pvd	Peripheral vascular disease	1-digit numeric code: 0 No 1 Yes
copd	Chonic obstructive pulmonary disease	1-digit numeric code: 0 No 1 Yes
cancer	Cancer	1-digit numeric code: 0 No 1 Yes
hypercho	Hypercholesterolemia	1-digit numeric code: 0 No 1 Yes
hyperten	Hypertension	1-digit numeric code: 0 No 1 Yes
smoker	Smoker	1-digit numeric code: 0 No 1 Yes
<b>Complications:</b>		
occdur	Occlusion during surgery	1-digit numeric code: 0 No 1 Yes
occaft	Occlusion after surgery	1-digit numeric code: 0 No 1 Yes
mi	Peri-op MI	1-digit numeric code: 0 No 1 Yes
cardiaca	Peri-op Cardiac Arrest	1-digit numeric code: 0 No 1 Yes
tia	Transient ischemic attack	1-digit numeric code: 0 No 1 Yes
cva	Stroke	1-digit numeric code: 0 No 1 Yes
transfus	Transfusion required	1-digit numeric code: 0 No 1 Yes

The pre-operative SF-12™ data are contained in the file [preSF12.sas7bdat](#). The data correspond to the following variables:

Variable Name	Description	Format / Codes
IDNO	Study ID number	4-digit numeric
MF1	Mental Function Score	numeric
PF1	Physical Function Score	numeric

The post-operative SF-12™ data are contained in the file [pstSF12.sas7bdat](#). The data correspond to the following variables:

Variable Name	Description	Format / Codes
IDNO	Study ID number	4-digit numeric
MF2	Mental Function Score	numeric
PF2	Physical Function Score	numeric