|  |  |
| --- | --- |
|  | **August, 2019** |
|  | **PMH Biostatistics** |

|  |
| --- |
| **[Training MANUAL]** |
|  |

**Table of Contents**

1. **Introduction**
2. **Administrative**
3. **Statistical**
4. **Important Resources**
5. **Conclusion**

**Introduction**

The purpose of this training manual is to provide new staff members within the Biostatistics Department a resource to assist them in making the transition to their new role as professional biostatistician on clinical (and translational) research at Princes Margaret Cancer Centre.

**Administrative**

**New employees should be introduced to the following administrative details by the department manager or designate as part of their orientation to the department:**

**(**See T:\Dept Admin\New EmployeeInformation foradditional up-to-date information)

1. Introduction to UHN (University Health Network), PMH/OCI (Princess Margaret Cancer Centre/Ontario Cancer Institute), RSS (Research Solutions & Services), CCRU (Cancer Clinical Research Unit)
2. Review of job description
3. Review of Standard Operating Procedures
4. Research and Corporate Intranet
   * **Research:** [**http://intranet.uhnresearch.ca/**](http://intranet.uhnresearch.ca/)
   * **Corporate:** <http://intranet.uhn.ca/applications/inews/default.aspx>
5. Virtual Library and PMH Library
   * **Virtual Library:** <https://guides.hsict.library.utoronto.ca/UHNvirtuallibrary>
   * **PMH Library:** Princess Margaret Cancer Centre/Ontario Cancer Institute Health Sciences Library (5th floor)
6. Important phone numbers
   * **Emergency:** 5555
   * **RIS:** (Research Information Systems) (For computer issues): 2321 or 18-7777 or 416-581-7777
     + email: [helpdesk@uhnresearch.ca](mailto:helpdesk@uhnresearch.ca).
     + RIS office: PMH 8-131
   * **SIMS:** (Shared Information Management Services) (For phone issues): 14-4357
   * See the staff directory folder in the share drive for all department member extensions, office locations, and emails.
   * To dial non-UHN numbers dial 9-phone number
7. Contacting UHN staff in peripheral buildings
   * **PMCRT/Mars:** Dial 18 – extension
   * **TGH:** Dial 14 – extension
   * **TWH:** Dial 13 – extension
   * **PMH/Hydro:** Dial 16 – extension
8. ID Badge Request
   * A photo ID security card is required to be worn at all times while working/on UHN property
   * To receive your badge, after a manager’s request has been submitted go to the card issuing office at TGH Eaton South Basement Room 426A across the hall from the main security office (M,T,W 7:30-12:00,13:00-15:30). Badges can be received Thursdays and Fridays at TWH
   * Your UHN badge acts as an elevator access card after hours. Just tap the card on the card reader in the elevator before selecting a floor.
   * Accessing some floors in the PMDT (Princess Margaret Discovery Tower) or Hydro building will require additional badges. **See :\Dept Admin\New Employee Information\information for new employees**, for more information
9. Key to office
   * If there are available keys in the department you will be provided with one
   * If there are no keys available you will need to apply for one at:[http://intranet.uhn.ca/departments/security\_operations/key\_lock\_request/form.asp](http://intranet.uhn.ca/departments/security_operations/key_lock_request/form.asp%20%20)
   * Make sure to lock the main door and turn off all lights if you are the last person in the office
10. Office / workstation location
11. Office / workstation telephone number and long distance code
    * Ask for long distance code if required
12. Computer username and Research Information Systems training
13. Accessible printers and location
    * Photocopier on 7th and 9th floors (Ask for code if needed)
    * Black and white printer in department
14. Important department files
    * For any concerns speak to a department Senior Analyst.

* Time Log
  + Refer to activity codes
  + Location on shared drive
  + Description
  + Uses
* Project status file
  + Description
  + Coding a new project
  + What do we mean by project?

1. Appliances available for use
   * Microwave for use on the 7th floor
2. Set up meeting with department programmer
3. Set up meeting with mentor (if applicable)
4. Additional resources can be found in the share drive under T:\Dept Admin\New Employee Information

**Standard Operating Procedures**

Currently there are SOPs for clinical trials, observational studies, and translational studies. (Add locations of SOPs)

**Committees**

Within the PMH Biostatistics department there are a number of committees which keep the department running efficiently. At some point you may be asked to join one of these committees, or if you are interested in joining a specific committee feel free to contact the committee head. Below is a brief description of each committee.

* Standard Operating Procedures (SOP) committee

The purpose of the SOP committee is to develop/revise standards appropriate for biostatisticians and analysts in our department. All standards shall promote the efficient and consistent operation of our department.

* Knowledge Transfer Round (KTR) committee

The KTR committee arranges information sessions and round table discussions presented by members of the department or outside guest speakers. They also keep track of seminars and webinars which may be of interest to members of the department

* Database committee

The database committee works to maintain, build, clean, transfer, and output databases using MS Access, REDcap, Excel, Medidata Rave, and third party applications (eg. Oracle, SQL server, or application from NCI United States). The database committee also provides guidelines to users about properly recording data values in excel, and training for databases built in the department.

* Computation and Programming Standardization Committee

The functions of this committee are to generate, improve, and standardize computational functions and macros for biostatistics research. It aims to improve the efficiency and quality of research activities.

* Supply committee

The supply and internal communication committee functions to improve the work *e*nvironment of the department. From ordering supplies, equipment and services to monitoring the usage and maintaining of them. The committee will provide access and guidance for the department members as such needs arise.  It aims to create a sense of smoothness in our daily activities. The specifics are:

1. Regular/minor supplies
2. New equipment and large supplies
3. Requests/forms for internal (UHN) help/service

**Computing**

**Computing is an essential tool within the Biostatistics Department. A new employee orientation should delve into our computer configuration and point out statistical software available. Finally, when should the Research Information Systems Department be called upon for assistance?**

1. **Introduction to computer drives** (The specific drive locations change often; ask a department Senior Analyst)
   * Hard Drives
   * Personal Network Drive
   * Shared Drives
2. **List of available software**
   * SAS for data cleaning and analysis
   * R for data cleaning and analysis
   * PASS for sample size and power calculations

***Talk to a department*** Senior Analyst ***if additional software is required***

**Computing resources**

**R:** <http://www.r-project.org/> and <http://cran.r-project.org/doc/manuals/R-intro.pdf>

**SAS**: <http://support.sas.com/> (offers free and paid online course) and <http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_913/base_step_10071.pdf>

**PASS:** <http://www.ncss.com>

1. **ReportRX**
   * The reportRx system is used with Rstudio to create statistical reports as word documents and pdf files. Using the system results in easy to create, visually appealing, and consistent reports. Using the system allows for all members of the department to use the same formatting.
   * The R package is maintained by the Computation and Programming Standardization Committee within the department
   * See the ReportRX files in the share drive for more information on installation and usage

4. Computational function library

Should include both SAS and R function library

1. **Structuring your directories** 
   * All files should be stored in a simple consistent manner
     + First by PI
     + Then by Project Code within PI
     + Each project code directory should be further divided, suggested folders include
       - Analysis
       - Data
       - Results/Outcomes
       - Reports
       - Further subfolders by dates
     + Adding dates/version numbers in file names is an easy way to keep track of changes to a project

**Statistical**

**The biostatistician is called upon to provide researchers with design and analysis advice. This invariably includes interaction with researchers, trying to determine what questions the researcher is asking, and how best to answer these questions using an appropriate statistical framework. The biostatistician will usually be called upon to provide a written report of his/her findings, using non-technical language. Below are some considerations for the new biostatistician.**

1. **Effective elements of statistical consulting**

* Perspective from Sir David Cox <http://www.ssc.ca/en/statistical-resources/some-general-remarks-consulting>
* If possible collaborate (i.e., work with an investigator over a period of time) rather than consult (i.e., some occasional discussion of very specific statistical issues with the investigator).
* Be interested in the subject matter involved.
* Aim to use the terminology of the subject matter field where it differs from common statistical usage.
* If collaborating, go to subject matter seminars from time to time, and read journals in the field.
* Discreetly determine how much understanding of statistical issues the investigator has. Mechanical use of significance tests to confirm overwhelming effects for example is a bad sign.
* Frequently review what is being done to check that the statistical analysis addresses the correct questions. This may help the investigator clarify thinking as well as protect against the most common error in statistical work – answering the wrong question.
* Aim, if feasible, to see some raw data, to understand the measurement processes involved, and to have some appreciation of the general quality of the data.
* Enquire into aspects of the study design that might have a bearing on the appropriate analysis.
* Begin with very simple methods.
* If possible, end with simple methods.
* Since preliminary ideas for analysis often do not work the first time, be prepared to do modifications.
* Take considerable care over presentation of conclusions.
* If your work is to be acknowledged in a paper or a report, ask firmly to see what is written before it is submitted.
* If you feel you should have been a co-author and have not been invited to be, pause for a few days. If, on reflection, you still feel the same, speak quietly to the friendliest of the investigators pointing out, assuming it is true, that you have spent a lot of time and thought on the work.
* Occasionally; very rarely one hopes; be prepared to say that the data are incapable of throwing useful light on the issues involved.
* Find a good balance between thinking things out for yourself and obtaining advice from statistical colleagues (and, of course, therefore, finding time to help them in return).
* If more than ten per cent of what you do ends up by being directly useful, you are doing well.
* If the investigator begins by saying he has a trivial little problem which they are sure you will be able to sort out immediately, don’t altogether believe him!
* For developing statistical consulting skills see <http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_C186_SHARPLES.pdf>
  + See also Statistical Consulting (Cabrera and McDougall, 2002)

1. **Data manipulation - preliminary analysis**

* If an investigator is going to be providing you with a database, it is important to give the investigator a list of formatting requirements for the data. This will save you from spending a lot of time cleaning the data yourself. For some examples of formatting requirements see the Data entry specifications file (Location of database committee document)
* During the course of your analysis an investigator may ask you to make changes to the data (eg. additional exclusions, changing a last follow up date). Keeping a log of all of these changes and asking for confirmation from the investigator will prevent future confusion. This is especially important in studies with multiple investigators or with complex data changes.
* Check consistency between source data file and imported SAS/R data set (e.g. format difference)
* Check for duplicate observations
* Consistency checks including date sequencing checks (e.g. diagnosis date comes after date of birth, last follow up date before diagnosis date)
* Range checks (identifies implausible or extreme values)
* Logical checks (e.g. Is the subject in a clinical trial? If so, which? If answer to question 1 is “No” then no trial should be identified in question 2)
* Missing values
* Violations of assumptions (e.g. Normality)
* For multiple data source, check the identity system, overlap of the samples, and inconsistency of the variable names before merging
* Check inconsistencies between last follow up date and patient death date

1. **Sample size considerations and power analysis**

* How many patients do I need? This is probably the most common request from a clinical researcher. The answer, of course, depends on several items including the size of the effect and the magnitude of the Type I and II errors.
* PASS is an effective software for calculating the sample size in a variety of situations

1. **Study designs encountered**

The most common study types encountered by biostatisticians are observational studies, translational studies, and clinical trials.

1. Observational

* Many of our projects fall under this category. Often these are retrospective reviews with a hypothesis generating objective. Observational studies may have a host of potential biases that the researcher/statistician need to be cognizant of. A good reference on this subject is “Bias in Analytic Research” by D. Sackett in *J. Chronic Diseases*, 1979, pp 51 – 63.
* Case-control
  + These are retrospective in nature. For more details and a list of advantages and disadvantages please refer to Designing Clinical Research (3rd edition) S. Hulley, S. Cummings, W. Browner, D. Grady and T. Newman.
* Cohort
  + These are prospective in nature. For more details and a list of advantages and disadvantages please refer to Designing Clinical Research (3rd edition) S. Hulley, S. Cummings, W. Browner, D. Grady and T. Newman.
* Retrospective Cohort
  + Such as char review study

1. Translational

Translational research is a process of determining a treatment solely on the basis of molecular biological characteristics to enhance health outcomes. Therefore, the design includes both clinical as well as biological characteristics such as genetics, DNA, SNP etc. This type of data, requires strong collaboration with clinicians as well as other scientists (eg bioinformaticians, Pathologists) to put the data into context and thoroughly explain limitations.

**Resources for statistical genetics**

* + Introductory genetics - Gonick L, Wheelis M (1991). Cartoon guide to genetics.
  + Human Heredity, Principles and Issues. Michael R Cummings. Brooks/Cole Publishing (6th Ed., 2003)
  + Introduction to statistical genetics - Ziegler A and Konig IR (2--6). A statistical approach to genetic epidemiology. Wiley-VCH Verlag GmbH & Co. KGaA.

1. Clinical Trials
   * Parallel Group Randomized Controlled Trial
     + These represent the gold standard in determining evidence for or against a particular treatment.

* Early Phase Trial
  + 3+3 Phase I design and two-stage one arm Phase II design.

**Resources for clinical trials**

* + Peto *et al.* *British Journal of Cancer* (Part I and II)
  + Fundamentals of Clinical Trials (Freidman, Furberg, DeMets)
  + Clinical Trials: A Practical Approach (Pocock)
  + Clinical Trials: A Methodological Perspective (Piantadosi)
  + See shared drive **ICH E9 Notes** for Statistical Principals for Clinical Trials

1. **Common Analyses**
   * Linear
     + See Multiple Regression: A Primer by P. Allison
   * Logistic
     + See Logistic regression using the SAS System by P. Allison
   * Survival
     + See Modeling Survival Data in Medical Research by D. Collett
     + See Applied Survival Analysis: Regression Modeling of Time to Event Data by D. Hosmer and S. Lemeshow
     + See Competing Risks: A Practical Perspective by M. Pintilie
   * Repeated measures
     + See Applied Longitudinal Analysis by G. Fitzmaurice, N. Laird and J. Ware
     + See Applied Longitudinal Data Analysis by J. Singer and J. Willett
     + See Analysis of Longitudinal Data by P. Diggle, P. Heagerty, K-Y Liang and S. Zeger

1. **Writing an effective statistical report**

The Introduction (or Presentation, or Background)

* + Should contain a brief presentation of the data, and the setting in which the data were collected. Further, it should contain a description of the problems or hypotheses to be addressed in the analysis. The Introduction must be written in a non-technical language that can be read and understood by non-statisticians. Statistical terminology and concepts should be avoided, if possible.

The Methods and results (or Statistical analysis)

* + Should contain a description of the model and a check of the model assumptions. It should describe the statistical procedures and methods which are used in the analysis and it should present the relevant figures and numerical results from the analysis. The Methods and results should contain enough details about the procedures and methods that are used, to allow other statisticians to repeat the analysis.

The Discussion(or Conclusion)

* + Should contain a discussion of the results of the statistical analysis in the Methods and results, relating to the problems or hypotheses posed in the Introduction. It should also discuss any detected problems about the data (for example, outliers) or the model (for example, lack of fit). The Discussion must be written in a non-technical language that can be read and understood by non-statisticians. Statistical terminology and concepts should be avoided, if possible.

1. **Common statistical errors** 
   * Not defining each variable in measurable terms
   * Not providing the level of measurement of each variable
   * Dividing continuous data into ordinal categories without explaining why or how the categories were created
   * Using the mean and standard deviation to describe continuous data that are not Normally distributed
   * Using the standard error of the mean as a descriptive statistic
   * Reporting only P values for results
   * Not confirming that the assumptions of statistical tests were met
   * Interpreting non-statistically significant results as “negative” when they are, in fact, inconclusive
   * Not reporting whether or how adjustments were made for multiple hypothesis tests
   * Confusing statistical significance with biologic importance
   * A few words on what a statistical report should not include. It should not include any computer output-as that belongs in an appendix. It should not include a long list of tables and figures with no explanations-the report should be written in prose, including only the most relevant tables and figures. And finally, it should not include too many pages-a concise and precise report is much preferable to a long and rambling report.

**Important Resources**

**During orientation the new biostatistician should be introduced to all the internal resources at their disposal. These include 1) resources to help them understand cancer terminology, treatments and other pertinent aspects of the numerous diseases that fall into the cancer domain and 2) statistical resources such as available books, journal articles and important statistical guidelines.**

**Cancer resources**

* + PMH library
  + Patient library
  + On-line resources (NCI etc.)
  + NCI web site: <http://www.cancer.gov/>

**Biostatistics library (Sign out form on printer)**

* + Encyclopedia of Biostatistics
  + Statistical texts
  + See shared drive
  + SAS manuals (version 8)

**Books in the department** (including personal libraries)

* + See shared drive

**In-house library of journal articles**

* + See filing cabinet near fax machine in 10-502B. Subject matter is in alphabetical order.

**CONSORT statement**

* + See [www.consort-statement.org/](http://www.consort-statement.org/)
  + CONSORT, which stands for Consolidated Standards of Reporting Trials, encompasses various initiatives developed by the CONSORT Group to alleviate the problems arising from inadequate reporting of randomized controlled trials (RCTs).
  + The main product of CONSORT is the [CONSORT Statement](http://www.consort-statement.org/consort-statement/overview0/), which is an evidence-based, minimum set of recommendations for reporting RCTs. It offers a standard way for authors to prepare reports of trial findings, facilitating their complete and transparent reporting, and aiding their critical appraisal and interpretation.
  + The CONSORT Statement comprises a 25-item [checklist](http://www.consort-statement.org/consort-statement/overview0/#checklist) and a [flow diagram](http://www.consort-statement.org/consort-statement/overview0/#flow), along with some brief descriptive text. The checklist items focus on reporting how the trial was designed, analyzed, and interpreted; the flow diagram displays the progress of all participants through the trial.

**ICH guidelines**

* + See <http://www.ich.org/>
  + The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) is unique in bringing together the regulatory authorities and pharmaceutical industry of Europe, Japan and the US to discuss scientific and technical aspects of drug registration. Since its inception in 1990, ICH has evolved, through its ICH Global Cooperation Group, to respond to the increasingly global face of drug development, so that the benefits of international harmonization for better global health can be realized worldwide. ICH's mission is to achieve greater harmonization to ensure that safe, effective, and high quality medicines are developed and registered in the most resource-efficient manner.

**Criteria for authorship** – International Committee of Medical Journal Editors

* + Newly practicing statisticians are usually reluctant to discuss authorship with their clinical colleagues. Ideally, authorship should be discussed at the beginning of the project, but often it isn’t discussed at all. Thankfully statisticians are usually included as co-authors. This is because design and/or analysis contribution are often given as criteria for authorship, based on significant intellectual contribution.

**Conclusion**

This manual is intended to help you understand the workings of the biostatistics department and provide you with statistical guidelines and recourses. However, you may require further clarifications, or have questions about things not covered in this manual. For any questions you have feel free to approach any member of the department and they will be glad to assist you.