

**SEMINARS
IN APPLIED STATISTICS
FOR RADIATION CYTOGENETICS
AND BIODOSIMETRY**

Volodymyr Vinnikov
S.P. Grigoriev Institute
for Medical Radiology and Oncology
of the National Academy
of Medical Science of Ukraine



SEMINAR I. INTRODUCTION

DISCLAIMER

These seminars are not for professionals in statistics or mathematics.
In these Seminars I present my personal opinions and approaches.
I apologize for possible mistakes.

ACKNOWLEDGMENTS

I'm very grateful to my teachers – David Lloyd and Alan Edwards of the NRPB UK.

I learned a lot from my friends – Dr. Elizabeth Ainsbury of the UK Health Security Agency and Prof. Manuel Higuera Hernández of the Universidad de la Rioja, Spain.

WHERE THE INSTRUCTIONS CAN BE FOUND?

The IAEA Manual on Biodosimetry contains references to all useful publications.

However, there are only few articles, presenting simple algorithms, which can be reproduced using a calculator and which are suitable for a biomedical researcher, who doesn't have special mathematical education.

During the IAEA CRP E3.50.08 and E.3.50.10 I talked to colleagues from different labs located in different countries. The average knowledge of statistics is below the level, required for understanding formulas included into the IAEA Manual.

**Cytogenetic Dosimetry:
Applications in
Preparedness for and
Response to Radiation
Emergencies**



SEMINAR I. CONTENTS

- ❖ Typical tasks for statistical data treatment in Radiation Cytogenetics and Biodosimetry.
- ❖ Overview of relevant statistical methods.
- ❖ Role and place of the applied statistics in Radiation Cytogenetics and Biodosimetry.
- ❖ Content of subsequent seminars.



TYPICAL TASKS FOR STATISTICAL DATA TREATMENT IN RADIATION CYTOGENETICS AND BIODOSIMETRY

Main Categories of Tasks

- 1. Presentation and formalization of data: results of assays or supportive information**
- 2. Comparison of datasets: experimental points, groups, etc.**
- 3. Obtaining numerical values for parameters of biophysical models**
- 4. Revealing hidden links or unclear relationships**

TYPICAL TASKS FOR STATISTICAL DATA TREATMENT IN RADIATION CYTOGENETICS AND BIODOSIMETRY

Practical examples:

- **Characterization of studied groups: patients' data or cytogenetic data**
- **Testing for randomness: biomarkers among cells or breaks among chromosomes**
- **Generating a dose response *in vitro***
- **Detection of genotoxic exposure**
- **Radiation dose estimation from dosimetric biomarker yield**
- **Expertise conclusions for court cases with maximum confidence**
- **Intercomparisons**
- **Studying time-effect relationship**
- **Validation of new biomarkers**
- **Linking biomarker yield to health effects**

OVERVIEW OF RELEVANT STATISTICAL METHODS

Most demanded categories of statistical methods:

- **Variation analysis**
- **Analysis of qualitative (alternative) parameters**
- **Non-parametric methods**
- **Correlation analysis**
- **Regression analysis**
- **Dispersion analysis**
- **Probability analysis**

OVERVIEW OF RELEVANT STATISTICAL METHODS

Choosing statistical methods to solve particular tasks:

- Testing inter-patient homogeneity of biomarker yield inside a studied group ← Chi-squared test.
- Setting a distribution model for a formal description of the group biomarker yield ← Chi-squared test.
- Analysis of painted chromosomes involvement into aberrations ← Chi-squared test.
- Revealing fragile sites in banded chromosomes ← Chi-squared test.
- Testing biomarker's among cell distribution for randomness ← Papworth's u-test.

OVERVIEW OF RELEVANT STATISTICAL METHODS

Choosing statistical methods to solve particular tasks:

- Comparison of biomarker yields in independent groups ← Students unpaired t-test; Fisher's exact test.
- Comparison of biomarker yields in a group or individual with repeated measures ← Students paired t-test.
- Revealing the effect of a qualitative factor on biomarker yield: ← Chi-squared test; Student's t-test; Fisher's exact test; Odds ratio; Risk estimation; Dispersion analysis (ANOVA).
- Detecting the effect of a quantitative factor on biomarker yield: ← Correlation analysis.

OVERVIEW OF RELEVANT STATISTICAL METHODS

Choosing statistical methods to solve particular tasks:

- **Generating a dose response curve ← Iteratively reweighted least square method (IRLS).**
- **Formalization of time-effect relationship ← IRLS.**
- **Radiation dose estimation via LQ equation ← Conventional algebra.**
- **Conclusions for court cases ← Probability analysis (Bayesian or Bayesian-like techniques).**
- **Intercomparisons ← z-test; reproducibility standard deviation.**
- **Validation of new biomarkers ← Chi-squared test.**

ROLE AND PLACE OF THE APPLIED STATISTICS IN RADIATION CYTOGENETICS AND BIODOSIMETRY

The best approach is to assign a statistician during planning our research. Correctly formulated questions contribute 90% to the correctness of the answers, and thus to the overall success of the study, e.g.:

- How many patients must be included into a study?
- How many experimental points will be optimal for setting a dose response curve?
- How can we design an experiment or study, if we don't have enough resources to evaluate the impact of each factor separately?
- Which statistical criteria we must apply for data treatment?

ROLE AND PLACE OF THE APPLIED STATISTICS IN RADIATION CYTOGENETICS AND BIODOSIMETRY

The most typical questions to the statistician **AFTER** the completion of data gathering in our study:

- We collected a large piece of rubbish instead of proper data! We don't know what to do with it !!! **PLEASE, HELP!!!!**
- We don't see any clear patterns in our datasets. Results are chaotic. Are there some hidden correlations? **PLEASE, HELP!!!!**
- The scientific journal rejected our manuscript due to poor statistics. What we did wrong? **PLEASE, HELP!!!!**

CONTENTS OF THE SEMINAR PROGRAMME

Seminar II. Descriptive statistics for the presentation of research data in Radiation Cytogenetics and Biodosimetry.

- Types of data.
- Variation statistics: distributions, mean, mode, median, standard deviation, dispersion, standard error, quartiles, confidence intervals.
- “Normalization” of data.
- Qualitative data: alternatives and frequencies.
- Cytogenetic damage yield: individual data versus group data.

CONTENTS OF THE SEMINAR PROGRAMME

Seminar III. Comparison of datasets. Parametric and non-parametric methods. Assessment of the effect of acting factor: Detection versus measurement. Student's t-test. Chi-squared test. Fisher's exact test. Relative risk and odds ratio. Most widespread non-parametric methods. ANOVA.

Seminar IV. Correlation analysis. Regression analysis: Least squares method. Goodness of fit. Dose response and other applications of regression analysis in Radiation Cytogenetics and Biodosimetry. Radiation dose estimation in biodosimetry. Software packages: CABAS versus Dose Estimate.

CONTENTS OF THE SEMINAR PROGRAMME

Seminar V. Simple biophysical models used in Radiation Cytogenetics and Biodosimetry. Special statistical methods for experiments or intercomparisons. Testing and validation of new biomarkers: sensitivity, specificity, overall efficacy (informativity).

Seminar VI. “Odds and sods”: Non-conventional methods and approaches. Multiple regressions. Cluster analysis. Bayesian analysis. Non-Poisson distributions and mixed distributions. Novelty and prospective statistical techniques.

CONCLUSIONS

What we expect from this course?

- Refreshing our basic knowledge in statistical methods, which we use in cytogenetic biodosimetry
- Better understanding of the overall methodology of data treatment in radiation cytogenetics, and how it is linked to the biophysics behind cytogenetic processes
- A guidance on which statistical method should be applied to a particular dataset in order to answer a particular question
- How to formulate the work tasks for professional mathematicians or statisticians, who have to perform data processing beyond our capabilities

HOMework:

PREPARE A TABLE

“TASKS IN BIODOSIMETRY ↔ STATISTICAL METHODS”.

FILL IN THE COLUMN “TASKS”.

UPDATE IT EVERY DAY.

REGISTER THE TIME, WHEN YOU START TO HATE IT.