

Prostate Clinical Outlook (PCO) Technical Manual

Table of Contents

1	FUNCTIONAL REQUIREMENTS	3
1.1	USER REQUIREMENTS	3
1.2	DATA FLOW DIAGRAMS	7
1.3	FUNCTIONAL REQUIREMENTS	7
2	OTHER REQUIREMENTS	10
2.1	INTERFACE REQUIREMENTS	10
2.2	HARDWARE INTERFACES	10
2.3	DATA CONVERSION REQUIREMENTS	10
2.4	HARDWARE/SOFTWARE REQUIREMENTS	10
2.5	OPERATIONAL REQUIREMENTS	11
2.6	DATA RETENTION	11
2.7	VALIDATION RULES	11
3	TEST CODES	12
	APPENDIX A - GLOSSARY	15

1 FUNCTIONAL REQUIREMENTS

1.1 User Requirements

0) Pre-Requisites: Preparing MAC machine which has the R-Studio, MySQL and R Packages listed in the 4.4 Hardware/Software Requirements.

1. Operating System: Mac OS X or Linux based OS
2. Database: MySQL Database(initial userid/passwd/dbname) : shiny/shiny(CDSS))
3. R installation(ref:<https://cran.r-project.org/>))
4. Pandoc installation(ref:<http://pandoc.org/installing.html>)

1) Setup: After unzip the setup file for the PCO on MAC machine, it needs to setup a root directory configuration. Finally make the PCO running script as a background process when the MAC starts. The particular setup such as background process depending on the machine and operating system version that here not described.

1. download PCO source to local file system
2. mkdir -p /srv/shiny-server/PCO
3. cp -rf [PCO download path]/* /srv/shiny-server/PCO/
4. cd /srv/shiny-server/PCO
5. Rscript ./install_libraries.R
6. sh ./PCO
7. connect to the URL: <http://localhost:3838> or <http://127.0.0.1:3838>

2) Login: User log in the MAC machine, Future system will have a web based log in UI.

3) Turn on web-based UI: Simply turn on the machine and run the web browser will show the front page of the PCO. If user could not see the PCO screen then try this URL (<http://localhost:3838>) on the web browser. The current version of PCO was tested successfully using modern browser such as Safari, Chrome Browser, and Mozilla Firefox.

4) Data input UI: There is input form on the screen, user can change patients' information: age, clinical stage, Gleason score, and PSA. The outcome graph, on the right side of the screen, changes itself instantly based on the input patients' data.

Doctor's Name
Anatoly Dritschilo

Next Patient

Patient's Name
Patient Name

Age
60

PSA
10

Clinical Stage
T1a
T1b
T1c
T2a
T2b
T2c
T3

Gleason Score
2
3
4
5
6
7=3+4
7=4+3
8
9
10

Calculate

Reference Patients Info.
Relapse free analysis was based on 809 patients
Quality of life analysis was based on 809 patients

Graph Min #
Minimum number of patients in a time slot
0 5 10

Check this and Push Calculate button for reloading data(it takes around 30sec.)
☐ Reloading ON/OFF

☒ First Load

Figure 27 Patients' information

a) Age: patient's age at diagnosis – in the Figure 15 the age was considered as a 60s age category. The PCO instantly gathered the same age category data from the existing data.

b) Clinical stage: 8 clinical stage (T1a, T1b, T1c, T2a, T2b, T2c, T3) – There were 2 patients in the T3x category. 2 of patients data is not enough to show the patient specific data. The T3x category will be considered as T2c category that shows T2c + T3x 2 patient's outcomes on the screen. However, If patient's clinical stage is T2c that will only gather T2c data to show outcomes.

c) Gleason score: Sum value of Gleason score (2, 3, 4, 5, 6, 7=3+4, 7=4+3, 8, 9, 10) – Gleason score and PSA were calculated into risk category (Ch.2. Methodology, Table 5). The PCO gathered same risk category data from the existing patients' data.

d) PSA:

e) Graph Min #: It filtered data by the number of patients that makes the likely outcome plots on the screen.

Every change in the input form will activate analysis script in the PCO that instantly changes patients' specific outcome on the screen within 5 seconds.

5) View likely outcomes: The likely outcome screen has survival curve and QOL plots.

a) Survival curve: BCR-free survival curve

The red line presents all existing patients' survival curve and the blue line presents patient specific survival curve. Those small vertical lines along with horizontal line means censored number of patients. Because of this graph had BCR-free survival patients data, it was not affected the number of patients filter in the patient's information input form. The strata table had number of patients, for each days of time point, calculated by eliminating the number of censored patient.

1. Relapse Free Survival Probability

(* Relapse: Biochemical Relapse)

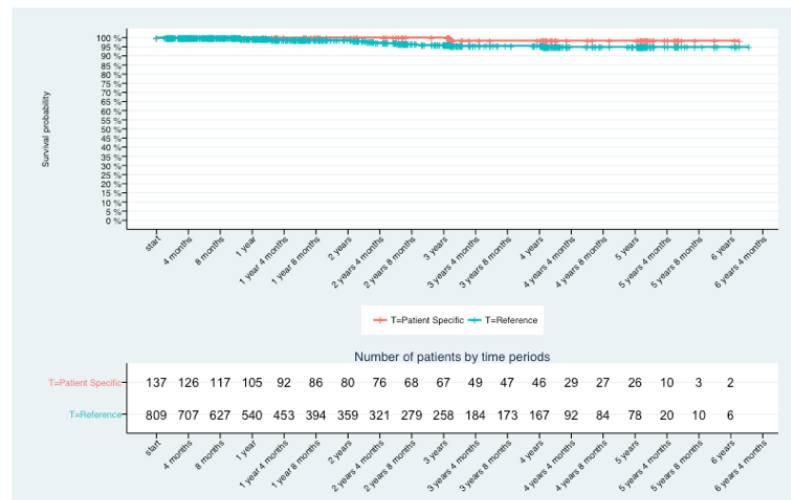


Figure 28 Disease free survival curve

b) QOL plots: QOL plots have results 1) Rectal QOL, 2) Risk of rectal bleeding, 3) Urinary QOL, 4) Risk of urinary incontinence, 5) Risk of urinary irritation, 6) Risk of urinary bleeding, 7) Sexual QOL, 8) Risk of erectile dysfunction

In the following plot has an example of QOL (Rectal QOL). The red line represents all existing patients' rectal QOL trend, higher value indicated better status than lower value, while the blue line represents patient specific trend that is gather data from existing patients' who has same age, clinical stage and risk category shown in Ch.2 Methodology.

The risk plots show same colors of section, the red line for all existing patients and the blue line is patient specific data. However, the meaning of values in the risk plots have opposite meaning to the QOL value, lower value indicated better status than higher value. Because of the QOL plots showed satisfaction or positive outcome values while the risk plots show negative risk outcome. That was why QOL plots and risk plots had different meaning of values.

2. Rectal Quality of Life

2.1 Rectal Quality of Life Overall

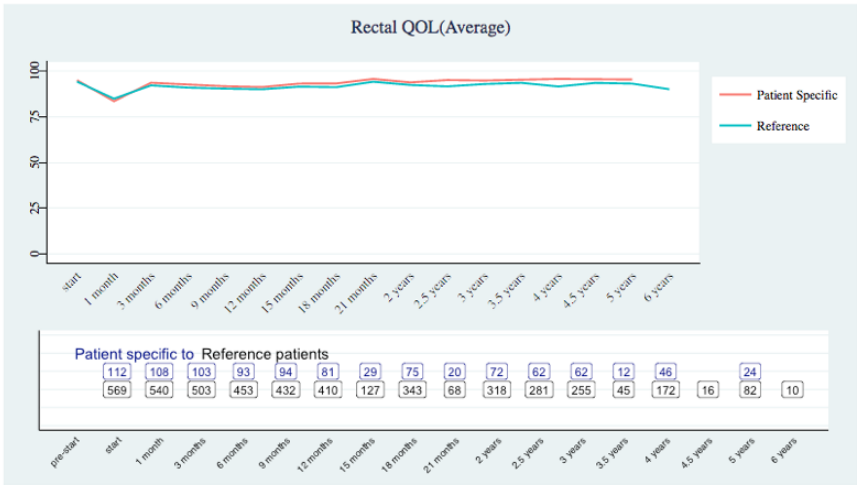


Figure 29 Example of Rectal QOL

6) **Generating the report:** PDF report file has patient’s information and outlook of risk free profile. – For the security reason the PCO was not connected to any network or intranet. All report could be printed by local connected print which was not connected to any network.

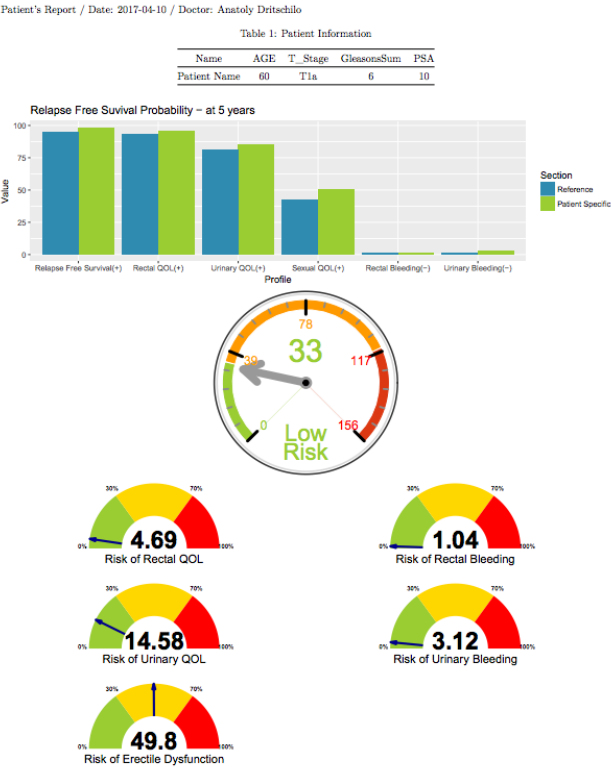


Figure 30 Example of report

1.2 Data Flow Diagrams

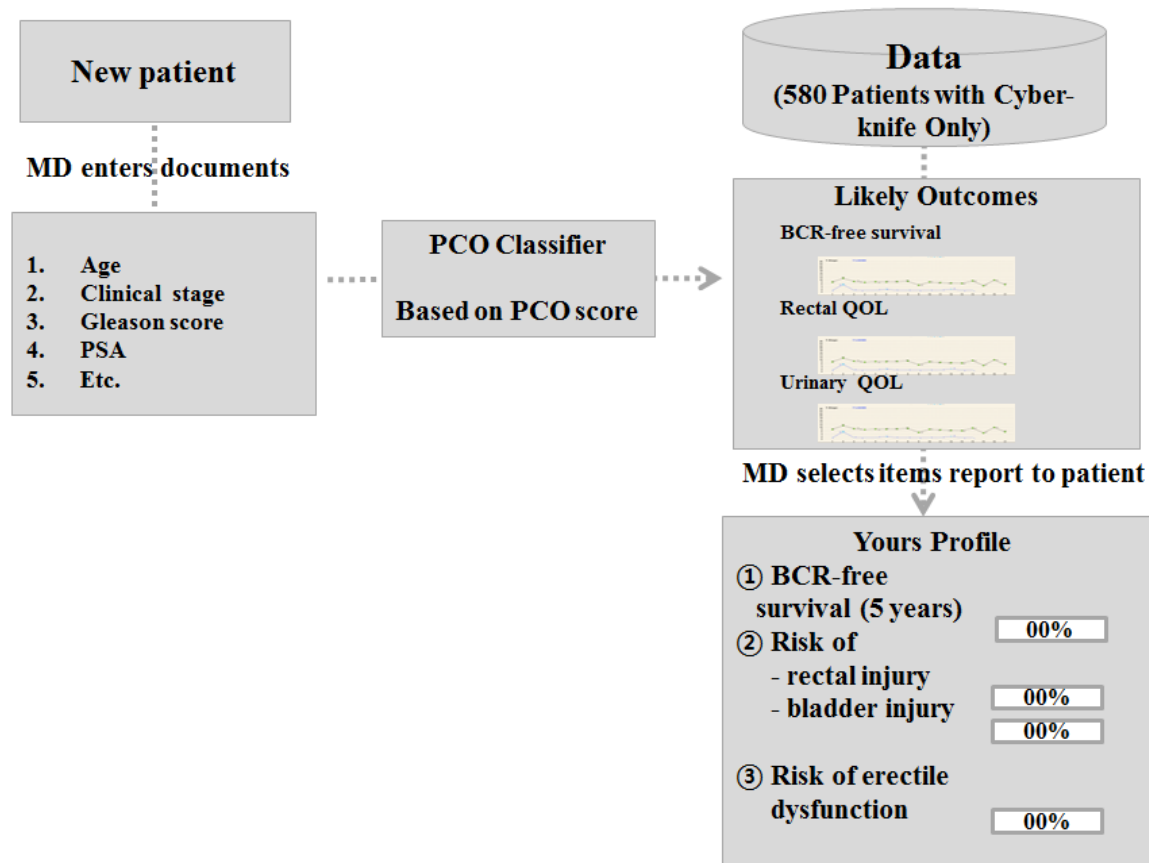


Figure 31 Data Flow Diagram

1.3 Functional Requirements

1.3.1 Data Acquisition

Source data could be composed as a data from Excel file (.xlsx) or comma-delimited file format (.csv). It is loaded into the MySQL database server which was installed in the MAC PC. The origin data is already scrubbed with cleansing criteria. Once the database has cleansing data, it resides on the system memory for fast access. The PCO did not aware whether the data was redacted or not. The original raw data must be redacted before move into the system directory.

1.3.2 Data Preprocessing

According to the Ch.2. Methodology in this document, R analysis engine preprocesses the raw data using the categorizing criteria. It makes character string value into number value and continuous number value into categorized value. If user adds new data or makes changes existing

data, the data preprocessing function, in the PCO, automatically repeat the process for new or changed data.

The data preprocessing function activated in following conditions.

- a. Check reload on button on the PCO screen with web browser.

The PCO shows data loading status on the top or bottom line of the web browser screen.

1.3.3 Web Interface

Users may insert new patient's information such as age, at diagnosis clinical stage, Gleason score, and PSA level. It translates the input data to age at diagnosis, clinical stage, Gleason score, and PSA group to select same group data from the existing data. It analyzes the selected data and show likely outcomes' plots. User also may create profile report file by the button 'REPORT'. All these function present the new outcome and report screen along with user data input changed. All web interface functions were developed by R scripts. Developers of the PCO only need to know about the R scripts. It reduced cost for development and management of the system, because the system manager of developer doesn't need to know about a Linux system, HTML or Java-scripts.

1.3.4 Data Analysis

The analytics algorithm, powered by program R, is installed in the PCO. It analyzes new patient's likely outcomes such as probability of disease-free survival years and risk-free survival years after radiation therapy.

1.3.5 Data Management

The data from the hospital, which is anonymized, was stored in the MySQL database in the MAC PC. User can access the data through the authenticated by the MySQL database server. Once the data was loaded into the database the original raw data does not required anymore, because everything is in the server-side, that the raw data can be safely removed for security reason. All the result and reports for specific patients are not stored to the MySQL database, it is only can be stored as a file such as PDF, WORD, and HTML. The latest version of PCO generate only PDF files for download and email functions.

1.3.6 User Authentication

User authentication is performed by OS authentication. If the system extended to server version, then it needs to use log-in web interface for the authentication. Also it needs user verification function for the log-in interface.

1.3.7 Audit Trail

If user views and makes reports, then all these activities are recorded to database. Administrator can manage the log in the database.

2 OTHER REQUIREMENTS

2.1 Interface Requirements

2.2 Hardware Interfaces

The PCO's both front-end and back-end is on the same machine. It is a standalone system that is not connected to outside network for the security reason.

2.3 Data Conversion Requirements

When the PCO starts, the local patient raw data will be loaded into the PCO. It preprocesses the raw data automatically by the pre-configured conditions. Users do not have to handle the data manually the system handle it.

2.4 Hardware/Software Requirements

The PCO works on Linux based Operating Systems (OS). The application server runs on the R language and Shiny software. It is optimized to the MAC OSX now, but compatible with any Linux OS.

The PCO was developed with OSX El-Capitan and MySQL5.

The PCO used R language in the analysis scripts and the R/Shiny web server architecture in the front/back end.

Table 25 Version of required R packages

Package	Version	Package	Version	Package	Version
DT	0.2	Nlme	3.1-128	shiny	0.14.1
formattable	0.2.0.1	openxlsx	3.0.0	shinyjs	0.7
GGally	1.2.0	partykit	1.1-1	shinythemes	1.1.1
ggplot2	2.1.0	Plotly	4.4.2	stringr	1.1.0
ggthemes	3.2.0	rmarkdown	1.1	survival	2.40-1
grid	3.3.1	RMySQL	0.10.9	survminer	0.2.3
knitr	1.14	Scales	0.4.1	xtable	1.8-2
mailR	0.4.1	sendmailR	1.2-1	V8	1.4

2.5 Operational Requirements

It has easy to understandable web interface. Simple input interface for new patient's data. An analyzed result presented by sophisticated plots and tables. By the report button, easily make the report as various formats such as PDF, WORD, and HTML.

User can access the application code, raw data or MySQL database, however, incorrect manipulation can cause a malfunction of the application.

The PCO has 3 directories such as Shiny, App-1 and Data. Shiny is the root directory. Sub directory App-1 has R scripts for web application and statistical analysis functions. Sub directory Data has raw data. Manual adjustment of any file or directory in the root directory and sub directories is not permitted.

2.6 Data Retention

The PCO follows IRB regulation of the Georgetown University Medical Center. The system shall retain application information for any year user wants unless the storage is corrupted.

2.7 Validation Rules

The PCO system validates the likely outcomes with the results of R analysis scripts and MySQL queries to compare result numbers between database and analysis results such as BCR free survival rate, risk scores and risk free probabilities.

If some of the likely outcome has wrong presentation number on it, then you can check it with MySQL data and analyze it with R analysis scripts. This way helps developers and statistician to be confident with their results of analysis and presentation.

3 Test Codes

Run test code for data management, visualization, and report generation.

If it shows error on the output file UNIT_TEST_*.txt, please see the test target file in the output file “Unit Test For: [Test Target File]”.

The test codes have representative functions for the PCO. When all test has done without error, the PCO will work properly.

```
$ cd /srv/shiny-server/PCO/TestCode
```

```
$ sh DataManagementTest.sh
```

```
$ cat UNIT_TEST_DATAMANAGEMENT.txt
```

```
=====
= PCO UNIT TEST for Data Mamangement;Data loading and preprocessing =
=====
= PCO UNIT TEST for Data Mamangement: STEP 1 =
=====
[1] TRUE

[Done] Unit Test For:  mysql_init_data.R
=====
= PCO UNIT TEST for Data Mamangement: STEP 2 =
=====
[1] TRUE

[Done] Unit Test For:  cleansing.R
=====
= PCO UNIT TEST for Data Mamangement: STEP 3 =
=====
[1] TRUE

[Done] Unit Test For:  SAVE_SURVIVAL_DATA2.R
=====
= PCO UNIT TEST for Data Mamangement: STEP 4 =
=====
[1] TRUE

[Done] Unit Test For:  EPIC_score.R
=====
= PCO UNIT TEST for Data Mamangement: STEP 5 =
=====
[1] TRUE

[Done] Unit Test For:  QOL_score.R
=====
```

```
$ sh VisualizationTest.sh
```

```
$ cat UNIT_TEST_VISUALIZATION.txt
```

```
=====
= PCO UNIT TEST for Data Visualization;Visualization of Likely outcome=
=====
= PCO UNIT TEST for Data Visualization: STEP 1 =
=====
[1] TRUE

[Done] Unit Test For:  GAUGE_CHART.R
=====
=====
= PCO UNIT TEST for Data Visualization: STEP 2 =
=====
[1] TRUE
=====
Warning on:  OverallVSPProfile.R
Here's the original warning message:
We recommend that you use the dev version of ggplot2 with `ggplotly()`
Install it with: `devtools::install_github('hadley/ggplot2')`
[Done] Unit Test For:  OverallVSPProfile.R
=====
=====
= PCO UNIT TEST for Data Visualization: STEP 3 =
=====
[1] TRUE

[Done] Unit Test For:  DiseaseFree_Risk.R
=====
```

```
$ sh ReportTest.sh
```

```
$ cat UNIT_TEST_REPORT.txt
```

```
=====
= PCO UNIT TEST for Data Report;Report of Likely outcome=
=====
= PCO UNIT TEST for Data Report: STEP 1 =
=====
[1] TRUE

processing file: report.Rmd
|.....| 9%
  inline R code fragments

|.....| 18%
label: table1 (with options)
List of 1
 $ echo: logi FALSE

|.....| 27%
  ordinary text without R code

|.....| 36%
label: profile1 (with options)
List of 8
 $ collapse : logi TRUE
 $ echo     : logi FALSE
 $ warning  : logi FALSE
 $ error    : logi FALSE
 $ prompt   : logi FALSE
 $ message  : logi FALSE
 $ fig.width: num 10
 $ fig.height: num 3

|.....| 45%
  ordinary text without R code

|.....| 55%
label: profile2 (with options)
```

<output screen picture continues on next page>

```
|.....| 91%
label: riskprofile (with options)
List of 9
$ collapse : logi TRUE
$ echo     : logi FALSE
$ warning  : logi FALSE
$ error    : logi FALSE
$ prompt   : logi FALSE
$ message  : logi FALSE
$ fig.width : num 8
$ fig.height : num 4
$ fig.align : chr "center"

Loading required package: gridExtra
|.....| 100%
ordinary text without R code

output file: report.knit.md

/usr/local/bin/pandoc +RTS -K512m -RTS report.utf8.md --to latex --from markdown+autolink_bare_uris+ascii_identifiers+tex_math_single_backslash --output report.pdf --template /Library/Frameworks/R.framework/Versions/3.3/Resources/library/rmarkdown/rmd/latex/default-1.17.0.2.tex --highlight-style tango --latex-engine /Library/TeX/texbin/pdflatex --variable graphics=yes

Output created: report.pdf

[Done] Unit Test For: report.Rmd
=====
```

APPENDIX A - GLOSSARY

Biochemical recurrence (BCR)

Clinical Decision Support System (CDSS)

National Comprehensive Cancer Network (NCCN)

Expanded Prostate Cancer Index Composite (EPIC)-26

Health-related quality of life (HRQOL)

Institutional Review Board (IRB)

Memorial Sloan Kettering Cancer Center (MSKCC)

Prostate-specific antigen (PSA)

Prostate Clinical Outlook (PCO)

Quality of life (QOL)

Receiver-operating curve (ROC)

Stereotactic body radiation therapy (SBRT)

The American Society of Therapeutic Radiology and Oncology (ASTRO)

The concordance index (c-index)

TNM classification: Tumor (T), lymph nodes (N) and secondary cancer (metastases: M)

User interface (UI)