2019 - 2020

Dr. Fazıl Küçük Faculty of Medicine, EMU

Year 1

Biostatistics Course

INSTRUCTOR: Assist. Prof. Dr. İlke Akçay

ilke.akcay@emu.edu.tr

Brief information about the biostatistics course in EMU Medical Faculty

- Year 1 Biostatistics courses:
  - Committee 1 -> 6 hours
  - Committee 2 -> 10 hours
  - Committee 3 -> 12 hours

- **Note:** You will have approximately one question in committee exams for every hour spent at the lectures

Brief information about the biostatistics course in EMU Medical Faculty

- Reading/Studying Materials
  - Lecture Notes
    - Will be provided at the end of each lecture
    - Warning: During the lectures no video recordings please!

Brief information about the biostatistics course in EMU Medical Faculty

- Reading/Studying Materials
  - Books
    - Biostatistics: Basic Concepts and Methodology for the Health Sciences, 10th Ed. Wayne W. Daniel, Chad L. Cross
    - Introduction to Biostatistics For Health Sciences, Micheal R. Chernick, Robert H. Friss,Willey
    - Medical Statistics at a Glance, Aviva Petrie, Caroline Sabin, Blackwell Science, London 2003
    - Practical Statistics For Medical Research, Douglas Altman, Chapman & Hall, London 1995
    - Basic Statistics: A Primer for the Biomedical Sciences, Olive Jean Dunn, Virginia A. Clark, 2009
    - Medical Statistics: A textbook for the Health Sciences, David Machin, Michael J. Campbell, Stephen J. Walters, 2007

Topics of Y1C1 biostatistics course

- What is statistics and biostatistics?
- Statistics in medical research
- Designing Research
- Types of Data
- Describing data with numbers
- Describing data with graphics

Today's Topics

- What is statistics and biostatistics?
- Statistics in medical research

What is statistics and biostatistics?

Statistics

- We live in the information age
- Information

how it is obtained

how it is analyzed

how it is interpreted

Statistics

- **Statistics** is a field of study concerned with:
  - 1) The design of research/experiments or sampling procedures
  - 2) collection, organization, summarization, and analysis of **data**
  - 3) Drawing of inferences about a body of data when only a part of the data is observed
- Behind it, there is theoretical statistics
- Statisticians try to interpret and communicate the results to someone else as the situation demands.

Some Basic Concepts of Statistics

- Population-Sample
  - Sampling methods
  - Determination of the sample size
- Data
  - Sources of data
  - Types of data
    - Nominal/Ordinal/Interval/Scale
  - Concept of ‘variable’
    - Qualitative/quantitative variables
    - Discrete/continuous random variables
    - Dependent/Independent variables

Population vs. Sample

- **Population** is the largest collection of values of a random variable for which we have an interest at a particular time.

Example:

  - The weights of all the children enrolled in a certain elementary school.
  - Populations may be finite or infinite.

- **Sample** is a part of a population.

Example

  - The weights of only a fraction of these children.

Data

- The raw material of statistics is data.
- We may define data as values, which result from the process of counting or from taking a measurement.
  - Quantitative (numbers)
  - Qualitative (observation)

Data

Example:

- When a hospital administrator counts the number of patients discharged from the hospital on a given day (counting)
- When a nurse takes a patient's temperature (measurement)

Sources of Data
1. Routinely kept records

Example:

- Hospital medical records contain immense amounts of information on patients
  - Age, gender, weight, test results, number of incidences of a specific disease, etc.
- Hospital accounting records contain a wealth of data on the facility's business activities

Sources of Data
2. Surveys

- The source may be a survey, if the data needed is about answering certain questions.

Example:

- If the administrator of a clinic wishes to obtain information regarding the mode of transportation used by patients to visit the clinic, then a survey may be conducted among patients to obtain this information.
- For doing research about the awareness of society on diseases/effects of some habits, etc.

Sources of Data
3. Experiments

- Frequently the data needed to answer a question are available only as the result of an experiment.

Example:

- If a doctor/researcher wishes to test the effectiveness of a candidate HIV vaccination
  - Design an experiment with two groups: control/experiment
  - Analyze the results
- **NOTE:** Researchers have to be very careful on ethical issues when conducting experimental research.

Sources of Data
4. External Sources

- The data needed to answer a question may already exist in the form of published reports, commercially available data banks, or the research literature, i.e. someone else has already asked the same question.

Some Basic Concepts of Statistics

- Descriptive Statistics
  - Organization, summarization and descriptive analysis of data
- Inferential Statistics
  - Conducting hypothesis tests to get some inferences
    - Test the hypothesis -> how significantly we can reject or fail to reject our hypothesis?

"STATISTICS IS THE GRAMMAR OF SCIENCE."

KARL PEARSON

"STATISTICS: THE MATHEMATICAL THEORY OF IGNORANCE."

MORRIS KLINE

I CAN PROVE ANYTHING BY STATISTICS EXCEPT THE TRUTH.

George Canning
English Statesman

QUOTEHD.COM

DEMOCRACY IS AN ABUSE OF STATISTICS

JORGE LUIS BORGES

STATISTICS MEAN NEVER HAVING TO SAY YOU'RE CERTAIN

Why we need statistics?

- Statistical methods are required to ensure that data are interpreted correctly and that apparent relationships are meaningful (or "significant") and not simply chance occurrences.

Why we need statistics?

- **Example 1: Weather Forecasts**

Do you watch the weather forecast sometime during the day?

How do you use that information?

Have you ever heard the forecaster talk about weather models?

These computer models are built using statistics that compare prior weather conditions with current weather to predict future weather.

Why we need statistics?

- **Example 2: Predicting Disease**

Lots of times on the news reports, statistics about a disease are reported. If the reporter simply reports the number of people who either have the disease or who have died from it, it's an interesting fact but it might not mean much to your life. But when statistics become involved, you have a better idea of how that disease may affect you.

For example, studies have shown that 85 to 95 percent of lung cancers are smoking related. The statistic should tell you that almost all lung cancers are related to smoking and that if you want to have a good chance of avoiding lung cancer, you shouldn't smoke.

Why we need statistics?

- **Example 3: Political Campaigns**

Whenever there's an election, the news organizations consult their models when they try to predict who the winner is. Candidates consult voter polls to determine where and how they campaign. Statistics play a part in who your elected government officials will be.

Why we need statistics?

- **Example 4: Stock Market**

Another topic that you hear a lot about in the news is the stock market. Stock analysts also use statistical computer models to forecast what is happening in the economy.

- For more examples, you may visit
  - http://www.mathworksheetscenter.com/mathtips/statsareimportant.html

Statistics

- The tools of statistics are employed in many fields: business, education, psychology, agriculture, economics, … etc.

Biostatistics

- When the data analyzed are derived from the biological science and medicine, we use the term biostatistics to distinguish this particular application of statistical tools and concepts.

Statistics in medical research

Statistics in medical sciences

- Statistics are an essential part of medical research
- Journals and magazines for physicians are full of statistical material, as well as the findings of research studies.
- Researchers use statistical tests to determine results from experiments, clinical trials of medicine and symptoms of diseases.
- The use of statistics in medicine provides generalizations for the public to better understand their risks for certain diseases, such as links between certain behaviors and heart disease or cancer.

Statistics in medical sciences

- Descriptive statistics show the portion of a population with a disease, for example.
- Inferential statistics help determine causes to diseases, and relationship between factors and diseases.
- Researchers in the pharmaceutical, forensic and biological sciences all use statistics to relay information about health and medicine.

Eur J Epidemiol (2013) 28:169–180
DOI 10.1007/s10654-013-9771-5

DIABETES MELLITUS

Twelve-year trends in the prevalence and risk factors of diabetes and prediabetes in Turkish adults

Ilhan Satman · Beyhan Omer · Yildiz Tutuncu · Sibel Kalaca · Selda Gedik · Nevin Dinccag · Kubilay Karsidag · Sema Genc · Aysegul Telci · Bulent Canbaz · Fulya Turker · Temel Yilmaz · Bekir Cakir · Jaakko Tuomilehto

Received: 24 September 2012 / Accepted: 24 January 2013 / Published online: 14 February 2013

Abstract There is concern about an emerging diabetes epidemic in Turkey. We aimed to determine the prevalence of diagnosed and undiagnosed diabetes, prediabetes and their 12-year trends and to identify risk factors for diabetes in the adult Turkish population. A cross-sectional, population-based survey, 'TURDEP-II' included 26,499 randomly sampled adults aged ≥ 20 years (response rate: 87 %). Fasting glucose and biochemical parameters were measured in all; then a OGTT was performed to identify prevalence to the TURDEP-I population (performed in 1997–98) was 13.7 % (if same diagnostic definition was applied diabetes prevalence is calculated 11.4 %). The prevalence of isolated-IFG and impaired glucose tolerance (IGT), and combined prediabetes was 14.7, 7.9, and 8.2 %, respectively; and that of obesity 36 % and hypertension 31.4 %. Compared to TURDEP-I; the rate of increase for diabetes: 90 %, IGT: 106 %, obesity: 40 % and central obesity: 35 %, but hypertension decreased by 11 % during

Statistical methods

All analyses were performed using SPSS for Windows (version 19.0; SPSS/IBM, Chicago, IL). The χ^2, the student's t, and ANCOVA tests were used when appropriate. Pearson's, or Spearman's CVs, ORs, and 95 % CIs were calculated. A p value <0.05 was considered statistically significant. The prevalence of diabetes and prediabetes was estimated by 5-year age groups for both genders separately. Logarithmic transformations of nonhomogenously distributed factors were used. Variables that were associated with diabetes in the univariate analysis were included in the multiple logistic regression (backward elimination) models in men and women, separately.

To generate nationally and internationally-comparable results, the age-standardized prevalence was calculated using the 'TURDEP-I', 'TurkStat-2009', 'WHO's new World' and 'European' populations as standards [6, 13, 23, 24].

IJC
International Journal of Cancer

Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012

Jacques Ferlay[1], Isabelle Soerjomataram[1], Rajesh Dikshit[2], Sultan Eser[3], Colin Mathers[4], Marise Rebelo[5], Donald Maxwell Parkin[6], David Forman[1] and Freddie Bray[1]

[1] Section of Cancer Surveillance, International Agency for Research on Cancer, Lyon, France
[2] Tata Memorial Hospital, Mumbai, India
[3] Izmir Cancer Registry, Izmir Hub, Izmir & Hacettepe University Institute of Public Health, Ankara, Turkey
[4] Department of Measurement and Health Information Systems, WHO, Geneva, Switzerland
[5] Department of Surveillance and Cancer Information, Brazilian National Cancer Institute, Ministry of Health, Rio de Janeiro, Brazil
[6] Clinical Trial Service Unit & Epidemiological Studies Unit, University of Oxford, Oxford, United Kingdom

Estimates of the worldwide incidence and mortality from 27 major cancers and for all cancers combined for 2012 are now available in the GLOBOCAN series of the International Agency for Research on Cancer. We review the sources and methods used in compiling the national cancer incidence and mortality estimates, and briefly describe the key results by cancer site and in 20 large "areas" of the world. Overall, there were 14.1 million new cases and 8.2 million deaths in 2012. The most commonly diagnosed cancers were lung (1.82 million), breast (1.67 million), and colorectal (1.36 million); the most common causes of cancer death were lung cancer (1.6 million deaths), liver cancer (745,000 deaths), and stomach cancer (723,000 deaths).

Methods of estimation

Cancer incidence and mortality rates for 2012 by sex and for 10 age groups (0–14, 15–39, 40–44, 45–49,. . ., 75 and over) are estimated for the 184 countries or territories of the world having a total population greater than 200,000.[4] Results are

5. When recent mortality data were not available from national sources, national mortality was estimated from national incidence estimates using modelled survival as described in Appendix B (83 countries)

Random fluctuations in the predicted age-specific incidence and mortality rates were smoothed using a *loess* function, a locally weighted regression, by country, sex and cancer site. Estimates for the 20 world regions (Fig. 1) were obtained by the population-weighted average of the incidence and mortality rates of the component countries. These rates were applied to the corresponding population for the region for 2012 to obtain the estimated numbers of new cancer cases and deaths in 2012. The rates were age-standardized (ASRs per 100,000 person-years) using the direct method and the World standard population as proposed by Segi[17] and

IJC
International Journal of Cancer

worldwide: Sources, methods and major patterns in GLOBOCAN 2012

Jacques Ferlay[1], Isabelle Soerjomataram[1], Rajesh Diksh
Donald Maxwell Parkin[6], David Forman[1] and Freddie Br

[1] Section of Cancer Surveillance, International Agency for Research on
[2] Tata Memorial Hospital, Mumbai, India
, Turkey
of Health, Rio de Janeiro, Brazil
ingdom

d for all cancers combined for 2012 are now
Cancer. We review the sources and methods
briefly describe the key results by cancer site
ses and 8.2 million deaths in 2012. The most
nd colorectal (1.36 million); the most common
5,000 deaths), and stomach cancer (723,000

Grass pollen immunotherapy for seasonal rhinitis and asthma: A randomized, controlled trial

Samantha M. Walker, PhD,[a] Giovanni B. Pajno, MD,[b] Marcia Torres Lima, MD,[a] Duncan R. Wilson, MD,[a] and Stephen R. Durham, MD[a] *London, United Kingdom, and Messina, Italy*

Background: Grass pollen immunotherapy significantly reduces hay fever symptoms and medication requirements. Effects on seasonal asthma are less clear, and concerns over safety persist.
Objective: The goal of this study was to assess the effects of grass pollen immunotherapy on symptoms, bronchial hyperresponsiveness, and quality of life in seasonal rhinitis and asthma.
Methods: Forty-four patients with severe summer hay fever (of whom 36 reported seasonal chest symptoms and 28 had seasonal bronchial hyperresponsiveness) participated in a randomized, double-blind, placebo-controlled, parallel group study. After symptom monitoring for one summer, participants received injections of a depot grass pollen vaccine (n = 22) or matched placebo injections (n = 22) in a rapid updosing cluster regimen for 4 weeks, followed by monthly injections for 2 years. Outcome measures included hay fever symptoms and

Abbreviations used

PC_{20}:	Provocation concentration producing a 20% fall in FEV_1
RQLQ:	Rhinoconjunctivitis Quality of Life Questionnaire
SQ-U:	Standard quality units

We previously showed that grass pollen immunotherapy improved symptoms and reduced medication requirements in patients with severe summer hay fever.[1] We also showed that continuing immunotherapy for 3 to 4 years results in prolonged clinical remission for at least 3 years after discontinuation.[2,3]

Grass pollen immunotherapy for seasonal rhinitis and asthma: A randomized, controlled trial

Samantha M. Walker, PhD,[a] Giovanni B. Pajno, MD,[b] Marcia Torres Lima, MD,[a] Duncan R. Wilson, MD,[a] and Stephen R. Durham, MD[a] *London, United Kingdom, and Messina, Italy*

Methods: Forty-four patients with severe summer hay fever (of whom 36 reported seasonal chest symptoms and 28 had seasonal bronchial hyperresponsiveness) participated in a randomized, double-blind, placebo-controlled, parallel group study. After symptom monitoring for one summer, participants received injections of a depot grass pollen vaccine (n = 22) or matched placebo injections (n = 22) in a rapid updosing cluster regimen for 4 weeks, followed by monthly injections for 2 years. Outcome measures included hay fever symptoms and medication use, health-related quality of life, and measurements of nonspecific bronchial responsiveness.

Grass pollen immunotherapy for seasonal rhinitis and asthma: A randomized, controlled trial

Results: Significant reductions were observed in the immunotherapy group compared with the placebo group in hay fever symptoms (49%, 15%; *P* = .01), medication scores (80%, 18%; *P* = .007), and seasonal chest symptoms (90%, 11%; *P* < .05). Impairment of overall quality of life (mean score of 7 domains) during the pollen season was less in the immunotherapy group than in the placebo group (median difference [95% CI], 0.8 [0.18-1.5]; *P* = .02). During the pollen season there was no change in airway methacholine PC_{20} (provocation concentration producing a 20% fall in FEV_1) in the immunotherapy-treated group (*P* = .5), compared with an almost 3 doubling-dose decrease in the placebo-treated group (*P* = .01, between-group difference). There were no significant local or systemic side effects during the study.

Hypertensive diseases of pregnancy and risk of hypertension and stroke in later life: results from cohort study

Brenda J Wilson, M Stuart Watson, Gordon J Prescott, Sarah Sunderland, Doris M Campbell, Philip Hannaford, W Cairns S Smith

Abstract

Objective To examine the association between hypertensive diseases of pregnancy (gestational hypertension and pre-eclampsia) and the development of circulatory diseases in later life.
Design Cohort study of women who had pre-eclampsia during their first singleton pregnancy. Two comparison groups were matched for age and year of delivery, one with gestational hypertension and one with no history of raised blood pressure.
Setting Maternity services in the Grampian region of Scotland.
Participants Women selected from the Aberdeen maternity and neonatal databank who were resident in Aberdeen and who delivered a first, live singleton from 1951 to 1970.

disability. In Britain angina may be as common in women as in men.[3] Most of the aetiological research in women has concentrated on risk factors that are shared by men—for example, smoking, social class, family history, obesity, and lipid and clotting factors. Women, however, are exposed to several possible risk factors that are sex specific, including pregnancy, menopause, hysterectomy, and the use of exogenous hormones. Vital statistics collected in England and Wales between 1938 and 1960 suggest that pregnancy has an effect as parous women have higher mortality from hypertension, ischaemic and degenerative heart disease, and cerebrovascular disease than nulliparous women.[4]

During pregnancy, many women are affected by hypertensive problems, especially in the first pregnancy.[5] Such problems fall into four categories: chronic

Department of Public Health, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD
Brenda J Wilson *senior lecturer*
M Stuart Watson *lecturer*
Gordon J Prescott *research fellow in medical statistics*
Sarah Sunderland *research nurse*
W Cairns S Smith *professor*
Department of Obstetrics and Gynaecology,

Hypertensive diseases of pregnancy and risk of ... fe: results from cohort

... Sunderland, Doris M Campbell,

Sample size

The expected frequency of outcome measures varies for different cardiovascular conditions in women in Scotland, from 0.5% for electrocardiographic evidence of myocardial infarction to over 8% for angina defined by Rose questionnaire criteria.[3] The frequency of hypertension varies depending on the definition used. One study of women aged 40-59 years in Scotland identified 7% with a diastolic pressure greater than 100 mm Hg.[20] For an outcome present in 5% of the control group, a sample size of 932 subjects per group would detect a doubling of risk, with 80% power at the 5% level of significance. On this basis, we included all women with pre-eclampsia or eclampsia (likely to be the group with fewest potential members) in the cohort and randomly sampled control and gestational hypertension groups (matched for age and year of delivery) of the same size as this group.

Statistical analysis

We used χ^2 and t tests as appropriate to test for baseline differences between the three groups, and, for all analyses, we used the control group as reference for comparisons with each of the two "exposed" groups (gestational hypertension and pre-eclampsia or eclampsia). For the questionnaire and clinical examination data, we computed odds ratios and used logistic regression methods to adjust for the influences of age, body mass index, and smoking behaviour at follow up.

Participants Women selected from the Aberdeen maternity and neonatal databank who were resident in Aberdeen and who delivered a first, live singleton from 1951 to 1970.

...effect as parous women have higher mortality ...pertension, ischaemic and degenerative heart disease, and cerebrovascular disease than nulliparous women.[4]

During pregnancy, many women are affected by hypertensive problems, especially in the first pregnancy.[2] ...

medical statistics
Sarah Sunderland
research nurse
W Cairns S Smith
professor
Department of Obstetrics and Gynaecology,

Use of Statistics in Research

PLANNING / literature review, research hypothesis, sample size etc.

DESIGN

EXECUTION (data collection)

DATA PROCESSING

DATA ANALYSIS

PRESENTATION

INTERPRETATION

PUBLICATION

The general sequence of steps in a research project

Each step may have substeps

Next Lecture

Designing Research
Types of Data