Ramakrishna HK

Medical Statistics

For Beginners

Springer

Medical Statistics

Ramakrishna HK

Medical Statistics

For Beginners

Springer

Ramakrishna HK
Subbaiah Institute of Medical Sciences
Shivamogga
Karnataka
India

ISBN 978-981-10-1922-7 ISBN 978-981-10-1923-4 (eBook)
DOI 10.1007/978-981-10-1923-4

Library of Congress Control Number: 2016953631

© Springer Science+Business Media Singapore 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.
The registered company address is 152 Beach Road, #22-06/08 Gateway East, Singapore 189721, Singapore

This book is dedicated
To my parents
Keshavamurthy and Bhagyalakshmi, who taught me elementary knowledge of how to handle numbers,

To my wife
Dr. Swarnalatha MC, who gave valuable suggestions, is a constant source of encouragement, and tolerated my odd working hours,

To my children
Manu and Ajay, who made my life worth living.

Foreword

I feel privileged to write this foreword to the book on *Medical Statistics for Beginners* by Dr. H. K. Ramakrishna.

His introduction to the book, sharing his personal experience and journey as an author, is a "must read." Today, the world is really flat—one's effort is the only limiting factor. Dr. Ramakrishna has shown what can be done with focused effort. Where you come from and what your official designation is are completely irrelevant, as far as accomplishments are concerned. Dr Ramakrishna deserves high praise for this work and he should be taken as a "role model" by all surgeons.

Statistics is considered a "dry" subject. But it is essential for us to understand at least the basics of statistics to function well as surgeons. This book meets this requirement. The language is simple. The approach is direct and practical. All the principles are explained in a succinct manner.

One can see from the screenshots that everything is worked out using basic computer tools that are universally available. Several free resources available on the Internet are introduced. With this, every one of us can do the things that are shown in the book. The links to statistical calculators that he has himself developed are a great "value add."

The chapters on "Designing a Study Clinical Trial or Dissertation," "Evidence-Based Medicine" (EBM), and "Writing an Article for Journals" are very apt. After all, the purpose of statistics is to help us evolve into "scientific surgeons." Every one of us must endeavor to practice EBM and share our vast experience and add to the scientific knowledge by formally publishing our work.

Finally, I do hope that this book inspires all of us in this country to diligently document our work and publish more often. We fall woefully short of other countries—even small South-East Asian countries—in the field of scientific publication! We cannot afford to let the status quo continue.

I wish this book every success. It is a landmark effort in the field of surgical publication.

Bangalore, India K Lakshman, MS, FRCS

Opinion

"Author's style of writing is easy to read and understand. He gives meaningful examples to explain each new idea and also explains the methodology which at times is quite complex and painstaking. Suggestion to use the computer liberally encourages the reader to learn easy ways of making the complex calculations. The overview of the Biostatistics is nicely covered.

I believe that this short book may be a nice book for the medical students and postgraduates to get to know the various statistical terms and methods for them to use in their academic career."

Dr. R. D. Prabhu FRCS

Contents

1 Introduction 1

2 My Journey from a Rural Surgeon to an Author 3

3 Understanding Biostatistics, Probability, and Tests of Significance 9
Skewed Distribution Curve 17

4 Understanding Basic Statistical Terms 21
Measures of Central Tendency 22
Measures of Dispersion or Spread 26
Variance and Standard Deviation 27
Confidence Interval and Confidence Level 31
Errors Type 1 and Type 2 (Alpha and Beta, Respectively) 31
Factors Increasing Type II Error (False Negativity) 32
Conditional Probability 32
Independence 32
Normal or Gaussian Distribution and Skewed Distribution 33

5 Tests of Significance 35
Chi-Square Test or Simply Chi Test 36
Fisher's Exact Test 40
Formula for Fisher's Test 40
! Is the Symbol for Factorial 40
Student's *T* Test or Gosset's Test 45
One- and Two-Tailed Tests 48
Paired *T* Tests 49
ANOVA 51
Which Test to Use? 51
Mann–Whitney–Wilcoxon (MWW) Test 54
Summary 56
Conclusions 56

6 Other Commonly Used Concepts 67
ANalysis Of Variance (ANOVA) 67
Rank Test: Wilcoxon Signed-Rank Test 69

Risk and Risk Ratio 70
Odds 71
Odds Ratio 71
Relative Risk Reduction (RRR), Absolute Risk Reduction (ARR), and Number Needed to Treat (NNT) 71
Correlation: Relation Between Two Factors 71
Positive Correlation. 72
Negative Correlation 72
Correlation Coefficient (*r*) 73
Regression 73
Linear Regression 74
Other Types of Regressions 76
Logistic Regression. 76
Multiple Regression: To Predict an Outcome 76
The Poisson Regression (Distribution) 77
Cox Regression 77
Survival Analysis 78
Kaplan–Meier Estimator or Survival Graph 79
Multivariate Analysis 79
Power of a Study 80

7 Designing a Study/Clinical Trial/Dissertation, Etc. 85
General Considerations 85
Concerns While Designing a Trial 87
Negative Results 88
Teamwork 89
Phases of Trials: Applies to New Drugs 89
Some Commonly Used Terms 89
Parallel Studies and Crossover Study 90
Planning and Pilot Study. 90
Sample Size. 92
Pilot Study. 93
Collection and Compilation of Data 94
Data 98
Types of Study 98
Data Collection 99
Bias 99
Sampling 101
Sampling Errors and Nonsampling Errors 102
Data Analysis and Inference 103
Presentation of Data and Results 105
Tables 105
Bar Chart 107
Pie Chart 113
Line Chart 114

Area Chart . . . 115
Doughnut Chart . . . 116
Histogram . . . 117
Difference Between Histogram and Bar Chart . . . 120
Clinical Audit . . . 120
Clinical Audit Versus Research . . . 120
Clinical Audit Is an Ongoing Process . . . 121
Problems with Audit . . . 122

8 Writing an Article for Journals . . . 129
Clear Message . . . 130
Type of the Articles . . . 131
Quality Indicators of a Journal . . . 132
Indexing . . . 132
Impact Factor . . . 132
Uniform Requirement for Manuscript Submission . . . 133
What Happens After Submission? . . . 142
Common Reasons for Rejection of Article or Article Sent Back for Changes (Problems in the Article) . . . 143
Best Chance of Acceptance . . . 143
Plagiarism . . . 144
Peer Review . . . 145
Copyright . . . 146
Literature Review . . . 146
Scholarly Databases . . . 149
Limitations and Issues . . . 151
Open-Access Journals . . . 152
Publish or Perish? . . . 152

9 Evidence-Based Medicine . . . 161
What Is EBM? . . . 161
Why Do We Need EBM? . . . 161
How? . . . 162
The Five-Step Approach to EBM . . . 165
Benefits of EBM . . . 165
Limitations of EBM . . . 166
Conclusions . . . 166

10 Model Example . . . 167
Postoperative Management Protocol to Be Followed . . . 168
Results . . . 170
Paper . . . 171

Index . . . 179

About the Author

Dr. H. K. Ramakrishna, MBBS, MS, DNB, FMAS, FRCS
Consultant Surgeon (General & Laparoscopic)
Professor of Surgery, Subbaiah Institute of Medical Sciences

Dr. H. K. Ramakrishna completed MBBS from Government Medical College, Bellary, in 1989 and MS in General Surgery at Karnataka Medical College, Hubli, in 1992, DNB (Diplomate of National Board) in General Surgery in 1993, and got his FRCS (Glasgow) in 2016. He started first upper GI endoscopy center in 1995 and first laparoscopic surgery center in 2001 in Bhadravathi, a rural area. He started mobile laparoscopic surgery unit in 2005 to extend the benefits of laparoscopic surgery to nearby rural places. The unit covered a distance of about 60 km radius.

He has to his credit several articles published in various journals. He is also a reviewer for many journals like *BMJ Case Reports* and *World Journal of Gastroenterology*.

He got many best paper awards in the conferences. He has given many guest lectures in conferences both at state and national levels. He has convened many symposia in state conferences. He has been judge for award papers in conferences. He is in charge of the web site of KSCASI (www.kscasi.com).

He is keenly interested in medical statistics.

1 Introduction

I completed my master's degree in general surgery in 1992. When I read articles in the journals or watched presentation of research articles in conferences, I felt I should conduct some studies and write an article to a journal or present papers. But it appeared to be an impossible task, as statistical terms were jargons to me. The moment numbers came, I used to skip and jump to read the next sentence.

While writing postgraduate dissertation, I found difficulty in handling data like numbers, graphs, and tables and presenting my findings in a proper way. I used to hear and read tests like Student's T tests, one-tailed or two-tailed tests, chi-square test, Fisher's exact test, "Normal" curve, etc. If you take exact English meanings of words in these, you get confused. When to apply which test? How to do a test? How to interpret the results? How to write a conclusion? These are the questions every postgraduate faced. They ran to a statistician who didn't have any medical knowledge but can handle the numbers. He puts the data in some table or graph in his own way, which may or may not be appropriate to the context. I always wished to have sufficient knowledge of statistics so that I myself can present the data and graphs which will be a perfect combination of statistics and medical knowledge.

When I interacted with medical college teachers and some practicing in private, they also expressed the same problem. They had a lot of materials but were unable to convert it into an article or a research paper.

Why is this? It is because of not knowing enough about medical statistics. Medical statistics methodology and concepts are not sufficiently taught in our medical college curriculum. Slowly I started learning statistics and gave a few lectures on this topic in conferences and also conducted a symposium in one of our state conferences. After the talk a lot of interesting discussions and interactions by the audience followed. There I understood that a lot of doctors want to write an article or present a paper but cannot, because of the phobia for the presentation of their data in a proper form. Then I thought, I should write a book on medical statistics and put it in a very simple way with all sorts of commonly faced examples so that every doctor or medical student would be able to understand what he/she reads in the journals and also be able to present his/her own data in dissertation or paper or article writing.

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_1

This book gives basic ideas in a very simple and very easy to understand way. Today, the Internet and computers have made our job a lot easier than it was during our times. I intend to discuss how to utilize these also. For more complex data and more detailed explanations, other recommended books are always there.

Though many books on statistics are available, they are with data other than medical and so may not be helpful for doctors. Also they are in great details and often not understandable for a medical man. Medical men lose interest quickly in reading these books as they contain lot of nonmedical examples.

The use of computers and Internet in calculations, doing tests of significance, collection of information, article search, and presenting the data are neither taught well in medical colleges nor in these books.

Presentation in a simple way with lots of medical journal article like examples helps to understand the concept and to present your own data.

The book also deals a little bit about how to design and conduct clinical trials and how to write an article to a journal. These concepts help a postgraduate student in his/her dissertation work and a young doctor to present a paper or write an article.

No single book gives all these concepts.

The book is beneficial to other branches of biological sciences like dental, veterinary, agriculture, etc. which have a lot in common with medical field in the methodology.

Finally, this book is intended for beginners. For more detailed discussion and information, the reader is advised to refer to standard books on statistics and appropriate web sites.

2 My Journey from a Rural Surgeon to an Author

I am writing something of my own journey from being an ordinary rural surgeon to becoming the author of this book. I thought this is appropriate here because many of the readers would have been at a similar, if not the same, situation as I was in 1993. But with efforts, each and every one of them is capable of becoming an author. I hope this account will stimulate them to undertake some research or study and write articles for the journals on their own. For medical teachers, this will help to guide their postgraduate students.

I started my profession as a freelance consultant surgeon in 1993 in a small taluk place, almost a rural place with limited facility and resources. From the beginning of the practice, I noticed I've very limited access of gaining new knowledge, which my urban counterparts are enjoying. My only source was our specialty national journal, *Indian Journal of Surgery*. While reading the journal, I always dreamt of writing an article and becoming an author. I am sure many of the readers of this book also will have the same feeling. But how? From where to get the material for writing the article? How to get the information for writing the discussion part in the article? All these appeared like a *mission impossible*. But my passion remained. Then I joined medical college. The situation was not much different there also. My colleagues also faced the same problem as I.

> Put off one day:
> And ten days will pass.
>
> A Korean Proverb

I kept on postponing the dream without a clue how to start. Ten years passed!! Now I have learnt this lesson. If I have a dream, I should start collecting resources and start working right away.

In 1993, I got a call from one of the senior surgeons of my region Dr RD Prabhu, asking me to write an article for the *Indian Journal of Surgery*. Today, I remember him as my guide and starting point of my journey. I hesitated to accept the offer. He encouraged me to write something out of experience. I wrote not one but two articles! And both were published in the same issue of *Indian Journal of Surgery* ((**1**) **Perspectives of Rural Surgeons - Taking Newer Technologies to the Rural**

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_2

Patients and (2) Challenges in Rural Surgery – A Difficult Case of Acute Intestinal Obstruction Managed in a Rural Set-up. Both were published in Dec. 2003 issue of the *Indian Journal of Surgery*). It is rare for any author to have two articles published in one issue of a journal.

Thus the journey started. Another surgeon, Dr Ramesh N, who read this article, invited me to participate in a symposium in a state conference of our surgeons' association. I was asked to talk about how to use technology and the Internet to improve rural surgeons' profession. While preparing for the talk, it struck me to use the Internet for collecting information.

> The more you practice what you know,
> The more will you know what to practice.
>
> Unknown

I started using computer and access the Internet for various articles. But the problem remained as many of the article's full text are not accessible without subscription or payment.

In the meanwhile many more offers came to me to give guest lectures on rather unconventional topics for CMEs. I gave lecture in our national conference on Computers and Surgical Audit. I could gather all the information to prepare for the talk from the Internet. In another CME I talked on designing a clinical trial. The journey continued.

In one of the conferences, I heard an oration by one of the best orators I have heard, Dr Lakshman K. While we are all presented and informed by "experts" that we must always use a tissue-separating mesh (a variety of newer meshes available, which are costlier by 15–20 times than conventional polypropylene mesh), Dr Lakshman presented data and concluded there is no sufficient data at present to support this view of experts. When the audience asked him how to defend if a complication occurs and patient takes the surgeon to the court of law, the orator answered: "answer is evidence-based medicine." Inspired by this oration, I started collecting articles on the topic with the help of Dr Lakshman and did a meta-analysis on complications of intraperitoneal meshes: conventional polypropylene mesh and the newer meshes. Based on the information and data, I wrote an article with Dr Lakshman as the coauthor. It was accepted in *Indian Journal of Surgery* at the first attempt and was published without any corrections (H. K. Ramakrishna and K. Lakshman. **Intra Peritoneal Polypropylene Mesh and Newer Meshes in Ventral Hernia Repair: What EBM Says? Indian J Surg. 2013 Oct; 75(5): 346–351**). While preparing this article, I found that many seemingly difficult tests of significance can be done easily by using calculator on many web sites, and the best part is they are free. So what is preventing us to write an article?

I also presented this data as a paper and got Dr Mahadevan's Best Paper Award in our state conference.

With all these experiences, I gained sufficient confidence. I dared to request our association to give me a chance to convene a symposium on clinical trials and surgical audit in the state conference. It was accepted. I could find resource persons to talk about surgical audit, writing a journal article, evidence-based medicine, etc. But I could not

find a surgeon who could talk on biostatistics and tests of significance. So I decided to take up the topic myself. I realized how difficult for a medical person to digest statistics. The main reason for this is that it is not taught sufficiently in the medical curriculum. Another reason is that there are hardly any books, which explain these concepts in simple words and with examples from medical field. Medical men when they read the journals, they do not critically analyze the numbers and results. They just accept whatever conclusions the authors write. The authors collect data themselves but they do not analyze them: they just hand it over to statisticians and present in the journals whatever analysis the statistician provides as it is. This leads to a problem in interpretation sometimes. As an example, I have intentionally given data from an imaginary trial (see example on paired *T* test given in the end of the chapter on tests of significance) to show how we can be mislead while interpreting the conclusions. To draw conclusions one needs knowledge of statistics as well as clinical knowledge. In high-quality articles, authors take very professional dedicated medical statisticians with experience in medical field, and the articles are thoroughly analyzed before writing conclusions. It is difficult to find any faults. In some articles of our journal, I could find faults and wrote a few letters to the editors about them: They were also published. I feel these flaws are because of too much reliance on statisticians who do not have medical knowledge, and authors do not have sufficient knowledge of biostatistics.

I wished there is a book which helps a beginner, starting from basics of biostatistics to design a trial to converting the information to write an article or a paper. I searched for such a book but failed.

> The best way to predict the future is to invent it.
>
> Alan Kay

Here I learnt another lesson. Instead of waiting for someone to write a book covering all these topics, it is better I write one. It appeared an uphill task initially. But as somebody puts it ….

> You will never know what can be done until you try it.

So I started to try.

> A task well begun is half done.

I just gathered all my previous attempts of various papers, publications, guest lectures, symposia, etc. in one place. Soon I realized that I have already done half of the job.

In this book I have tried to explain the concepts in simple words with simple examples so that beginner will understand. But it is not that simple. No pain, no gain. Reader has to spend some time to read slowly and repeatedly until he is sure of having understood the concept.

> Tell Me: And I forget.
> Teach me: And I remember.
> Involve me: And I learn.
>
> Benjamin Franklin

The reader should try to imagine similar different clinical situations he/she comes across. He/she should try to apply their knowledge of biostatistics while reading journal to see if the data presented and conclusions mentioned are correct. To begin with, he/she should design some "mock trials" or imaginary studies and articles. He/she has to go back to high school days and remember how they used to solve exercises given at the end of the chapter. This goes a long way in understanding the concept and solving similar problems in similar situations. In difficult problems or situations, the Internet and books are always there to get more information and solve the problem.

Another thing I wish to mention here. We are going more and more toward paperless era. Many journals now do not give any hard copies: they are just online. Learning the Internet, searching for required data and information is neither a luxury nor required only for a researcher: but it is an essential part of the profession of the doctors. There is no way but to go with technology. **Believe me I have not used a single piece of paper in writing this book**. Technology is not only saving money but also helping to preserve the environment.

In WhatsApp I saw a joke.

What is the similarity between a spouse and a mobile phone? **A better option is available once you finalize and are committed**. But if you keep waiting for a better model, you will never marry or buy a mobile phone. So do not wait for a perfect collection of data. No one is perfect. Start your attempt to design a study or writing an article with whatever data you have. With repeated attempts the quality gradually improves.

Warren Buffet, one of world's richest man and is well known for his investment wisdom, said "***While employing a person, look for three things in him: Intelligence, energy and integrity. If he lacks third one don't even bothered about the first two***." Just reject him. Because, if he lacks the third one, the first two will kill you with much more power. While publishing, present only true data, even if it sounds absurd. Explanation can always follow.

Another saying goes like this…:

- Successful people do not work hard: they work SMART.
- They do not do different work: they work differently.

Successful authors do not have different resources: but they use their resources differently. To be successful, you need not have all the facilities, but you can successfully work with available facilities, provided they are properly utilized.

A negative thinker sees difficulty in every opportunity, whereas a positive thinker sees an opportunity in every difficulty. When faced with a difficulty, it is an opportunity in disguise: think how to convert the difficulty into an opportunity.

Mind always thinks complicated answers: not simple ones.

Try to answer this question:

Mathew works in a vegetable store. The shop address is 13, Victory Lane, Calcutta. He is obese. His height is 5 ft 3 inches only. His waist is 44 inches. He wears shoes size 8. The question is "What does he weigh?"

I am sure everybody is thinking how to calculate the weight based on the data given. The answer is too simple, provided you keep your mind open: vegetables. Observe the question: it is "what" and not "how much." So read the question carefully and understand the question properly. Try to find simple answers.

While writing, keep your language simple. It should convey directly what you want to tell. Do not assume the listener will understand the hidden meaning, even if it looks quite obvious. What is obvious to you may not be so for everyone. Do not confuse the reader by using rather difficult words in an attempt to impress.

To drive this message, I would quote a joke (even though this is a book on a serious topic) as I feel it is appropriate here. An angel told a 45-year-old married man to ask for a wish: and she would grant it. He wished his 40-year-old wife to be 25 years younger than him, expecting the angel to make his wife 20 years old. For the man this is too obvious. The angel misunderstood him (or the angel understood him and wanted to punish him?) and she made the man 65 years old!! (He should have made his request simple and direct by asking the angel to make his wife 20 years old.) Always make clear and complete straightforward statements in the article. Different readers interpret the same sentence in different ways, if there is ambiguity in statements.

Give attention to spellings and grammar while writing. Journals expect it. Spelling mistake can lead to a different meaning than intended. When a person was asked to write what is democracy, he wrote: **Buy** the people, **Far** the people, and **Off** the people. Pronunciation is the same (by the people, for the people, and of the people), but the meaning is altogether different. Even computer spell-check cannot detect such mistakes.

While writing conclusions:

- Do not jump to the conclusion.
- Cross check and make sure you understood correctly.
- If necessary with further questioning.

Again a joke as an example to drive home this message may be mentioned here. A woman asked a man "I am feeling lonely, want to go out have a drink and enjoy life. Are you free?" He jumped to the wrong conclusion and said hurriedly "yes," thinking of a nice outing. She said "Fine, then. Please look after my kids." Here the man misinterpreted her words and jumped hurriedly into the wrong conclusion.

A single (or a few) conclusion backed by a strong data is better than many conclusions backed by few or no data. Sometimes a beginner wrongly thinks that if he writes many conclusions, chances of acceptance are higher. It only leads to "verbal diarrhea with mental constipation."

Publications may not give you any remuneration but your colleagues will recognize you, and you get professional satisfaction.

To end I would like to repeat the Korean proverb mentioned earlier:

Put off one day: And ten days will pass.

Let this not happen to you: start designing your own trials, start documenting your results, start writing articles, and start publishing.

Finally, having got many things from this world, we must give something to the world. I saw this photo on the Internet many years back. Now I do not know from where I got it, to cite reference, but I thought it is worth sharing with you.

Wish you best of luck.

Understanding Biostatistics, Probability, and Tests of Significance

3

Learning Objectives
To understand and find answers to….
What is biostatistics?
Why its knowledge is required for medical men?
What is probability?
What is *P* value?
Why *P* value is kept significant at 0.05?
What are tests of significance?
How to interpret the results of the tests?

Errors of using inadequate data are much less than those using no data at all.
Charles Babbage (1792–1871)

The importance of using data in medical research is stressed by Charles Babbage way back in the nineteenth century. If we present some results or conclusions without giving the data in terms of numbers or statistics, we will be committing great errors. It is better to give some data to support the conclusions even if it seems inadequate. Gradually sufficient data will accumulate to support or disprove the hypothesis. This highlights the importance of collecting data, documentation, analysis of numbers, inference of the analysis, presenting the results, etc. Medical men need to know some basic aspects of statistics. During the course of treating patients, we would have gained some experience. Unless we document it, analyze the data documented, draw conclusions, and publish the results, the invaluable experience perishes along with us. The medical world would not have developed to the present proportions without research and publications. After learning medicine from these experienced and published books, we must also contribute something to the medical field and continue the legacy.

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_3

Since the concepts of medical statistics apply to all branches of biological sciences, we can call it ***BIOSTATISTICS*** also.

Biostatistics can be defined as "***the science of theory and methodology for acquisition and use of quantitative evidence in biomedical research***".

Biostatistics or medical statistics deals with the use of statistical methods to analyze data and various tests of significance. We collect data by studying a population. Then we compile the data in a systematic way. Then we apply statistical methods to analyze the data. Here we come across various statistical terms, statistical tests, etc.

The **theory of probability** is a very important concept of the biostatistics. In the journals while comparing the results of two groups, we come across statements like *difference statistically not significant* or *difference is significant at P value of 0.05 or probability value of 0.5*. What does it mean? Let us understand it by studying a few example tables.

	No. of patients	Infection control	Percentage
Drug A	100	86	86
Drug B	100	50	50
Total	200	136	

In the above example, a study is conducted to compare the efficacy of two drugs in controlling infection. A group of 200 patients with the same type of infection are equally divided into two groups (of 100 patients each). To one group, Drug A is given and to the other group, Drug B is given. At the end of treatment, data are collected and results are arranged in the form of a table as above. Now, if you look at the table, you will say easily Drug A is superior since it has controlled infection in 86 % of cases in comparison to 50 % cases by Drug B. Here it is easy to interpret the data because both groups have the same number of patients, and the difference in the results is large.

In many study designs where we use random assignment of patients to the groups, we may not get equal numbers in both the groups for various reasons. Also, the difference in the results may not be as large. For example, consider this table.

	No. of patients	Infection control	Percentage
Drug A	42	16	38.09
Drug B	10	5	50
Total	52	21	

There is nothing more deceptive than an obvious fact.
Sir Arthur Conan Doyle

You think Drug B is a better one because it controls infection in 50 % of patients in comparison with 38.09 % of Drug A. Wait!! Do not hurry to the conclusion! If a statistical test of significance is applied, it proves your conclusion is wrong. If you calculate the *P* value or probability value for the above table, it is 0.74, which is greater than 0.05. For the difference to be significant, *P* value should be less than 0.05. So, the correct conclusion is "***there is no statistically significant difference in the efficacy between the two drugs***". We shall postpone the questions like how to apply test, which test, why not significant, how we arrive at the *P* value of 0.05, etc. for the time being. But **why do we say results are not significant even though there is a striking difference of about 12 %***?*

It is because of the probability that the results could have been due to a phenomenon of "**chance factor**" or, to put it in better terms, because of **natural variation**. This chance factor should not be more than 5 % for the results to be significant. That is to say, confidence in the conclusion should be greater than 95 % for the result to be significant.

Then ***what is this natural variation***?

Every medical doctor knows that the results in treatment of medical conditions cannot be predicted accurately or guaranteed. For example, when you give the same drug to a group of patients, in the same dose to treat the same condition, you will find that some patients improve completely, some partially, and some not at all! Also, there are many conditions which improve spontaneously. If you give a drug during this period of spontaneous improvement (or natural healing), it is possible that the improvement in the disease will be wrongly attributed to the efficacy of the drug. Every doctor is also aware of placebo effect. Placebo effect is an improvement or a result not attributable to the effects of the drug given. Sometimes patient improves even when you give an inactive ingredient, for example, a "tonic" improves weakness or appetite. Results may be because of psychological factors. On the other hand, we also come across cases where a dose of oral polio vaccine has produced death occasionally. If the vaccine were to be banned because of one such result, it would have been a great loss for mankind, because millions of children are protected from polio by the use of the vaccine. Some patients may have other factors which can affect the results. Many more reasons are there. All these can be because of the so-called ***natural variation***. So, a trial or a systematic study is required to properly decide and infer about the result.

In order to understand this concept of natural variation, we shall do an experiment in a very simple way, which everybody can understand. Let us toss a coin ten times. Let us ask a question: how many times you expect a head? This is a very simple question which everybody can understand and answer. We all know that if we toss a coin ten times, the probability of getting heads is five times. But, in the actual experiment, this may not be true. You too can try this. We may get six heads. We may get seven or eight or even nine heads. You try to do this experiment, and you will be surprised to see you won't get five heads quite a number of times! That doesn't mean you have any supernatural power to get more number of heads or the coin is *unfair*. In everyday observation, we know that the chances of a pregnant woman delivering male and female baby are 50:50. Yet, many couples have only

sons or only daughters. If they have eight children, it may not be four sons and four daughters. Yet we say, the probability is 50:50. MS Dhoni may win the toss in 5 out of 6 matches in one series. In another series, he may win only 1 out of 6, while the expected frequency of toss wins is three wins out of six matches in both the series. We attribute such results to what is known as **natural variation**. Then what is the use of prediction? How to confirm the results of probability are true?

To understand this, we have to repeat our coin experiment 100 times. I've done this experiment 100 times with each set consisting of ten tosses and recorded the result how many heads I got in each set.

No. of heads	No. of sets
0	0
1	0
2	1
3	3
4	19
5	52
6	18
7	6
8	0
9	1
Total	**100 sets**

I got a "perfect result" or "expected result" of five heads in only 52 sets of tossing. You can see from the table that six times I got seven heads and once I got nine heads. This is perfectly possible when you keep on repeating the experiment. If I get nine heads on the first occasion itself by chance (or better we call it **natural variation**) and that is the only set I have tried, I may wrongly attribute the result to my supernatural power! As science people we shouldn't do that. We should test before arriving at conclusions. So now I hope you have understood the concept of **natural variation**.

What is the "**expected result**" for the above experiment? The table for expected result for my experiment should have been like this.

No. of heads	No. of sets
0	0
1	0
2	0
3	0
4	0
5	100
6	0
7	0
8	0
9	0
Total	**100**

I should get five heads out of ten tosses in all 100 sets. Now, we have two tables, one actual, another expected. If you apply a test of significance (like Student *T* test, we shall discuss how to do that in a later chapter) and calculate the probability, you will find that the *P* value is 0.5, which is greater than 0.05. Hence, the conclusion is *there is no statistically significant difference between the results of these two tables*. Thus, you can disprove this supernatural power and infer that the experiment was fair and results were within the expected limits.

So now we know that in an actual experiment, it is possible to get nine or even ten heads out of ten tosses even though the expected probability of heads is 5. But the chances are less than 5 % (less than 0.05) of experiments. To put it in other words, we can confidently say in 95 % of cases, such results will not be got. This can be called "***confidence limit of 95 %***".

In the coin experiment, testing and conclusion was easy as we know the expected frequency (which is 5). In clinical settings we do not know the expected frequency. So, we need a **control group** to compare the results. We use this control group to calculate the expected frequency. The results in the study group should differ from the control group significantly. We keep the significance level at probability value or *P* value of less than 0.05. It means the result (significant difference) could have occurred by chance in less than 5 % cases. To put it in other words, our confidence in the results is more than 95 %. Then only we accept the results that there is difference between the groups.

We use various tests of significance to find out this probability or *P* value.

Then the next question "**why do we consider that the results are significant at *P* < 0.05?**"

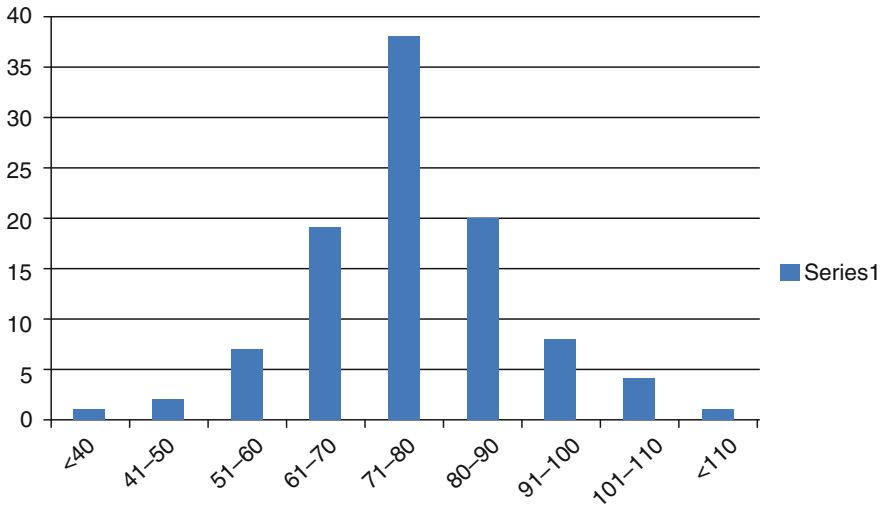
In order to understand this concept, we should know something about "normal curve." Here the word normal doesn't give the same meaning as we use in medicine. It does not mean other types of curves are "abnormal." It is just a statistical word indicating a type of distribution. If an observation or value doesn't fall under a normal curve, it doesn't mean that it is an abnormal value. So probably a better term is "reference interval." In a normal curve, the observations fall around the mean (of all observation) symmetrically. Most of the medical parameters follow a normal distribution curve. The data which give a *normal distribution curve* may be called **parametric** data.

For example, let us collect data of pulse rate of 100 men and arrange them in ascending order. We may get a data like this.

39,48,49,51,53,54,54,55,58,59,61,61,62,62,62,62,62,62,63,64,64,65,67,68,68,69,69,69,70,71,71,71,71,71,71,72,72,72,72,73,73,73,74,74,74,74,75,75,75,75,75,76,76,76,76,76,78,78,78,78,79,79,79,79,80,80,80,81,81,82,82,82,83,83,83,83,85,85,86,86,87,87,87,88,88,89,90,92,93,93,94,96,98,99,100,102,104,104,106,111.

We may not get any idea from these numbers about how the pulse rate is distributed among the normal persons. So, we shall rearrange the data in the form of a table, counting how many of the observations fall in a particular range.

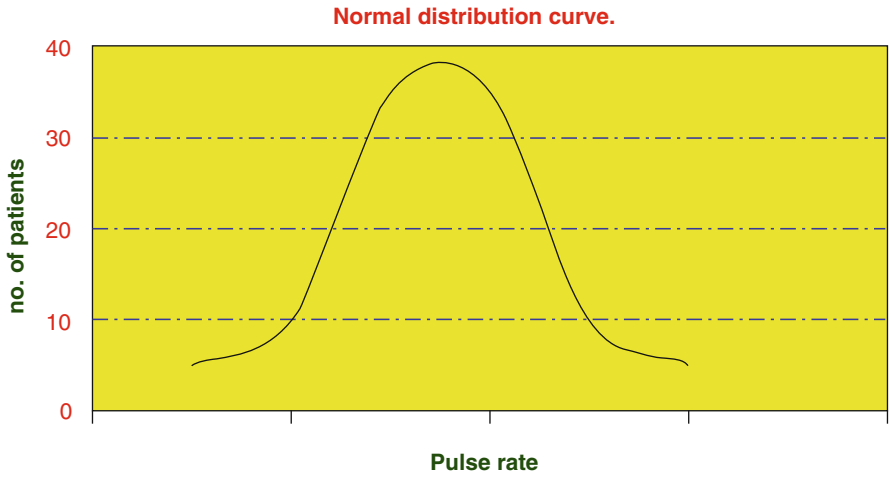
<40	1
41–50	2
51–60	7
61–70	19
71–80	38
81–90	20
91–100	8
101–110	4
>110	1



Graph 3.1 Bar chart: pulse rate frequency

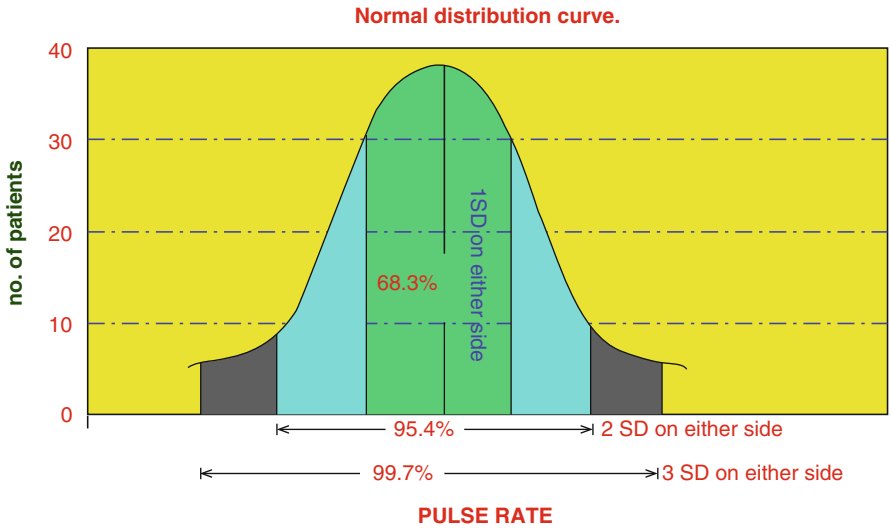
Looking at the graph, you can immediately infer that most of the men have pulse rate in the range of 71–80. This type of chart is known as **bar chart**. The more the height of the bar, the more the frequency of observations falls in that range. This data represents a typical **normal distribution**. Normal distribution only means a type of distribution as it is already mentioned above and *normal* is nothing to do with the meaning of normal that we use in medicine.

If the same graph is represented in a smooth line, we get an elliptical-shaped curve like this.



Graph 3.2 Pulse rate frequency expressed as a line graph

Now we have arrived at a *normal distribution curve*. This is a typical normal curve. The mean, median, and mode all coincide. The maximum no. of observations falls in the center. It's a symmetrical approximately elliptical-shaped curve.



Graph 3.3 Normal distribution curve with its characters

A *normal distribution curve* has certain characters. The central line is the mean value. If we take two **standard deviations (SDs)** (standard deviation is discussed in Chap. 3 on ***Understanding Basic Statistical Terms***) on either side of the mean, it covers 95 % of the area under the curve: 95 % of the observations fall under this area. Of the remaining 5 %, 2.5 % each of the observations fall on either side of the

two SD areas (called the **tails of the data**). That is to say, 5 % fall outside the area of two SDs. So we can call two SDs as 95 % confidence range. It means "the confidence of an observation of falling within the area is 95 %." If an observation falls within these two SD areas, we assume that it is due to chance. Any observation falling outside this area is considered to differ from the observed or expected data. The result is assumed to differ significantly. So, for any difference to be significant, it should be outside the area of two SDs, i.e., beyond two SDs from the mean. To put it in other words, if the probability is less than 5 % or 0.05 or 1 in 20, it is significant. All these values are the same. That is why in a test of significance, if the P value is less than 0.05, its result is considered as significant. So, remember $P<0.05$ is significant in medical statistics. The lesser the value of P, the stronger the validity of the result. The principles of tests of significance, theory of null hypothesis, etc. are dealt in the chapter on ***Tests of Significance***.

Here, I must stress and make it clear that the value of $P<0.05$ is **arbitrary**. We can fix significance level at even lesser level, e.g., at $P \leq 0.005$. Then confidence level increases to 99.5 and the conclusions on the results will be even stronger and more reliable. But it carries a risk of ignoring the true results to an extent (**increases false-negative rates**). It is more difficult to get evidence. Even though result is significant at $P=0.05$ level, it will be discarded as nonsignificant at 0.005 level. The benefits of the drug or a procedure may be ignored. Hence, we need to strike a balance, and for practical purposes, we take the results significant at $P \leq 0.05$. While presenting the results, it is better to mention what is the significance level that we have fixed while carrying out the tests of significance.

When samples are drawn repeatedly from the same population, the mean of each sample differs from other samples, although they are expected to be the same. This happens because only a part of the population (sample) instead of the whole population is measured. This error (which is due to sampling) is called **sampling error** or in common terms results of chance. So, some difference is to be expected always. But this difference is due to chance. When there are two or more groups showing the difference, a question arises: "**Is this difference due to chance or real? Or greater than the expected chance?**" In other words, is it likely to be a true (real) difference in the population mean? That is, the result is significant.

Let us continue with the example of pulse rate in normal men. We already have a table on pulse rate data of normal men. Now, suppose there is a condition where you expect pulse rate would increase (e.g., anemia, appendicitis, the use of a drug, etc.). You collect data from three men with that condition and find the pulse rate in three men as 86, 92, and 96. You can argue that all these rates were found in normal men also as seen in the table as they fall within the two SD ranges of the curve. How to prove that the condition is associated with increase in the pulse rate? We need to have data of more patients with the same condition and create a table like this.

58,60,72,76,78,78,78,81,81,82,82,86,86,87,88,88,88,89,90,92,92,92,93,93,95,95,96,99,99,100,100,114.

If you observe carefully, all these numbers are found in the ***NORMAL*** pulse table also. But if you apply Student's *T* test on these two tables (Tables 3.7 and 3.5), we get the *P* value as 0.000864, which is less than 0.05. Hence, the result is significant and we can conclude that the condition is associated with increased pulse rate.

I did a meta-analysis on the reported articles on intraperitoneal use of polypropylene mesh and newer meshes (which are costlier by 20–25 folds). I found the data on complications summarized in Table 3.1. There are 12 cases of infection out of 719 in polypropylene group versus 29 infections out of 1762 cases of newer mesh group. Similar data on other complications are given in the table. Based on these numbers, how to conclude which mesh is better? By applying tests of significance, it can be shown that there is no statistically significant difference in the incidence of most of the complications between the polypropylene mesh and the newer meshes. This example illustrates how to interpret the data in a journal (Ramakrishna H.K., Lakshman K. **Intraperitoneal polypropylene mesh and newer meshes in ventral hernia repair: what EBM says?**. *Indian J. Surg.* 2013;75:346–351).

Table 3.1 Summary of results

	N	Infection	Fistulization	Sinus formation	Adhesions	Recurrence	Seroma
PPM	719	12	1	7	7	21	40
Newer mesh	1762	29	2	12	4	76	33
Statistical test used		Chi-square test	Fisher's exact test	Chi-square test	Fisher's exact test	Chi-square test	Chi-square test
P value		0.967	1	0.452	0.018	0.117	<0.0001
Significance		NS	NS	NS	Significant	NS	Significant

So, now we have understood biostatistics, probability, *P* value, why *P* value is kept significant at 0.05, what the tests of significance are, and how to interpret the results of the tests.

Skewed Distribution Curve

In contrast to a normal distribution curve, in **skewed distribution curve**, majority of observations fall on to one side of the mean. The mean, median, and mode do not coincide. The curve is not bilaterally symmetrical, rather asymmetrical. The data which give skewed distribution curve are called **nonparametric**. For example, serum levels of amylase and lipase in acute pancreatitis follow a skewed distribution

if plotted against time. If blood glucose levels are recorded after a meal in a particular patient, we may get the data as in Table 3.2.

Table 3.2 Blood glucose levels at various time intervals after food

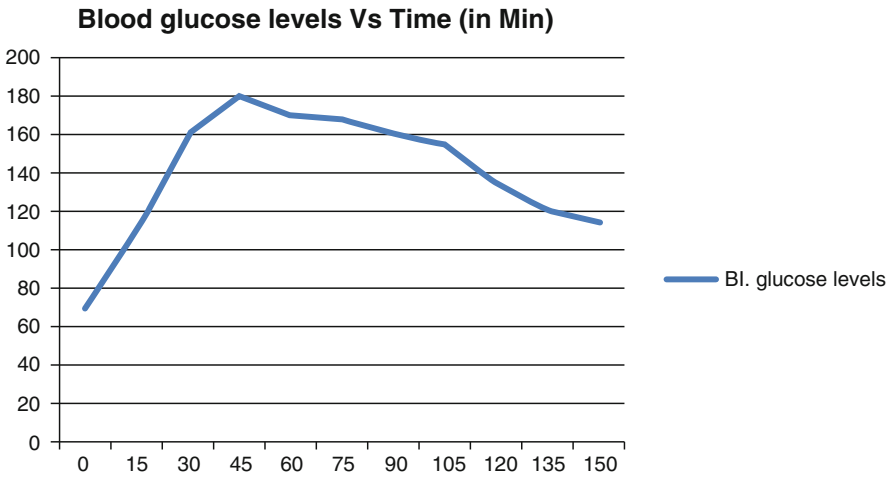
Time (min)	Bl. glucose levels
0	70
15	110
30	160
45	180
60	170
75	168
90	160
105	155
120	135
135	120
150	115

If mean, median, and mode are calculated on this data, we can see that they do not coincide but differ significantly.

Inferences from Table 3.2

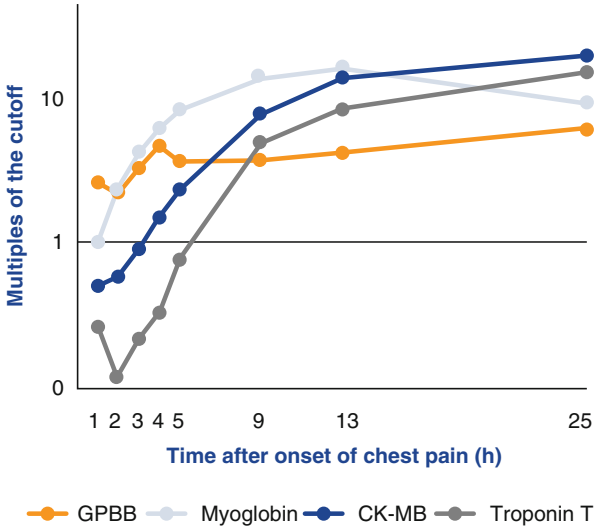
Mean	140
Median	155
Mode	160

A graph depicts the distribution better. So we call this type of curve a skewed distributions curve. More observations fall on the right side of the mean.



Graph 3.4 Examples of skewed distribution curve

Another example of skewed or nonparametric data is serum levels of cardiac enzymes after an attack of acute chest pain in myocardial infarction (downloaded from the Internet https://en.wikipedia.org/wiki/Cardiac_marker#/media/File:Cardiac Marker Comparison2.gif).



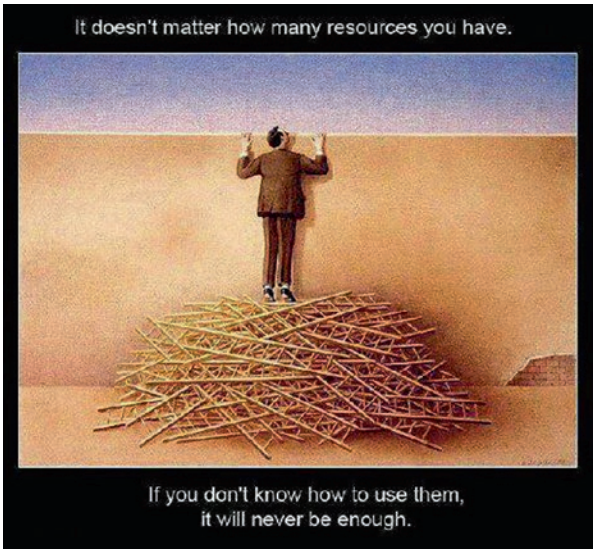
Graph 3.5 Examples of skewed distribution curve

Understanding Basic Statistical Terms

4

Learning Objectives
Measures of central tendency
Measures of dispersion
Confidence interval and confidence level
Sampling errors
Conditional probability
Independence

I saw this photo on the Internet many years back. Now I do not know from where I got it, to cite the reference, but I thought it is worth sharing with you. No matter how much of information and resources you possess if you do not know how to use them, you will land up like this…:



© Springer Science+Business Media Singapore 2017
H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_4

Actually all he needs is a single ladder: and he has plenty of them. Still he is struggling to see. This happens because he does not know how to use his resources. This is what happens for many of us. We have plenty of materials but it is all lost in the course of time as we do not document them in a systematic way and analyze and publish the results or inference. We need knowledge of at least basic statistics for this purpose.

Statistical inference can be defined as the *process of generating conclusions about a population from a noisy sample.* Without statistical inference we're simply living within our data. With statistical inference, we're trying to generate new knowledge.

If we simply collect some data and look at it, we may get any useful information merely from the numbers. If we express the data in some statistical terms, we get some useful information or results with which we can predict similar outcome on some other data. For example, we shall consider data of pulse rate of 75 healthy men and arrange them in ascending order.

Table 4.1 Pulse rates of 75 normal adults males

68,68,69,69,70,71,71,71,71,71,71,72,72,72,72,73,73,73,74,74,74,74
75,75,75,75,75,76,76,76,76,76,78,78,78,78,79,79,79,79,80,80,80,81,81
82,82,82,83,83,83,83,85,85,86,87,87,87,88,88,89,90,92,93,93,94,96
98, 99,100,102,104,104,106,111

As I said earlier, if we just look at these numbers, it doesn't mean much. If I state that the average pulse rate is 81, it is more understandable to medical men. It is a useful information. This is one type of statistical inference. Likewise we come across many statistical terminologies while reading journals, some of which we shall discuss in this chapter.

Measures of Central Tendency

Mean, median, and mode are the measures of central tendency.

Mean is simply the average of all observations. We calculate the mean by adding all the observations and dividing the total by the number of observations. In the above example, $68+68+69+\ldots+106+111=6101$ is the total. There are 75 men. So the mean is $6101/75=81.34$.

Median is the central observation when the observations are arranged in ascending or descending order. For simplicity consider the example data 53,54,54,55,58,59,61,61,62. The observations should be either in ascending or descending order, if not already arranged. Then the central, in this example 58, the fifth one, is the median. If there are even numbers of observations, there will be no single central number, but two. In such cases, the average of

the two central numbers will be the median. For example, consider this data: 53,54,54,55,58,59,61,61,62,64. Here, the median will be average of 58 and 59, which are the two central observations. So the median is 58 + 59 divided by 2, i.e., 58.5. Median can be viewed as a number which separates the higher half from the lower half.

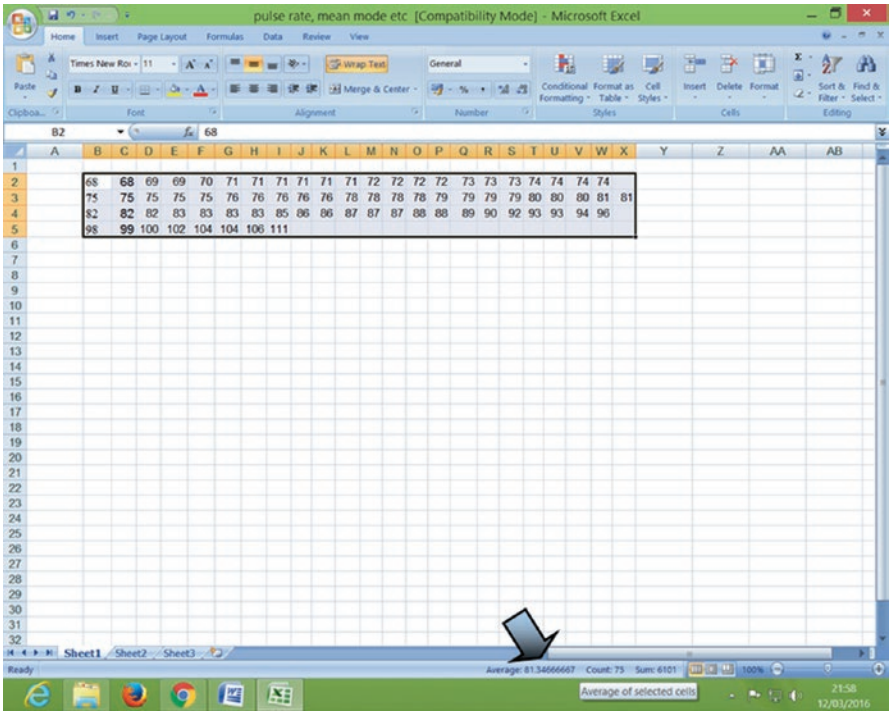
Mode is simply the most frequently occurring observation. In the above example table, 71 occurs six times (more frequent than any other number). So the mode is 71. If there are two numbers which occur in the same frequency, then there are two *modes*. If all numbers are appearing only once, then there is no *mode*.

From the above table, we can draw certain ***inferences***:

1. The expected pulse rate in an adult male is around 81.
2. About 50 % of the men have pulse rate of less than 79.
3. The commonest pulse rate found among men is 71.

These give us some idea on pulse rate.

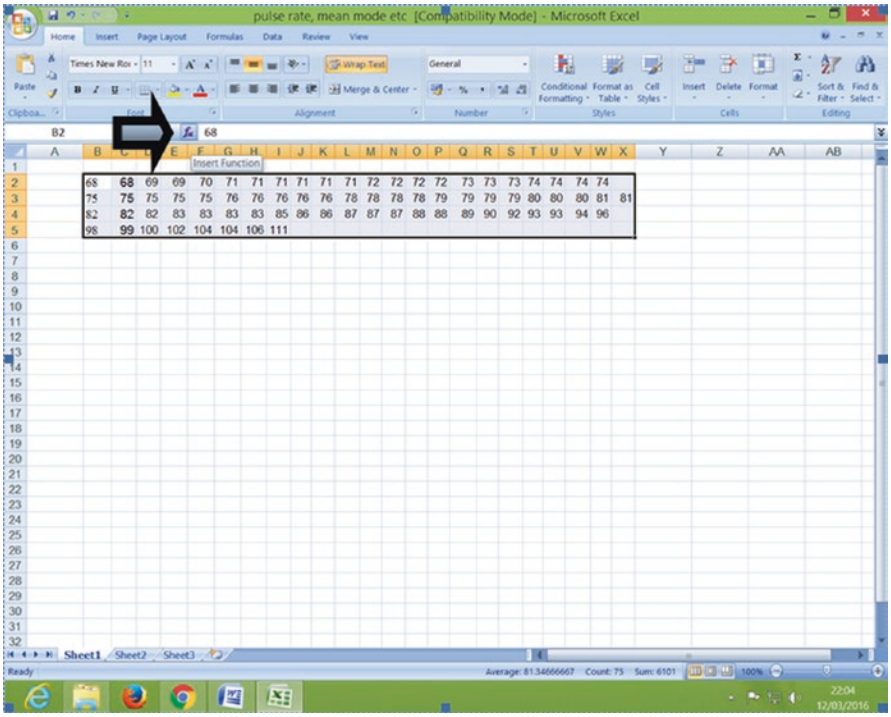
MS Excel or WPS Spreadsheet can calculate these parameters easily. First we have to enter the data in the cells of Excel sheet. Select the data by using shift and arrow keys or by using the mouse.



Data entered in MS Excel sheet

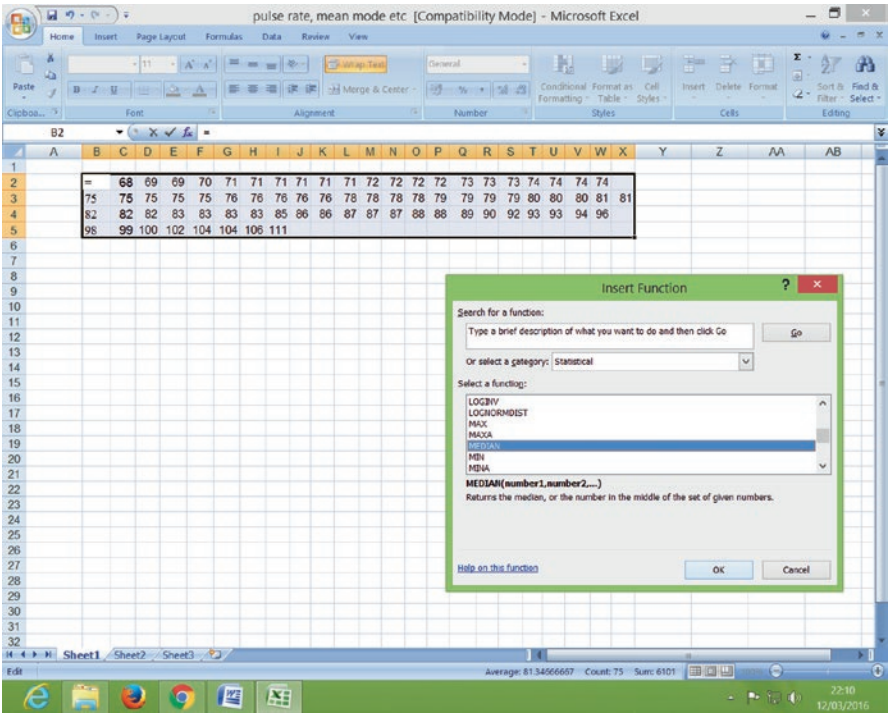
Excel is already showing the average or mean (arrow). It is also showing the total number of observations and the sum of all these numbers.

To calculate the median, select an empty cell by clicking over an empty cell. Click on Fx button at the formula bar (arrow).



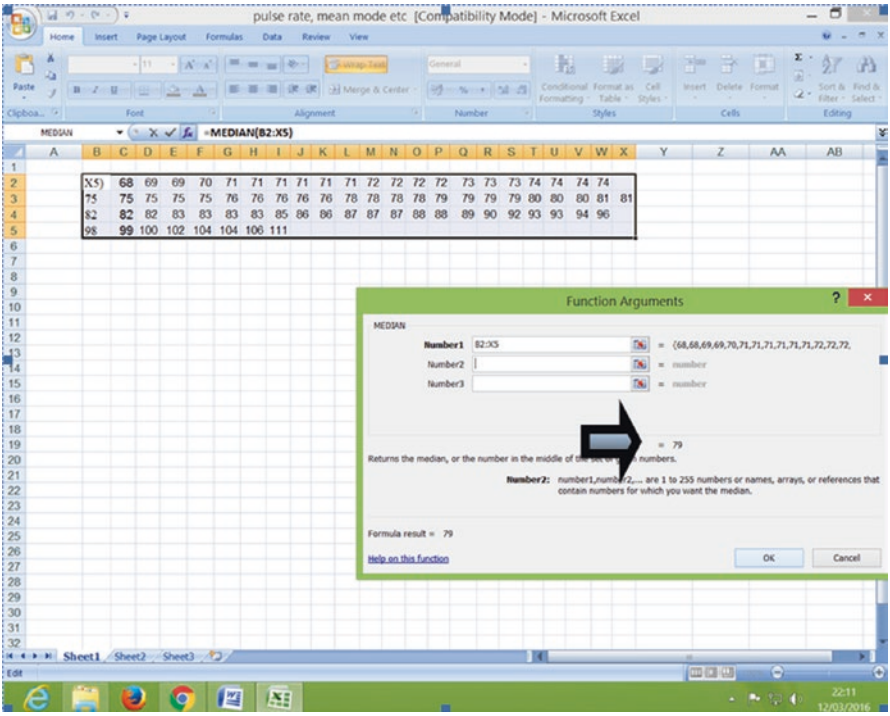
Click on Fx (*arrow*)

It opens a window. Select "category" as "statistical." It shows the list of statistical functions to be calculated. Select median (or mode or any other function you want).



Select “category” as “statistical” and function “Median”

And click OK.



Type the cell range or simply select the cells which contain the data

In this window, click on the box next to Number 1 and type the cell range or simply select the cells which contain the data. It shows the result immediately (arrow).

Similarly we can easily calculate many statistical functions using the computer.

Measures of Dispersion or Spread

Range **(the smallest number, the largest number)**
Mean deviation: $\mathrm{MD} = \frac{1}{N}\sum_{i=1}^{N}\lvert x_i - \bar{x}\rvert,$
Standard deviation: $\sigma = \text{sq root}\{S(x-x^{-})^2/n\}$

Range It indicates the range within which the data is spread. For example, in the table above, range is 68–111.

Mean Deviation Deviation is how much each data deviates from the mean. In the above Table 4.1, considering mean as 81, the deviation for "the data 68" is

81–68 = 13. Similarly for "the data 111," deviation is 30. Here we ignore the negative sign (81–111 = −30). Likewise if we find deviations for all the data and find the mean of these deviations (by adding all deviations and then dividing this total by the number of observations), we get *Mean Deviation*.

For simplicity we shall take data of only ten observations (Table 4.2). The table shows how to calculate the mean deviation manually.

Table 4.2 Calculating Mean Deviation

	Mean = 94	No. of observations = 10
Data	Deviation (mean-data value)	Ignoring negative value
88	94–88	6
89	94–89	5
90	94–90	4
92	94–92	2
92	94–92	2
94	94–94	0
96	94–96	2
96	94–96	2
100	94–100	6
103	94–103	9
	Sum	38
	Mean Deviation is 38/10 = 3.8	

If the mean is a fraction, it is difficult to calculate the deviation for a large number of data. Also, as we ignore negative sign for some data which are higher than mean, it is difficult to put it mathematically. In medical statistics standard deviation is more often used.

Variance and Standard Deviation

These are (in contrast to mean, median, and mode which are measures of central tendency) measures of spread.

Standard deviation (SD) is *e*xpressed as σ (Greek letter sigma). Formula: $\boldsymbol{\sigma}$ = sq root$\{S(x-x^{-})^2/n\}$.

It expresses how much the individual data deviates from the mean or, in other words, how the data is spread.

Calculating SD is a five-step process.

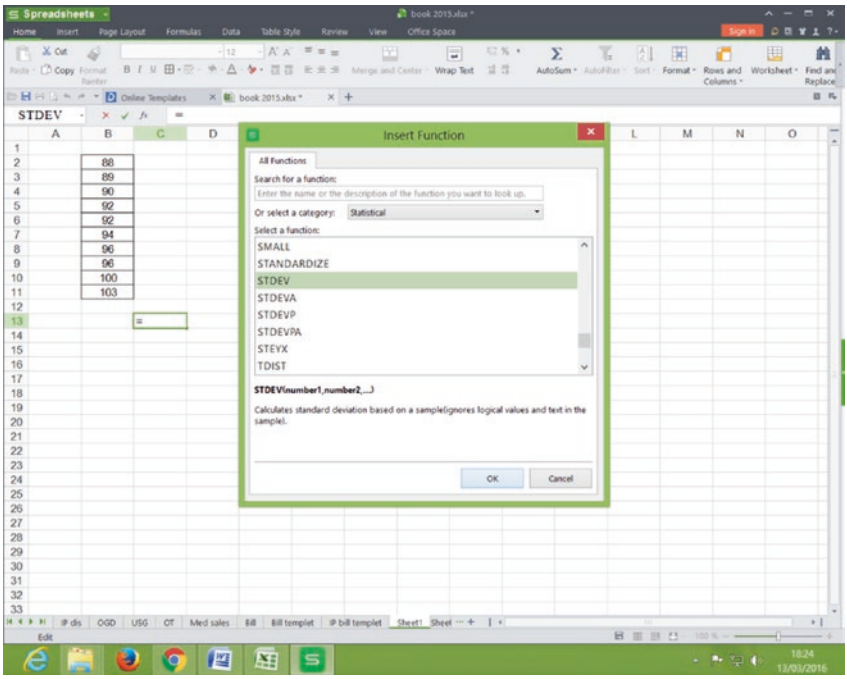
1. Find the mean.
2. Find deviation for each observation (deducting each data value from the mean).
3. Square this deviation.
4. Find the mean of these squared deviations (variance).
5. Take the square root of this mean value, that is, the SD.

To calculate SD for the above data:

Calculation of variance and standard deviation

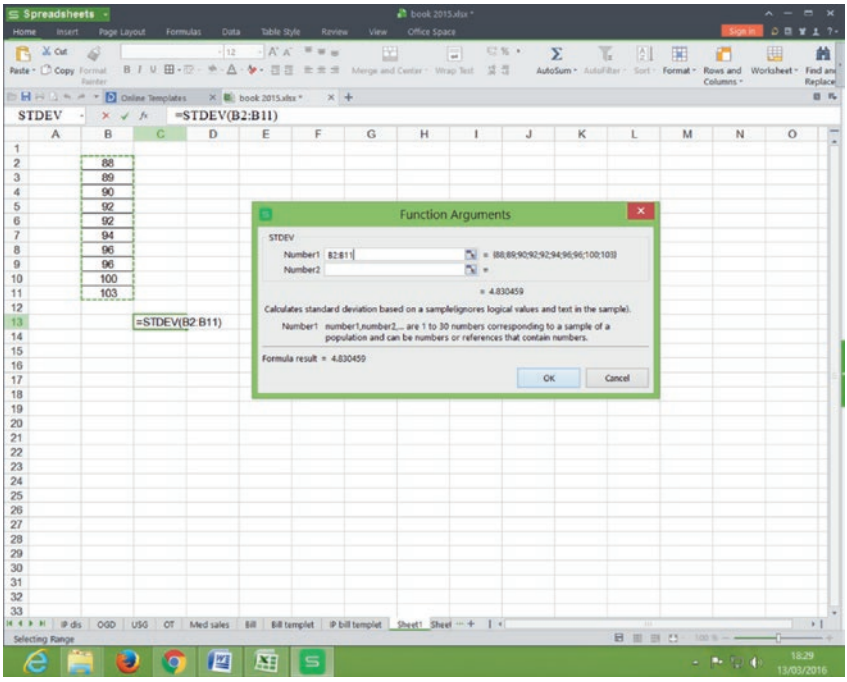
	Mean = 94	No. of observations = 10	
Data	Deviation (mean-data value)	Deviation	Square of deviation
88	94–88	6	36
89	94–89	5	25
90	94–90	4	16
92	94–92	2	4
92	94–92	2	4
94	94–94	0	0
96	94–96	2	4
96	94–96	2	4
100	94–100	6	36
103	94–103	9	81
		Sum	210
Mean of square of deviation is 210/10 = 21. This is the **variance**			
	SD = sq root of 21	= 4.58	

SD can be calculated easily on computers. This time I've shown it on WPS Spreadsheet. The same can be done with Excel also. Data is entered in the spreadsheet. Select an empty cell. Click on Fx. Function window opens. In the category select statistical and, in the function, select STDEV (standard deviation).



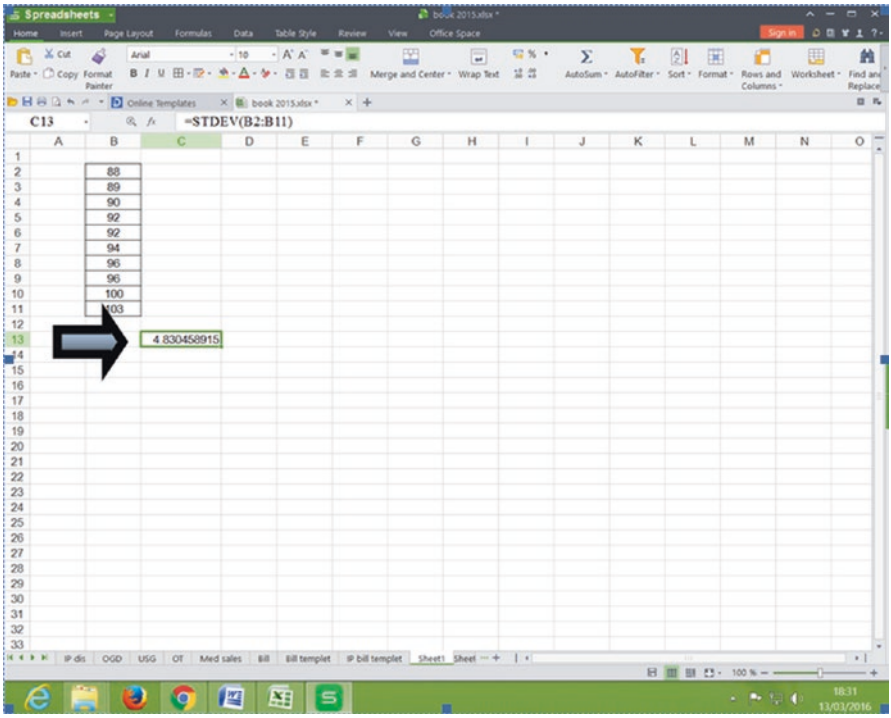
Calculation of SD with the help of MS Excel sheet

Click OK.



Data is entered. Click on Fx. Select Category as “Statistical”. Select function as “STDEV” (standard deviation)

Enter the cell range in the window next to Number 1 (or simply select the cells containing data). Click OK.



Enter cell range in the box (or simple select the cells containing data). Click OK

You have the answer: 4.83.

This is somewhat different from our manual calculation. This is because Excel or Spreadsheet uses ***sample SD***. If our data represented the whole population, we must use the total/n to find variance. This is called ***population SD***. In our example, population variance is 210/10. If data is a sample of a larger population, we have to use the total/n-1 to find variance. In our example, sample variance is 210/9 (SD = square root of variance, i.e., 210/9 = $\sqrt{}$23.33 or 4.83). This is the method computer uses.

From here onward, we shall not discuss manual calculations as it is difficult and unnecessary. We shall see how calculations can be done on computers easily.

I have created a ***stats calculator*** module on Excel sheet to easily calculate statistical parameters like mean, median, mode, standard deviation, confidence interval, etc. To use them, open the following link with your browser https://drive.google.com/open?id=0B4uZKhNcSM7cWTBnUVhOQ3NuRlE. Download and open the file as Excel files to use them. Explanation regarding how to use is given in the Excel sheet itself.

Confidence Interval and Confidence Level

Confidence intervals consist of a range of values (interval) that act as good estimates of the unknown population parameter. It is easier to understand with an example. For example, we have 10,000 men group. We want to estimate average pulse rate of this population. Since it is difficult to count pulse rate of all these men, we take a random sample of 100 men, count the pulse rate, record, and find the average. Let us say we get average pulse rate 70/min. We assume that 70 is the average pulse rate of 10,000 men group also. If we repeat the job by choosing different samples of 100 men, we may get different averages. It may be 65 or 75 or 80 or any other value. Since we do not know the population average, we cannot say which one is the correct value. So we calculate the confidence interval. The population average falls within this confidence interval (range). If confidence interval is 50–85 with an average of 70, it means that the population's (10,000 men group) average pulse rate falls within 50–85.

This concept holds well for not only average but for many other statistical parameters.

Confidence level is an indicator of how many times the population average falls within the confidence interval, if we calculate the average on different samples from the same population. In medical statistics usually confidence level is fixed at 95 %. For example, if we say "At confidence level of 95 %, confidence interval of average pulse rate of men is 60–90," it means that if we draw repeated samples (say 100 times) and find the average pulse, it falls in the range of 60–90 in 95 % of samples. In other words, there is still a possibility of the parameter falling out side this range in 5 % of cases.

Confidence interval at 95 % confidence level is calculated by taking two standard deviations on either side of the average (confidence interval = (average-2 SD) to (average + 2 SD) for 95 % confidence level). For example, if average is 70 and SD is 10, then confidence interval is (70–10) to (70 + 10), i.e., 60–80.

Errors Type 1 and Type 2 (Alpha and Beta, Respectively)

In statistical hypothesis testing, we come across certain types of errors. Sometimes there is no effect but the test shows the effect. This is **false-positive or type 1 error**. There is some effect but we fail to detect it by our testing. This is ***false-negative error or type 2 error***.

Example

FNAC is done on a group of patients to detect thyroid carcinoma. FNAC is reported as malignancy in 100 patients and nonmalignant in another 100 patients. When specimen is subjected to histopathology after surgical excision, the following results are obtained. So we have four possibilities.

Pre-operative FNAC correlation with post-operative HPE report

	FNAC: malignancy	FNAC: nonmalignant
Histopathology report: malignant	95	7
Histopathology report: nonmalignant	5	93
	$N = 100$	$N = 100$

Considering HPR is the final and confirmatory in detecting malignancy, we can see from above table that:

1. FNAC has detected malignancy in 95 % of cases. **True positive or sensitivity rate of the test or positive predictive value**.
2. FNAC has wrongly detected malignancy in 5 % cases. These cases are reported as malignancy in FNAC but these cases are actually nonmalignant. These are false-positive cases or ***type 1 error***. FNAC has correctly ruled out malignancy in 93 % of cases. **True negative or specificity rate of the test or negative predictive value**.
3. FNAC has wrongly ruled out malignancy in 7 % cases. These cases were actually malignant but FNAC has missed the diagnosis. In other words, FNAC has 7 % false negativity. This is *type 2 error.*

A false-negative test report gives false assurance to the treating doctor and patient and hence results in no or inadequate treatment. A false-positive test result on the other hand produces tension and panic and may result in overtreatment.

Sensitivity rate of a test is its ability to pick up the correct diagnosis.

Specificity rate of a test is its ability to rule out the diagnosis correctly.

Factors Increasing Type II Error (False Negativity)

1. Sample size: the larger the sample size, the lesser will be the type 2 error.
2. Lower the significance level: keeping significance low leads to a higher error rate. In other words, higher levels of standards will lead to increasing chances of missing some positive results.
3. Effect size: for a small effect size, error rate is more (means higher chances of missing a rare condition or a rare complication).

Conditional Probability

If we apply a condition to a probability, the probability increases. This is called *conditional probability*. For example, let us say the probability that a woman develops carcinoma breast is 0.001. If you apply a condition like the mother had carcinoma breast, then probability of a woman developing carcinoma breast increases to 0.01. To give another example, if one of the parents has diabetes mellitus, 25 % of the offsprings will develop diabetes. If both the parents have diabetes, 50 % of the offsprings will develop diabetes.

Independence

Two variables are said to be independent, if occurrence of one does not affect the probability of the other variable. For example, let us consider the sex of the two successive babies born in a hospital as the two variables, they are independent variables

because the sex of a baby born does not affect the sex of the next baby. If the probability of developing hernia is 10 % in a series, then the probability of developing hernia after the present surgery is 10 % if the immediate previous patient in the series had developed hernia. It remains 10 % even if the previous patient had not developed hernia. That is to say probability will not increase or decrease irrespective of hernia that occurred or not in the previous case. So we say the events are independent.

Let us consider mortality and infection as the two variables. If study finds that probability of mortality is not affected irrespective of infection occurs or not, then these two are independent variables. On the other hand, if study finds that probability of mortality increases whenever infection occurs, then these two are dependent variables.

Normal or Gaussian Distribution and Skewed Distribution

These are discussed in Chap. 2.

Exercises:

1. Ages of the patients recruited to a study are given in Table 4.3. Find mean, median, mode, standard deviation, and range for the data using the ***stats calculator*** module which can be downloaded from https://drive.google.com/open?id=0B4uZKhNcSM7cWTBnUVhOQ3NuRlE.

   Can you calculate yourself manually?

Table 4.3 Age of the patients

28,36,20,17,50,55,29,51,49,30,48,35,60,18,44,25,35,50,39,30,40,29,28,46,29

2. A study was conducted to evaluate the ultrasound scanning in the diagnosis of acute appendicitis. Ultrasound scanning was done on all the patients undergoing surgery. Results are given in Table 4.4. Post surgery, all the specimens are subjected to histopathology to confirm the diagnosis of acute appendicitis. Considering histopathology as gold standard in the diagnosis, calculate the negative predictive value of ultrasound scanning in the diagnosis of acute appendicitis.

   What is the sensitivity and specificity of ultrasound scanning in the diagnosis of acute appendicitis?

Table 4.4 USG diagnosis of appendicitis correlation with post operative HPE report

	USG: appendicitis	USG: no appendicitis	Total
HPR: appendicitis	112	31	143
HPR: no appendicitis	18	12	30
Total	130	43	173

I have created a module for calculating sensitivity and specificity module on Excel sheet to easily calculate these statistical parameters. To use them, open the following link with your browser https://drive.google.com/open?id=0B4uZKhNcSM7cWTBnUVhOQ3NuRlE. Download and open the file as Excel files to use them. Explanation regarding how to use is given in the Excel sheet itself.

3. Suppose the risk of developing myocardial infarction is 20 % if the patient is having diabetes mellitus. If the patient is a smoker, the risk increases to 35 %. What statistical concept is applied here?

4. In a gambling game of tossing the coin, consecutively four times head occurred. Observing this, a gambler bets on tails. What are the chances of him winning now?

(Answer to exercise 3 is conditional probability. Answer to exercise 4 is 50 %. Events of flipping the coin are independent. The probability is 50 % only irrespective of how many times heads appeared previously.)

5. The birth weights of consecutively born 25 babies are given in Table 4.5. Based on the data, define confidence interval at with 95 % confidence level.

Table 4.5

Birth weights of the babies
2.5
2.3
3.2
3.0
1.6
1.8
2.6
2.5
3.6
1.9
2.8
1.9
3.1
2.9
1.6
1.9
2.8
2.9
3.1
3.6

Clue, find mean and standard deviation. CI = mean ± 2 SD

Tests of Significance

5

Learning Objectives
Chi test
Fisher's test
T test
Paired *T* test
Mann–Whitney–Wilcoxon (*MWW*) *test*
Finding *P* value using computer-/Internet-based *calculators*
R module
When to use which test?
Examples to understand the above concepts and some related terms

Medical Science is Science of uncertainty and Art of Probability.
William Osler

We have discussed why tests of significance are required in Chap. 3. Their basic concepts and phenomenon of *natural variation* are also discussed. These tests first assume that there is no significant difference between the two groups. This we call *null hypothesis*. Then calculate the probability or *P* value. If *P* or probability value is more than 0.05, the null hypothesis is accepted. That means there is no significant difference between the groups. If *P* value is less than 0.05, null hypothesis is rejected and alternate hypothesis is accepted. That means there is significant difference between the two groups. The lesser the *P* value, the more significant is the difference.

In this chapter we shall discuss three tests of significance in a little more detail.

1. Chi-square (χ^2) test or simply chi test
2. Student's *T* test or simply Student's test
3. Fisher's exact test or simply Fisher's test

These three tests are commonly used in medical statistics on various types of data.

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_5

Chi-Square Test or Simply Chi Test

Sometimes the term goodness of fit or test of homogeneity is also used for this test.

Chi test is used to test whether the difference between the two proportions is significant or not. It calculates the probability. It also tests independence of two categorical variables. As it is already mentioned, we take the result significant if the probability is less than 5 % or *P* less than 0.05.

For this type of data, chi-square test is useful.

Effectivity of drugs A and B in controlling infection

	No. of patients	Infection	Percentage
Drug A	42	16	38.09
Drug B	10	5	50
Total	52	21	

This type of data we commonly come across in many of the studies. We may be comparing a surgery with another type of surgery. For example, highly selective vagotomy is compared with truncal vagotomy. In case-control studies, the study group is compared with the control or placebo group. The following table shows another example:

Table 5.1 Incidence of recurrence of hernia in hernia repairs with and without mesh

	No. of patients	No. of recurrences (at 2 years)
Mesh hernioplasty	1274	11
Herniorrhaphy without mesh	1756	68

These are typical 2 × 2 tables (two rows and two columns of data). Chi test can also be applied for 2 × 3 or 3 × 3 tables or for tables with higher numbers of rows and columns.

There are certain conditions to apply chi test:

1. All values should be actual numbers.
2. You cannot use percentages or averages.
3. The value in all the cells should be five or more*. (For smaller sample sizes, Fisher's test is applied.)
4. The sample should have been randomly drawn.
5. Variable should be *categorical***.

(***Yates' correction** is applied to improve the accuracy when numbers are small.

***Categorical Variable*: Variables can be *categorical* or *continuous*. Categorical variables can take only certain value. For example, survival or death, male or female, infection or no infection, recurrence or no recurrence, etc. There are no in-between values.

Continuous Variables: Continuous variables are the variables which can take infinite values. For example, blood glucose levels,can be 90, 90.1, 90.2, 90.3, etc.,

and age can be 55, 55.1, 55.2, 56, 58, etc. In between two values, also it can take any number of values.

Continuous variable can be converted to categorical variable by creating categories. For example, age, we can create categories such as </=10, 11–20, 21–30, 31–40, etc. Now, age can take only one of these categories, So take only one of the values.)

If these criteria are not satisfied, the test cannot be applied or the results are not valid. For example, you cannot apply chi test for this table even though the data is the same as the above table written in a different way, because the table contains percentage.

Incidence of recurrence of hernia in hernia repairs with and without mesh (recurrence expressed as percentage)

	No. of patients	No. of recurrences (at 2 years) (%)
Mesh hernioplasty	1274	0.86
Herniorrhaphy without mesh	1756	3.87

To understand chi test, we should understand certain terminologies.

Observed Frequency It is the data presented in the table we got from the study. For example, in Table 5.1, 1274, 11, 1756, and 68 are the observed frequencies.

Expected Frequency It is the frequency we expect if there is no significant difference between the groups (null hypothesis is true). This can be calculated by pooling the results. To understand it better, we shall rewrite Table 5.1 by adding a column and a row for the total (See Table 5.2).

Table 5.2 Incidence of recurrence of hernia in hernia repairs with and without mesh (recurrence expressed as total)

	No. of no recurrences	No. of recurrences (at 2 years)	Total
Mesh hernioplasty	1263	11	1274
Herniorrhaphy without mesh	1688	68	1756
Total	2951	79	3030

Now, there are 79 (2.6 %) recurrences out of the total 3030 (sample size). If there is no difference between the two groups, we expect recurrences in the same percentage in both groups. For mesh group this would be 2.6 % of 1274 = 33 cases (rounding off to the nearest whole number). Another way to calculate is 1274*79/3030 = 33 (2.6 %) recurrences for 1274 cases. But observed frequency is 11 which is less than the expected. Similarly for nonmesh group, we expect 1756*79/3030 = 46 recurrences. But the observed frequency is 68 which is more than the expected. So, mesh group has less than the expected recurrences and nonmesh group has more than the expected. The question is, can this result be out of **natural variation** (discussed in Chap. 2) or significant? Chi test decides it by calculating the *P* value or probability

of the result. If we find P value less than 0.05 or less than 5 % probability, then we take the result as significant.

Degree of freedom: It is given by the formula:

$$\mathrm{DF} = (\text{no. of rows in the table} - 1) \times (\text{no. of columns in the table} - 1)$$

For Table 5.2 (consider only the data and ignore the total, headings, etc.), it is a 2×2 table. So the degree of freedom is $(2–1) \times (2–1) = 1$.

If we look at the actual calculation to find chi value and P value, it is very cumbersome for doctors. The reader can refer to statistics books for details how to calculate, if interested. Doctors usually do not do the calculations and seek the help of statisticians. But, if we understand the principles, it can be easily done on the web with what are known as calculators (Fig. 5.1). For example, this is one such calculator for chi test (http://vassarstats.net/newcs.html).

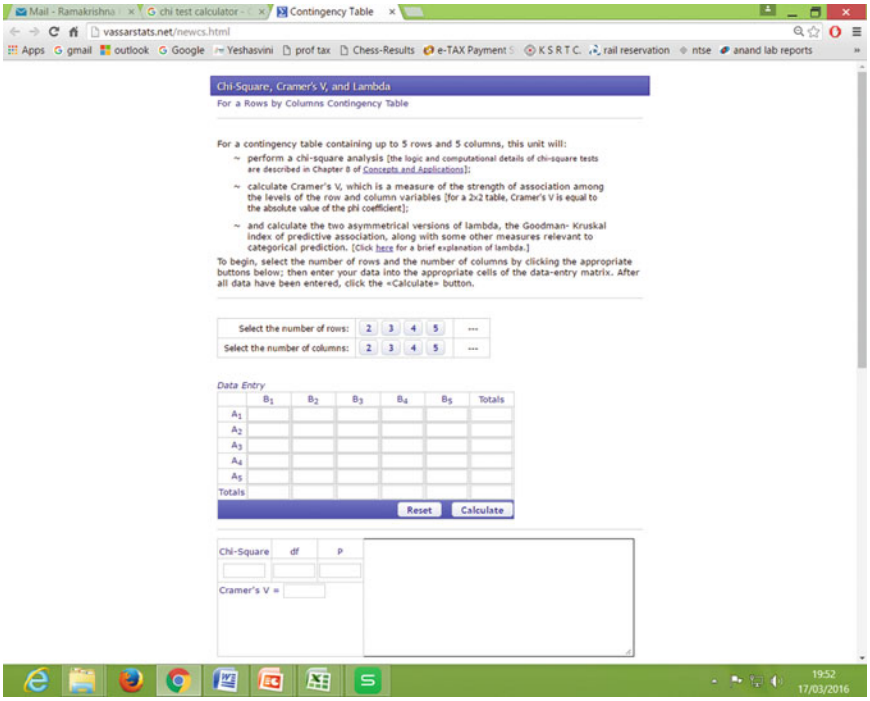


Fig. 5.1 Vassarstat calculator

There are some descriptions that we need not bother for our present requirements. What we have to know for calculation is that we have to enter the number of rows and the number of columns in our table. Select 2×2 for our Table 5.2 (arrow) (Fig. 5.2).

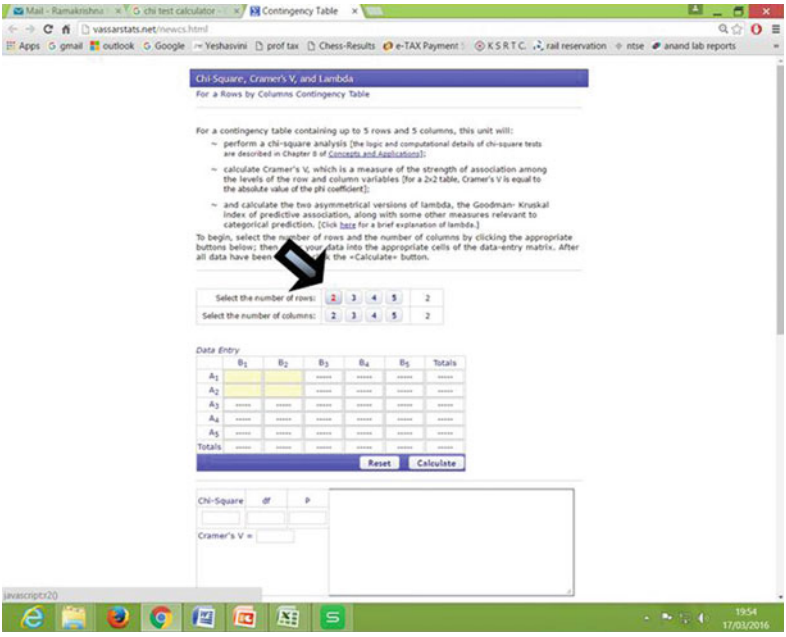


Fig. 5.2 Vassarstat Calculator: Number of rows and columns selected

In the data entry area, our data is entered (arrow) (Fig. 5.3). Click on calculate. We get the result: $P<0.0001$, highly significant. If this is practiced once on the web site, it will be better understood.

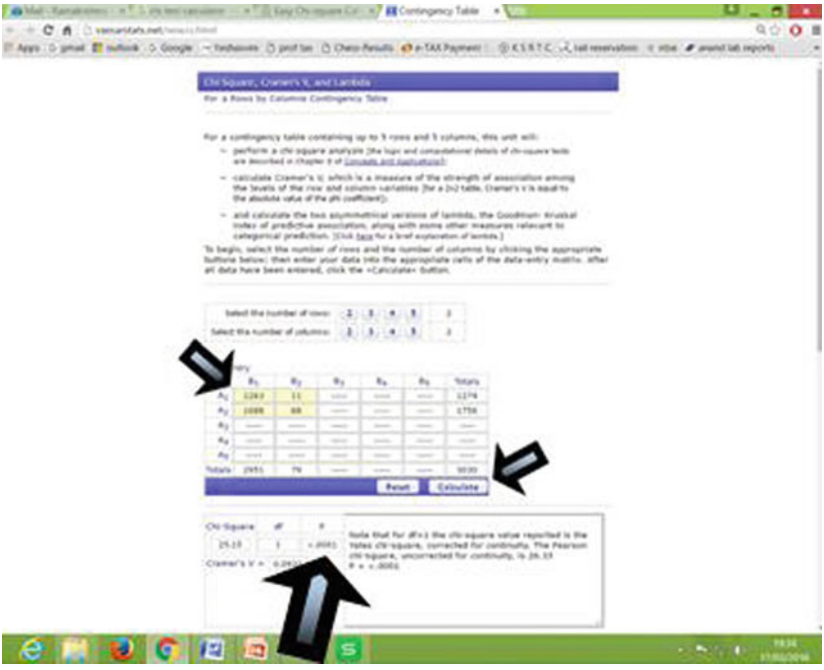


Fig. 5.3 Vassarstat Calculator: Data entered, result is shown

Fisher's Exact Test

When the frequency is small, we cannot use chi test. In such cases we must use Fisher's exact test or simply Fisher's test. This test is used in similar situations where chi test is used but the frequency is small. It is already mentioned that chi test is invalid if the value in each cell is less than 5 for a 2×2 table. In such cases Fisher's test is useful.

For example, the incidence of fistula after intraperitoneal placement of polypropylene mesh and newer mesh is shown in Table 5.3. For this large number of cases, we have only one and two fistulae. So we cannot use chi test.

Table 5.3 Incidence of fistulisation in PPM and newer mesh

Mesh type	*N*	Fistulization
PPM	719	1
Newer mesh	1762	2

The problem with Fisher test is it is difficult to calculate when the number is large as it uses factorials in the calculations as can be seen in the formula below: 5 factorial = 5×4×3×2×1.

Formula for Fisher's Test

$$P = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,n!}$$

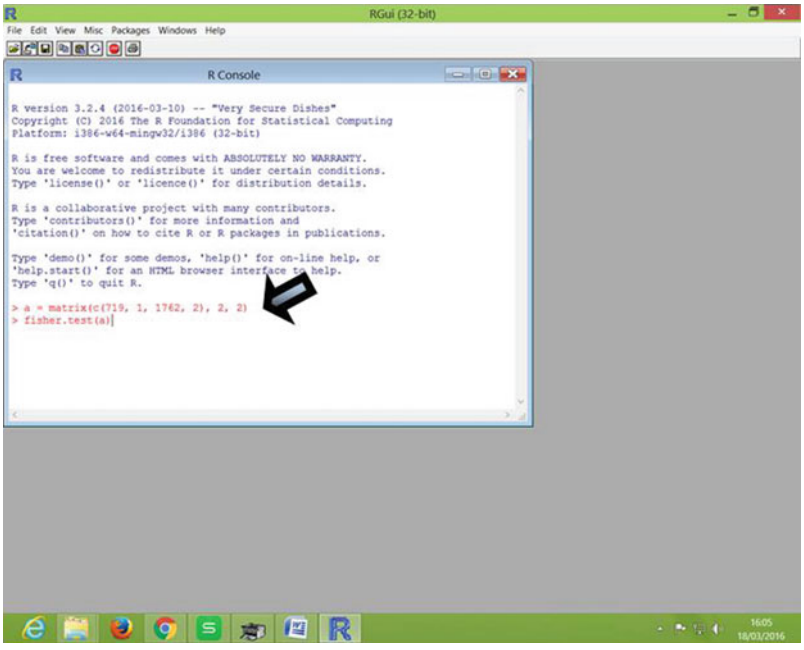
! Is the Symbol for Factorial

You can imagine how difficult to calculate 900 factorial. However, computer modules are now available which can handle larger numbers and that makes the job easy. For example, we shall calculate *P* value using Fisher's test for this table.

This can be done on R Programme [R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/]. R Programme is a free download and does not need any license to use. It can be downloaded from https://cran.r-project.org/.

R module: download

This is known as R console.



R module: R console

[redacted]

It has a help menu and you need to study and learn how to write an argument. This is the argument for Fisher's test. I've entered the data. 2,2 implies it is a 2×2 table. These are our numbers. Type Fisher test (a) and press enter.

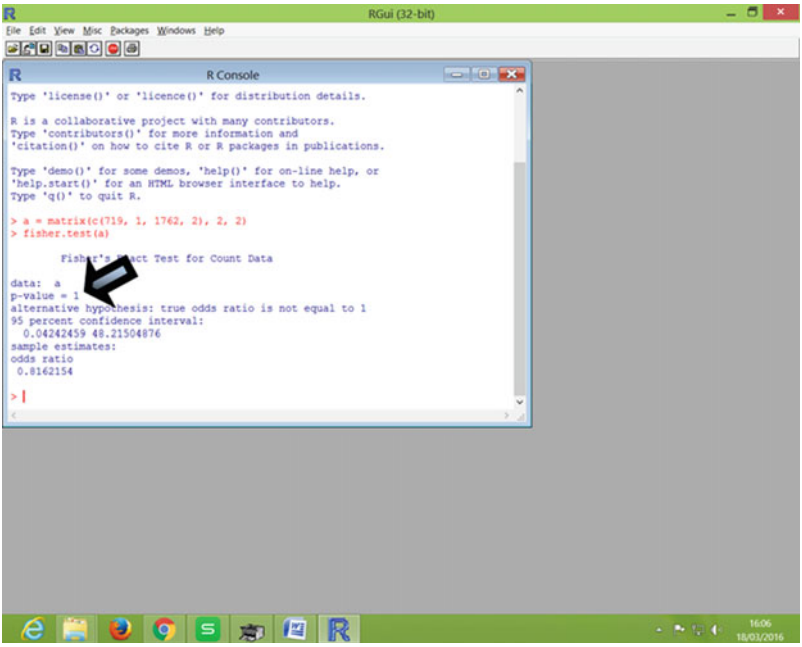


Fig. 5.4 Fisher's test using *R* module

That's all. You have *P* value = 1. As it is more than 0.05, the difference is not significant. This is a freeware and you can download freely from this web page. For further details regarding statistical tests, the reader is advised to visit their web site, http://www.ats.ucla.edu/stat. The readers can try asking questions at CrossValidated, http://stats.stackexchange.com, a question-and-answer board for people interested in statistics.

If downloading and studying regarding how to enter an argument in *R* module is difficult, there are online calculators. They can guide the user online. They are quite simple to use. There is one such online calculator available at http://www.socscistatistics.com/tests/fisher/Default2.aspx (Figs. 5.4–5.8).

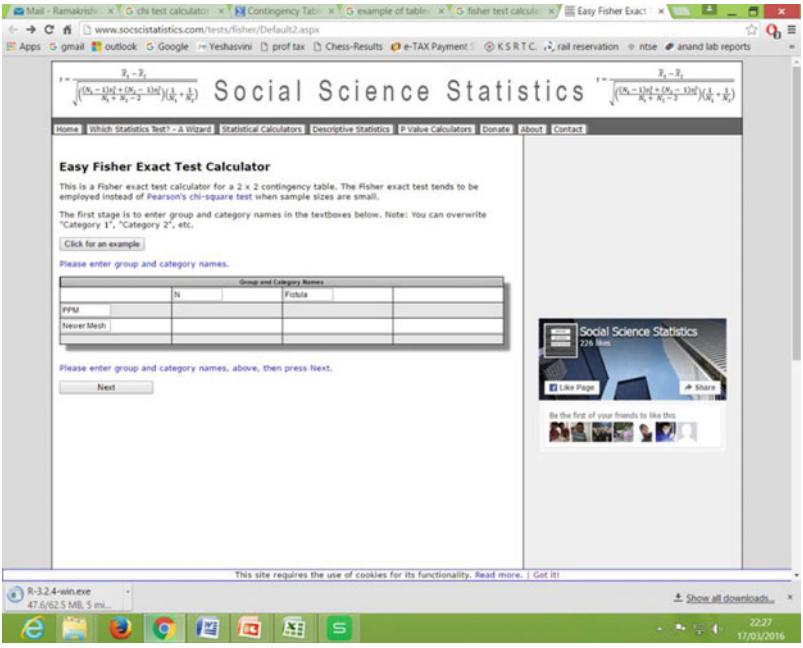


Fig. 5.5 Fisher's test using online calculator

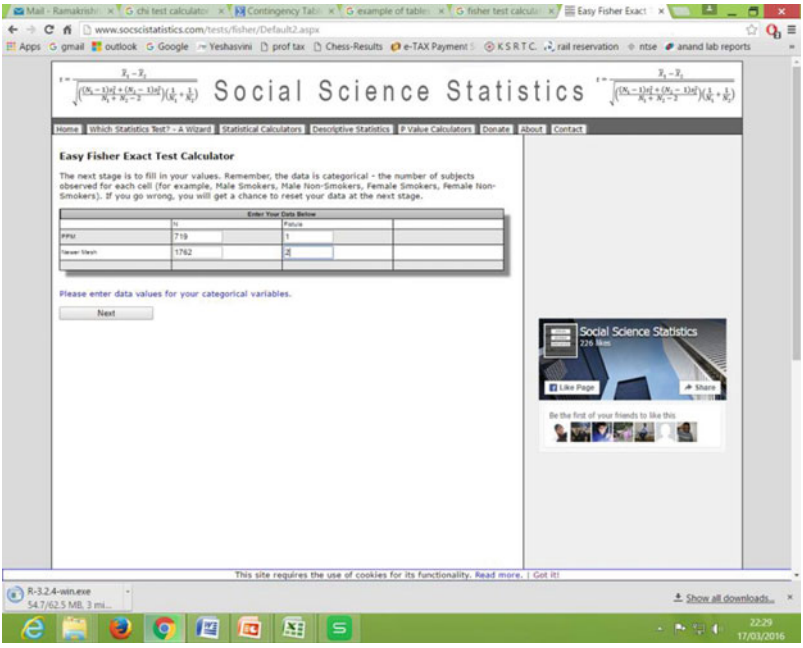


Fig. 5.6 Data entered on the calculator

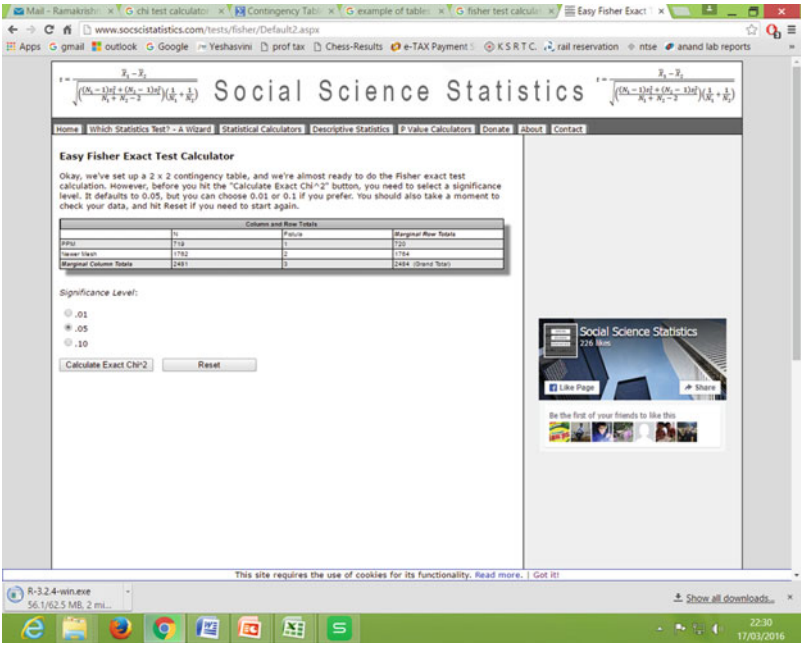


Fig. 5.7 Significant level selected as 0.05

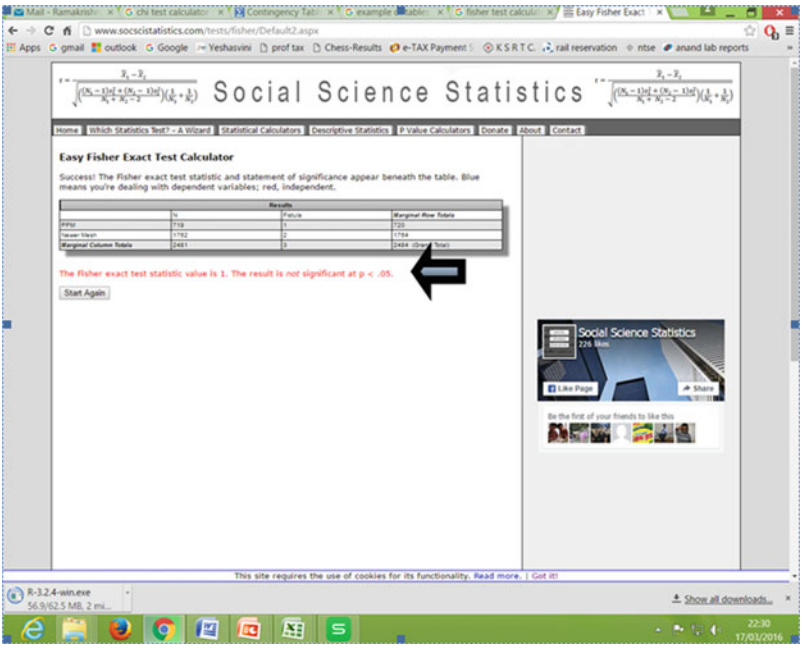


Fig. 5.8 Calculator gives the result ($P = 1$: not significant)

Enter the row and column heading Click ***NEXT***.
Enter the data and Click NEXT.
Select the significance level (here 0.05 is selected). Click calculate exact chi^2.
The *P* value is 1. The result is not significant at $P = 0.05$.

Student's *T* Test or Gosset's Test

Another important test in biostatistics is Gosset's Student's *T* test or simply *T* test. This is used to calculate the probability of *two normal distribution curves* being the same or different. We need two sets of data to compare. The distribution should be *normal*.

Procedure
$t = x1 – x2/A \times B$.
$A = n1 \times n2/n1 \times n2$.
$B = [(n1–1)\ S12 + (n2–1)\ S22]/n1 + n2–2$.
Compare *t* value with the table.

That's the formula to calculate *T* value. We need not bother about difficult calculations. We can do it easily using Excel program as I explained earlier, or still easier by using web-based online calculators.

Table 5.4 No. of days from the onset of pain to surgery and perforation rate in cases of acute appendicitis

	Perforated cases (days)	Non-perforated cases (days)
Case 1	2.5	3.0
Case 2	3.6	0.6
Case 3	2.5	3.5
Case 4	3.8	1.1
Case 5	3.2	1.2
Case 6	4.3	2.2
Case 7	3.0	1.1
Case 8	5.0	1.9
	Mean 3.49	Mean 1.82

T test is suitable for this type of data. The data are continuous variables (meaning, it can take infinite number of values: see above under chi test). So chi test or Fisher test cannot be used. We have two sets of data. This table shows the time since pain to surgery in perforated and non-perforated appendicitis. The same data can also be presented like this (mean ± standard deviation).

Mean +/- Standard deviation for the Table 5.4

Cases	Mean ± SD
Perforated cases	3.49 ± 0.06
Non-perforated cases	1.82 ± 0.05

The question is, *if there is a delay in the operation*, *is there an increased rate of perforation*? In other words, *is there a significant difference in the time since pain to surgery in non-perforated and perorated appendicitis*? Student's *T* test can find the *P* value to answer this question.

In case of control studies, we have two groups of patients. It is essential to prove that there is no statistically significant difference in the age distribution between the two groups. Otherwise one may argue that the results are due to different age pattern of the group. For example, if you are comparing mesh hernioplasty and nonmesh repair of hernia, you may be able to show superior results of mesh hernioplasty in terms of lower recurrence rate. But, if the mesh group contains majority of young patients and the nonmesh group is predominantly aged patients, then critics may say the superior result is due to the fact that study group has younger patients, because it is well known that older age is a risk factor for recurrence. Then the whole exercise of the study has gone waste. So, it is better to apply *T* test to prove that both groups are similar in terms of age.

How to do it? For simplicity, I've taken only ten patients in each group. We can have more number.

Age of the patients ($N = 10$)

Study group	Control group
62	65
55	68
72	70
40	65
42	55
45	46
35	42
39	38
29	37
28	35

We shall open a web calculator for *T* test. This is the web site of Social Science Statistics [http://www.socscistatistics.com/tests/studentttest/Default2.aspx] (Fig. 5.9).

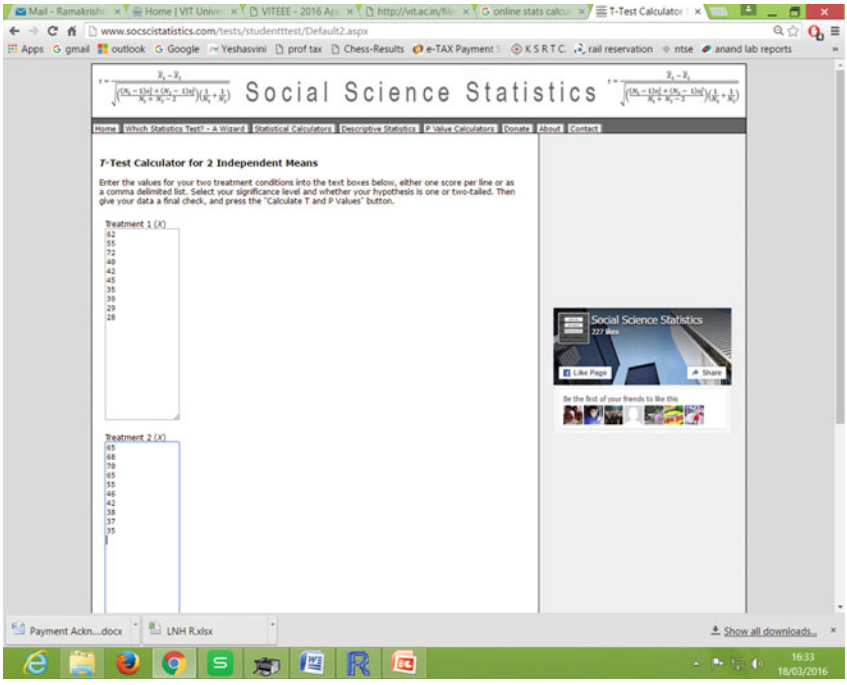
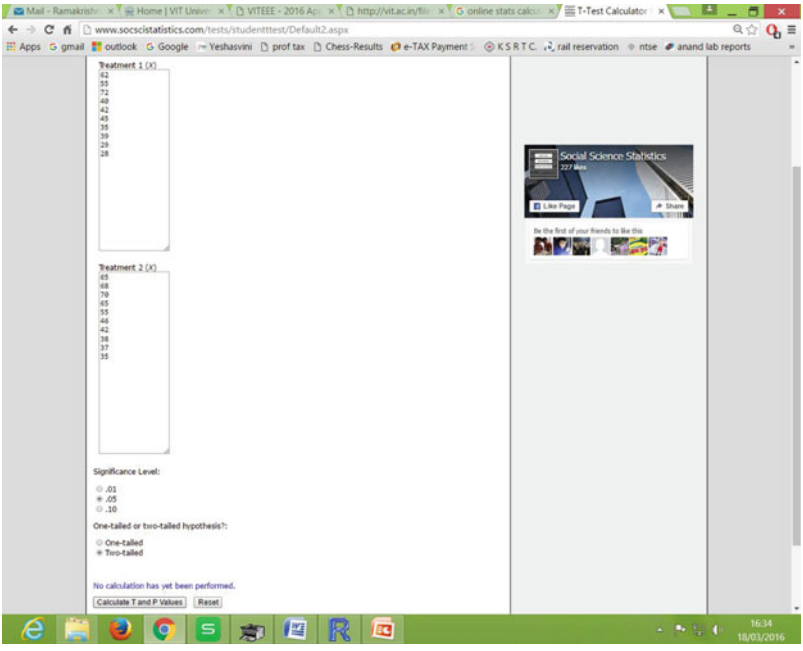


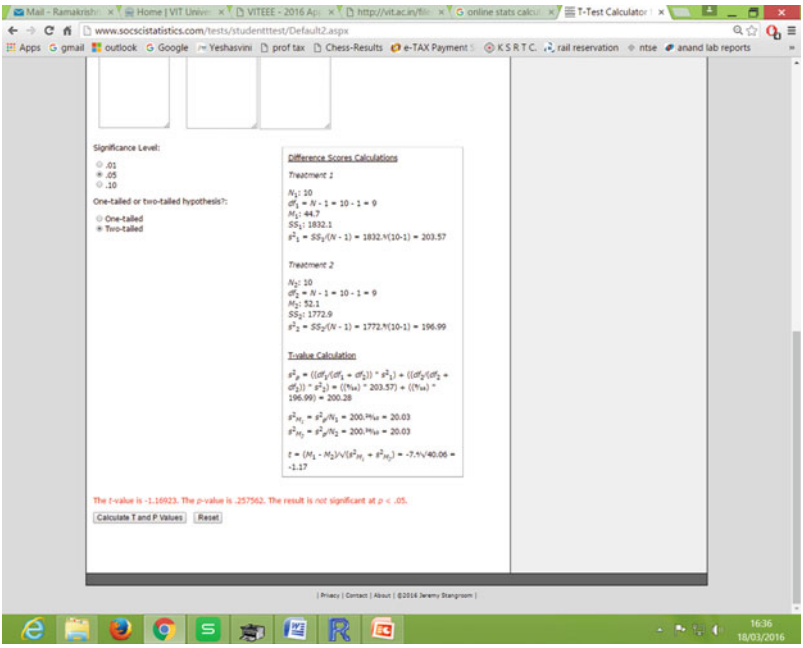
Fig. 5.9 Internet based calculator for *T* test

There are two boxes where we have to enter the data. I've entered the data of the study group in the first box and the data of the control group in the second box.



Data entered in the boxes provided. Significance level selected as 0.05. Two tailed test selected

We need to give two more conditions. I've chosen significance level as 0.05 and opted for two-tailed test [what are one- and two-tailed tests will be discussed subsequently]. Click on calculate.

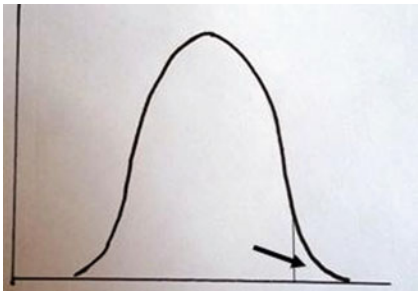
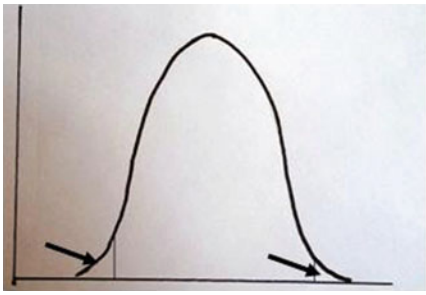


Result is shown in red fonts

We have the result: $P = 1.169$. So there is no significant difference in the age distribution between the groups. In other words, both groups are comparable with respect to age.

One- and Two-Tailed Tests

One-tailed test is used when data falling on only one side of the distribution (tail of the curve) is considered. If data falling on both sides to be considered, then two-tailed test is to be used.

One tailed test	Two tailed test
	

One-tailed test: data falling on one side of the curve (tail of the curve) is taken. The area represents 0.05 for $P=0.05$	Two-tailed test: data falling on both sides of the curve (tails of the curve) are taken. The area represents 0.025 on each side (a total of 0.05) for $P=0.05$

Example 1
In comparing stapled hemorrhoidectomy and conventional hemorrhoidectomy, with respect to patient's satisfaction and complications, two types of results are possible:

1. Stapled hemorrhoidectomy is better than conventional hemorrhoidectomy.
2. Conventional hemorrhoidectomy is better than stapled hemorrhoidectomy.

In the study, both these results are important to decide which procedure is superior. So use a two-tailed test.

Example 2
In comparing intraperitoneal mesh repair and nonmesh repair with respect to intra-abdominal adhesion formation, mesh group has "fewer adhesions" makes no meaning. It is obvious that repair with intraperitoneal mesh cannot produce *fewer adhesions* than repair without mesh. We are interested only in verifying if "mesh" group has the same incidence of adhesion as "nonmesh" group or it produces more incidences. "Fewer incidences of adhesions" is not an issue. So use a one-tailed test.

Paired *T* Tests

Paired *T* tests are used when there are two sets of observations for each subject. Other conditions of *T* test are the same. To apply paired *T* test, both groups should have equal number of observations. An example for this is preoperative and postoperative weights after weight-reducing surgery. Another example is to study the effectivity of lateral pancreaticojejunostomy in relieving pain in chronic pancreatitis. For each patient, the pain score before surgery and after surgery is recorded (Table 5.5).

Table 5.5 Pain score (Visual analog score)

Patient no.	Before surgery	After surgery
1	6	5
2	8	4
3	6	6
4	5	3
5	6	4
6	8	7
7	7	7
8	8	6
9	9	7
10	6	4

Now, apply paired T test. To calculate the paired T test, using internet based calculator, visit the following web site: http://www.physics.csbsju.edu/stats/Paired_t-test_NROW_form.html (Fig. 5.10).

In the box for the number of items, enter the number of observations, in this case, the number of patients (10).

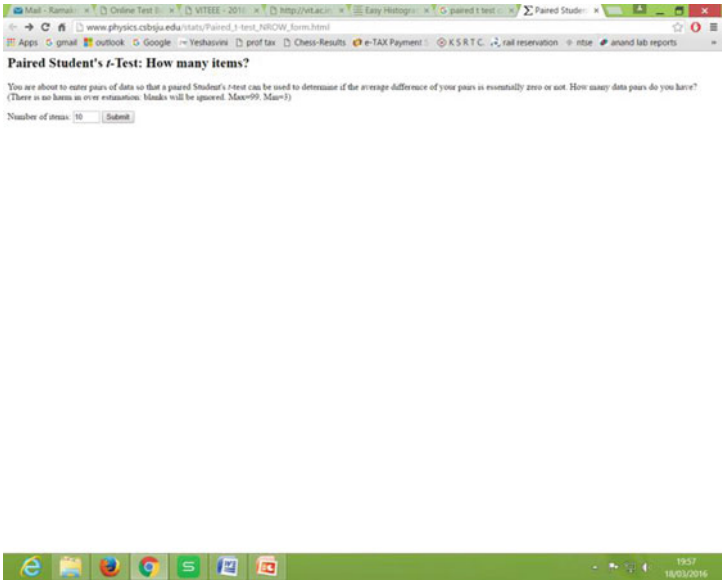


Fig. 5.10 Paired T test

Click ***SUBMIT*** button.

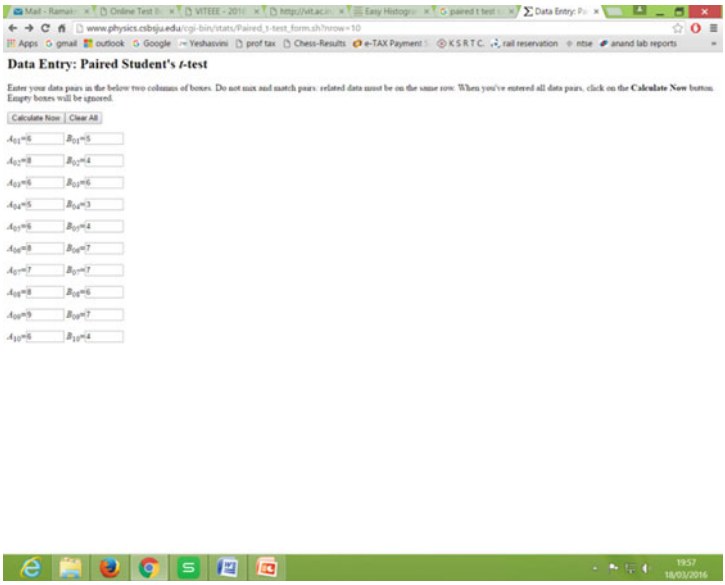


Fig. 5.11 Paired T test: data entered

In the boxes A01 to B01, enter first patient data pair. Similarly, the data of patient No. 2–10 are entered. Click on ***CALCULATE NOW*** (Fig. 5.11).

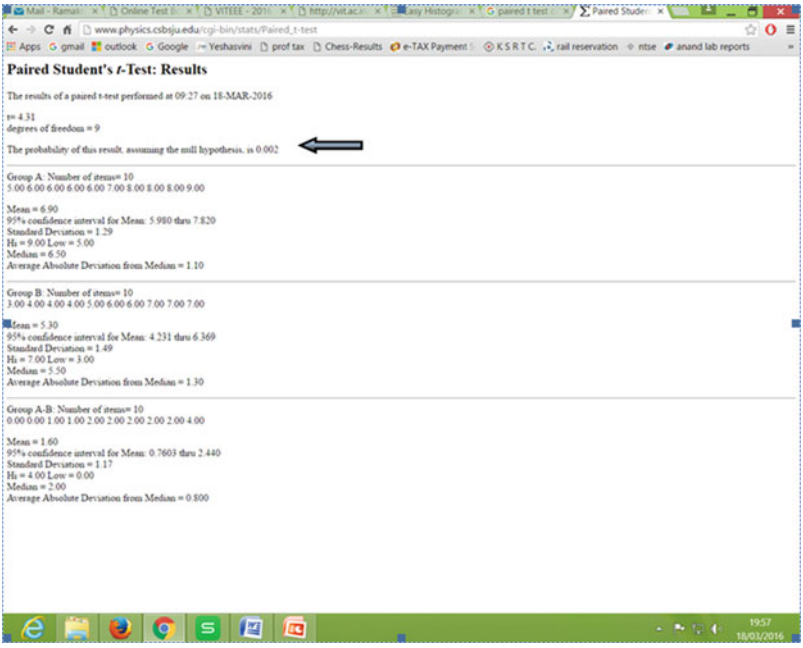


Fig. 5.12 Result (*Arrow*)

We have *P* value (arrow) 0.002. Hence, the difference in pain relief is significant. In other words, lateral pancreaticojejunostomy significantly reduces pain in chronic pancreatitis under the set conditions.

ANOVA

This is a difficult concept to discuss and full details are beyond the scope of this book. However, the basic details are discussed in the next chapter.

Which Test to Use?

A beginner often finds it difficult to decide which test is to be used for the data under consideration. Web sites have solutions to this also. You need to understand certain terminologies for using this *which test to use? wizard.* The explanation is also available on the same page (http://www.socscistatistics.com/tests/what_stats_test_wizard.aspx) (Fig. 5.13).

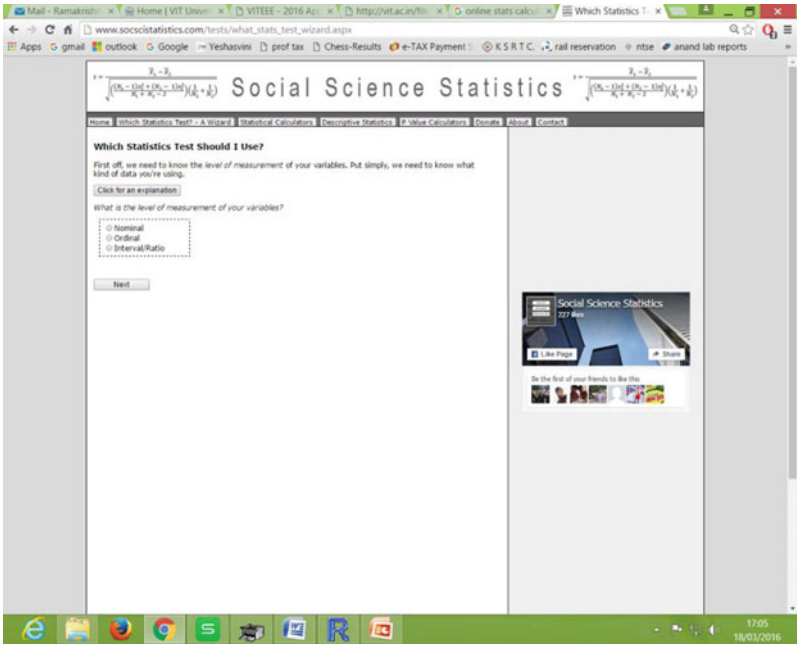


Fig. 5.13 Which test to use? On line help

Once we understand our data, if necessary by clicking on explanation, we have to select the proper option and click ***NEXT***. It takes to the next step, and after a few questions about the data, the wizard suggests the best test for our data.

For example, we shall see which test to be used for the data of Table 5.1 above. Here, for first query option is "Nominal", for second query option is "More than one nominal variable". Once we select the options, the wizard suggested chi test. It also suggested alternative tests, Fisher exact test, and Z test for two population proportions (Figs. 5.12–5.16).

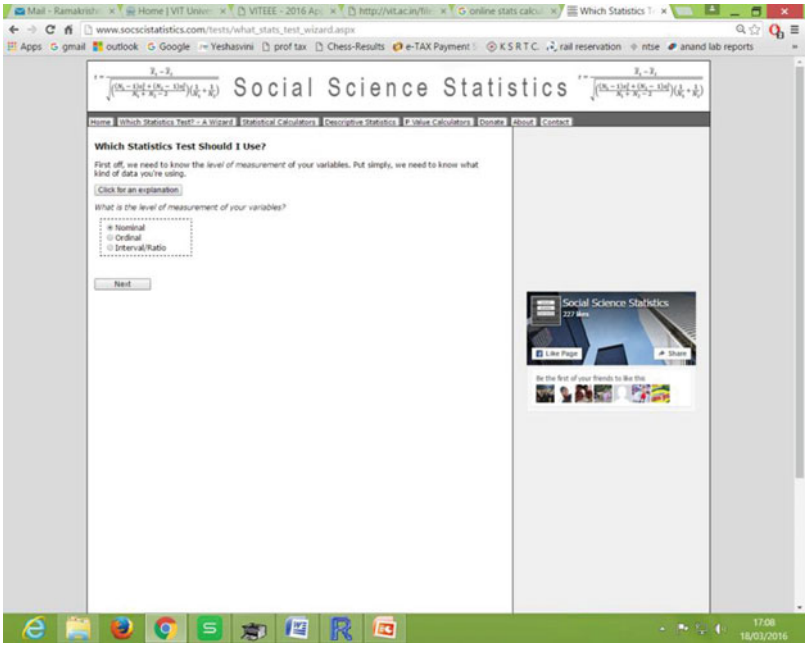


Fig. 5.14 Data type selected (Nominal)

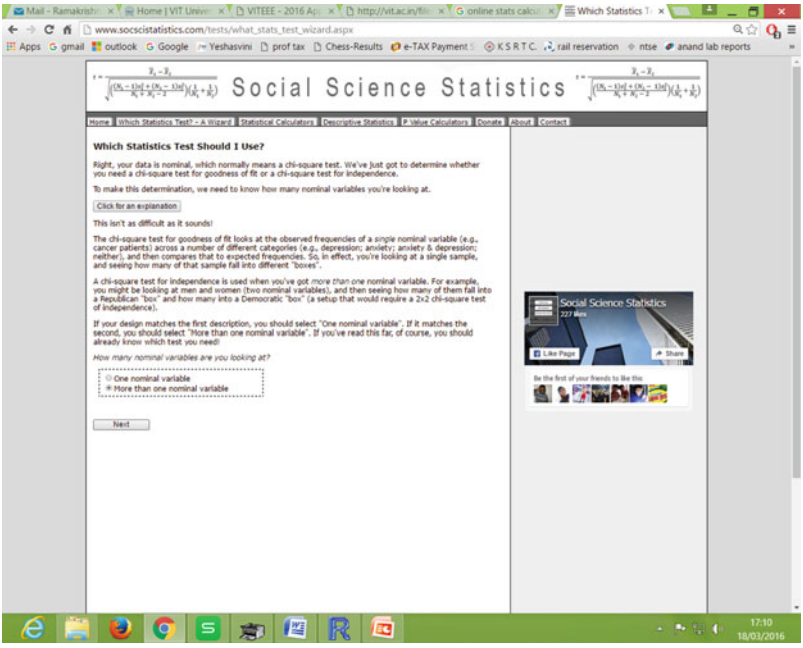


Fig. 5.15 Variables selected

Fig. 5.16 Web site suggested Fisher's Exact Test

Mann–Whitney–Wilcoxon (MWW) Test

One of the prerequisites for using *T* test is that the data should follow a *normal distribution curve*. If the data follow a *skewed distribution curve* (*nonparametric*), Mann–Whitney test can be used (for an explanation on *normal distribution curve* and *skewed distribution curve*, see Chap. 3). It can be used if certain conditions are satisfied:

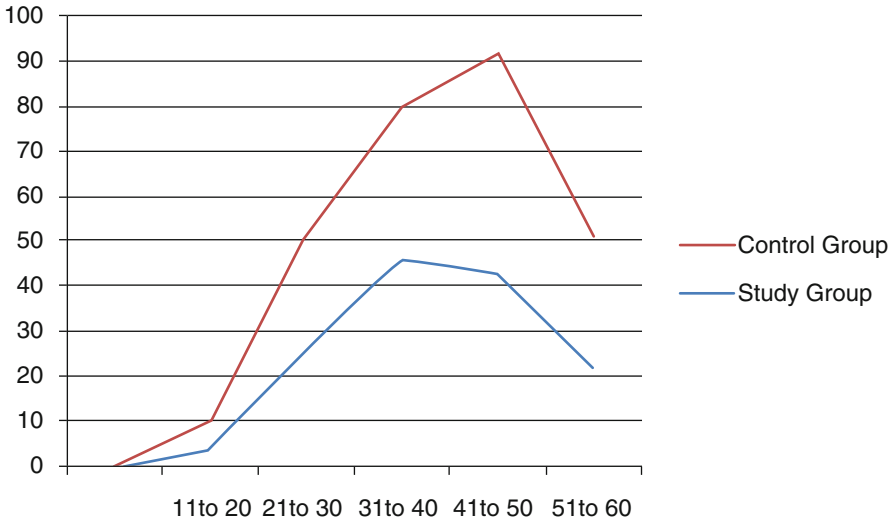
1. Two samples should be drawn from the same population.
2. Data are independent.
3. Data are ordinal for Rank Test (can be ranked higher or lower. Blood sugar levels 145 and 167 can be ranked: 145 is lesser than 167. The data are *ordinal*).
4. If the difference between the consecutive observations is not assumed to be equal, *T* test cannot be used. MWW test can be used. Under this condition the data follows *skewed distribution or nonparametric distribution curve*.

Example

The age distribution of the study group and control group of patients undergoing mesh hernioplasty and herniorrhaphy without mesh is shown in Table 5.6, and a graph plotted using this data is shown on Graph 1.

Table 5.6 Age distribution of the study group and control group of patients

Age in years	Study group	Control group
11–20	4	6
21–30	26	25
31–40	46	34
41–50	43	48
51–60	22	29
Total	141	142



Graph 1 Age distribution of the study group and control group of patients

It can be seen that the data are not distributed *normally*. The data are following a skewed distribution curve or, in other words, data are *nonparametric*. So we cannot use Student's *T* test, but *Mann–Whitney–Wilcoxon* (*MWW*) *test* can be used to compare these two groups.

Which test to be used for different data conditions

Test	Conditions
T test	Two sets of data Continuous variables Normal distribution (parametric) data Data can be expressed as mean ± SD
Paired *T* test	Two sets of data for each observation Other conditions the same as *T* test
Mann–Winey–Wilcoxon (MWW) test	Nonparametric data Other conditions the same as *T* test
ANOVA	More than two sets of data Other conditions the same as *T* test
Chi test	Two or more sets of data Categorical variables Parametric data Data are numbers and not percentages Minimum of five observations in each of the cells of the table
Fisher's test	No minimum number of observations Other conditions the same as chi test

Summary

Summary of Tests

Conclusions

Chi test, *T* test, and Fisher test are the most frequently used tests in medical statistics.

It is important for every clinician to understand what the tests of significance are and why these tests should be applied.

How to do the test is comparatively easy, once the concepts and which test to use under which conditions are understood. The Internet- and computer-based calculators are very useful in calculating *P* value.

There are many other tests. If we know the basics, we can always find a way to find solutions by further reading standard statistics books and with the help of web sites.

Examples and Self-Tests: All Fiction Examples

Question 1

Subarachnoid block using spinal needles to inject drug into the subarachnoid space is practiced in lower abdominal surgeries. Postoperatively many patients complain of headache (spinal headache). Is the incidence of headache related to the size of the needle used?

Setting: In a hospital 286 patients undergoing elective lower abdominal surgery were randomly assigned to two groups.

Intervention: In group A, 24 G spinal needle was used to give a subarachnoid block. In group B 26 G needle was used.

The results showed that out of 159 patients in group A 22 patients developed postspinal headache. In group B, out of 127 patients, seven patients developed postspinal headache.

Questions and Tasks

(a) Present the data in the form of a table.
(b) What type of variables are you dealing with?
(c) Which test you wish to apply?
(d) What alternate test can you think of?
(e) What is the *P* value?
(f) How do you interpret the results?

Question 2

In laparoscopic hernioplasty, fixation of mesh is a controversial issue. Many argue that fixation of mesh is unnecessary. Others advocate fixing the mesh. Fixation may be done with mechanical devices like tackers or with glue. A researcher designed a study to evaluate if one method is superior to the other. He randomly assigned patients to three groups. In the first group, mesh was not fixed. In the second group of patients, mesh was fixed with a mechanical device. In the third group of patients, glue was used to fix the mesh. Other details of the procedure, mesh size, etc. were similar in all the patients. Recurrence was assessed after 2 years of follow-up. The results were tabulated (Table 5.7).

Table 5.7 Incidence of recurrence of hernia after laparoscopic hernioplasty with various ways of mesh fixation

Group	Recurrence	No recurrence
No fixation	527	7
Mechanical fixation	632	6
Glue fixation	554	5

Which statistical test of significance can be used?
How you would do it?
What is the degree of freedom for this table?

Question 3

Two surgeons (surgeon A and surgeon B) are experts in thyroidectomy working in a hospital doing a large number of thyroidectomies. Their operative statistics show the recurrent laryngeal nerve injury rates as shown in the table (Table 5.8).

Table 5.8 Recurrent laryngeal nerve injury rates of two surgeons

	No. of surgeries done	Rec. N injuries	Percentage
Surgeon A	178	2	1.12
Surgeon B	293	6	1.71

It appears from the table that surgeon A is more competent than surgeon B in thyroidectomy as his complication rate is less than surgeon B. Is it so? How do you decide based on the data furnished?

Which test do you use and why?

Question 4

Local application of a drug (X gel) in the form of a gel is claimed by the company that it helps in faster healing of the ulcer. In order to test the claim, a dermatologist used the cream for a selected type of ulcer. He chose patients with healthy posttraumatic ulcers without infection with a size of 4–5 cm. He measured the area of the ulcers and assigned the patients randomly into study group and control group. For control group patients, he used conventional wet dressings. For study group patients, he used X gel. All patients did not have any factors delaying the healing like diabetes mellitus, peripheral vascular diseases, infection, chronic venous insufficiency, etc. After 2 weeks he measured the area of the ulcers and tabulated the results. The dermatologist assumes that the data is parametric (Tables 5.9 and 5.10).

Table 5.9 Ulcer size: control group

Control group $N=12$		All in sq cm	
	Initial area	Area after 2 weeks	Decrease in size
Patient 1	18.3	10.3	8.0
Patient 2	16.7	12.6	4.1
Patient 3	19.5	16.0	3.5
Patient 4	19.9	15.3	4.6
Patient 5	22.3	17.3	5.0
Patient 6	24.8	20.3	4.5
Patient 7	24.7	21.0	3.7
Patient 8	16.9	10.2	6.7
Patient 9	18.7	12.0	6.7
Patient 10	16.8	13.5	3.3
Patient 11	19.7	15.4	4.3
Patient 12	20.6	16.7	3.9

Table 5.10 Ulcer size: Study group

Study group $N=10$			
	Initial area	Area after 2 weeks	Decrease in size
Patient 1	19.6	14.2	5.4
Patient 2	16.9	10.6	6.3
Patient 3	19.4	12.9	6.5
Patient 4	22.5	18.1	4.4
Patient 5	23.6	17.3	6.3
Patient 6	23.8	18.6	5.2
Patient 7	15.3	10.3	5.0
Patient 8	19.4	14.2	5.2
Patient 9	15.6	11.0	4.6
Patient 10	18.2	12.4	5.8

What type of variables you are dealing with?
What type of study is this?
How will you ensure the groups of patients are comparable?
Which test would you do to see if the claim of X gel is valid?
If there were to be no control group, how do you test?
How do you interpret the results? Is it correct to apply paired *T* test?

Question 5
If the data were to be nonparametric, how will you proceed to test the claim?

Explanations: Question 1
The table for the data is given in Table 5.11.

Table 5.11 Incidence of spinal headache

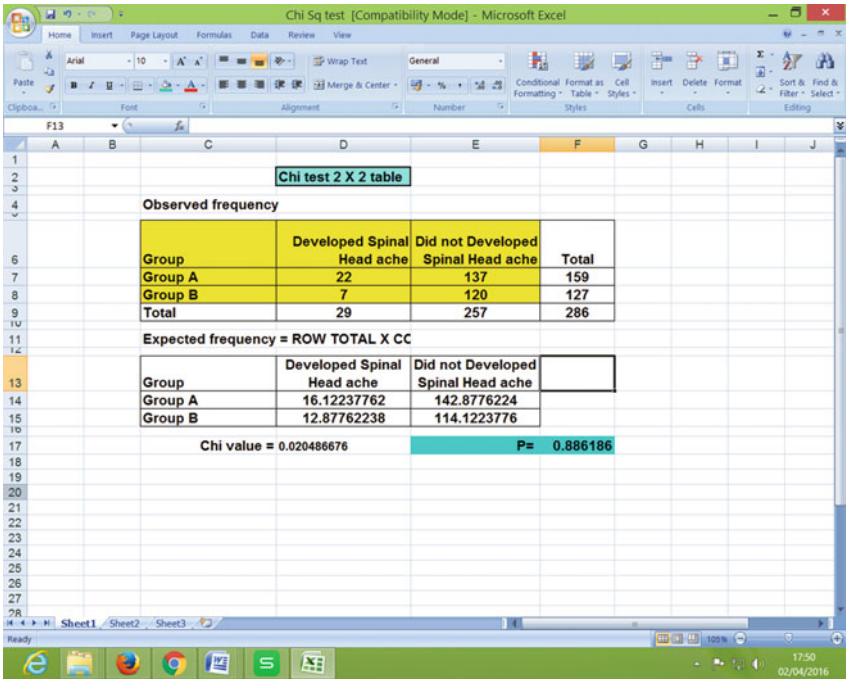
Group	Developed spinal headache	Total
Group A	22	159
Group B	7	127
Total	29	286

These are categorical variables. Outcome data can fall in only one of the two categories: developed headache or did not develop headache.

Chi test can be applied. Alternatively, Fisher's test can also be used.

The P value is 0.89.

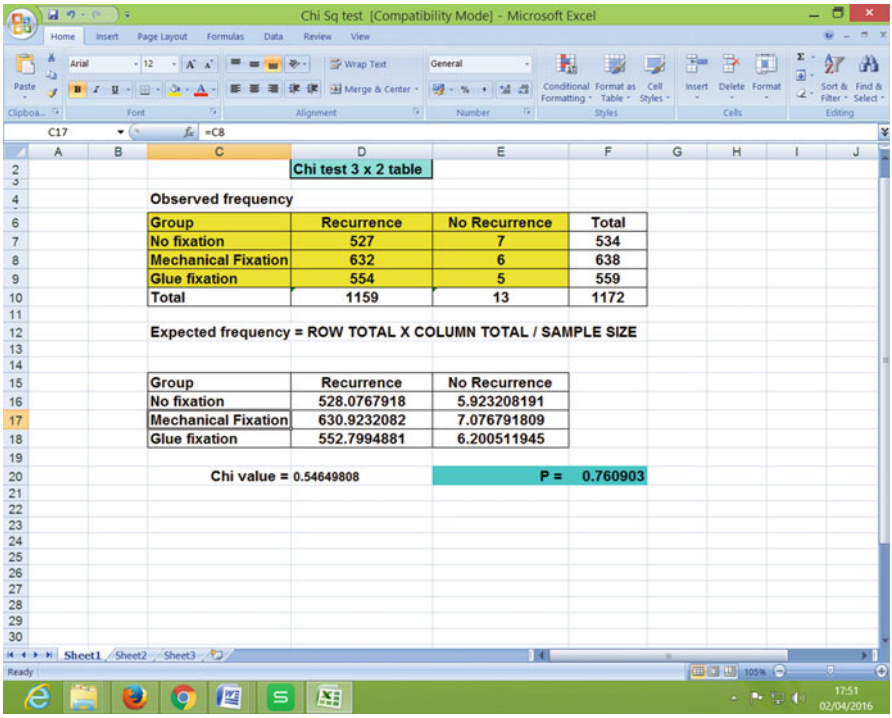
Interpretation: There is no significant difference in the incidence of postspinal headache whether 24 G or 26 G spinal needle was used.



Chi test calculation using MS Excel sheet

[redacted]

Explanations: Question 2
Chi test can be applied for this data also. It is a 3×2 table.
The degree of freedom is 2 {(3–1) * (2–1)=2*1=2}.



Chi test 3 x 2 table

Observed frequency

Group	Recurrence	No Recurrence	Total
No fixation	527	7	534
Mechanical Fixation	632	6	638
Glue fixation	554	5	559
Total	1159	13	1172

Expected frequency = ROW TOTAL X COLUMN TOTAL / SAMPLE SIZE

Group	Recurrence	No Recurrence
No fixation	528.0767918	5.923208191
Mechanical Fixation	630.9232082	7.076791809
Glue fixation	552.7994881	6.200511945

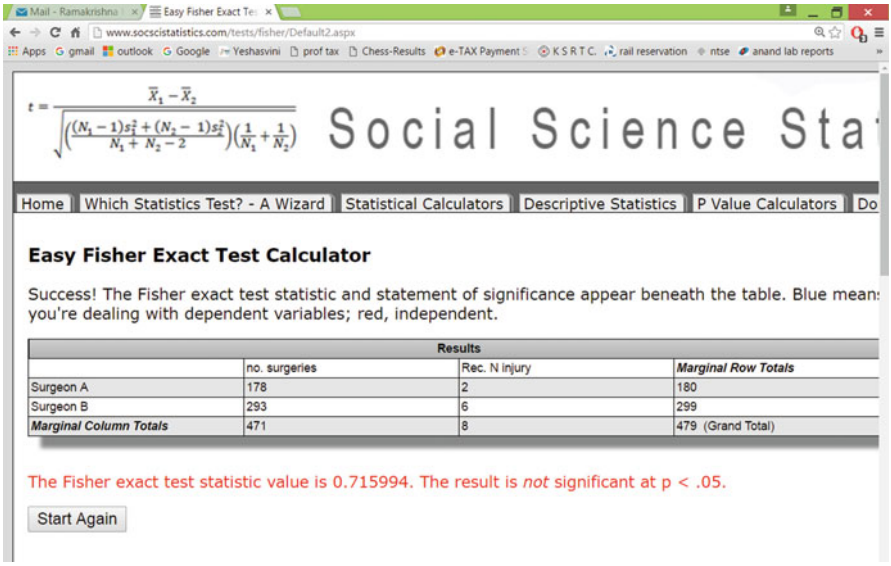
Chi value = 0.54649808 P = 0.760903

Chi test calculation using MS Excel sheet

Explanations: Question 3

Since the data value is small (2), it is better to use Fisher's test.

As P value is 0.71, which is greater than 0.05, the results are not significant. In other words, there is no significant difference in the competence of the two surgeons.



Internet based calculator for Fisher test

Explanations: Question 4

The data shows continuous variable, since the area can take any number of values.

It is a randomized control trial.

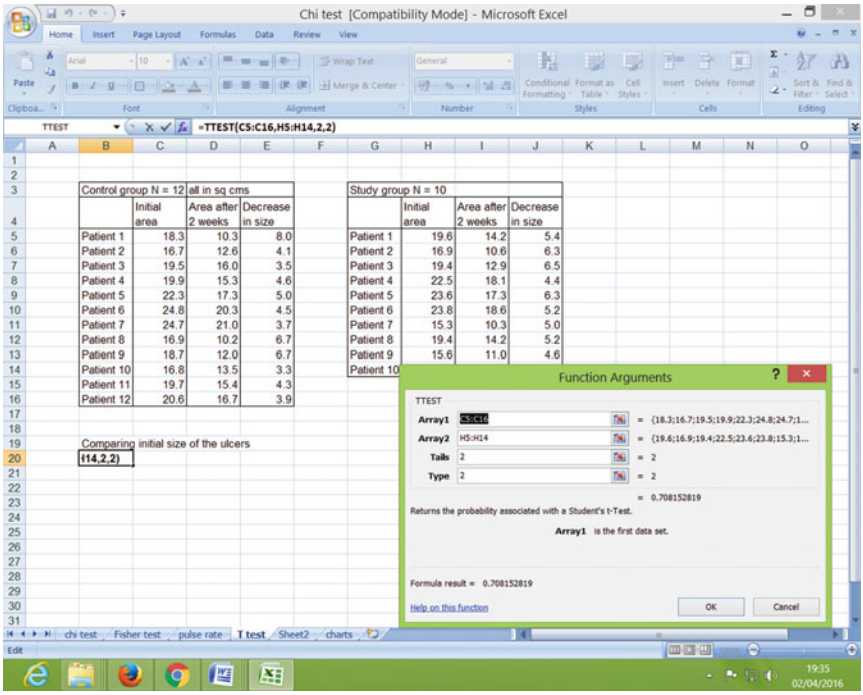
To ensure that the two groups are comparable, we have to compare the initial area of the ulcers of the two groups and find the P value.

Initial size of the ulcer in control group

Control group	Initial area
Patient 1	18.3
Patient 2	16.7
Patient 3	19.5
Patient 4	19.9
Patient 5	22.3
Patient 6	24.8
Patient 7	24.7
Patient 8	16.9
Patient 9	18.7
Patient 10	16.8
Patient 11	19.7
Patient 12	20.6

Initial size of the ulcer in study group

Study group	Initial area
Patient 1	19.6
Patient 2	16.9
Patient 3	19.4
Patient 4	22.5
Patient 5	23.6
Patient 6	23.8
Patient 7	15.3
Patient 8	19.4
Patient 9	15.6
Patient 10	18.2



T test calculation using MS Excel sheet

By applying T test for this data, we get $P = 0.708$.

To see the efficacy claim of X gel, we have to compare the "decrease in size" column of the two groups.

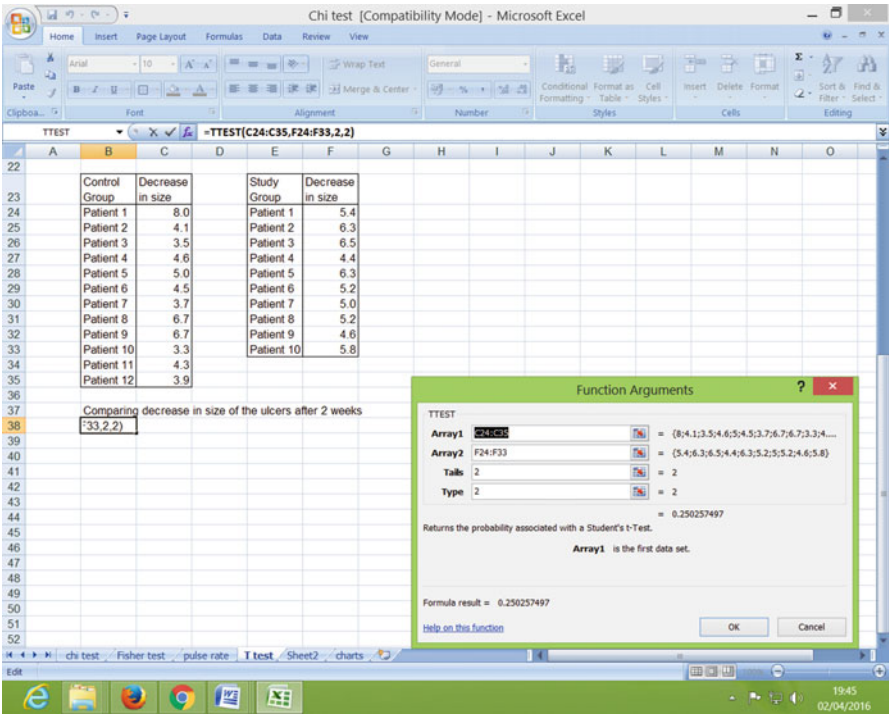
Decrease in the size of the ulcer in the control group

Control group	Decrease in size
Patient 1	8.0
Patient 2	4.1
Patient 3	3.5
Patient 4	4.6
Patient 5	5.0
Patient 6	4.5
Patient 7	3.7
Patient 8	6.7
Patient 9	6.7
Patient 10	3.3
Patient 11	4.3
Patient 12	3.9

Decrease in the size of the ulcer in the study group

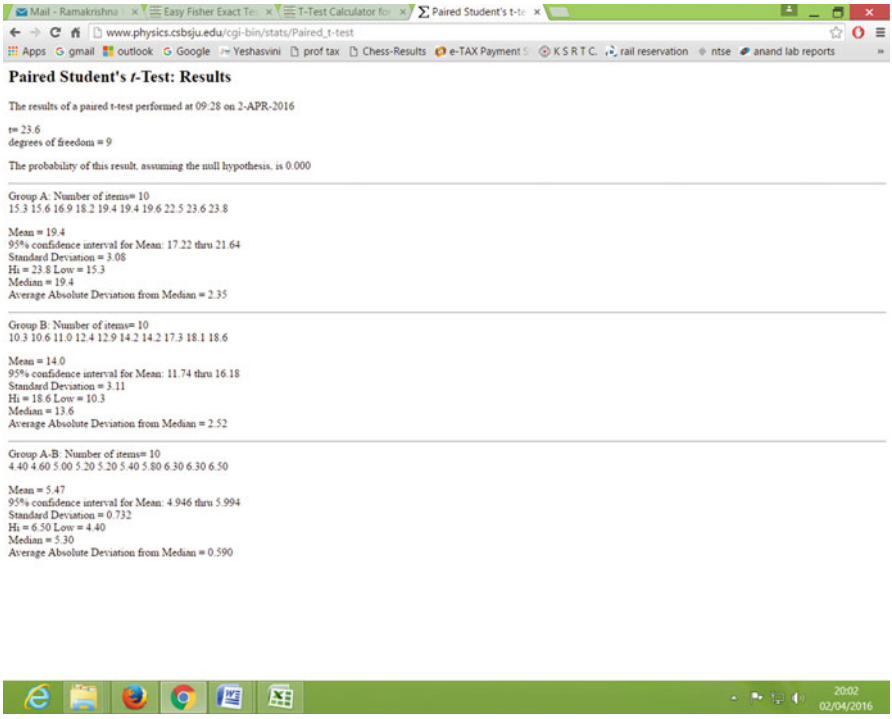
Study group	Decrease in size
Patient 1	5.4
Patient 2	6.3
Patient 3	6.5
Patient 4	4.4
Patient 5	6.3
Patient 6	5.2
Patient 7	5.0
Patient 8	5.2
Patient 9	4.6
Patient 10	5.8

By applying T test for the data of these two groups, we get $P = 0.25$: hence the difference in the results between the groups is statistically not significant. The dermatologist concludes by saying X gel does not hasten the healing of the ulcers under the said conditions.



T test calculation using MS Excel sheet: data entry

If there were to be no control group, it is tempting to use paired *T* test for the data of the study group alone. As explained in the main text, this data satisfies the conditions for a paired *T* test. There are two sets of data for each patient. The data is assumed to be parametric.



Paired *T* test using internet based calculator

P is 0: it is highly significant. The probability of a null hypothesis being true is 0.

But *T* test showed result is not significant. How to explain these seemingly contradictory results?

We have to understand that ulcers can heal by natural process. If paired *T* test is applied for the data of control group similarly, again, we get $P=0$. This is because of the natural healing process of the ulcers.

Now, it is important to see if the rate of healing by using X gel is faster than the natural process for the claim to be valid. That is the importance of control study. If the pharmaceutical company shows the results of the study group alone and claims the drug is useful, we will be misled by the results unless we know how to interpret the results.

Observe the difference between the data presented in the main text to explain paired *T* test and the present data. In the text data, there was no natural process to pain relief. So whatever the results we got should be due to the drug. Hence, the result is valid. In the present data, we need a control as there is a natural process by which similar results can be expected.

Explanations: Question 5

If the data were to be nonparametric, *Mann–Whitney–Wilcoxon* (*MWW*) *test* can be used to compare these two groups.

Other Commonly Used Concepts

6

Learning Objectives
To understand:
ANOVA
Rank test
Various risk ratios and odds
Correlation
Various types of regressions
Examples and exercises to understand the above concepts

ANalysis **O**f **VA**riance (ANOVA)

ANOVA tests the difference between means of two or more groups. In other words, it tests whether the means of multiple groups are equal or not. Although it tests the difference in the means, it is called analysis of variance because it does the test by looking at the variances of the data.

If there are only two groups, *T* test can be used. If there are multiple groups, *T* test can be applied to each pair of groups individually. Other conditions of *T* test are the same, like the data should follow:

(a) *Normal distribution curve*
(b) Independent data
(c) Homogenous variance

But when multiple *T* tests are applied, the chances of type 1 error (false positive) are increased. The alternative in these cases is to use ANOVA.

ANOVA is a combination of many concepts and is used in several settings. So it is difficult to define and explain the concept of ANOVA. For detailed description

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_6

and application, the readers are advised to refer to standard statistical textbooks or consult medical statistics experts. Some basic concepts are explained here.

Table 6.1 Increase in BP in after intubation in control and two study groups

	A	B	C
A	x	AB	AC
B	Same as AB (BA and AB are the same combination)	x	BC
C	Same as AC	Same as BC	x

In the example in Table 6.1, there are three variables: so three combinations are possible (e.g., if A, B, and C are the variables, AB, AC, and BC are the possible combinations). If there are four variables, six combinations are possible (e.g., if A, B, C, and D are the variables, AB, AC, AD, BC, BD, and CD are the possible combinations). If *T* test is to be applied, it has to be applied individually three times (or, in case of four variables, six times). When multiple tests are used, type 1 errors (false positive) are magnified, and resulting conclusions may be wrong. ANOVA generalizes *T* test to more than two groups.

Actual calculation is complicated and beyond imagination of the beginner. Suffice it to say, a *P* value less than 0.05 suggests ***that at least one group is significantly different from the rest***.

ANOVA test can be done on R module.

> *T* test is used when comparing two groups. ANOVA is used when three or more groups are to be compared. In fact, if ANOVA is applied to two groups, it yields the same result as *T* test.

Example 1

There will be an increase in systolic blood pressure during endotracheal intubation. This increase can have deleterious effects on the cardiovascular system. Two drugs, Drug A and Drug B, are used to prevent this increase in BP. An investigator wanted to test the beneficial effects of these drugs. Patients undergoing endotracheal intubations were randomly divided into three groups, namely, control group, Drug A group, Drug B group. BP was recorded at 5 min of endotracheal intubation. The increase in blood pressure (the difference between BP prior to endotracheal intubation and at 5 min of endotracheal intubation) was tabulated.

Increase in BP with Drugs A and B

Group	Increase in BP
Control ($N=98$)	40 ± 12
Drug A ($N=122$)	29 ± 11
Drug B ($N=101$)	25 ± 14

For this type of data, ANOVA is appropriate.

Did You Know This?
The term variance was introduced by an evolutionary biologist, Ronald Fisher, who developed statistical models for ANOVA

Rank Test: Wilcoxon Signed-Rank Test

In very simple words, rank test is used to make a skewed distribution (nonparametric data) curve to a normal curve. It is a paired test and is used as an alternative to paired *T* test when data are nonparametric. Worked examples can be read on http://users.sussex.ac.uk/~grahamh/RM1web/WilcoxonExample2008.pdf.

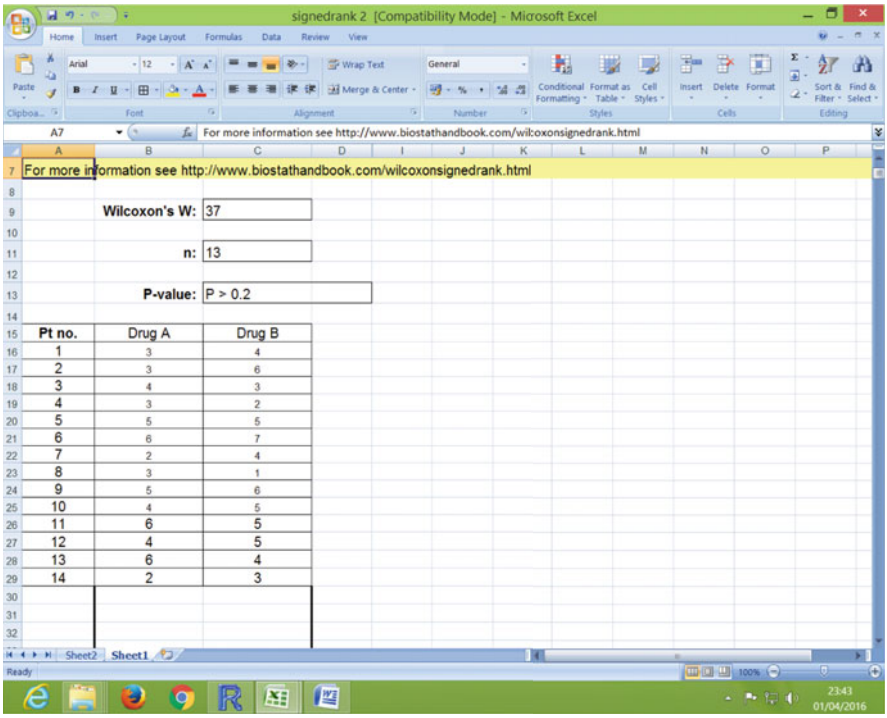
Example 2
In patients with chronic pain, two drugs are used to decrease the intensity of pain. Each drug is given on different day when no other analgesics are used. Drugs are tested on the same 14 patients. For each patient, the first visual analog score is recorded and the drug is given. Visual analog score is recorded again after 2 h. The difference in the score is recorded as "decrease in VAS." So for each patient, we have two data (paired): one for Drug A and another for Drug B. The results are given in the table. There were reasons to believe that data are nonparametric. Hence, Wilcoxon signed-rank test is to be used in this type of data.

Rank test

	Decrease in VAS for Drug A	Decrease in VAS for Drug B
1	3	4
2	3	6
3	4	3
4	3	2
5	5	5
6	6	7
7	2	4
8	3	1
9	5	6
10	4	5
11	6	5
12	4	5
13	6	4
14	2	3

For calculation, spreadsheet-based calculators have been designed (for more information, see http://www.biostathandbook.com/wilcoxonsignedrank.html) (Ref. McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed). Sparky House Publishing, Baltimore, Maryland).

The data is entered in the appropriate cells as paired values. It shows *P* value as >0.2. Hence, result is not significant. There is no significant difference in the efficacy of these two drugs.



Calculation of Wilcoxon-Rank Test on MS Excel sheet

Risk and Risk Ratio

Risk is the probability that an event will occur. For example, the risk of developing a recurrence after a hernioplasty is 2 % means if 100 patients undergo hernioplasty operation, two of them will develop a recurrence eventually.

Risk ratio is calculated when two groups are compared. For example, the risk of recurrence after mesh hernioplasty is 2 % and that after herniorrhaphy without mesh is 10 %. The risk ratio is 2/10 = 5. This means the patients who undergo herniorrhaphy without mesh are at fivefolds higher risk than patients who undergo hernioplasty with mesh. (Please note that the risk of recurrence is still 10 % for the group: But five times when compared with hernioplasty group.) If the risk ratio is 1, that means both the groups have the same risk.

Odds

It is the ratio of an event happening to not happening.

For example, in the above example of 2 % recurrence after hernioplasty, the odds ratio is 2/98 = 0.0204. (Two recurrences mean 98 no recurrences.)

Odds Ratio

It is the ratio of odds of the study group to odds of the control group. For example, a new drug (Drug X) is being studied for adverse drug reactions. The data revealed that the study group where the Drug X was used had mortality of 6 %, and the control group had a mortality of 1 %. Then **odds ratio** is odds of mortality of the study group divided by the mortality of odds of the control group.

Odds ratio = odds of the study group/odds of the control group
=6/94 divided by 1/99 = 6.316.

Relative Risk Reduction (RRR), Absolute Risk Reduction (ARR), and Number Needed to Treat (NNT)

To understand these concepts, consider this example. This example is also useful to highlight why we should know how to interpret the data and statistical terms. Sometimes the pharmaceutical companies use the term ***relative risk reduction***. If we do not know the proper interpretation, we will be misled to overrate the efficacy of the drug and write their drugs. Suppose there is a condition which has a mortality of 3 in 10,000, and a particular drug is shown to reduce this mortality to 2 per 10,000. Then the relative risk reduction is 33 %. (Reduced mortality is divided by the mortality: here reduction is 1 and mortality is 3. Hence RRR = 1/3 or 33 %.) Pharmaceutical company may hide the other details and show only this line in bold highlighted letters "the drug reduces relative risk by 33 %." This 33 % looks very impressive, and if we do not know how to interpret the result, we may recommend this drug to our patients. If we analyze the actual data and not the conclusion, we will see that **absolute risk reduction** is only 0.01 %, because it has reduced the mortality rate 1 in 10,000. (ARR is the reduced mortality divided by the total number of patients: 1/10,000 = 0.01 %.) That is to say, we have to give this drug to 10,000 patients to reduce the mortality by 1. In other words, the **number needed to treat** is 10,000: NTT = 100/ARR.

Correlation: Relation Between Two Factors

If two parameters have a linear relationship, there is ***correlation*** between them. For example, height and weight in children correlates with each other, which means as one parameter changes, the other also changes. The relationship may be positive or negative.

Positive Correlation

In positive correlation, if the value of one parameter increases, the value of the other parameter also increases, and vice versa. For example, a delay in operating on cases of acute appendicitis results in higher perforation rate. In other words, if the time since pain to surgery increases, perforation rate also increases (Fig. 6.1). Lower body mass index (BMI) associated with lower death from cardiac arrest is another example for positive correlation. But it does not mean the second variable always exists whenever the first variable is present. For example, lower body mass index (BMI) associated with lower death from cardiac arrest does not mean that cardiac arrest *will not* occur in patients with lower BMI.

> The correlation indicates only a relationship: but it does not indicate cause–effect relationship. The second parameter may or may not be the cause for the first or vice versa.

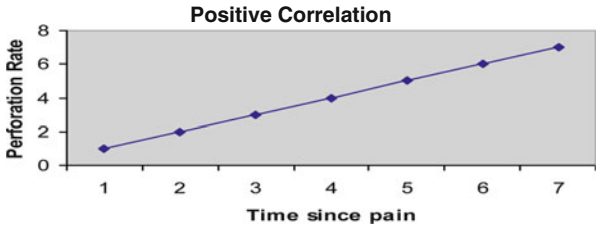


Fig. 6.1 Relation between time since pain and perforation rate in acute appendicitis

Negative Correlation

The correlation may be negative also. If negative laparotomy rate and perforation rate data are collected from different series of study on acute appendicitis and plotted as graph, the following type of relationship may be found. As the negative laparotomy rate increases in a series, the perforation rate in that series decreases. So perforation rate and negative laparotomy are said to be having a **negative correlation** (Fig. 6.2). To give another example, higher socioeconomic status populations have lesser deaths due to infection.

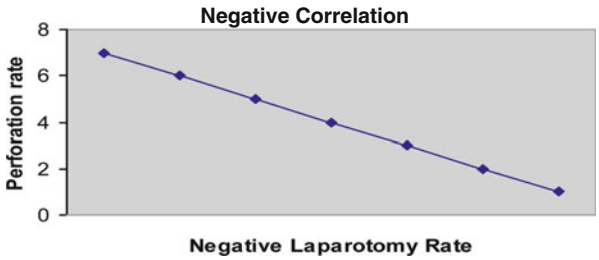


Fig. 6.2 Negative laparotomy and perforation in acute appendicitis

Think over it:
Consider smoking and the incidence of lung cancer as two variables. Do they correlate? If so what type?

Correlation Coefficient (*r*)

Correlation coefficient measures the strength of relationship, that is, how strong or weak the relationship is. If there is a perfect relationship, the coefficient will be 1 (+1 if positive correlation and −1 if negative correlation). If there is no relation at all, then the coefficient will be 0. Depending upon the strength of relationship, the coefficient varies between −1 to +1. If the correlation coefficient value is nearer to 0, the relationship is weak. If the correlation coefficient value is away from 0, that is, nearer to the extremes, the stronger will be the relation. If the correlation coefficient value is away from 0, that is, nearer to the extremes, the relationship will be strong.

Pearson's correlation coefficient is used for parametric data (following normal distribution curve), and ***Spearman correlation coefficient*** is used for non-parametric data.

The ***significance of correlation*** also depends upon the sample size. If the sample size is large, even a lesser degree of correlation is also significant, and for a small sample size, even a higher degree of correlation may or may not be significant.

There may be **nonlinear correlation** where correlation coefficient is small (indicating a weak relationship) but association may be strong. They are not revealed because association is not linear.

Correlation is not an all-or-none phenomenon. There may be multiple factors correlating with a variable.

Regression

The idea is similar to and sometimes confused with correlation. It is important to clarify the difference between correlation and regression. Correlation only indicates the strength of the relationship between two factors or parameters. Regression quantifies the relationship. Regression is used only when there is cause–effect relationship. It can quantify the relation: that is to say, once regression is applied, one parameter can be calculated if the other parameter is known.

Linear Regression

Linear regression is used to analyze continuous relationships. When two dependent data are plotted as graph, ***linear regression*** calculates the best line through the data. If one of the parameters is known, the other can be calculated using this line.

For example, the data of height and weight of infants are recorded (Table 6.2).

Table 6.2 Table showing weight (kg) vs height (cm) in infants

Weight in kgs	Height in cm
2.5	46
3.4	51
4.4	54
5.1	57
5.6	60
6.1	61
6.4	62
6.7	63
7	65
7.2	66
7.4	67
7.5	69

This data is plotted as a scatter graph with weight against height.

Weight is marked on *x*-axis (horizontal) of the graph and height is marked in *y*-axis (vertical). Each dot represents a pair of data. When a number of data are entered, a scatter graph is obtained (Fig. 6.3).

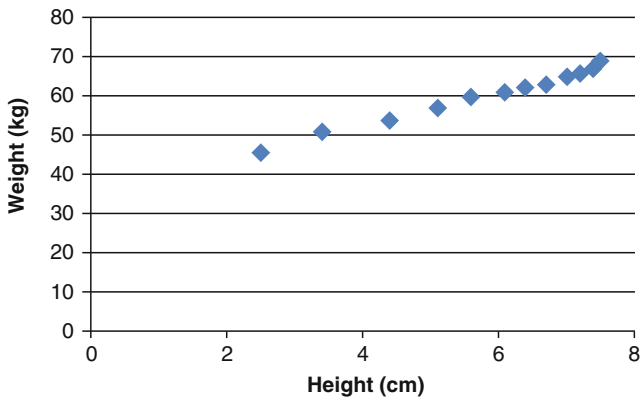


Fig. 6.3 Graph plotted using the data from Table 6.2 showing weight (kg) vs height (cm) in infants

If a straight line which can best fit all the data is drawn, it will be the regression. The slope represents ***regression coefficient*** (Fig. 6.4).

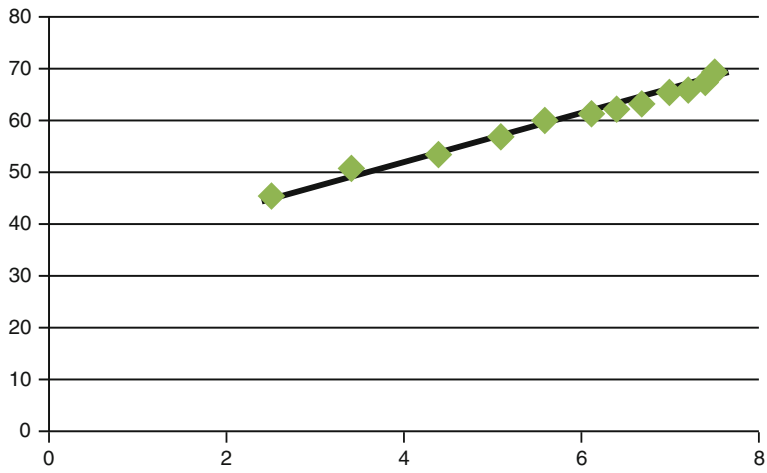


Fig. 6.4 Best fitting straight line drawn: The slope is the regression co-efficient. Weight (kg) vs height (cm)

If one of the parameter is known, the other can be calculated using this line. For example, for 4 kg weight, the height would be 51 cm (Fig. 6.5).

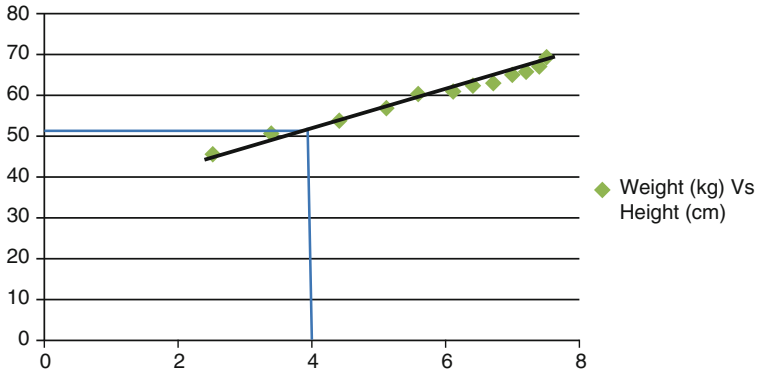


Fig. 6.5 Calculation of weight when height is known or vice versa. Weight (kg) vs height (cm)

Similarly if height is known, weight can be calculated by drawing lines horizontally and vertically.

We know that for a linear graph, $X = a + bY$. So, with appropriate calculations, it is possible to calculate the height for weight (or vice versa) mathematically without referring to graph.

> It is important to restrict calculations within the data range. Calculations should not be extended beyond the range. In the above example, we must not try to calculate height for 2 kg and less or 8 kg and more.

Other Types of Regressions

Logistic Regression

It is a statistical method of analyzing variables (one or more) to predict whether an outcome falls into a category or not. In other words, it predicts ***dichotomous outcome*** like survival or death, male or female, recurrence or no recurrence, wound gets infected or no infection, etc. Dependent variable is categorical. For example, let us assume there is a condition where mortality is related to age. Observational data is presented in Table 6.3.

Table 6.3 Age as a predictor of survival (in a particular condition): Logistic regression applicable

Age group (in years)	Average no. of survivors in a group of 100 patients	Probability of surviving
11–20	21	21
21–30	26	26
31–40	8	8
41–50	2	2
51–60	2	2
61+	1	1

This data may be plotted as scatter graph as in the above example. From the graph, the probability of survival or death can be predicted for a patient when the age is known.

Multiple Regression: To Predict an Outcome

Multiple regression is similar to linear regression. Here the outcome (dependent variable or target variable) is predicted depending upon two more input variables. Interpretation of the results of multiple regression is complex and difficult as multiple variables are involved and different variables may have different degrees of influence

on the outcome, for example, predicting the 5-year survival of a cancer patient depending upon the *TNM* status of the patient. Here, *T* status, *N* status, and *M* status are the three independent variables. Predicting the 5-year survival rate is the outcome.

The Poisson Regression (Distribution)

It's a discrete probability distribution and predicts the probability of a given number of events occurring in a fixed time interval.

Conditions:

1. The average rate is known.
2. Events are independent of the time since the last event.

Example 3

Let us assume that cleft lip incidence is 10 per year in a particular city. That means ten new cases of cleft lip are detected in a year. (Average rate is known =10/12 per month.)

If a new case of cleft lip is detected today, it does not have any relation to the time interval to detection of the next case of cleft lip. The next case may be detected on the same day, after 3 days, after 1 month, etc. So, when one case is detected today, it does not give any idea as to when the next case would be detected (events are independent). So both the conditions mentioned above are satisfied: Poisson regression is applicable.

With these data, we can predict the probability of the number of cleft lip cases detected in the month of, say, May.

Cox Regression

Cox regression is used for survival analysis. It calculates the time to certain event. For example, time to death or time to recurrence. Cox regression aims to estimate the hazard ratio. **Hazard ratio (HR)** is the ratio something (outcome) happening in one group to that of another group. For example, HR of death from lung cancer is 2 for smokers means the chances of an individual dying from lung cancer is twice if he is a smoker compared to nonsmoker. Based on HR, life expectancy can be calculated.

Correlation indicates only the relationship between two variables. Regression indicates cause–effect relationship. Correlation gives strength of the relationship. Regression quantifies the relationship. With regression, if one variable's value is known, the value of the other variable can be calculated.

Survival Analysis

Life table and plot of survival is a graph showing survival as the percentage of a population over time. Similar plot can be constructed for other events also. For example, the incidence of recurrence after hernioplasty at different time period can be plotted as a graph. Here recurrence is taken as the event instead of death. It is useful when different patients are followed up for different periods of time, and the event has not occurred in all the patients. For example, consider the patients with breast cancer being followed up for mortality, and the data is presented after 7 years. We do not have data of all patients regarding mortality, as many patients are still surviving. So it appears the data is incomplete and cannot be presented. However the data can be presented as survival analysis (Fig. 6.6).

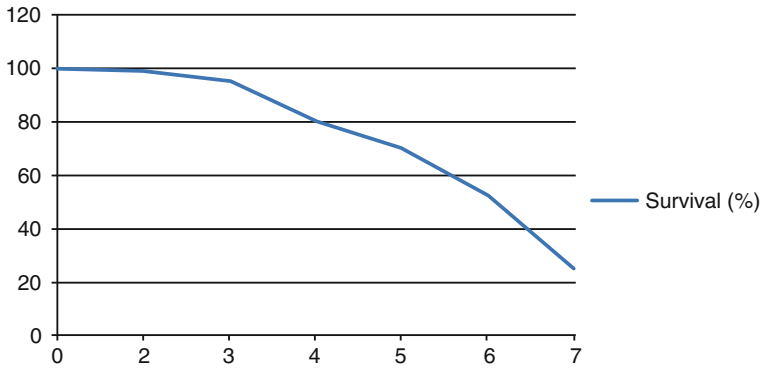
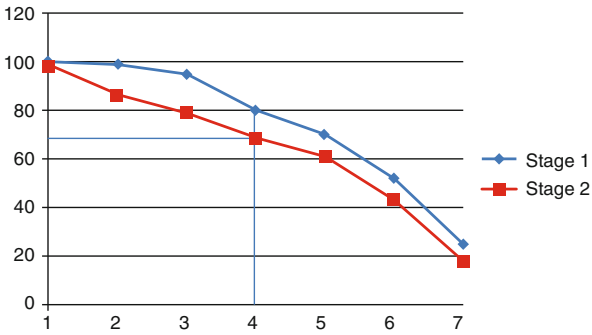


Fig. 6.6 Survival analysis graph. Survival (%)

When data for two categories (stage 1 and stage 2) are plotted, the ratio between the two can be used to predict **hazard ratio**. It can be seen on the graph (Fig. 6.7) that at 4 years, 80 % of the patients with stage 1 disease are surviving where as only 68 % of patients with stage 2 disease are surviving. Stage 1 to stage 2 hazard ratio is 80/68 = 1.18. It means that at 4 years patients with stage 2 disease are 1.18 times at 4 years more likely to die when compared to stage 1 disease (Fig. 6.7).

Fig. 6.7 Survival analysis graph showing survival for 2 conditions



Kaplan–Meier Estimator or Survival Graph

Example 4
HIV-positive patients have progressive mortality over a period of time. A new anti-HIV drug (Drug X) claims to improve the mortality rate. In a city 5000 positive patients are found and are randomly assigned to study group and control group. Study group patients received the Drug X. Control group patients received a placebo. When death occurs it is recorded with date. The dates are recorded when a patient enters the study or lost for follow-up or excluded from the study or dies. At the end of the study, the number of survivors is recorded.

For this type of data **Kaplan–Meier estimator or survival graph** can be used. It plots the percentage of survivors against time. The line has several small steps. Each point on the graph line has a corresponding point on *y*-axis showing the number (or percentage) of survivors and on *x*-axis showing the time in months or years. A number of similar graphs can be found on the web site. For example, in the following graph, the number of patients surviving at 15 months: 68 (approximately) for study group (upper line) and 78 control group (lower line), respectively (Fig. 6.8) (ref: http://www.kurtosis.co.uk/ideas/kaplan-meier/).

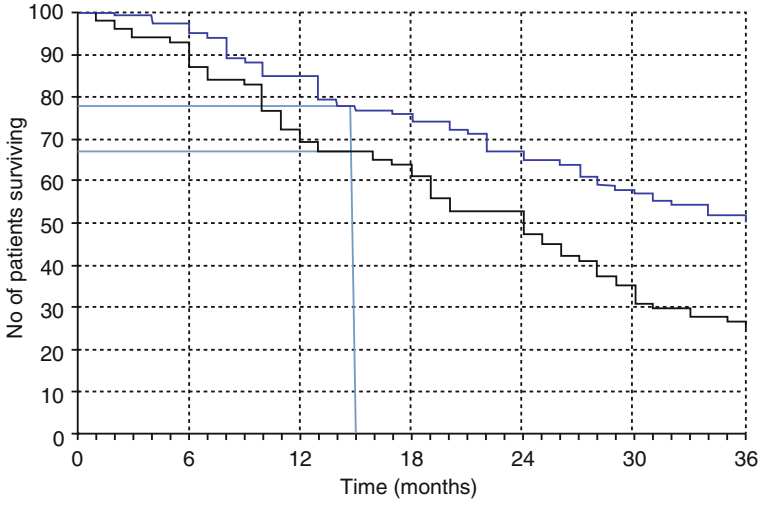


Fig. 6.8 Kaplan–Meier estimator or survival graph

Survival graph has a curved line. Kaplan–Meier estimator has a number of steps. At each step the proportion of survivors is plotted against time.

Multivariate Analysis

There may be multiple factors affecting the outcome. Then we have to quantify each of them in the order of importance. For example, in carcinoma breast a number of factors affect the outcome like estrogen receptors, nodal status, tumor size,

menopausal status, etc. Multivariate analysis is more realistic in a variety of medical conditions. It can be used for more complex data where univariate analysis is not possible. If univariate analysis is done, it ignores many variables, and outcome prediction is somewhat less accurate. On the con side, multivariate analysis is more complex and requires a larger number of observations to accurately predict the outcome. Interpretation and understanding the interpretation is more difficult. The aim of multivariate analysis is to predict the outcome based on some existing information. For example, if TNM status and grade of the tumor of a particular cancer are known, 5-year survival can be predicted. The second aim is to explain. For example, we can explain which variable out of four variables, is the most important variable (**factor analysis** is a type of multivariate analysis: A factor can have many variables. So a number of variables are taken together to form a factor to analyze), which is the most important variable that can be explained. Different methods (like MANOVA, logistic regression, multiple regression, ANCOVA, etc.) are applied for different types of data. A full explanation is beyond the scope of this book, which is written for a beginner. A few examples can help to understand the ideas behind the analysis and when it can be applied.

Examples where multivariate analysis is applicable
Example 5
To study the factors affecting weight reduction, researcher collects data on dietary habits, calorie intake per day, vegetarian or nonvegetarian, the type of work (sedentary or manual), and exercise habits. How do these data affect outcome (weight reduction) after an intervention?

Example 6
The risk of stroke and cardiac event is dependent upon many factors like hypercholesterolemia, diabetic or not, whether hypertensive or not, and smoker or not. A study as to how much these factors contribute individually to myocardial infarction has to consider all these variables in the analysis.

Multivariate analysis is not a single test. Different types of analysis are required for different types of data.
Multivariate analysis can quantify the importance of multiple factors in the order of importance.

Power of a Study

It is $1-\beta$. β value is the possibility of accepting hypothesis when the hypothesis is false. So as the power increases, this possibility (of accepting a false hypothesis) decreases. If β of a study is 0.1, then the power of that study is 0.9.

Exercises

1. Consider a hypothetical study: hepatitis B has a mortality of 3 % in 10 years. A new drug is discovered which was proved to reduce the mortality to 1 % of hepatitis B when given for 10 years. The risk of adverse reactions which could be life threatening at times is 0.5 % when the drug is taken for 1 year. Calculate the absolute rate reduction and the number needed to treat.
2. Subarachnoid block using spinal needles to inject drug into the subarachnoid space is practiced in lower abdominal surgeries. Postoperatively many patients complain of headache (spinal headache). Is the incidence of headache related to the size of the needle used?

   Setting: In a hospital 278 patients undergoing elective lower abdominal surgery were randomly assigned to two groups.

   Intervention: In group A, 24 G spinal needle was used to give a subarachnoid block. In group B 26 G needle was used. The results showed that out of 149 patients in group A, 22 patients developed postspinal headache. In group B, out of 129 patients, seven patients developed postspinal headache.

   Calculate the relative risk reduction, absolute risk reduction, and number needed to treat.
3. There is a question whether the incidence of incisional hernia is more after midline incision than after paramedian incision. To decide the issue, a prospective randomized study was conducted.

   Settings: Patients undergoing elective upper abdominal surgeries between Jan 2011 and Dec 2013 were recruited to the study. They were randomly assigned to midline group and paramedian group. The same surgeon operated on all patients. Suture material, antibiotic protocol, and all other factors were similar. Patients were followed up for 2 years from the date of operation. All patients were carefully examined to see if there is an incisional hernia and results were recorded.

   Results: Results are given in Table 6.4.

Table 6.4 Incidence of incisional hernia after 2 types of incisions

Group	Incisional hernia	Total
Midline group	12	178
Paramedian group	6	152
Total	18	330

   Calculate odds and odds ratio.

4. A study was undertaken to see the correlation between the practice of perioperative ceftriaxone (antibiotic) usage and resistant strains of bacteria grown from the pus of wound infection. Ten hospitals in a city were chosen for the study. In each hospital the percentage of patients receiving ceftriaxone was calculated. Cultures from each hospital were recorded separately. Positive culture which showed bacterial resistance to ceftriaxone is also recorded. From this data, the percentage of resistant cultures is calculated. The data are presented in the table. The data are recorded over a period of 1 year (Table 6.5).

Table 6.5 Exercise 4

	Percentage of surgeries where perioperative ceftriaxone was given	Positive cultures	No. of positive culture with resistance to ceftriaxone	% of resistant cultures
Hospital 1	10	38	5	13.15
Hospital 2	20	60	10	16.67
Hospital 3	30	55	9	16.36
Hospital 4	40	48	9	18.75
Hospital 5	50	35	8	22.86
Hospital 6	60	22	7	31.82
Hospital 7	70	67	27	40.30
Hospital 8	80	42	23	54.76
Hospital 9	90	51	32	62.75
Hospital 10	100	31	25	80.65

A. Construct a graph to see if these two variables correlate with each other. What type of correlation do you see?

B. From this model, calculate the percentage of resistant cultures in the 11th hospital where ceftriaxone is used in 35 % of patients.

Answer to 1.

RRR = 2/3 = 66.66

ARR = 3–1 = 2 %

NTT = 100/2 = 50

Answer to 2. See Table 6.6

Table 6.6 Answer to Question 2

Group	Developed spinal headache	Did not develop spinal headache	Total	%
Group A	22	137	159	13.836
Group B	7	120	127	5.5118
Total	29	257	286	
RRR = 13.43812437 (=13.836–5.5118/13.836) **ARR = 8.324666964** (=13.836–5.5118) **NTT = 12.01249256** (=100/ARR)				
Risk ratio = 2.510332435 (=13.836/5.5118)				
Interpretation: group A has 2.51 times higher risk of developing postspinal headache				

Answer to 3.

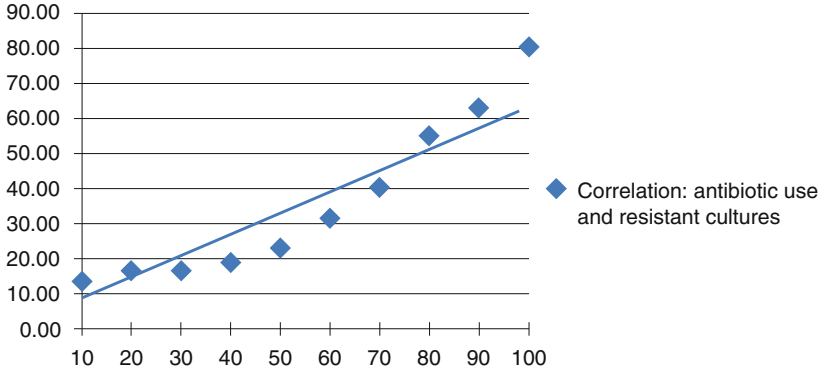
Answer to Question 3

Group	Incisional hernia	No incisional hernia	Total
Midline group	12	166	178
Paramedian group	6	146	152
Total	18	312	330

Odds for midline group = 12/166 = 0.0722
Odds for paramedian group = 6/146 = 0.0410
Odds ratio = 0.0722/0.0410 = 1.759

The odds ratio > 1 indicates that the rate of incisional hernia is more in midline group. (Odds ratio = 1 indicates there is no difference in the rate of event.)

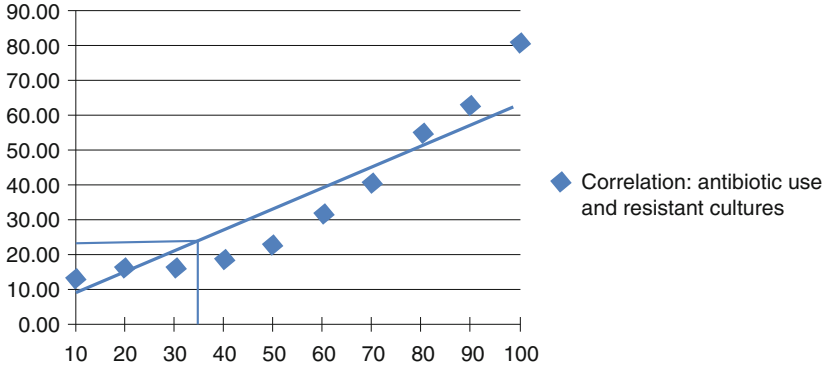
Answer to 4.



Answer to 4A

Horizontal axis shows the percentage of patients receiving ceftriaxone in different hospitals. Vertical axis shows the percentage of culture that is resistant to ceftriaxone. Each dot represents one hospital with its percentage of resistant cultures. We can see a positive correlation.

Observe the smaller line drawn vertically at 35 on horizontal axis (*x*-axis). From the point where it touches the graph line, a horizontal line is drawn to touch vertical axis line (y-axis). The value on *y*-axis corresponds to 24. So, for 35 % ceftriaxone use, the percentage of resistant cultures expected is 24 (approximately).



Answer to 4B

Designing a Study/Clinical Trial/ Dissertation, Etc.

7

I don't teach my children.
I create condition for them to learn.

Albert Einstein

Learning Objectives
To give an over view of clinical trial
Phases of trials for drugs
Steps of clinical trial
Principles of designing a trial
Presentation of data
Principles of clinical audit
Principles of mass screening

General Considerations

Clinical trial is defined as a study on human beings designed to test a device or drug or a procedure.

Clinical trials are required to answer a clinical problem whether a treatment or surgery is superior to another or a particular drug is effective in a particular condition. Sometimes the drug or the procedure already exists but we want to know new things about it. Also, trials are required to convince the government regulatory authorities for projects or launching a new drug.

Let us consider a few examples of why we need trials and audit in our day-to-day practice. In the 1990s there were strong recommendations for hormone replacement therapy (HRT) to mitigate the postmenopausal symptoms and to prevent osteoporosis. Today, there is clear evidence that HRT should not be used as a routine, because of VTE complications. So the current practice is not to use HRT. We got this information because of good clinical trials.

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_7

As undergraduates we read Halsted operation as the standard operation for carcinoma breast. Today no surgeon performs this operation. Why? There were good trials by which we came to know that these practices were wrong, and we have better options. We need trials to get these data.

This knowledge of design of trials and biostatistics helps us to analyze and interpret the presented data correctly. While reading journals or explanation by medical representatives, we come across terms like ***relative risk reduction, absolute rate reduction, and number needed to treat***. To understand these concepts, consider this example. This example is also useful to highlight why should we know how to interpret the data and statistical terms. Sometimes the pharmaceutical companies use the term ***relative risk reduction***. If we do not know the proper interpretation, we will be misled to overrate the efficacy of the drug and prescribe their drugs. Suppose there is a condition which has mortality of 3 in 10,000 and a particular drug is shown to reduce this mortality to 2 per 10,000. Then relative risk reduction is 33 %. Pharmaceutical company may hide the other details and show only this line in bold highlighted letters "The drug reduces relative risk by 33 %." This 33 % looks very impressive, and if we do not know how to interpret the result, we may recommend this drug to our patients. If we analyze the actual data and not the conclusion, we will see that **absolute risk reduction** is only 0.01 %, because it has reduced the mortality rate by 1 in 10,000. That is to say, we have to give this drug to 10,000 patients unnecessarily to reduce mortality by 1. This number is called ***number needed to treat***. If the risk of adverse effects is about 5–10 %, this is actually increasing the risk of adverse drug reactions in the other 9999 patients. So we will be doing more harm than good. This is like saying "*You are 33 % less likely to die from a tree landing on your head if you wear a helmet all the day.*" Ridiculous, is it not? That is because chances of tree landing your head in normal circumstances are so remote that reduction of 33 % or even 75 % of this remote possibility makes no sense. The results presented by two different terms can have opposing effects on prescribing decision of the clinician. Consider the statement "*The drug has relative risk reduction of 33 %.*" Fantastic results: the clinician considers using this drug. The same data can also be presented in another statement "*Number needed to treat is 10,000 to get benefit and save 1 patient.*" The clinician considers against using this drug. Irony is that both the statements are based on the same data.

Incidence of a Disease The number of new patients affected by the disease per 100 population in 1 year. For example, the incidence of HIV positivity is two means, and two new HIV-positive cases are detected per 100 populations in a year. If in a geographical area with a population of 15,000, *six* new cases of cleft palate are detected; then the incidence of cleft palate in that area is $6 \times 100/15{,}000 = 0.04$.

Prevalence of a Disease The number of patients having the disease per 100 population at a given point of time. For example, the prevalence of HIV in a state is 6 % means 6 % of the population tests HIV test positive at that particular time. If HIV test is done on 10,000 patients, test will be positive 600 patients.

Concerns While Designing a Trial

When we think of a trial, we need to address certain concerns.

The study should be **ethical**. We should not cause harm to any patient just for the sake of conducting a study. Human beings are not experimental animals. You cannot wait and watch the patient without giving blood transfusion when he is massively bleeding, just because you want to study hemodynamic changes after a massive bleed.

The study should be **scientifically valid**. Flimsy ideas cannot be tested in clinical studies. There should be sufficient preclinical data to show that the procedure or drug may be useful in a clinical situation.

Integrity of the investigator is very important. He should not give cooked-up data and conclusions for his personal benefit. If the surgeon starts a study thinking that a particular procedure is beneficial but results from the study indicative of the contrary, he should not hesitate to conclude that the procedure is not useful. If he gives wrong conclusions deliberately, he will be doing a great disservice to the society and mankind.

> Worry more about your conscience than your reputation.
> Because your conscience is what you are and your reputation is what others think of you.
> Unknown (selected from a WhatsApp message circulated in the group)

There are medicolegal and regulatory concerns in many areas and subjects. These things are to be addressed appropriately.

Cost We must also estimate the cost of the study and ensure adequate resources before starting a study.

Approval of local ethical committee is to be obtained.

The World Medical Association has developed the Declaration of Helsinki as a statement of ethical principles to provide guidance to physicians and other participants in medical research involving human subjects. This is a web page of the declaration. You can visit this page on the Internet for further details (http://www.wma.net/en/20activities/10ethics/10helsinki/).

World Medical Association web site

Negative Results

> Good judgment comes from experience.
> And often experience comes from a bad judgment.
>
> Rita Mae Brown

When a negative result comes out of a trial, it must be reported as it is. Negative results are as important as positive results. For example (hypothetical examples are used with the sole purpose of clear understanding: statements are not to be used for clinical practice), a study is designed to see the beneficial effects of *flush therapy* for ureteric calculi. (Flush therapy is giving IV fluids rapidly and injecting IV frusemide to produce large quantity of urine to flush out the calculus.) Expected (positive) result was majority of the patients will be benefitted. Suppose the study results showed that many patients developed complications, and it is actually harmful to the patients. Researcher may stop the study and he is not keen on publishing the results. But the results must be reported and published as it is. It helps another researcher not to repeat similar studies. To quote another classical example, studies were conducted to find beneficial effects of routine hormone replacement therapy (HRT) in postmenopausal women. But studies showed more mortality in the study group who received HRT because of higher incidence of thromboembolism. These reported results lead to recommendation against routine use of HRT in postmenopausal women.

Teamwork

Modern sophisticated clinical studies are complex and are a teamwork of:

- Medical doctors
- Biostatisticians
- Data managers
- Monitors
- IT specialists
- Data analysts

Phases of Trials: Applies to New Drugs

Before a drug is accepted for general use, it undergoes various phases of trials. Initially a lot of preclinical tests and animal studies would be done to test the efficacy and safety. If the drug is found to be useful, then clinical trial is started to further evaluate the drug.

Phase 1 trial starts after preclinical studies in animal models. The main aim is to find safety. Usually, trial is done on healthy volunteers. There is no blinding or controls.

Phase 2 trial starts if the drug passes phase 1. The drug is tested on patients with disease to find out efficacy and more safety data in small number of patients.

Phase 3 Once the drug passes phase 2 trial, it is subjected to a randomized controlled double-blind study. Here, more number of patients are recruited. Many times it is a multicentric study. We get more data on safety. More often drugs fail to pass this phase of research to get into market. If it passes, then authorities approve the drug for general use in public.

Phase 4 studies explore further indications and efficacy. By this time the drug is already in the market. Many times, side effects found at this stage may result in withdrawal of the drug from the market. Cisapride was withdrawn after finding its cardiac side effects in this phase. In a recent article (BMJ 2016;352:i1541 http://dx.doi.org/10.1136/bmj.i1541), pioglitazone is reported to be associated with increased risk of bladder cancer. The drug is already in the market for more than 15 years. Such reports come after the use of the drug in general population. Such long-term studies are required to find some of the risks associated with the drug.

Some Commonly Used Terms

Effect Size It is the quantitative expression of something happening. For example, in the statement "the routine mammogram detects carcinoma breast in asymptomatic women is 1 %," the effect size is 1 % or 1 in 100.

Parallel Studies and Crossover Study

Parallel studies are the studies by drawing different samples from the same population. For example, we have 10,000 men group. We are estimating average pulse. We sample 100 men, count the pulse rate, record, and find the average. We then take another sample of 100 men different from the first and do the same. These two would become parallel studies.

In ***crossover study***, first we study a sample for a factor. After a period of time, the same sample group is studied for another factor. For example, we want to test analgesic effects of Drugs A and B in a certain group of patients with terminal cancer pain. We divide the group into study and control (placebo) groups. Initially we use Drug A in study group for certain period of time, say 1 week. Then for the next 1 week, we use Drug B, rest all remains same. We have to record VAS at the end of 1 week and at the end of 2 weeks compared with control group.

Designing a study or clinical trial or clinical audit or writing a dissertation has many principles in common. It consists of:

(a) **Planning and pilot study**
(b) **Collection and compilation of data**
(c) **Data analysis and inference**
(d) **Presentation of data and results**

Planning and Pilot Study

> If I had 9 hours to chop down a tree, I would spend the first 6 hours in sharpening my axe.
> Abraham Lincoln

First of all, you have to decide what you want to study. You have a hypothesis which you would like to test. Turn this hypothesis into a question. This question should be answerable. Based on this question, objectives of the study are fixed. For example, by theoretical considerations you hypothesize laparoscopic hernioplasty is superior to open hernioplasty in terms of recurrence. Turn it into a question: whether laparoscopic hernioplasty has a lesser recurrence rate? Then the study objective would be to compare the recurrence rates of these two procedures done in comparable groups of patients. You need to have two similar groups of patients. For one group, you do open hernioplasty and for the other group, laparoscopic hernioplasty. Fix a study period and protocol for follow-up. At the end of the study period, assess how many patients in each group have recurrences, and compare those using statistical tests. However this is an oversimplified case. In reality there are multiple factors to be considered.

The procedures should be safe and effective in treating the condition. Ethics do not allow a risky procedure to be tested when there is already safer alternative.

Now design the intervention. Write down the protocol in detail, which should include the aim of the study, the intervention, the inclusion and exclusion criteria, the outcome to be measured, the parameters, the definitions, etc. The same

procedure must be done for all the patients in the group. The protocol should be strictly followed. Following the same example, if you fix up the size of the mesh as 12×15 cm, you cannot use a 10×15 cm size mesh in any patient of the study. Because, if there is more recurrence, you cannot decide whether this increased recurrence is due the procedure or because of the smaller mesh. In the protocol you must write down in detail and specifically the procedure followed in the study.

You must decide beforehand what outcomes are to be measured. It may be one or more. In the above example, it is recurrence rate.

Control may be a placebo or another known procedure, usually an accepted and established procedure (gold standard). For example, while studying recurrence rate in laparoscopic hernioplasty, we compare it with Lichtenstein repair, which is considered as a gold standard method of hernioplasty.

Next you have to decide who the subjects are. Define their characteristics. Continuing with the above example, we need to decide whether we want to study primary hernia or include recurrent hernia also or unilateral vs. bilateral, etc. All the subjects must be comparable. If there are too many variables, it is difficult to analyze the data and come to conclusions. For example, if open surgery group has more of manual laborers or more of older age groups and laparoscopy group has more of sedentary class of patients or younger age groups, we may wrongly conclude that the laparoscopy is better, since manual laborers who lift heavy weights daily are known to have more recurrence. After you publish the data, critics may question the validity of the conclusions and say that the results are attributable to dissimilar demography of the patients. Your efforts, time, and resources are all wasted.

There must be strict criteria to include or exclude the particular patient from the study. For example, all patients with uncomplicated uni- or bilateral inguinal hernia are included, or all cases of obstructed hernia are excluded from the study, or patients with recurrent hernia are excluded. These criteria vary depending upon the objectives and can be one or more.

If the criteria are very narrow, it is easy to analyze and conclude as the patients are a homogenous group. However, the conclusions apply to only small set of population who satisfy the said criteria and cannot be extrapolated to other types of patients. Also, it is more difficult to recruit the patients as the number of patients available with the said criteria will be small. For example, if the inclusion criterion is the patients' age between 20 and 30 years, it is easy to analyze as the factors like BPH and COPD which can influence recurrence are not there, but it remains unanswered whether particular hernioplasty is superior in elderlies too.

On the other hand, if criteria are broad, it is easy to recruit the patients but difficult to analyze the results as patients will have many variables.

Typically inclusion criteria are like this:

1. The subjects should have the disease of interest (obviously, you cannot include a person without hernia in our example).
2. Subjects have certain amount or degree of disease. (Hernias have different sizes and complexities. Study should mention if all Nyhus types are included or only certain types are included in the study.)

3. Informed consent of the subject is very essential.
4. Any other specific criteria have to be predefined.

If the study is about medical treatment or drugs, following factors are to be considered. Patient must not be on active treatment, must not be allergic to the drug of intervention, must not be pregnant or breastfeeding or a child, etc., unless the purpose of the study is to examine the effects on pregnancy or breastfeeding or a child. If the patient is on some other treatment, its effect on the result cannot be known.

Once the population of the subjects to be studied is selected, the subjects are divided into two groups. Then study and control interventions are applied. Some studies make more than two groups.

Sample Size

Sample size is the number of patients required for the study. The larger the number, the better the reliability of the result. But a very large number makes it difficult to recruit the patients, involves more expenditure, and takes more time to complete. Hence, a balance is required.

Please refer to appropriate chapters and understand the following terms:
Type 1 errors and type 2 errors (alpha and beta)
Power of a study
Confidence interval and confidence level

Assigning the patient to a group can be random or nonrandom. Randomization is better and it avoids bias.

Then the predefined outcome is recorded. The outcome must be quantifiable to record and compare. We can apply statistical methods only to numbers. If the outcome is just improvement in the general condition or better quality of life, etc., we just cannot compare the results. So there should be a way to record the outcome in terms of numbers. For example, for assessing pain, recording pain as mild, moderate, or severe pain is not advisable. It is better to use visual analog scale to record the score as 1–10 numbers.

Points to ponder: attention must be paid to certain points.

1. While comparing two groups, the groups must be comparable in all other respects, e.g., age, risk factors, tumor staging, severity of the illness, etc.
2. Mathematical scoring systems useful in certain areas, e.g., ASA grading for risk stratification of surgical patients, Glasgow coma scale, visual analog scale for assessing severity of pain, etc.

3. There should be sufficient number of patients, because the possibility of error is high when the number of patients is small. On the other hand, you cannot have a very large number also because of cost factor and restricted resources. So, we have to strike a balance.
4. Account for loss of data, loss of follow-up, poor compliance, exclusion from the study because of a side effect or death, etc.
5. There should be a strict definition of terms and results to eliminate observer error. For example, if you are assessing hernia recurrence, define what a recurrence is, how it is assessed, whether only clinically or using any imaging modality, etc. If you are classifying recurrences as early recurrence and late recurrence, then define what "early" is, i.e., less than "X" months. The same method is to be applied to all the patients in all the groups.

The number of patients to be studied depends upon the frequency of effect: the smaller the effect, the more number is required, e.g., if operative mortality is 1 in 100, you need to study at least 100 (preferably ten times more) to see the improvement in mortality. A large number of patients, while ideal, costs more, takes more time, and is more difficult to complete the study. Hence, you have to consider the pros and cons, your resources, time available to complete the study, etc. to strike a balance and decide on the number.

While writing down the protocol, consider the following factors:

1. What is the aim of the study?
2. Is the selection of patients to groups randomized?
3. Are inclusion and exclusion criteria strictly applied?
4. Were patients, their clinicians, and study personnel "blinded" to the treatment?
5. Compare and ensure that the groups are similar.
6. Apart from the intervention, all the groups must receive the same treatment in all other aspects
7. The number of patients to be included.
8. Measured outcomes must be clinically relevant.
9. All clinically important outcomes must be considered.

Pilot Study

A **pilot study**, **pilot** project, or **pilot** experiment is a small-scale preliminary *study* conducted in order to evaluate feasibility, time, cost, adverse events, and effect size (statistical variability) in an attempt to predict an appropriate sample size and improve upon the *study* design prior to performance of a full-scale study (**Pilot**

experiment-Wikipedia, the free encyclopedia https://en.wikipedia.org/wiki/Pilot_experiment).

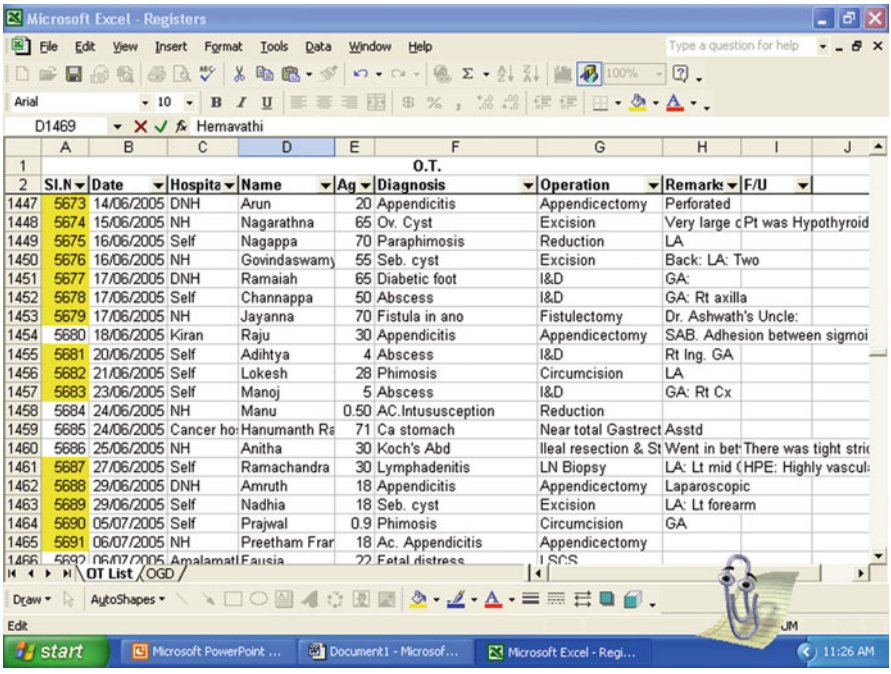
It is (often referred to as **feasibility studies***)* an attempt to avoid time and money being wasted on an inappropriate large study. We will know shortcomings/defects of the study design. We will know if any other variables/factors are to be added to outcome measurement, etc. Pilot study is conducted on different sample of the same population. Its duration may be short, say, 6 months or 1 year.

Collection and Compilation of Data

It starts from the beginning of the study. Initially patients' data are to be collected and then the data about the group to which it is assigned, outcome, side effects, complications, etc. To collect data, if patients are your own, you can use questionnaires or interviews. In multicentric studies, data from different centers are pooled together for analysis. Here, e-mail is very useful in exchange of data between the centers. For information and literature review, journals or various web sites on the Internet can be searched. After collecting the data, it is compiled systematically to create a database.

Although the data can be handled manually, it is cumbersome and time-consuming when large numbers of patients are there. There are certain softwares to handle database. MS Excel, WPS Spreadsheet, MS Access, etc. are examples of such softwares. Custom-made softwares are also possible to meet specific requirements.

This is an example to show how excel can be used. It is a spreadsheet and holds the data in rows and columns.

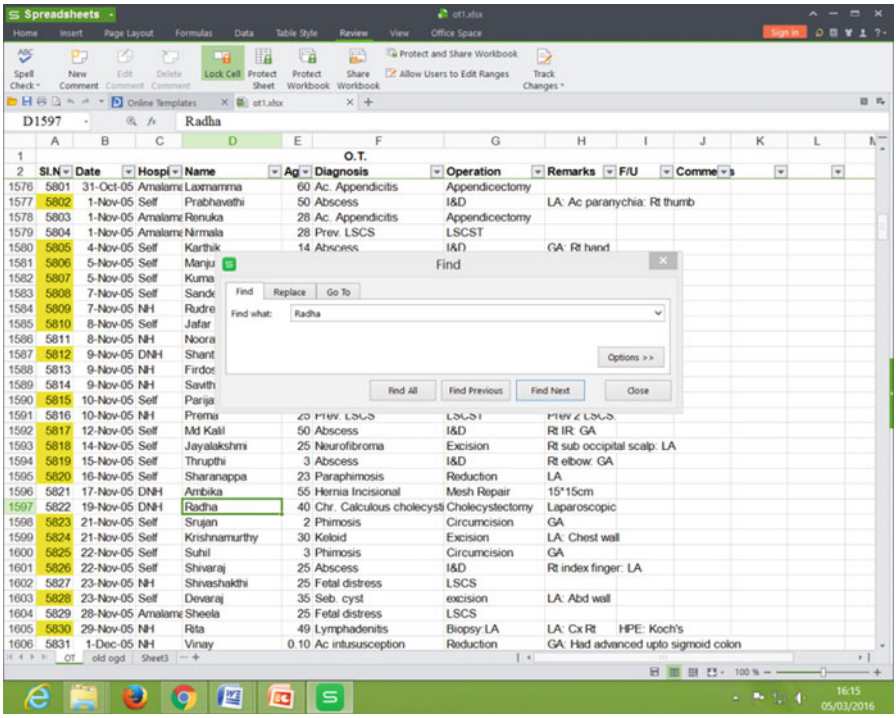


	Sl.N	Date	Hospita	Name	Ag	Diagnosis	Operation	Remarks	F/U
1447	5673	14/06/2005	DNH	Arun	20	Appendicitis	Appendicectomy	Perforated	
1448	5674	15/06/2005	NH	Nagarathna	65	Ov. Cyst	Excision	Very large c	Pt was Hypothyroid
1449	5675	16/06/2005	Self	Nagappa	70	Paraphimosis	Reduction	LA	
1450	5676	16/06/2005	NH	Govindaswamy	55	Seb. cyst	Excision	Back: LA: Two	
1451	5677	17/06/2005	DNH	Ramaiah	65	Diabetic foot	I&D	GA:	
1452	5678	17/06/2005	Self	Channappa	50	Abscess	I&D	GA: Rt axilla	
1453	5679	17/06/2005	NH	Jayanna	70	Fistula in ano	Fistulectomy	Dr. Ashwath's Uncle:	
1454	5680	18/06/2005	Kiran	Raju	30	Appendicitis	Appendicectomy	SAB. Adhesion between sigmoi	
1455	5681	20/06/2005	Self	Adihtya	4	Abscess	I&D	Rt Ing. GA	
1456	5682	21/06/2005	Self	Lokesh	28	Phimosis	Circumcision	LA	
1457	5683	23/06/2005	Self	Manoj	5	Abscess	I&D	GA: Rt Cx	
1458	5684	24/06/2005	NH	Manu	0.50	AC.Intususception	Reduction		
1459	5685	24/06/2005	Cancer ho	Hanumanth Ra	71	Ca stomach	Near total Gastrect	Asstd	
1460	5686	25/06/2005	NH	Anitha	30	Koch's Abd	Ileal resection & St	Went in bet	There was tight stric
1461	5687	27/06/2005	Self	Ramachandra	30	Lymphadenitis	LN Biopsy	LA: Lt mid (	HPE: Highly vascul
1462	5688	29/06/2005	DNH	Amruth	18	Appendicitis	Appendicectomy	Laparoscopic	
1463	5689	29/06/2005	Self	Nadhia	18	Seb. cyst	Excision	LA: Lt forearm	
1464	5690	05/07/2005	Self	Prajwal	0.9	Phimosis	Circumcision	GA	
1465	5691	06/07/2005	NH	Preetham Frar	18	Ac. Appendicitis	Appendicectomy		
1466	5692	06/07/2005	Amalamat	Fausia	22	Fetal distress	LSCS		

An example of database creation using MS Excel sheet

As many columns and rows as may be required are included in the database. Each column holds a certain type of data. Each row holds the data of a patient. It is important to give a unique ID number to each patient so that when there is more than one patient with the same name, we can identify them by their ID number. The advantage of this type of database is that we can retrieve data easily.

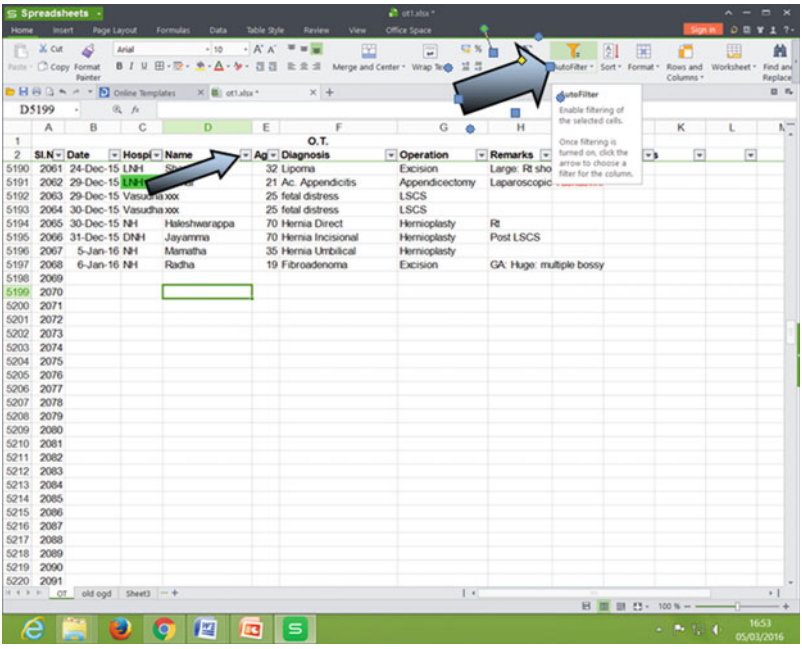
Suppose the data of the patient "Radha" needs to be searched: Open the database. Hold Ctrl key and press F key (Ctrl + F). A small window opens. In the space next to "find what," type *Radha*. Then click on "find next."



Searching for specific data from data base

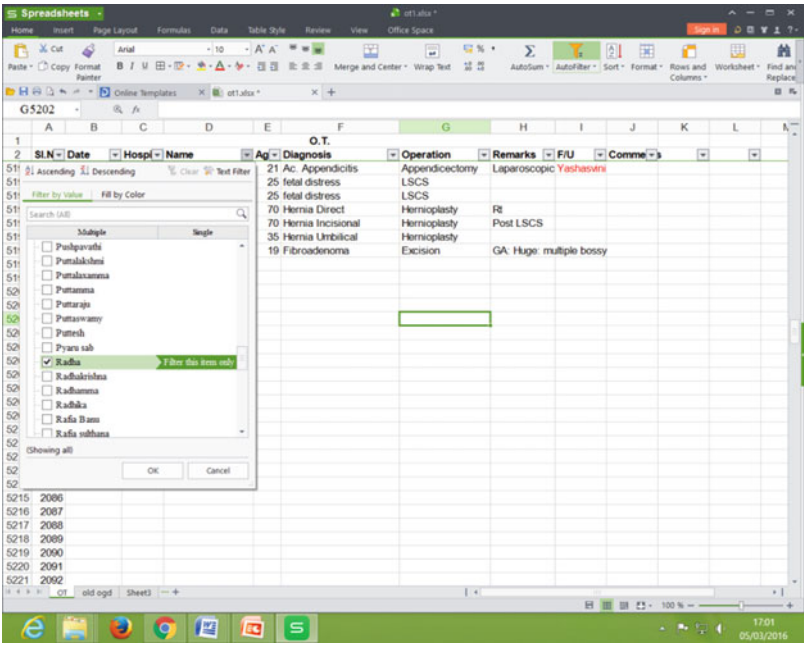
Immediately the cell containing patient's name Radha will be highlighted. If you keep on clicking the same tab, it highlights the next cell containing the same name. If there are no more cells with the same name, it simply shows the same cell.

Filters are useful to get only those rows of data in which we are interested. Filters can be inserted by clicking on the tab "autofilter" (bigger arrow). A downward-directed triangle is seen next to each heading (smaller arrow).



Use of "Filters" in Excel sheet

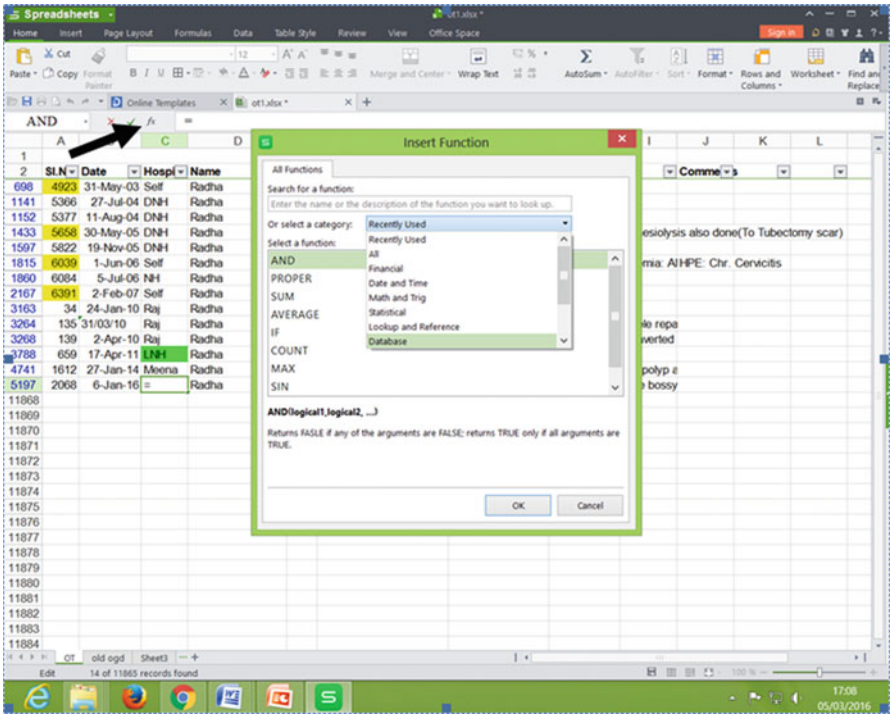
If you click on this triangle, a list of names (or data in the column) is shown. Select the one you are interested. Here the name Radha is selected. Click OK.



Use of "Filters" in Excel sheet-continued

Now the rows containing names as Radha only are displayed.

Complex calculations can be done easily on Spreadsheet/Excel by using formulae. These are discussed in appropriate chapters. There are several other functions which help us in handling data. A list can be seen by clicking on Fx button (arrow). There are about 235 functions in Excel. These are divided into nine categories like logical, mathematical, etc. All these calculations can be done easily.



Functions and calculation on Excel sheet

Likewise there are numerous useful features in Excel/WPS. The data can be arranged in ascending or descending order by a few clicks. Data can be *validated* for a column. For example, if data is *validated* as *number* for a particular column, then that column will take only numbers and will not take text data or any other types of data.

While creating database, take care to create all necessary fields and enter the data. Otherwise at a later stage, it becomes very difficult to add a column of field. For example, if a column for the field "Size of the mesh" is not created initially and the size of the mesh has to be added after 2 years of study, it is not possible to enter this data for all the patients as this data is not collected initially.

MS Access also can be used to create a database. It is essential to study this program to use it optimally. Custom-made softwares can be programmed with the help of IT professionals which will be very useful to cater to your needs.

Data

The data may be ***retrospective*** or ***prospective***. Retrospective data are collected from already existing previous data. For example, we want to study the perforation rate of acute appendicitis: we go to record room and search the case papers of patients with a diagnosis of acute appendicitis admitted over the previous 2-year period. Data of these patients are collected for study. This is ***retrospective study***.

In ***prospective study***, there are no patients to begin with. We start assigning the patients to the study group and collect the data as and when available. Taking the same example, whenever a patient with the diagnosis of acute appendicitis is admitted, we collect the data for the patient. This goes on till the end of the study period.

Types of Study

Open: Both the patient and the investigator know to which group the patient belongs (study or control group).

Single blind: Patient does not know but the investigator knows to which group the patient belongs.

Double blind: Both the patient and the investigator do not know to which group the patient belongs.

Study may also be classified as ***observational study*** and ***control study***. In observational study, we study and record various factors by observing the population or sample. We do not intervene in the process. The investigator does not have control over assigning the subjects to various groups. Usually they are done to assess cause-effect relation. Here ethical concerns are not there since we do not intervene in the treatment or natural process. Its value in the hierarchy of evidence is inferior to RCTs. The observational studies are of different types:

Case-control study
Cross-sectional studies
Longitudinal studies
Cohort study

In **case-control study**, we study two existing groups with different outcome and compare the results, usually for causal relationship. For example, a large population is divided into two groups: smokers and nonsmokers. We are not assigning the subjects into these groups. They already exist. Depending upon their habits, we just separate them into smoker or nonsmoker groups and study these groups. If carcinoma lung or COPD occurs at a higher frequency in smoker group, we conclude that the habit of smoking and carcinoma may have a causal relationship.

In **cross-sectional studies**, we study a particular population at a given point of time. For example, to know the incidence of HIV-positive status, we conduct HIV test on the population at a given point of time, and say X% of population are HIV positive. If we repeat the study after some time gap, we can know the trend of the disease over a period of time, with or without an intervention/treatment/preventive measure.

In **longitudinal study**, we study a population over a period of time for the same variable factors/parameters repeatedly. One of the uses is to know the natural history of the disease. For example, we select a group of patients with dementia. Study and record various parameters over a period of time, say 10 years.

Cohort study is a type of longitudinal study where a group of patients is closely monitored over a period of time.

Data Collection

Data may be collected by an independent data collector or by the involved clinicians. But it is important to avoid bias.

Bias

> People almost arrive at their beliefs not on the basis of proof,
> but on the basis of what they find attractive.
>
> Blaise Pascal

The greatest enemy of a scientific study is the bias. Investigator may assign better-risk patient to the group which he wants to push. If there is an unacceptable complication, investigator may hide the results or remove that particular patient from the study. Conclusions here can be unreliable and misleading.

These are certain measures to minimize the risk of bias.

Comparability of All Groups There is no point in comparing two procedures of cancer treatment if one group has the majority of stages 1 and 2 disease and the other has the majority of stages 3 and 4 disease, or one group has the majority of elderly and the other has the majority of younger patients. The statistical tests should be applied to all the groups in the study to prove that there is no bias, and all the groups are comparable in all aspects. For example, apply *T* test to table of age of the patients to show that there is no statistically significant difference in the age of patients in all the groups. This can eliminate the bias with respect to age. Both (or all) the groups are similar.

Measurement of Outcome Should be Objective and Not Subjective It is not enough to say patient had palliation after a palliative procedure, but should define what palliation is: what are the criteria to say that palliation is achieved. Record the results based on these criteria. It should be possible to express them in terms of

number. As an example it is already mentioned about "pain." Instead of just mentioning that patients had better pain relief, it is better to record the visual analog scale (VAS) score of pain before the starting of the procedure/intervention and record again after the procedure/intervention. Now paired *T* test can be applied on this data as they are expressed in terms of numbers.

Randomizing the patients while allotting to study or control group avoids bias as investigator doesn't have control over the allocation of the patients to different groups. Thus investigator's bias doesn't have an effect on the outcome. Random number table can be used for random assignment of patients. Computers can be used to assign the patients into the study and control group or for drawing the samples. As it is a machine, it avoids bias in selection. It works on random number table. Table 7.1 shows example of a set of random numbers.

Table 7.1 Random number table

103	748	124	254
623	458	16	413
576	680	742	279
670	156	882	798
362	837	978	433
558	663	281	736
695	343	851	574
974	98	776	27
9	579	177	694
107	313	821	700
832	199	611	587
585	630	885	891
49	286	867	170

The numbers are arranged haphazardly and cunningly without any logic or sequence. This is used to pick the patients randomly into different groups. Flipping a coin for heads or tails can also serve to randomly assign the patients to different groups.

Another important method is **blinding** the investigator and patients to intervention. If the investigator or the one who is recording the data does not know to which group (study or control) patient belongs, he will not have bias to avoid entering the complications or give "superior" results to one particular group. For example, if the investigator wants to conclude that Drug A is superior to Drug B, he will avoid entering the data on adverse events of Drug A. He will magnify the numbers of good effects. If he is blinded or if he is an independent person who does not know to which group the patients belong, he will not try to hide adverse events, which could have been in Drug B group. If he avoids, it gives superiority to Drug B, which he does not want. Ideally assessment should be by an independent person and not the principal investigator. As he is independent, he will not know whether

patient received intervention or not, he is not bothered by the conclusions, and he just records the results. If the patient comes to know that he is receiving only placebo, he may keep on complaining more. Psychological factors come into play. Blinding the patients to the intervention avoids the bias due to this factor. Blinding may be single where only patient doesn't know to which group he belongs. In double-blind studies, both patients and investigator will not know to which group the patient belongs.

Sampling

When the population is large, it is difficult, time-consuming, costly, and not practicable to study the whole population. Instead we take what is known as ***sample*** and study only the sampled population. Results obtained are then applied to the whole population. For example, to find the prevalence of carcinoma esophagus in the state of Karnataka, it is difficult to study many crores of the population. Instead, samples are taken (of say 1000 subjects in each group) randomly from different parts (may be 100 groups from different parts or even more). The higher the number, the better the reliability. If 100 groups with 1000 subjects in each are taken, there will be 100 × 1000 = 100,000 subjects. It is easier to study the sampled 100,000 subjects than many crores of population of the state of Karnataka for practical reasons. The results are then applied to the whole population. It is based on the statistical principle that the mean of the randomly drawn sample is nearly equal to the mean of the whole population.

An analogy can be given based on our examination system: suppose there are 1000 pages of information, the student is expected to learn. We want to know how much (% of information) the student knows. It is impracticable to ask all students to write all 1000 pages of information and evaluate this large volume. Instead question paper is set so that we ask questions about ten pages of information randomly. This is sampling: what % of this sampled information the student can reproduce is his marks. We assume that if he knows 60 % of the ten pages (questions sampled), then he knows 60 % of the 1000 pages. If you repeat this type of examination for a number of times with different samples of questions for the same student, he would score approximately the same % of marks every time. This is repeatability of the experiment.

Another example we are all familiar with is the opinion poll. The prediction of election results before the election in various newspapers. They sample the population from different parts of the country in random and ask the sampled population about their choice of the party. Then results of all the groups are pooled together to give the prediction. It is assumed that the opinion (or in this case, percentage of votes to various parties in the sampled population) of sampled population will be the opinion of the whole population since we have drawn samples randomly.

There are the three types of sampling:

- Simple random sampling
- Systemic random sampling
- Stratified random sampling

In sampling each unit/individual is assigned a number. Then individual is picked to be included in the study or control group by using random number table or any other method or randomization. If the selection is entirely random, it is called ***simple random sampling***.

In ***systemic random sampling***, only the first individual is picked up randomly. Then onward, all the other individuals are selected based on certain frequency, e.g., alternate individual or every fifth (or every tenth, every 15th, etc.) individual is selected.

In ***stratified random sampling***, certain groups are selected even though the groups are not distributed equally in the population, e.g., all Hindus in the state of Karnataka, age group of 20–30, etc.

Sampling Errors and Nonsampling Errors

There are certain errors arising out of samplings. Sampling errors are the errors in the averages (means) of the groups of the samples drawn from the same population. We assume that the mean of the randomly drawn sample of a population is equal to the mean of the whole population. So ideally, if we draw two random samples from a population, their means should be the same. But in reality, there may be significant variation. An example is what we see in preelection predictions published in newspapers. Although all the newspapers study the same population by drawing random samples, each newspaper predicts different percentages of votes for different parties. This happens because they study different samples from the same population. This phenomenon is due to **sampling errors**. Theoretically if all the newspapers study the whole population, then predictions of all the newspapers will be the identical. When the whole population is studied, no error occurs but error occurs when sampled population is studied. So we call this sampling error or error due to sampling method.

Nonsampling errors are not due to sampling methods but due to observer variation, inadequately calibrated instruments, incomplete coverage, etc. For example, one observer may brand a particular observation as mild pain but another observer brands the same severity as moderate pain. An evaluator may give 3 out of 5 marks for an answer but another evaluator may give 4 out of 5 marks for the same answer.

Sampling errors are caused by studying the sample instead of the whole population. For example, suppose we want to estimate average pulse rate of all the persons of a state. It is impractical to record pulse rate of the whole population. So we take samples from different regions of the state and record their pulse rate. Take the average of this record (X), and conclude that the average pulse rate of the people of the state is X.

But this may not be true value (error). If another investigator studies another set of samples, he may get a different value. This phenomenon is due to sample errors.

Bias increases the sampling errors. That's why it is important to avoid bias in sampling or assigning the subjects to various groups in the study.

Data Analysis and Inference

Having completed recording of data and creating the database, now it is time to analyze the data for various information, compare the results between the groups, apply tests of significance, and draw conclusions. Here we come across various terminologies and formulae. For complex data the help of a professional medical statistician may be needed.

The various tests of significance are discussed in a previous chapter. The test to be applied to each type of data is also discussed there.

A summary of various tests may be repeated here.

Summary of tests of significance

Test	Conditions
T test	Two sets of data
	Continuous variables
	Normal distribution (parametric) data
	Data can be expressed as mean± SD
Paired *T* test	Two sets of data for each observation
	Other conditions the same as *T* test
Mann-Whitney-Wilcoxon (MWW) test	Nonparametric data
	Other conditions the same as *T* test
ANOVA	More than two sets of data
	Other conditions the same as *T* test
Chi test	Two or more sets of data
	Categorical variables
	Parametric data
	Data are numbers and not percentages
	Minimum of five observations in each of the cells of the table
Fisher's test	No minimum of number of observations
	Other conditions the same as chi test

The data of different groups obtained from the study are presented in the form of a table. Then various tests of significance are applied to the data to see if there is statistically significant difference(s) between the groups. For medical statistics purposes, most of the studies keep significance level at $P<0.05$. The lesser the value of P, the more significant the difference.

Data may be collected from literature for comparing. Different series may give different results. Data are shown in the form of table or graph or any other appropriate form of presentation for easy comparison.

For example, complication rates in different series of intraperitoneal polypropylene mesh placement in the repair of ventral hernia are shown in the form of a table. Numbers in the bracket refer to references which should be quoted at the end of presentation.

PPM. Depicting the data in the form of a table

	Year	*N*	Infection	Fistulization	Sinus formation	Adhesions	Recurrence	Seroma
Vrijiland et al. [6]	2000	136	8	0	2	0		
Fuad Alkhouri et al. [7]	2011	123	4	0	0	2	6	1
Scripcariu V et al. [20]	2004	107	0	1	5	0	5	0
Juliane Bingener et al. [18]	2004	20	0	0	0	1	0	0
Kua KB et al. [19]	2002	25	0	0	0	1	3	0
Jitea N et al. [21]	2008	21	0	0	0	0	0	3
PK Chowbey et al. [16]	2000	202	0	0	0	0	2	36
Nihat Yavuz et al. [17]	2005	85	0	0	0	3	5	0
		719	12	1	7	7	21	40

Reproduced from Ramakrishna, H.K., Lakshman, K. Intraperitoneal polypropylene mesh and newer meshes in ventral hernia repair: what EBM says? Indian J. Surg. 2013;75:346–351

Conclusions are drawn based on the data presented. Usually conclusions are a repetition of the aims and objectives of the study. If the aim of the study is to find whether laparoscopic hernioplasty has lesser recurrence rate than open hernioplasty, the conclusion can be one of the following: "*laparoscopic hernioplasty has fewer recurrences*" or "*laparoscopic hernioplasty has more recurrences*" or "*there is no difference in the incidence of recurrence between the laparoscopic and open hernioplasties.*" Laparoscopic hernioplasty may be

superior in terms of less pain or early return to work but that was not the objective of the study. So, no conclusions should be drawn on those issues (unless those were also the objectives initially fixed and data are collected on those parameters also in the study).

If there is a difference between the groups, it should be explained, why? Explain what the importance of the findings and conclusions is. Explain how they make a difference in the management in clinical practice.

Presentation of Data and Results

Presenting the data and inference is both art and science. The same data can be presented in different ways. Some methods of presentation are more catchy and clear than others. Reader will immediately understand what is presented. So the best method of presentation has to be chosen for the particular situation.

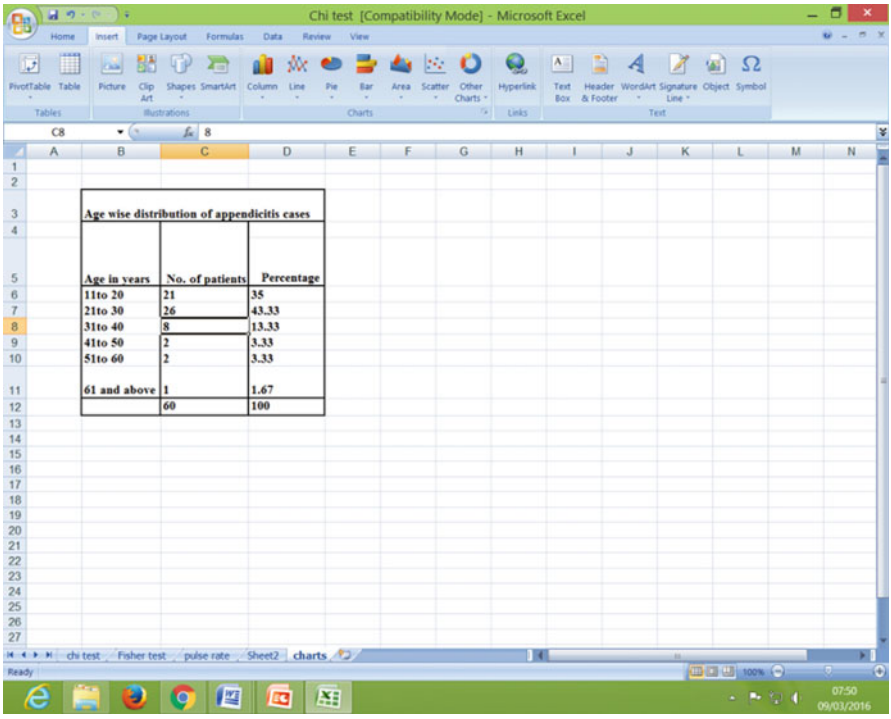
Tables

In a table the data are arranged in the form of rows and columns. A title is given to the table. The first row denotes different headings under which the data is classified. For example, the data on the incidence of appendicitis in different age groups may be presented as a table after assigning the patients to different age groups as follows.

Table 7.2 Age-wise distribution of appendicitis cases

Age in years	No. of patients	Percentage
21–30	26	43.33
31–40	8	13.33
41–50	2	3.33
51–60	2	3.33
61 and above	1	1.67
	60	**100**

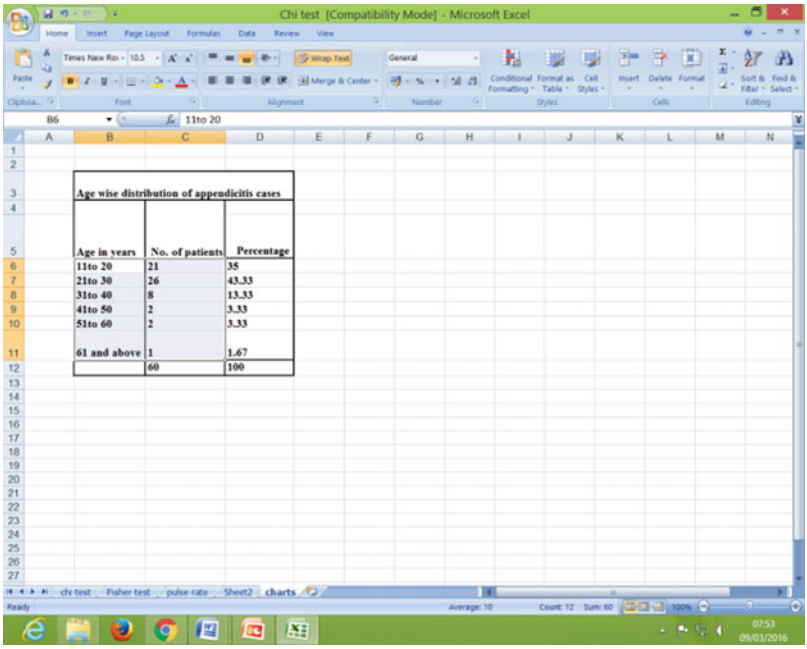
The same data can also be presented as a chart or graph. Graphs can be created manually or more easily with the help of computer. MS Excel has this feature. For example, to create a graph for the data of Table 7.2, enter the data in MS Excel.



Creating graph using MS Excel sheet

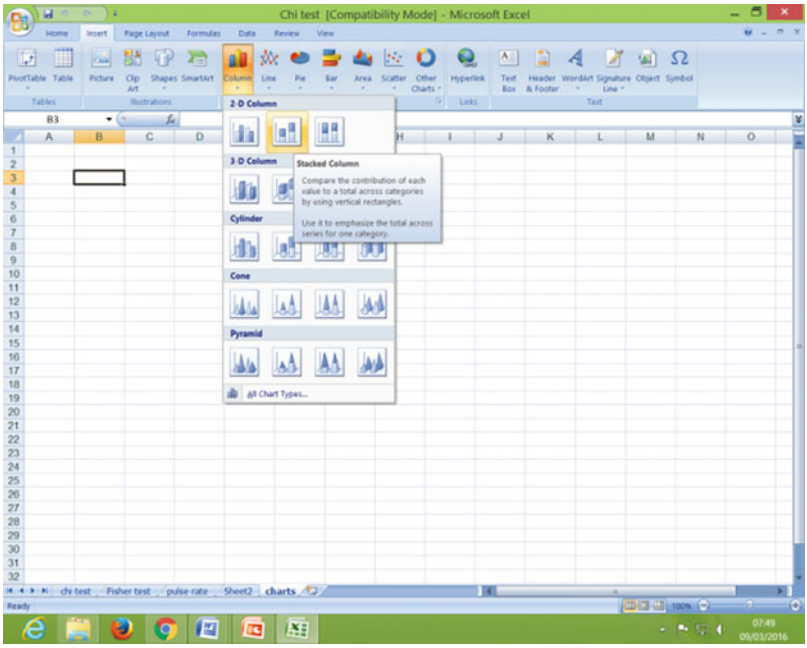
Select the cell containing the data. Select only cells containing data on age group and number of patients. If you select the whole table, you get a meaningless table. Computer has to be shown where the data is, on which graph is to be created.

Bar Chart



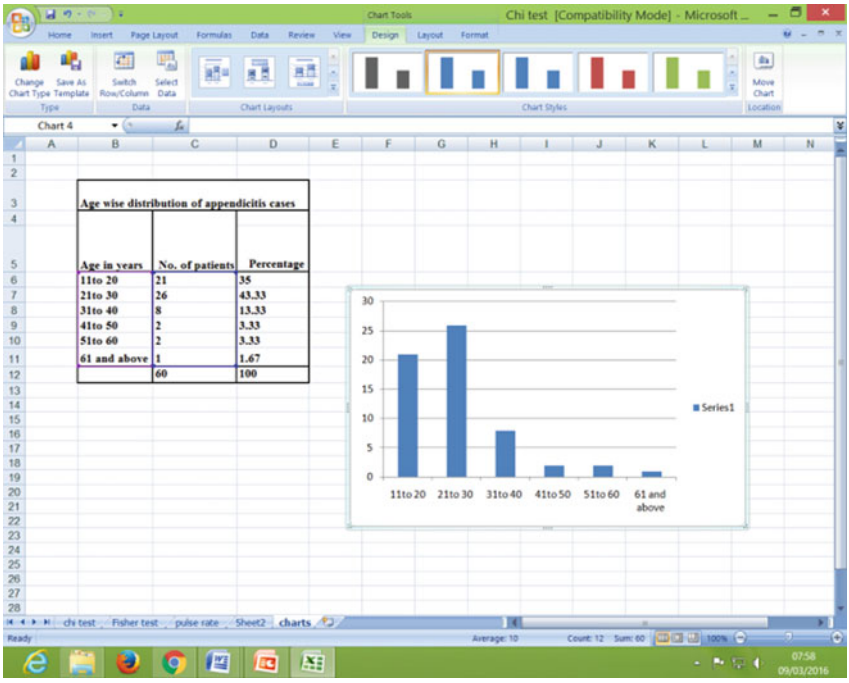
Select only data: Not entire sheet/details

Click on “INSERT”; it shows various options. Click on an appropriate design.



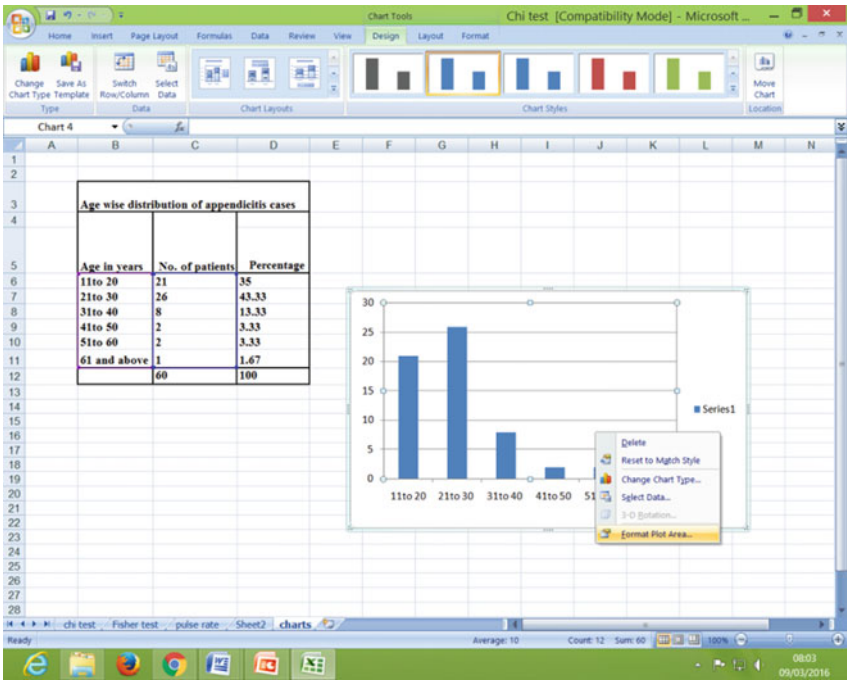
Types of graphs

We have a graph. It is as simple as that. Only thing we should know are where the options are and how to utilize them.



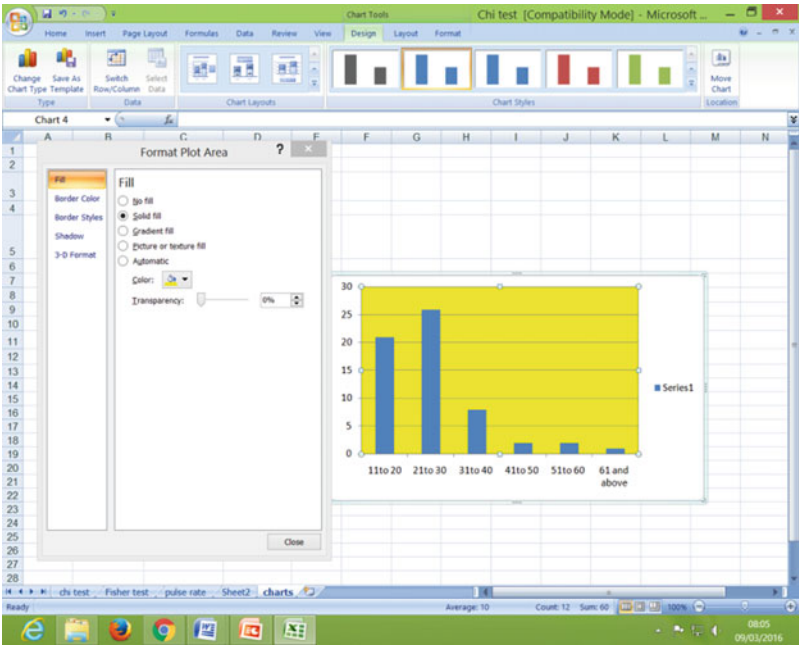
Bar chart

We can make it colorful also. Right click on the plot area (can also do the same on the chart area). You have the option format "plot area."

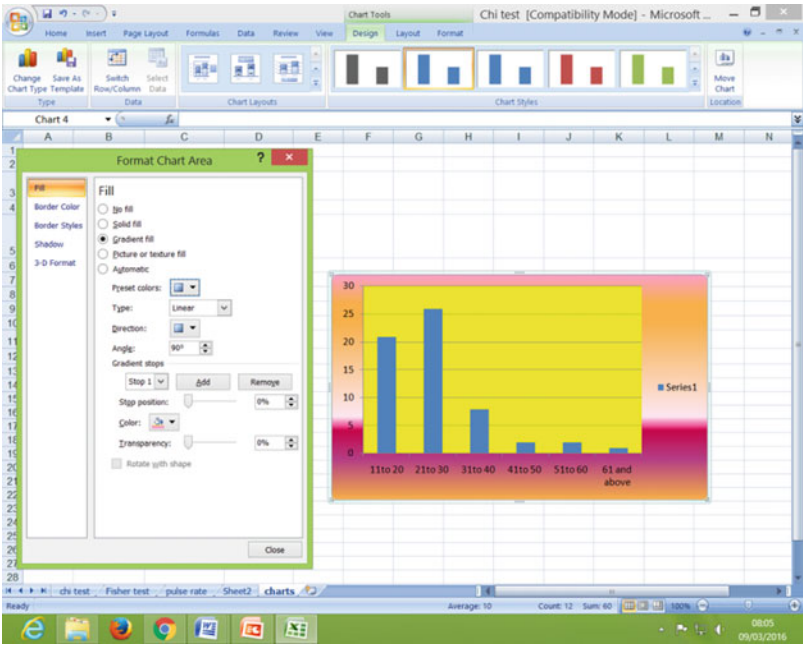


Making the graph colorful and adding details

Click over that we can get various options. Explore various possibilities to make it colorful.



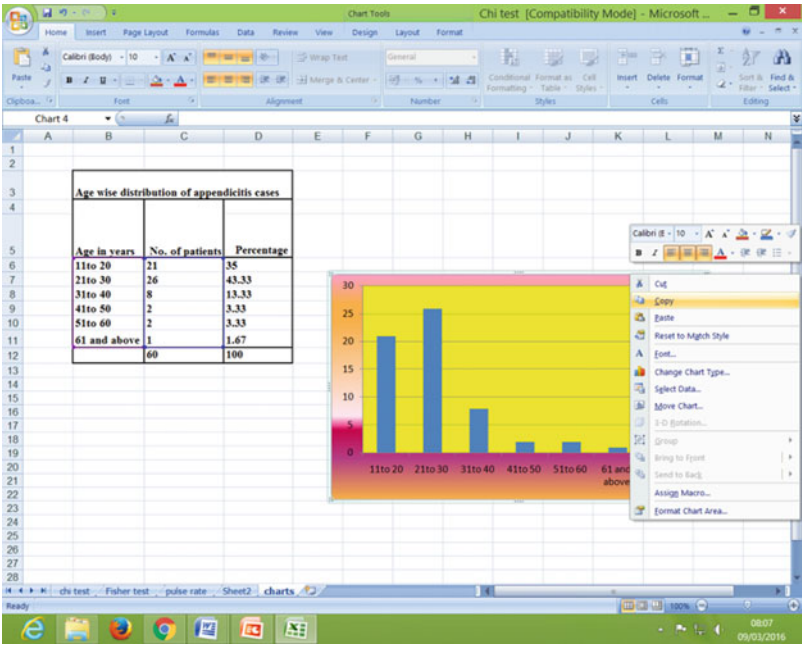
Color selection



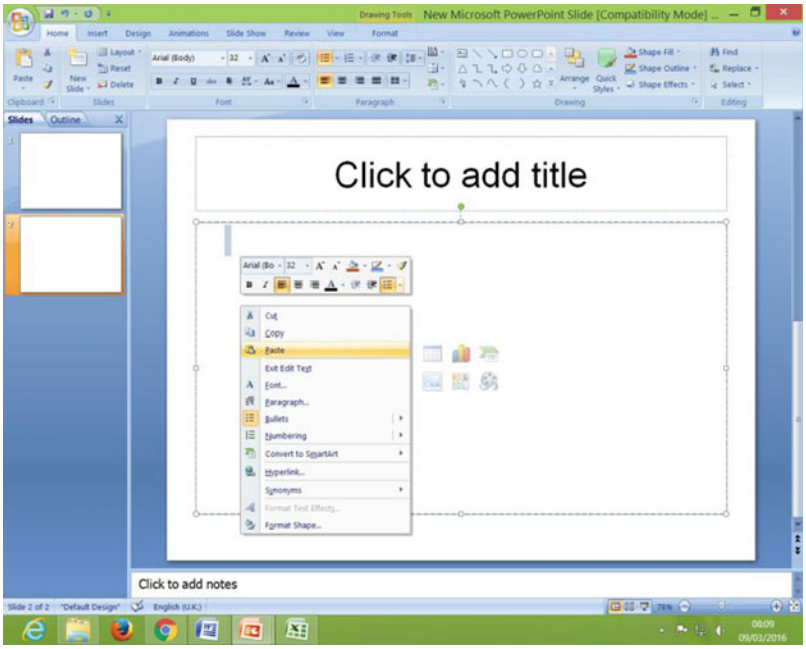
Graph ready

[redacted]

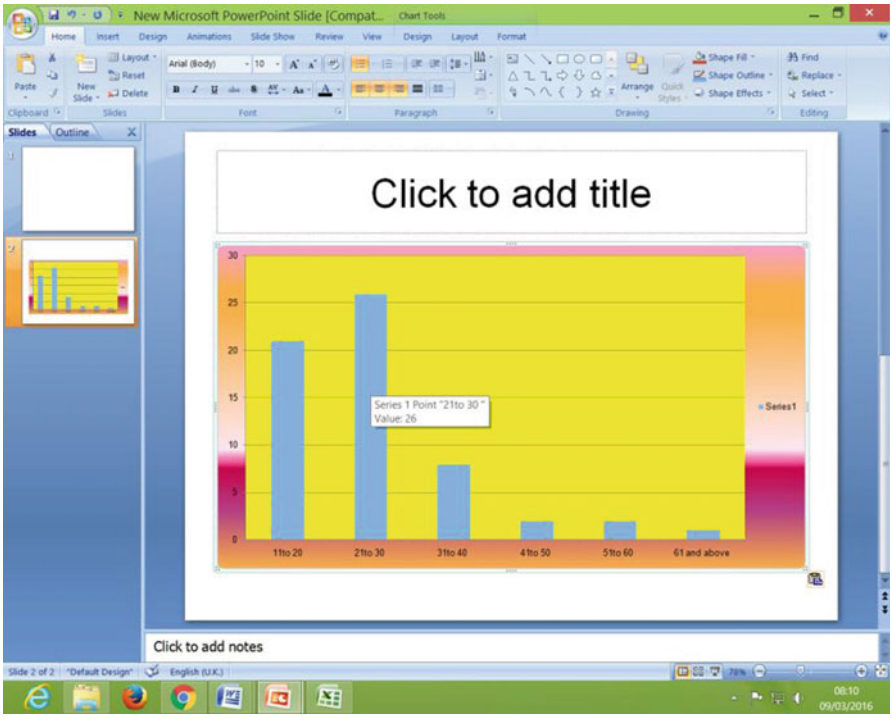
Right click on the border; it shows copy (or Ctrl + C) or cut (or Ctrl + X) option. We can copy the chart and can paste the graph on PowerPoint slide or Word document or any other format by right clicking to select option of paste (or Ctrl + V).



Graph copied



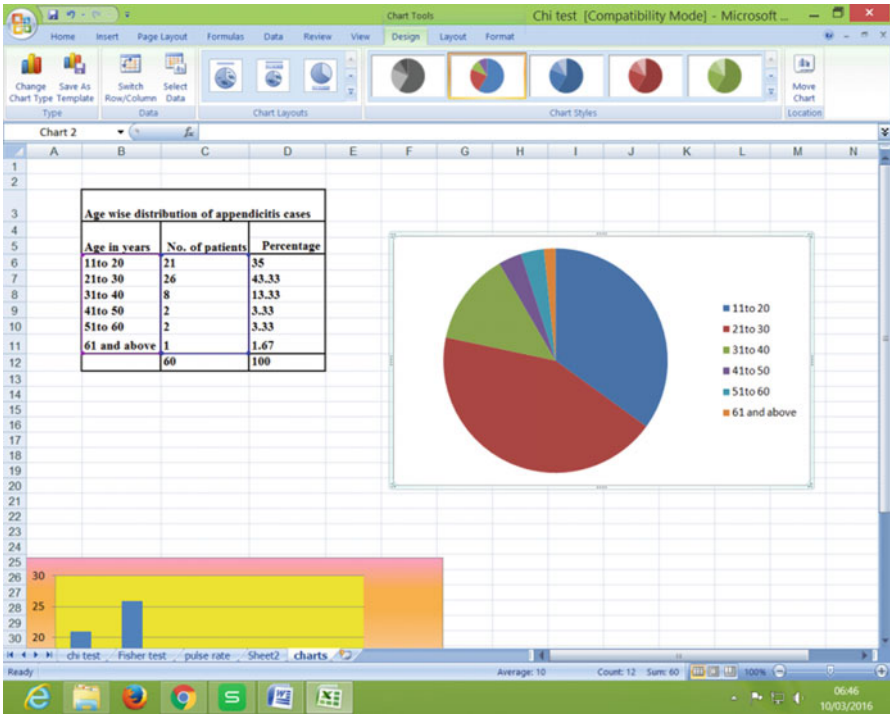
Graph can be copied and pasted on other programs like MS powerpoint, MS word etc...



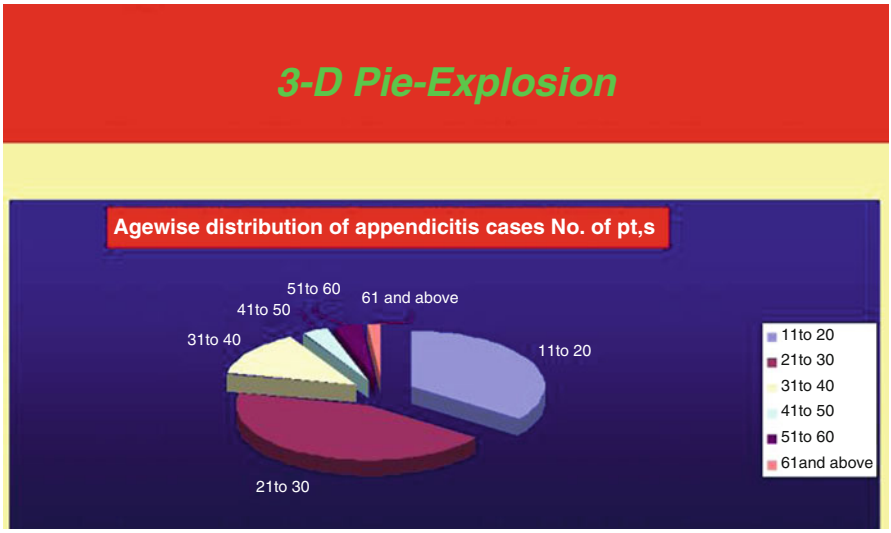
Powerpoint slide with the graph

The same data can also be represented in different ways to make it more informative.

Pie Chart



Same data can also be depicted as a Pie chart



3 D Pie explosion chart

Line Chart

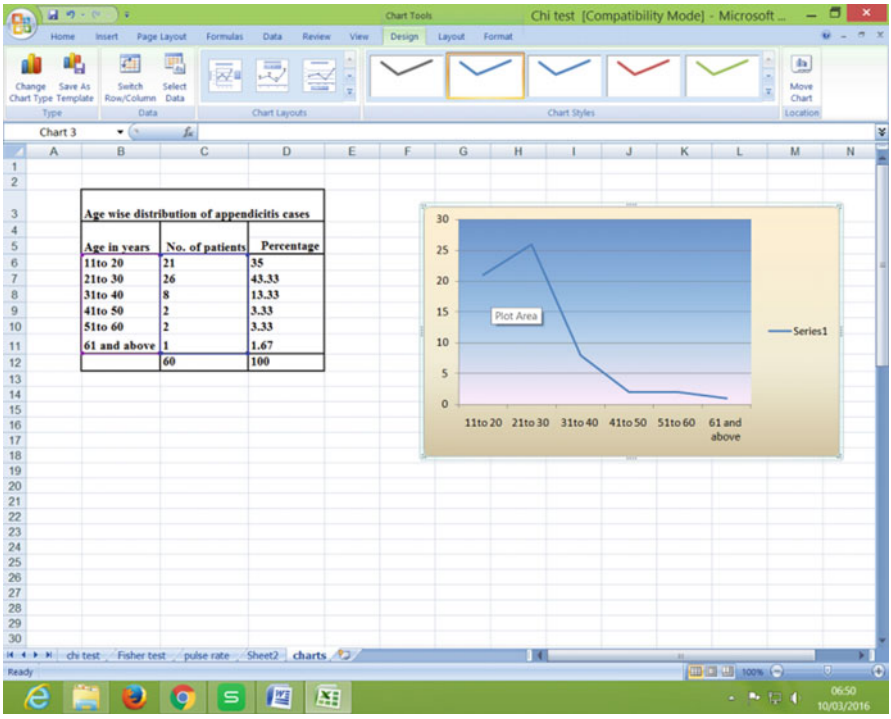
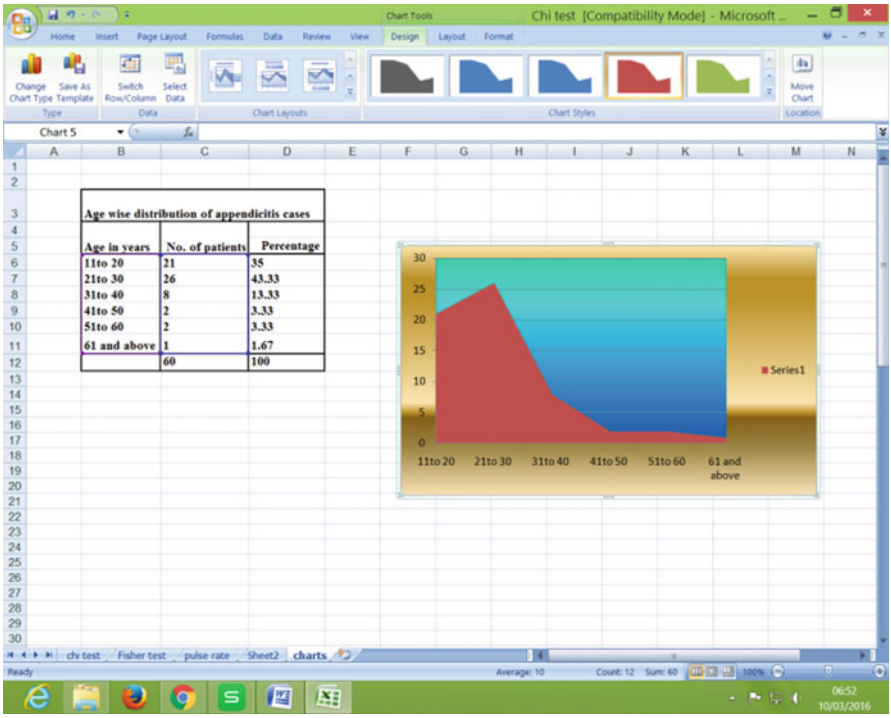
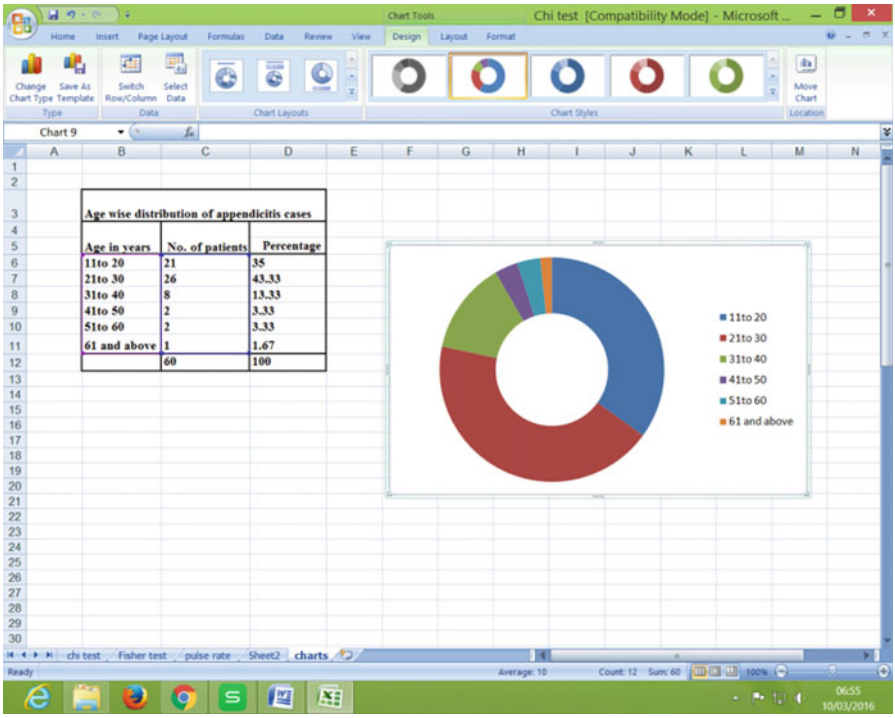


Chart Tools
Chi test [Compatibility Mode] - Microsoft ...
Home
Insert
Page Layout
Formulas
Data
Review
View
Design
Layout
Format
Change Chart Type
Save As Template
Type
Switch Row/Column
Select Data
Data
Chart Layouts
Chart Styles
Move Chart
Location
Chart 3
Age wise distribution of appendicitis cases
Age in years | No. of patients | Percentage
11to 20 | 21 | 35
21to 30 | 26 | 43.33
31to 40 | 8 | 13.33
41to 50 | 2 | 3.33
51to 60 | 2 | 3.33
61 and above | 1 | 1.67
60 | 100
30
25
20
15
10
5
0
Plot Area
Series1
11to 20
21to 30
31to 40
41to 50
51to 60
61 and above
chi test
Fisher test
pulse rate
Sheet2
charts
Ready
100%
06:50
10/03/2016

Area Chart

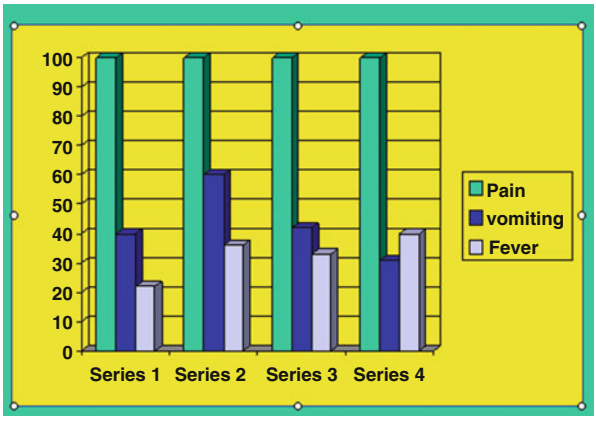

Chi test [Compatibility Mode] - Microsoft ...
Chart Tools
Chart 5
Age wise distribution of appendicitis cases
Age in years | No. of patients | Percentage
11to 20 | 21 | 35
21to 30 | 26 | 43.33
31to 40 | 8 | 13.33
41to 50 | 2 | 3.33
51to 60 | 2 | 3.33
61 and above | 1 | 1.67
60 | 100
30
25
20
15
10
5
0
11to 20
21to 30
31to 40
41to 50
51to 60
61 and above
Series1
chi test
Fisher test
pulse rate
Sheet2
charts

Doughnut Chart



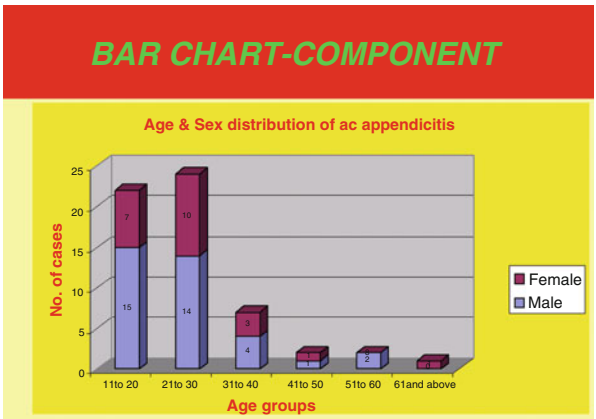
Multiple Bar Charts When two or more factors are to be depicted, multiple bar charts can be used. For example, to show relative incidence of various symptoms of acute appendicitis like pain, vomiting, and fever in different series, bar chart with multiple bars can be used.

Multiple bar charts



Component Bar Chart If two components of a group are to be shown, *component bar chart* can be used. In the following example, the lower component of the bar shows male patients, and the upper components show female patients.

Component bar chart



With good imagination and a little work, almost anything can be created using computers.

Histogram

Numerical data can be represented graphically as a histogram. Usually histogram is used for data containing continuous variables. For categorical data bar chart is used. It gives an idea of density of distribution of the data. Online histogram creators are available at various web sites.

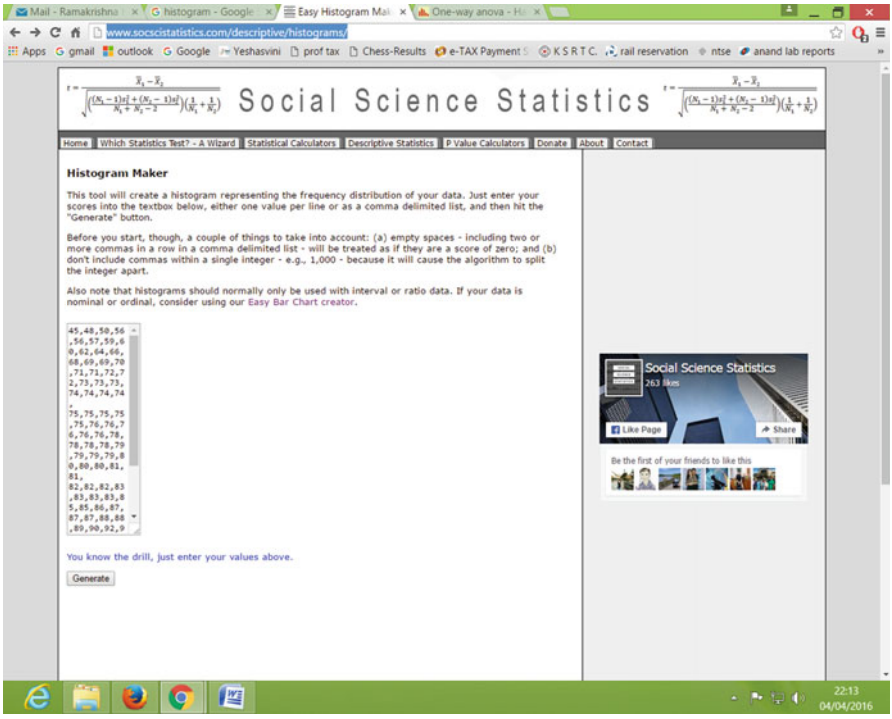
For example, one is on http://www.socscistatistics.com/descriptive/histograms/.

Let us consider the following data (Table 7.3) on pulse rate of healthy men and convert it to a histogram.

Table 7.3 Pulse rate of healthy men

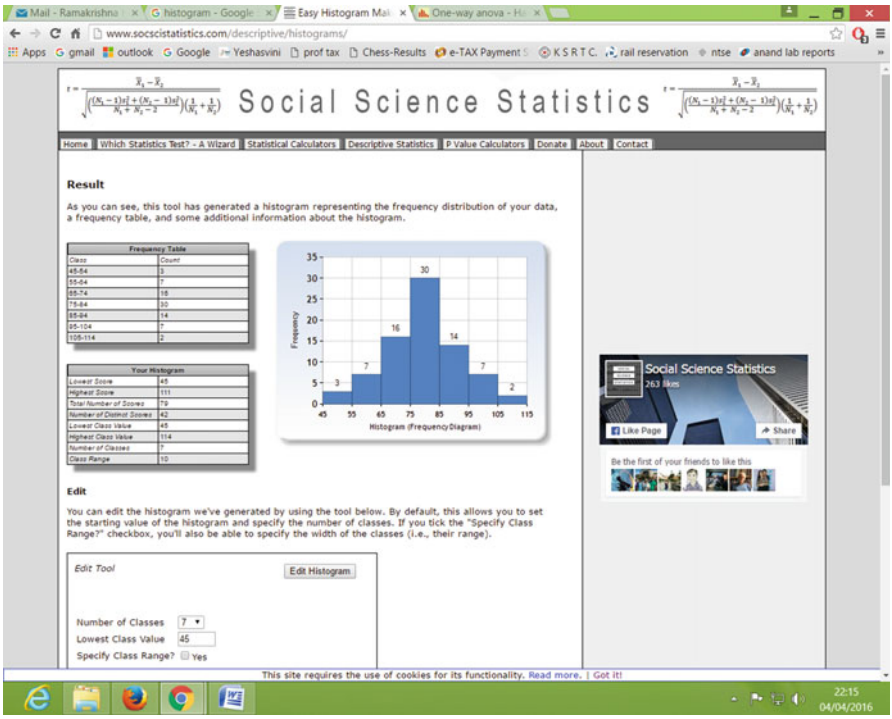
45,48,50,56,56,57,59,60,62,64,66,68,69,69,70,71,71,72,72,73,73,73,74,74,74,74,75,75,75,75,75,76,76,76,76,76,78,78,78,78,79,79,79,79,80,80,80,81,81,82,82,82,83,83,83,83,85,85,86,87,87,87,88,88,89,90,92,93,93,94,96,98,99,100,102,104,104,106,111

The web site mentioned above is opened and data are entered in the box. They can be entered one below the other or continuously separated by a comma.



Creating histogram web based application

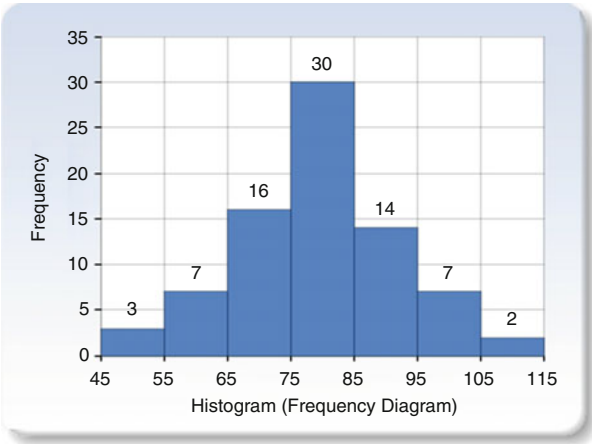
Click on Generate button. It creates a histogram. It can be further edited by following directions given.



Histogram

The graph can be saved in the computer from where it can be copied and pasted to slides, Word document, etc.

Histogram can be copied and pasted on any application



Difference Between Histogram and Bar Chart

A histogram is used to represent continuous variables, and a bar chart is used to represent categorical variables. In the later, there is discontinuity between the data. Each data falls into a category. The height of the bar represents the frequency in that category. In histogram, there is no gap. Variables can take any number of values. The area under the histogram represents the density (rather than the frequency) of data distribution.

Clinical Audit

Clinical audit can be defined as "a quality improvement process that seeks to improve patient care and outcomes through systematic review of care against explicit criteria and the implementation of change."

The word audit in the dictionary means ***to examine***. Clinical audit, any audit for that matter, is a process whereby the structure, process, and outcome are analyzed *critically* and adjustments made to bring about improvement. So, the end result should be to improve outcome for the patient.

Clinical Audit Versus Research

There are some similarities in the methodologies of research and clinical audit. However, they are not the same. Research involves ***acquisition of new knowledge,*** whereas audit aims at ***improvising the existing system***. Audit measures existing system against standards. To put it in simple words, ***research asks what is the right practice, whereas clinical audit asks are we doing the right practice***. Clinical auditing should be transparent. The aim should be only to improve the system and not to name, blame, and shame a particular clinician. Negative incentive results in not reporting an adverse event or hides substandard outcome or a complication or death. This in turn results in repetition of the same mistake in the system again and again.

Key Points

- Audit measures practice against performance.
- The audit cycle involves five stages: preparing for audit, selecting criteria, measuring performance level, making improvements, and sustaining improvements
- Choose audit topics based on high-risk, high-volume, or high-cost problems, on national clinical audits, national service frameworks, or guidelines from the National Institute for Health and Clinical Excellence.
- Derive standards from good-quality guidelines.
- Use action plans to overcome the local barriers to change and identify those responsible for service improvement.
- Repeat the audit to find out whether improvements in care have been implemented after the first audit.

(http://www.bmj.com/content/336/7655/1241)

Clinical audit can be used to scrutinize not only clinical parameters and outcome but also for scrutinizing other administrative parameters. For example, how long does a patient wait before seeing a consultant? Can this time be reduced by appropriate measures? It can be used to assess resources adequacy. For example, has the laparoscopic unit all the instruments and patient monitors for safe and successful conductance of routine laparoscopic surgeries?

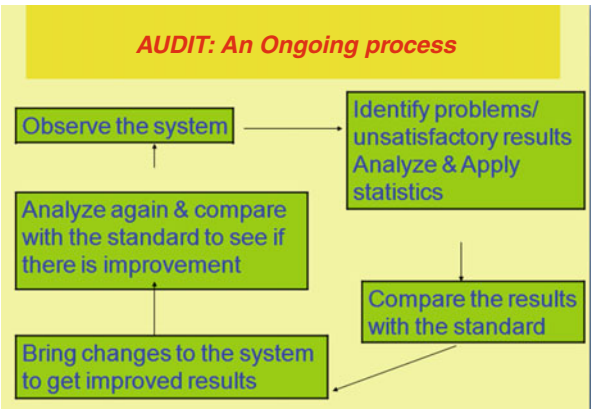
Why Audit?

1. Scrutiny creates responsibility (when there is a sense in the mind that every death is scrutinized by peers, one's mind undoubtedly concentrates on the best possible treatment: this improves the patients' care).
2. Results may be used to assist government to form policies.
3. Methodologies are useful for Post graduate students doing their dissertation.
4. Audit is now mandatory by certain bodies, e.g., RCS and GMC.

Ultimately, clinical audit improves the system.

Clinical Audit Is an Ongoing Process

Clinical auditing should be done continuously. Unlike a trial, which ends as soon as the study period is over, clinical audit is an ongoing process. It is a cycle: (1) Observe the system. (2) Identify problems or unsatisfactory results. (3) Analyze. (4) Apply statistics to compare with the standard. (5) Bring changes to the system to get improved results. (6) Analyze again. (7) Compare with the standard to see if there is improvement. This is known as an **audit cycle**.



Audit cycle

Problems with Audit

Inevitably everybody responsible have to work hard and extra for audit. If the job of the clinician is only to see patients, he can treat more patients. If audit is added, the time spent on the audit has to be taken out from this professional time. In many busy centers, clinicians may not have sufficient time for audit, especially in private practice. Many clinicians are not interested in audit as it not remunerating one. It only brings more responsibility and work without compensation in terms of money. However, in the long run improved quality of patient care and results improves the practice or profession (improved remuneration indirectly). Many clinicians view audit as a threat. They fear to report an adverse event or death or a complication or less than acceptable outcome. They think it will damage their reputation. That is why audit should not be a blame game. It should not be used to name, blame, and shame. Once the clinicians are assured of this, they come forward with more openness and report. They participate in audit actively. This will benefit all concerned in the system. For many small setups and small institutions, resource and finance may be a problem. Money is needed to bring improvements or changes. For example, if it is observed that the lack of monitor or defibrillator is causing many on the table deaths, then the hospital must purchase the equipment. It incurs expenditure. But it is a worthy investment.

Extra work
Lack of time
Lack of interest
Perception as a threat
Resource problem

Methodology

Methods are similar to clinical trials as discussed above. Four essential steps in the surgical audit are:

1. Collection of data.
2. Analysis of data using statistical methods (sometimes compared with results in the literature or standard).
3. Presenting the results with the evidence obtained from above.
4. Drawing conclusions. Framing recommendations for improvement. Then it enters the audit cycle mentioned above.

To summarize:

Clinical trials are intervention studies on human beings designed to test a drug, a device, or a technique and to gain new knowledge about a new or existing treatment.

To design a trial, there should be a hypothesis to test. Detailed protocol has to be written down and it should be followed strictly. It is essential to have accurate

definitions of the parameters and outcome to be measured to avoid ambiguity, observer errors, and bias.

The three essential steps are: (1) collection and compilation of data to create a data base, (2) analyzing the data using statistical methods, and (3) drawing conclusions based on the evidence obtained from above.

We should not ignore possibility of studies thinking that it is small or insufficient or not contributory to the scientific knowledge. We should try to give our best to improve the existing scientific knowledge without being bothered about it is enough or not. The more we do what we know, the more will we know what to do.

Clinical audit is an essential part of clinical practice of every doctor. Every doctor must know the basics of statistical analysis to read and analyze the articles and claims in the journals, all of which may not be entirely true.

Case Study and Simple Example

Example of an **imaginary** design of a trial will help in better understanding of the above concepts.

Suppose a new drug is invented which is claimed to help in passage of ureteric calculus.

Turn this into a question: Does the new drug help the patients to pass ureteric calculus?

So the subjects are patients with ureteric calculus.

Prospective study*:* All cases with a diagnosis of ureteric calculi on ultrasound scanning at XYZ hospital during January 2015 to December 2015 period were studied.

Sample size*:* A total of 268 patients are assigned to two groups randomly. The first group (study group) consisted of 143 patients, and the second group (control group) consisted of 125 patients.

Blinding*:* Each patient was given a code number. Both the patient and the investigator are blinded. Data were collected and recorded by an independent observer.

(So the study design is a **prospective randomized controlled double-blind study.)**

Inclusion Criteria:

1. All cases of ureteric calculi diagnosed by ultrasound scanning
2. Calculus size as measured by ultrasound scanning less than 8 mm
3. Age between 18 and 60 years

Exclusion Criteria:

1. Age less than 18 and more than 60 years
2. Calculus size more than 9 mm
3. Patients allergic to the drug
4. Pregnant and lactating women

Study Protocol: To study the group of patients, the Drug X was given in a dose of 0.4 mg at bedtime once daily along with analgesics. Patients were asked to drink a lot of water. To the control group of patients, a placebo tablet was given similarly (which looked similar to Drug X tablet but without active ingredient) along with the same analgesics. Patients were asked to drink a lot of water. All patients were instructed to watch urine (by means of a filter) for passage of calculus for a period of 1-month time.

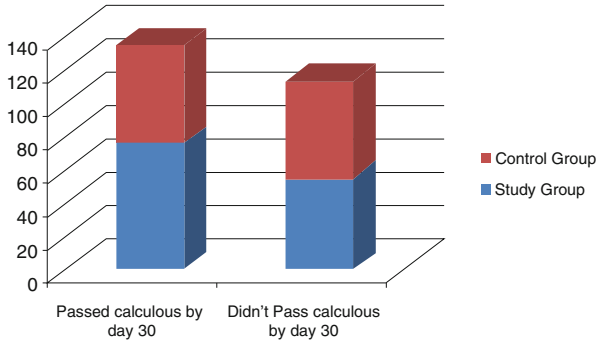
Follow-Up: Patients were followed up for a period of 30 days. In the study group, two patients were excluded from the study as they developed adverse effects to the drug. Another ten patients were lost for follow-up. In the control group, seven patients were lost for follow-up and one patient developed severe pain and was operated. These patients were excluded from the study. So, 131 patients in study group and 117 patients in control group (total of 248) were available for final analysis.

Data on these patients were collected.

Patient 1: Study group: Passed calculus on day 2
Patient 2: Study group: Didn't pass calculus till day 30
Patient 3: Control group: Passed calculus on day 4
Patient 4: Study group: Didn't pass calculus till day 30
.
.
.
Patient 248: Control group: Didn't pass calculus till day 30

Presentation of Data:

Passage of calculus by day 30



A table of summary was created.

Table showing passage of calculus by day 30

	Passed calculus by day 30	Didn't pass calculus by day 30	Total
Study group	76	55	131
Control group	59	58	117
Total	135	113	248

Analysis:

Chi test is applied to see if the difference in the result is statistically significant. Significance level is set to $P < 0.05$.

Chi-test result: $P = 0.2318$

Discussion: Ureteric calculi are known to pass out naturally when observed over a period of time. The Drug X claimed to increase this frequency of passage. In the study xxxx, the authors have reported frequency of passage of the calculus increases by 1.8 fold. In our study, we found that there is no statistically significant difference in the rate of passage of ureteric calculus between control and study groups.

Conclusion:

The drug does not have beneficial effect on passage of ureteric calculus of less than 8 mm size in the age group of 18–60 years.

(Since in the inclusion criteria, we fixed age group as 18–60 years and the size of the calculus as 8 mm or less, study results cannot be generalized to all age populations or all sizes of calculi. A single conclusion backed by adequate data is stronger than too many conclusions with inadequate or no data in the main text.)

How to write a model paper by using this data will be discussed in the Chap. 8, *Writing an Article for Journal.*

Mass Screening:

Mass screening is a method to diagnose a disease in the population of "normal" persons by a test or clinical examination. It is a part of public health campaign. Population consists of a large number of asymptomatic persons in whom there are no symptoms or signs of the disease. The aim is to diagnose a disease at an early stage so that it is better treated or its spread prevented. Detection of cancers at an early stage has better cure rates with treatment. If infective diseases are diagnosed early, its control is better and also measures to prevent the spread can be taken.

Though theoretically it appears to be beneficial, not all types of screening are beneficial. There are side effects or problems with screening. ***Overdiagnosis*** of the disease is possible. For example, if mass ECG is done to all population in an attempt to diagnose cardiac disease at an early stage, even minor clinically insignificant ECG changes may be inferred as "cardiac disease." It can produce a lot of stress in the "patients" resulting in cardiac neurosis. ***Underdiagnosis or missed diagnosis*** is the other end of the problem. If the diagnostic tool under consideration fails to identify or diagnose the disease and brands the patient as normal, both patient and the doctor fall into false sense of security. Required treatment may be withheld with its attendant problems.

So the tool used in the mass screening must have a high sensitivity and a reasonable level of specificity. High sensitivity is more important since once the disease is detected, other modality of diagnosis may be employed to confirm the diagnosis. For example, card test to detect HIV infection should have a very high sensitivity. It should not miss any patient with HIV infection. Even if some of the negative cases are diagnosed as positive (false positive, i.e., lesser level of specificity), it is acceptable. Because, these cases can be confirmed by a more specific test such as Western blot or some other tests which are highly specific.

Then why can't these highly specific tests be used in the first instance? Reasons may be multiple: they may be costly, time-consuming, technically difficult, less sensitive (higher false-negative rates), etc. So they are not suitable for mass screening.

Mass screening may be ***universal screening*** or ***screening of at-risk population (case finding)***. In universal screening all individuals in a category are screened. For example, all women more than 35 years are screened for carcinoma breast by mammogram. In at-risk screening, only those individuals at higher risk of developing a disease are screened. For example, women with family history of breast cancer (mother or sister) only are screened.

In **multiphasic screening** multiple diseases are screened by multiple tests simultaneously. For example, anemia, protein-energy malnutrition, and nyctalopia are screened simultaneously in schoolchildren.

Mass screening is not for diagnosing and starting a treatment. The aim is to identify those individuals who need further diagnostic tests. Most of the times, other diagnostic modalities are required to confirm the diagnosis before starting the treatment. For example, glucometer testing of random blood glucose level or urine sugar (cord method) is used as screening for diabetes mellitus. It is because they are quick, easy, and cheap. Test can be easily done by a nurse. But patients are not diagnosed as diabetic neither the treatment is started. Further testing is done in the form of fasting blood sugar and postprandial blood sugar to confirm the diagnosis. If these tests are normal, patients are just reassured that they do not have the disease and no treatment is started.

The initial *Wilson's criteria for screening published in 1968 by WHO is modified in 2008 and is available on the web site* http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2647421.

Synthesis of emerging screening criteria proposed over the past 40 years

- *The screening program should respond to a recognized need.*
- *The objectives of screening should be defined at the outset.*
- *There should be a defined target population.*
- *There should be scientific evidence of screening program effectiveness.*
- *The program should integrate education, testing, clinical services, and program management.*
- *There should be quality assurance, with mechanisms to minimize potential risks of screening.*
- *The program should ensure informed choice, confidentiality, and respect for autonomy.*
- *The program should promote equity and access to screening for the entire target population.*
- *Program evaluation should be planned from the outset.*
- *The overall benefits of screening should outweigh the harm.*

Common screening examples are screening for breast cancer (mammography), cervical cancer (PAP smear), colonic cancer (colonoscopy), etc. Screening is also done for some of the infections like tuberculosis (PPD), some metabolic disorders, newborn for hearing loss, diabetic retinopathy, etc.

Schoolchildren are screened for defective vision, dental caries, congenital heart diseases, etc.

Writing an Article for Journals

8

Not all who look at a journal are going to read even one article in it;
Writers must know therefore what turns a looker in to a reader
– J. W. Howie.

Learning Objectives
Highlight the importance of writing an article.
Discuss the qualities of a good article.
Qualities of a journal.
Know what a journal expects to publish an article.
Editorial process of publishing.
Flaws of too much pressure to publish.
Sources of information and how to search on the Internet.
Types of article.
Headings under which an article should be written.
A few examples of writing.

In previous chapters we have discussed how to design simple study for the purpose of research. Postgraduate dissertation is an attempt to teach the medical graduate the principles of research. Medical, or for that matter any scientific field, develops because of research. ***Research is complete only when the findings are made public through publications***. Other peers and the rest of the world should know the newer developments, and the benefits should be passed on to the general population at large.

To publish we need to write an article in a standard form which will be acceptable to the journals. We should present our data and conclusions of our study in a systematic way. Random writing results in rejection by the journals.

As stressed in the previous chapter, research starts with recognizing a problem. We begin with a question. On the basis of the question, fix aims and objectives. Then formulate a study protocol. Collect data, compile. Collect information from the literature. Compare. Draw conclusions. Now write everything in the form of a paper. These aspects are discussed in the previous chapter on Designing a Study.

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_8

Before beginning to write an article, read many articles already published, especially from the journal to which you want to send your article. This will give you an idea how to begin and what style or format the journal expects.

Some of the journals are available only online. Some journals are free to access while others are not accessible for free. They need subscription or the reader has to pay for accessing each article. These journals do not charge the authors. Open-access journals are free for the reader but charge the authors to publish. Charges vary. Authors have to visit the website or contact the journals' administration department beforehand to get all these informations.

Clear Message

There should be a clear message in the article worthy of reporting. Just writing for writing's sake or me too sake is not worth it. Simply repeating a message already published or known to the world does not merit acceptance. Laparoscopic cholecystectomy is cost-effective, decreases hospital stay, etc., and is well known to the world. An article on this topic would have been accepted many decades ago but not now. Similarly if you describe appendicectomy operation, nobody accepts it. These are all too well known. The study should have a clear question which is relevant to the current practice or has some controversy at present, and the study must give a clear answer to the question in the article. The topic must be new. Innovations attract journals and the chances of acceptance are high. If the findings in the article have something which can change the way we are practicing today, it merits acceptance. For example, the way we are managing a particular condition at present may be expensive. In one study authors used a very cheap nylon mosquito net mesh in the inguinal hernia repair instead of a current trend of expensive polypropylene mesh. The mesh hardly costs INR 15/–20/ compared to INR 1500/–2500. The article was accepted in the *Indian Journal of Surgery.* (Read Ravindranath R. Tongaonkar, Brahma V. Reddy, Virendra K. Mehta, Ningthoujam Somorjit Singh, Sanjay Shivade: Preliminary Multicentric Trial of Cheap Indigenous Mosquito-Net Cloth for Tension-free Hernia Repair: *Indian Journal of Surgery*, *Vol. 65*, *No. 1*, *Jan.-Feb. 2003*, *pp. 89–95.*) The study offered some cheaper alternative to the current practice, and it is useful to the public. Hence, it was accepted for the publication. Similarly, if newer alternative has better safety profile than existing method, it is also useful.

Rare and interesting cases are also published. A difficult case managed with restricted resources gives new ideas (e.g., read H. K. Ramakrishna. A Difficult Case of Acute Intestinal Obstruction Managed in a Rural Set-up. *Indian Journal of Surgery*, *Vol. 65*, *No. 1*, *Jan.-Feb. 2003*, *pp. 104–105*). A rare presentation of a common case is sometimes interesting.

Review articles are written after collecting information from a number of published reports. Each article may have only one or a few cases. When many articles are collected, we get varied presentation and management methods for a condition. Comprehensive information on a condition can be given. For example, the

following article gives an idea how a review article is written. (Read Ramakrishna, H. K. Intestinal duplication. *Indian Journal of Surgery* 70.6 (2008): 270–273.)

Meta-analysis is done by collecting data on a particular subject and analyzing the data to give new conclusions. Each published article gives small number of cases. When data from a number of articles are collected, we get more number of cases. So the conclusions will be more reliable. Also we get different opinions by different authors to compare. A meta-analysis of published reports on intraperitoneal use of polypropylene mesh is an example for this type of study (read HK Ramakrishna, K Lakshman. Intra peritoneal polypropylene mesh and newer meshes in Ventral Hernia Repair: What EBM Says? *Indian Journal of Surgery* 75 (5), 346–351).

Reading a number of similar types of articles published in various journals with message similar to the one in the study to be published helps to get an idea as to how to write the present article. The above examples serve this purpose.

Before writing an article, author should read instruction to author given in the journal web site. For example, http://bmjopen.bmj.com/site/about/guidelines.xhtml describes the guideline for submitting an article to BMJ. More or less similar guidelines are there for other journals as well. Still it is better to read the guidelines of the journal to which the author proposes to send the article.

Type of the Articles

Editorials
Original articles
Review articles
Case reports
How I do it?
Surgical techniques and innovations
Letter to editors
Images
Commentary
Etc.

Different journal may have different list. The authors should find an appropriate journal which suits their article. Some journals may not accept any case reports. Author has to find out whether his article type will be accepted by the journal he/she is considering. Otherwise it will lead to waste of time as after a period of time, the article comes back rejected.

Some important considerations (*many of which can be found in the instructions to authors or elsewhere on the journal web site*) *that can help guide the journal search* (and find an appropriate journal for the article) *include* (*for more details, visit http://www.tandfonline.com/doi/full/10.1185/03007995.2010.499344#*):

- *Rejection rate* (*which varies widely across journals*) (the number of articles rejected per 100 articles received by the journal)
- *Indexing* (*e.g.*, *through MEDLINE, etc.*)
- *Time to acceptance*; *time to publication*
- *Impact Factor*™ (*a measure of how frequently articles from a journal are cited*), e.g., the impact factor of *Indian Journal of Surgery* for 2014 is 0.353
- *Article length restrictions*
- *Types of articles typically published*
- *Acceptance of industry sponsorship*
- *Acceptance of acknowledged medical writing assistance*
- *Receptivity to pre-submission contact*
- *Opportunity to accept correspondence/feedback from readers*
- *Charges for pages*, *publication*, *color figures, or open access*
- *Expedited peer-review or publication services*

Quality Indicators of a Journal

Indexing

Indexing of a journal is considered to be a reflection of the quality of the journal. It means the journal is indexed in an indexing service. Index Medicus is one of the oldest and prominent indexing services. If a journal name finds its place in the index of Index Medicus, the journal is said to be an *indexed journal*. Now there are many other indexing agencies. For example, MEDLINE, PubMed, EMBASE, SCOPUS, EBSCO Publishing's Electronic Databases, SCIRUS, etc. (Yatan Pal Singh Balhara. Indexed journal: What does it mean? Lung India. 2012 Apr-Jun; 29(2): 193). It is debatable whether the quality of the articles in the indexed journals is better than the articles published in the non-indexed journals.

Impact Factor

The impact factor of a journal is the average frequency with which an article in the journal has been cited in a particular year. It is considered as a relative importance of the journal within the field. Roughly, it is the number of times indexed journals cite the articles published in the journal under consideration divided by number of citable articles published in that journal in a particular period of time (1 year). A detailed explanation has been given in Wikipedia regarding how the impact factor is calculated (https://en.wikipedia.org/wiki/Impact_factor). There are various websites which show the impact factors of various journals. Impact factor changes every year. The present impact factor of a journal is on the basis of previous year's data on the number of citations. For example, the impact factor for a particular journal for the year 2015 is published in the year 2016. It refers to articles of that journals cited by other indexed journals in the years 2013 and 2014 (2-year average). It is generally

considered that the higher the impact factor, the better is the quality of papers of that journal (because the articles are quoted more number of time by other authors).

But this may not be always true. There are criticisms for usage of impact factor as a measure of quality of a journal. Journals may take some measures with which impact factors can be boosted. For example, journals may publish more of review articles which have higher chances of getting cited and decline to publish case reports which are less likely to be cited. The journal may publish an article or an editorial citing its own articles. This increases its impact factor. Another factor that can influence impact factor is that the prospective authors may get influenced by a higher impact factor of a journal and have a tendency to cite those journals only in their articles. This tendency leads to a higher impact factor in the next year for those journals which have already higher impact factors (a positive vicious cycle).

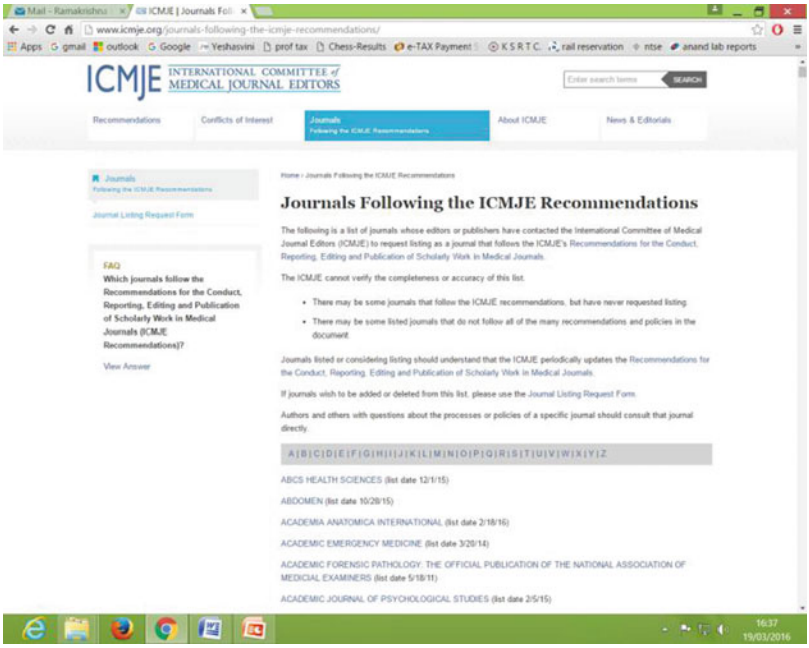
So the verdict is that impact factor, though *useful as a measure*, *is not an absolute indicator to the quality of a journal*.

Uniform Requirement for Manuscript Submission

Here it is important to mention that many journals follow "uniform requirements for manuscripts submitted to biomedical journals" laid down by the International committee of Medical Journal Editors (ICMJE), which is periodically revised by the committee.

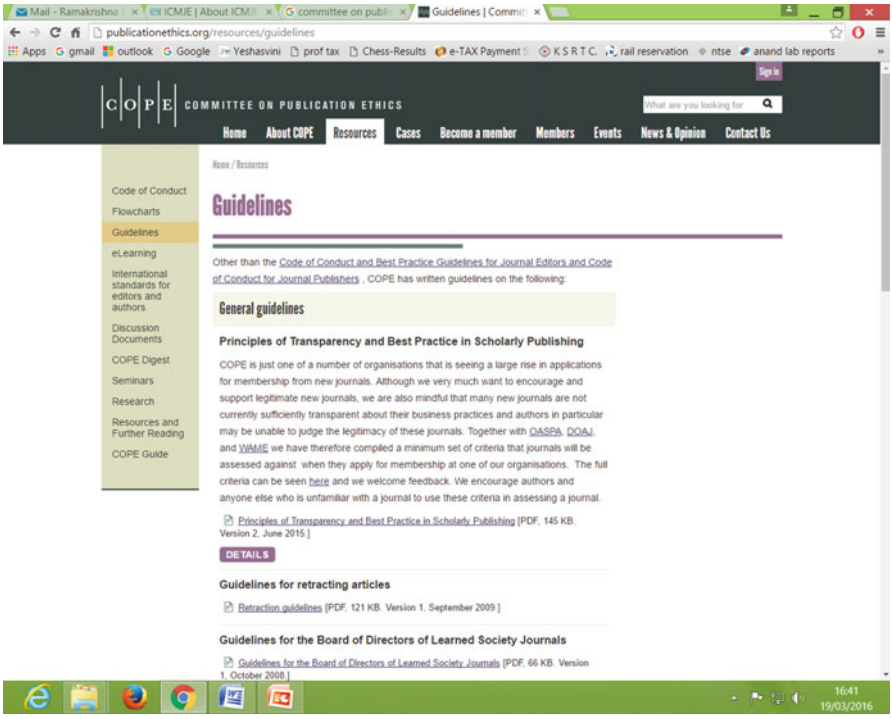
Uniform requirement for manuscript submission (http://www.icmje.org/about-icmje/faqs/icmje-recommendations/)

A list of journals which follow these recommendations can also be found in the web site.



List of journals following ICMJE recommendations (http://www.icmje.org/journals-following-the-icmje-recommendations/)

There are guidelines in the best practices in publishing available on the web site, for example, the Committee on Publishing Ethics (COPE).



Best practices in publishing guidelines (http://publicationethics.org/resources/guidelines)

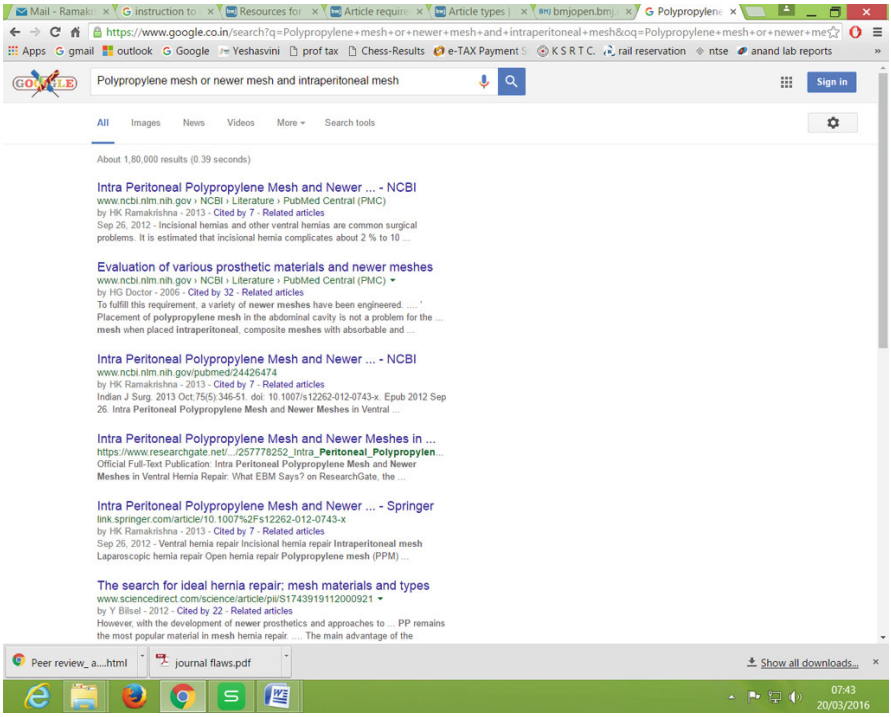
Headings in the Article

Title: It should be appropriate for the main article. It should reflect exactly what the article is about. It should denote what the reader can expect in the article. Avoid sensational words to sound more attractive like "world's first case of…," "a very rare case of…," "a 75 Kg tumor of…," etc. Title should be short and apt. A separate page should be dedicated for the title while submitting the article. Titles should not declare the conclusions or results of the study.

Authors: In a separate page, mention the names of all the authors and their details like their designation, the institute where they work, and their contact details (phone numbers, e-mail ID, etc.). Ensure that each author has contributed for the study. Names should not be mentioned simply because he/she is the department head or some prominent person. There are guidelines to be eligible as an author (http://www.bmj.com/about-bmj/resources-authors/article-submission/authorship-contributorship). Follow them. Each author will get credit and at the same time accountable for the part of the work he/she has done. The main author should know exactly the contributions of each of the coauthors and must have confidence in the integrity of all the coauthors. Out of them only one author should be mentioned as corresponding author. His name and contact details will appear in the journal for correspondence, if anybody wants to contact for further details on the study. A list

of contributors should be provided. These contributors may or may not be the authors.

Key Words: Write some key words for the article. They help in searching the article for the readers online after they are on publication. For example, if information on intraperitoneal use of polypropylene mesh or tissue-separating mesh (newer mesh) is needed, then type the key words "polypropylene mesh or newer mesh and intraperitoneal mesh" and search. Search results show this.



Search by Key words

This is because while writing article the author has given those "key words." The article will get a priority in a showup when those key words are used in the search.

Abstract: Though it comes first in the article, preferably it is best written in the end. Because, at that time only author knows what are the salient feature of the article. One or two paragraphs are written in a very brief and concise way to convey what the reader can expect in the article. Use about 200–250 words. It should not be too lengthy. More details can come in the introduction and discussion. It should create interest in the reader to read further. Mention aims and objectives of the study, current problem/knowledge, why the article is relevant, furnish salient data,

conclusion, remarks, etc. Do not use any undefined abbreviations. If any abbreviations are used, expanded form should be mentioned on the first usage.

It should touch upon the **Objective** of the study, **Design** of the study, **Setting** (where and under what settings the study is done), **How many patients** are involved, What is the **Intervention** (e.g., primary closure of the bladder instead of a conventional two-stage process , etc.), briefly the **Protocol**, **Main results**, **Conclusion/**s, how and under what **Circumstances** the results of the study is useful, how it can **Impact/**change the current practice, etc. These are only guidelines. All these heads may not apply in all cases.

Introduction: The idea is to get readers' attention. If this part is not written well, the reader will not bother to read the entire article. So author should pay sufficient attention to write a proper introduction. A brief description of the current knowledge, what are we missing in the current knowledge, what are the controversies, why the present study is important, how the findings of the study can affect the current knowledge/practice, etc. are to be written. Long historical reviews can be boring. The statements should be backed up by references of published articles/books. Quote statistics or published data with references (references should be mentioned at the end of the article). Never quote somebody else's data or statement as your own. Always give due credit to the original author. It is better to use inverted commas and italics while quoting other authors or books. And mention the reference of the article from where it is taken. Do not use "*sensational*" words what newspapers use. They do not sound good in scientific journals.

Materials and Methods: Under this heading everything about the material should be described in detail. Consider these factors (not necessarily all):

(a) Study period, for example, "all patients admitted from January 01, 2014, to January 01, 2016, are included in the study."
(b) Institution' name.
(c) Patients' description, for example, "all patients having ventral hernia," "all patients presenting with an ulcer in the foot," etc.
(d) How patients are grouped: randomly, by choice, as per patients' wish, etc.
(e) What type of a study: prospective or retrospective, open/single blind/double blind, case control, observational, case report, etc. (refer to Chapter 6 on Designing study for various types of study).
(f) Inclusion and exclusion criteria used to recruit the patients,for example, "all patients in the age group of 10–40 years are included," "patients with risk grade ASA 3 are excluded from the study," "patients allergic to ceftriaxone are excluded from the study" (if the study is on the effectivity of the said drug), etc.
(g) Is consent of the patients taken?
(h) Is ethical committee approval taken?
(i) Protocol: Describe in detail procedure followed. How the patients' data are recorded, the procedure, its technique (if it is operative procedure), detailed account of the intervention applied, etc.

(j) Write strict definition of the terms used for outcome measurement. Recording of the data should follow this definition strictly. For example, if recurrence after hernioplasty is the outcome, it must be mentioned how recurrence is assessed: whether only clinically or any imaging modality is used. If wound infection is the outcome measured, whether infection is documented by culture reports. If relief of pain is the outcome, what is the method used to assess the pain, etc.
(k) For case reports, this section does not apply. Instead, the description of the case with clinical presentation is explained.
(l) For review articles and meta-analysis, explain how the articles and other information are collected. Reading the examples given under "**clear message**" above gives an idea regarding how to write these paragraphs.

Results: Present the data obtained from the study in the form of tables, graphs, etc. This is discussed in Chapter 6 on Designing the Study. Then analyze the results. For case-control studies, apply tests of significance or other statistical methods to draw inferences. For review articles or meta-analysis, comparing with the data obtained from various sources of literature is done similarly. These are discussed in appropriate previous chapters. The presentation should be simple and the reader should be able to understand immediately. Do not describe the data presented in the table again in the text. Avoid presenting the same data again in another form. For example, once the data is presented in the form of a table and again in the form of graph or description in the text is to be avoided.

Results should state only facts and not opinions. They are mentioned in the discussion part.

Negative results, adverse effects, or unexpected results also should be reported.

Discussion: This section is for the analysis of the results. Any explanations for the results obtained or opinions on the results are to be mentioned. The study results are compared with the data from the available literature. If the study data defers, discuss possible factors for the difference. All the statements should be supported by the reference:

The implications of the results are to be stressed.
The limitations, if any, are to be mentioned.
What is the new finding in the study also needs to be stressed.
If study conclusions need to be confirmed by further studies, mention it.
Do not make any statements which are not supported by the data presented.

Conclusions: Usually conclusion/s is/are repetition of the aims and objectives. At the end of the study, the author should prove that the aims and objectives are fulfilled.

Conclusions should be logical outcome of the study design and the results obtained.

Avoid drawing too many conclusions. A single (or a few) conclusion supported with strong data is better than many conclusions without adequate data.

Do not give any conclusions for which there is no data in the main text.

Conclusions should not go beyond the scope of the study.

Acknowledgments: Acknowledge all people who have helped in preparing the article directly or indirectly. The people who do not meet the criteria of authors are mentioned under this section.

Disclosures: Competing interest, sponsorship or funding, financial/other relationship, etc. is to be declared at the end of the article.

References: It is a myth that the more the number of references, the better the chances of acceptance. Actually, properly quoted small number of references is more impressive.

Do not mention any reference unless its contents are utilized in the article.

Writing references in a proper style is very important.

Follow either Vancouver style or Harvard style* depending upon the journal's (to which the article is going to be submitted) requirement.

*Styles: The most common styles of references are the alphabetical (Harvard) and the Vancouver system.

In the *Harvard system*, a reference is to be mentioned in the following format: (name of the author, year of publication).

For example, "These cysts can occur anywhere along the alimentary tract from the mouth to the anus, although the ileum is the most frequently involved region (35 %) followed by the esophagus (19 %), jejunum (10 %), stomach (9 %), and colon (7 %) (Ramakrishna HK, 2008)."

In the end while writing the list of reference, reference is written in the alphabetical order of the authors' names.

Most medical journals use the *Vancouver system*. In the Vancouver system, the references quoted by serial numbers as superscript. The first reference is numbered as 1. The subsequent references are numbered as 2, 3, 4, etc. If the same reference comes at a later stage, it is quoted with the original allotted number. It is not given a new number.

For example, "These cysts can occur anywhere along the alimentary tract from the mouth to the anus, although the ileum is the most frequently involved region (35 %) followed by the esophagus (19 %), jejunum (10 %), stomach (9 %), and colon (7 %).[15]"

In the electronic form, these superscripts are hyperlinked. When clicked over the number, it takes us to the reference section to show the particular reference.

While writing references list, they are numbered in the order in which they appear in the article. So the article which is quoted first is mentioned first and so on.

There is a particular format to write the article quoted, e.g., “Ramakrishna HK. Intestinal duplication. Indian J Surg. 2008;70:270–273.”

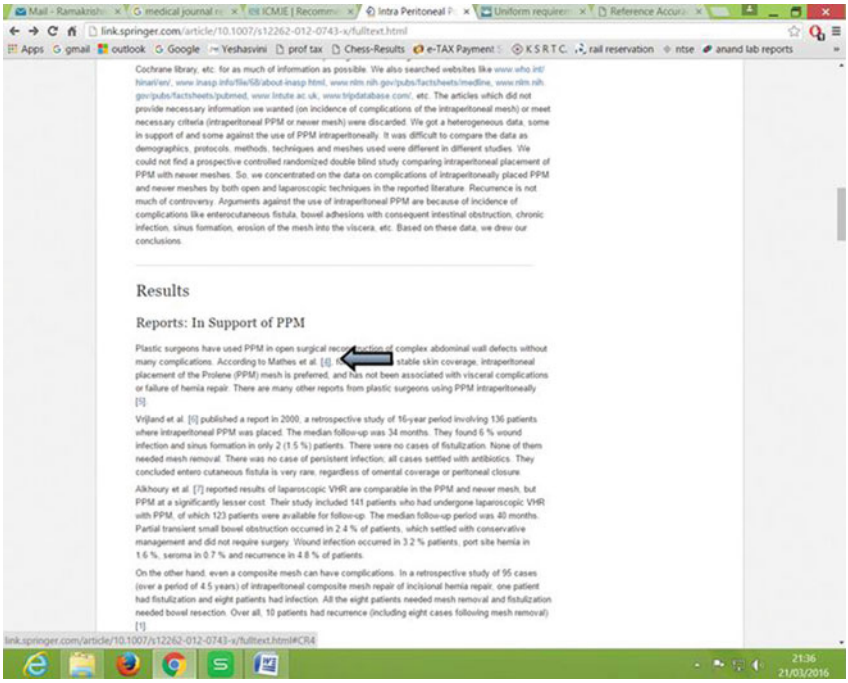
Observe the style of quoting. First author’s name is mentioned. Then the title of the article is mentioned. They are separated by a dot. Then journals abbreviated form is mentioned followed by year of publication. Then the sign, “;” then volume number, then the sign “;”, then page number range. The same pattern is to be followed for all articles while listing the references.

If there are only two or three authors, all the authors’ names can be mentioned. Sometimes there will be many authors and it is lengthy to mention all the names. In such situations only first author’s name is mentioned followed by et.al.

Abbreviated forms of journal names can be found on the appropriate websites.

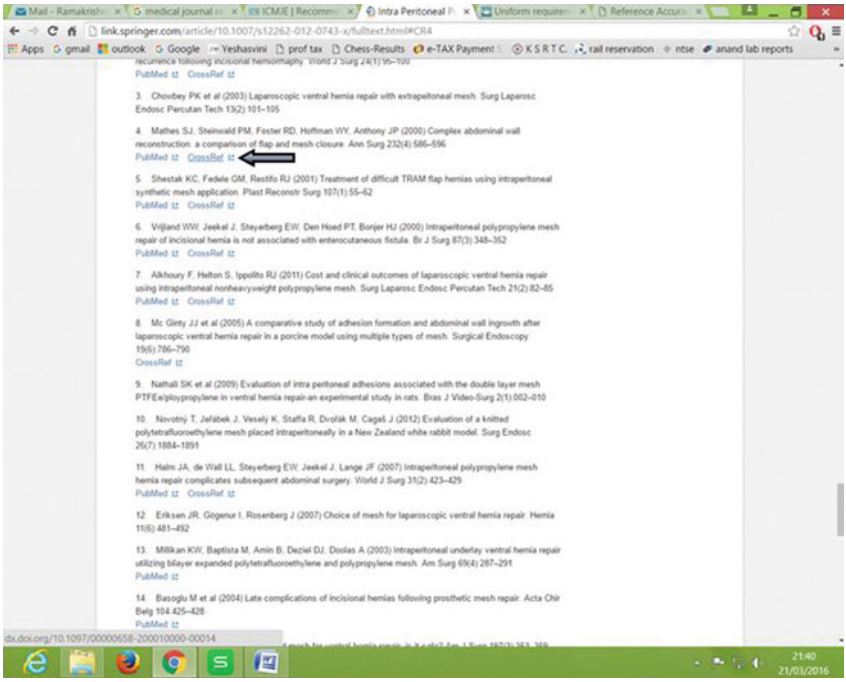
The advantage of writing references in a uniform format is that references can be cross-linked in electronic format of publication. It is easy to search the article and read that article also.

For example, while reading an article, a reference 4 is mentioned (arrow).



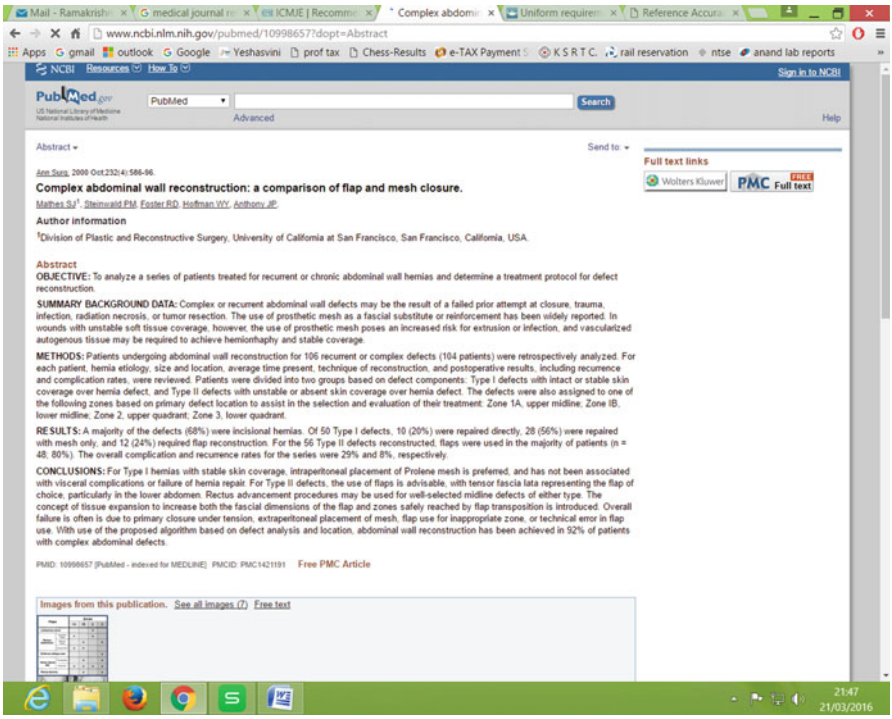
How to quote reference in the article?

When the reference number is clicked, it takes the screen to the reference section.



Citing the article under references. In electronic version, if reference in the article is clicked, it takes to reference section

There is a word “PubMed” (arrow). When clicked over the word, it takes the screen to the article quoted.



In the reference section if source is clicked, it takes you to the article source to display the article

Needless to say, all these operations require the computer to be connected to the Internet. Only abstract is shown if access to the full article requires payment or subscription.

Reference linking is the most useful feature of the electronic version of the article. If references are not written in a proper format, this feature cannot be used. That is why journals stress on proper format of reference.

What Happens After Submission?

(Further details can be read in the article P. F. Kotur. How to write a scientific article for a medical journal? Indian J Anaesth, *2002*;*46*(*1*);*21–25*.)

Editorial Process After the article is submitted to a journal, editor or one or more of the associate editors read and assess the quality of the article. They may reject the article at this stage, if found unsuitable for publication.

Peer Review If the article passes the scrutiny by the editor, the article is referred to the peers for review. Peers are experts in the field of the topic of the article. They have sufficient knowledge, experience, and interest. They review the article in detail for all aspects mentioned above. The review is confidential. Author will not know who is reviewing their article. After review, the article may be accepted as it is or rejected or returned to the author for major or minor revision. There may be suggestion for improvement.

Common Reasons for Rejection of Article or Article Sent Back for Changes (Problems in the Article)

No clear message.
Too much of information.
Too little information.
Inaccurate information.
Problem of structuring the article.
Missing information.
Grammatical errors.
Inadequate references.
Wrong format of writing references.
A similar article has been already published.

Resubmission The author has to resubmit the article, if returned for revision, after correcting the errors.

Process Repeats The above process repeats. If now found suitable, the article will be accepted: if not rejected.

Best Chance of Acceptance

- Clear and important message/learning point
- Interesting topic
- Relevant journal
- Geographical consideration
- Instruction to authors
- Diligent attention to references
- Attention to spelling/English grammar

Plagiarism

According to the Merriam-Webster online dictionary, plagiarism means (http://www.merriam-webster.com/dictionary/plagiarize):

- To steal and pass off (the ideas or words of another) as one's own
- To use (another's production) without crediting the source
- To commit literary theft
- To present an idea or product derived from an existing source as new and original

In other words, plagiarism is an act of fraud. It involves both stealing someone else's work and lying about it afterward. Reusing the same data to write multiple articles and sending to multiple journals are also a type of "self"-plagiarism (Source: http://www.plagiarism.org/plagiarism-101/what-is-plagiarism/).

Plagiarism is the practice of taking someone else's work or ideas and passing them off as one's own. It amounts to piracy or stealing. It is considered as academic dishonesty and unethical practice. *In some cases it can constitute copyright infringement* (https://en.wikipedia.org/wiki/Plagiarism). Legally it may not be punishable but ethically it is punishable (by the associations, professional bodies), e.g., by expulsion. It is still plagiarism even though there is no copyright. Many journals now use softwares to detect plagiarism.

An act of plagiarism can have several repercussions for the author, the journal in question, and the publication house as a whole. Sometimes, strict disciplinary action is also taken against the plagiarist (Natasha Das, Monica Panjabi. Plagiarism: Why is it such a big issue for medical writers? Perspect Clin Res. 2011 Apr-Jun; 2(2): 67–71). In the article the authors mentioned various types of plagiarisms:

Plagiarism of ideas
Plagiarism of text (direct plagiarism)
Mosaic plagiarism
Self-plagiarism

How to Avoid Plagiarism?

- Always acknowledge the original source.
- Use inverted commas ("*xxx*") and *Italics* for the quotes copied from another source and also cite the source.
- To explain a concept, read the information from various sources and describe it in your own words.
- Cite the references accurately. Inaccurate reference may also be considered as plagiarism.
- Do not present the same article to multiple journal (self-plagiarism). When one journal returns the article as rejected, then it may be tried in another journal.
- Some websites offer plagiarism check. They can be utilized to avoid unintentional plagiarism.

Peer Review

A peer is somebody having a type of expertise or competence in the field of the topic. The process of peer review is difficult to define. It is at the heart of every journal's processing of the article and decision for acceptance. Many sponsors or grant givers want to know whether the journal is peer reviewed. The status "peer reviewed" of a journal improves its value and gives a feeling to the reader that its quality is assured.

But there are also flaws in the process. The editor sends the article for review to a number of reviewers, and if majority of them recommend it to be accepted, the article is accepted. If the majority recommend rejection, it will be rejected. If 50 % recommend accept, then what? It will be sent for one person for tiebreak. Then it is something like flipping a coin for the decision.

On this discussion I cannot put it in better words than the author in his article on, peer review: a flawed process at the heart of science and journals (Richard Smith. Peer review: a flawed process at the heart of science and journals. J R Soc Med. 2006 Apr; 99(4): 178–182).

> That is why Robbie Fox, the great 20th century editor of the Lancet, who was no admirer of peer review, wondered whether anybody would notice if he were to swap the piles marked 'publish' and 'reject'. He also joked that the Lancet had a system of throwing a pile of papers down the stairs and publishing those that reached the bottom. He also wrote that when he was editor of the BMJ, he was challenged by two of the cleverest researchers in Britain to publish an issue of the journal comprised only of papers that had failed peer review and see if anybody noticed. He wrote back 'How do you know I haven't already done it?'

It shows that it all depends on the reviewers. When two reviewers have different opinions, how to say whose opinion is right? Even the well-informed and intelligent readers also cannot even make out by reading the article whether the article was peer reviewed or not.

The other flaws of peer review are (for detailed account, read the article Peer review: a flawed process at the heart of science and journals. J R Soc Med. 2006 Apr; 99(4): 178–182):

1. *It slows down the process of publishing.*
2. *There may be bias in the decision*: *there may be bias against articles written by women*.
3. *Abuse*: *When the reviewer gets the articles for the review*, *he first rejects it and then use the ideas to write his own article*.

Copyright

The creator of an original work has certain exclusive legal rights given by the law of the country. It is the intellectual property of the creator. These are known as copyright governed by copyright law. There may be a time limit and some limitation for the rights. Impingement of the copyrights is punishable under the law. If a book is copyrighted, no part of the information can be copied without the written prior permission of the person who holds the copyright.

Literature Review

For literature review, first, information is to be searched. Before electronic era, to get literature for review was a great task, consuming a lot of time. One had to go to the library to search several shelves for appropriate journals for days together, get photocopies, and read them to collect the information. In spite of the best efforts, gathered information may be incomplete.

Digitalization of the data and information has made this job easy. Now, thanks to the Internet and technology, it is quite easy to find appropriate information from various websites with the help of "search engines" sitting at any corner of the world. To search it is important to know how to use search words. In the article, journals give a few key words, which help a particular article to be picked up form a large database. Hence, these key words are to be chosen appropriately while writing the article.

Commonly used search engines are:

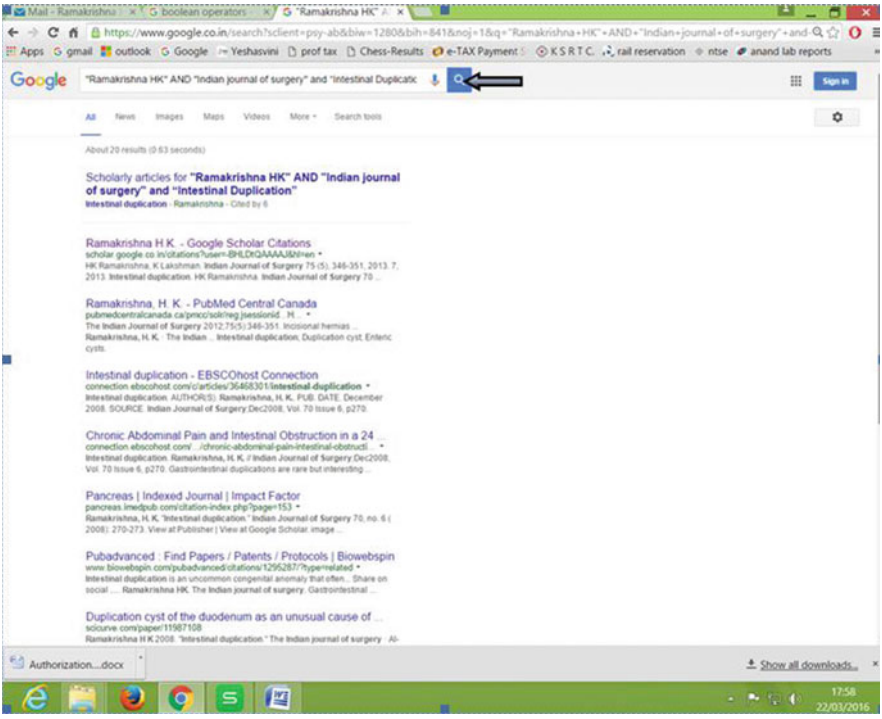
- "Google" search
- Google scholar
- Scholarly databases

While searching the Internet, certain tips help in getting specific information needed quickly.

Boolean operators are used to specify the type of search. They are ***and***, ***or***, ***not***, *or* ***and not.*** These are used as conjunctions with two key words. It is better to use more key words to narrow down information displayed. For example, in Google search, if "Ramakrishna HK" is typed as key word, it displays 26 pages of information. Majority of this information is unrelated to what is needed.

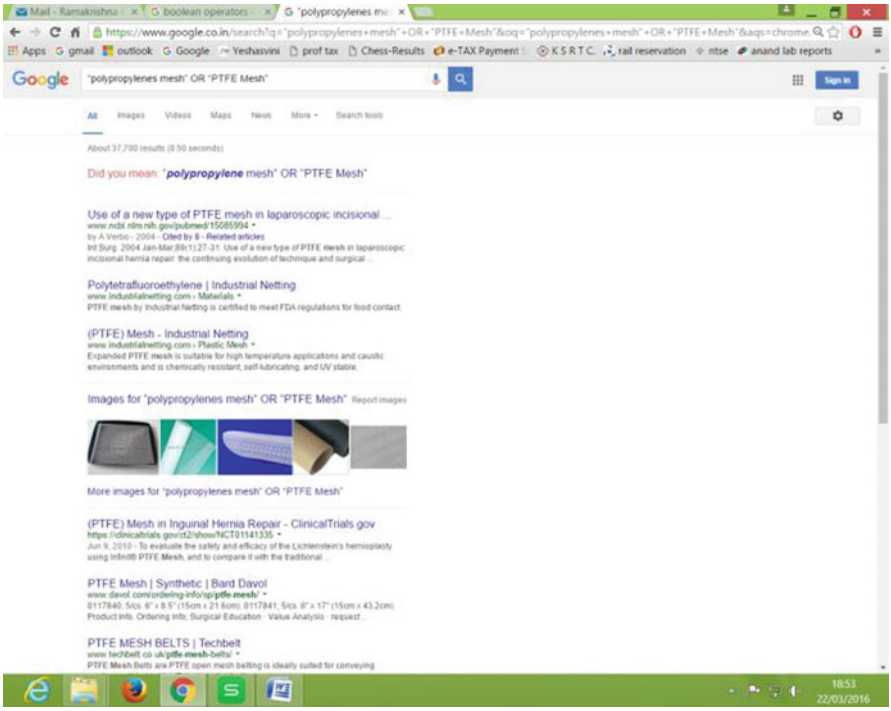
If "Ramakrishna HK" and "Indian journal of surgery" are typed as search words, it displays web pages which contain both these words. The information is now narrowed down to five pages, from where it is easier to find the information we want. It is important to note "inverted commas" for the search words. If only Ramakrishna HK Indian journal of surgery are typed, it displays web pages which contain all Ramakrishna, all of which may not be Ramakrishna HK or all Indian and may or

may not be Indian journal of surgery. Again more than 20 pages are displayed. So it is important to use inverted commas to contain two related words. It narrows down the display to only those pages which contain words **exactly matching** the phrase contained in the inverted commas. The search can still be narrowed if we know exactly what we are searching. For example, if we want to search for an article written by Ramakrishna HK in *Indian Journal of Surgery* on intestinal duplication, add the key word "intestinal duplication" to the above search; it displays only two pages, the article written by the author ***Ramakrishna HK*** in ***Indian Journal of Surgery*** on the topic ***intestinal duplication*** and related pages where this article is cited. There are only 11 links. So it is easy to search.



Boolean operator "and"

If "or" is used as conjunction between two key words, it displays results which contain either the first word or the second word. So it is a wider search. This is useful when information on two conditions is required. For example, "*polypropylene mesh*" *or* "*PTFE mesh*" is typed as key words, it shows the pages containing words either *polypropylene mesh or PTFE mesh*. This search should be used when information on both these types of meshes is required.



Boolean operator "OR"

When "not" (e.g., "xxx" *not* "yyy") operator is used, results containing the key word xxx are searched: then pages containing key word yyy are deleted. In the end, therefore pages containing the word xxx but not the word yyy are displayed.

Parenthesis () When some key words are used within parenthesis, other key words outside the parenthesis, first words within parenthesis, are searched. Then other conditions which are not enclosed are applied. For example, if we search with the by typing, ***("polypropylene mesh" OR "PTFE mesh") and "ventral hernia"***, it returns articles containing 1. polypropylene mesh and ventral hernia, 2. PTFE mesh and ventral hernia but does not show articles of ventral hernia if one of these two words ("*polypropylene mesh*" *OR* "*PTFE mesh*") within the parenthesis are not found in the article.

Scholarly Databases

There are many scholarly databases from where various articles and information can be searched. For example:

- MEDLINE
- PubMed
- ClinicalKey (previously MD Consult)
- HINARI
- Helinet
- ScienceDirect
- Ovid
- Publishing houses
- Medscape
- Cochrane library/database
- Etc.

MEDLINE® contains journal citations and abstracts for biomedical literature from around the world. PubMed® provides free access to MEDLINE and links to full-text articles when possible (https://www.nlm.nih.gov/bsd/pmresources.html).

PubMed is a free resource that is developed and maintained by the National Center for Biotechnology Information (NCBI), at the US National Library of Medicine (NLM), located at the National Institutes of Health (NIH). PubMed comprises over 25 million citations for biomedical literature from MEDLINE, life science journals, and online books. PubMed citations and abstracts include the fields of biomedicine and health, covering portions of the life sciences, behavioral sciences, chemical sciences, and bioengineering. PubMed also provides access to additional relevant websites and links to the other NCBI molecular biology resources (http://www.ncbi.nlm.nih.gov/pubmed).

ClinicalKey (previously known as MD Consult) is a comprehensive collection of medical and surgical resources in a variety of specialties designed to support evidence-based clinical care and clinical education. ClinicalKey is an upgraded and expanded version of MD Consult that offers full-text access to selected medical textbooks, medical journals, practice guidelines, drug information, patient handouts, and CME materials (http://www.lib.utexas.edu/indexes/titles.php?id=877).

HINARI Access to Research in Health Programme provides free or very low-cost online access to the major journals in biomedical and related social sciences to local, not-for-profit institutions in developing countries (http://www.who.int/hinari/about/en/).

Helinet is the digital library initiative at Rajiv Gandhi University of Health Sciences (RGUHS), Karnataka, India, is the first of its kind in the country in promoting e-learning culture and e-readiness preparedness for accessing huge amount of scholarly international medical e-journals and e-books. With the state-of-the-art infrastructures, RGUHS developed the digital library and information center for identifying, procuring, storing, processing, and disseminating the scholarly information resources in the field of health sciences by minimizing the cost of recurring

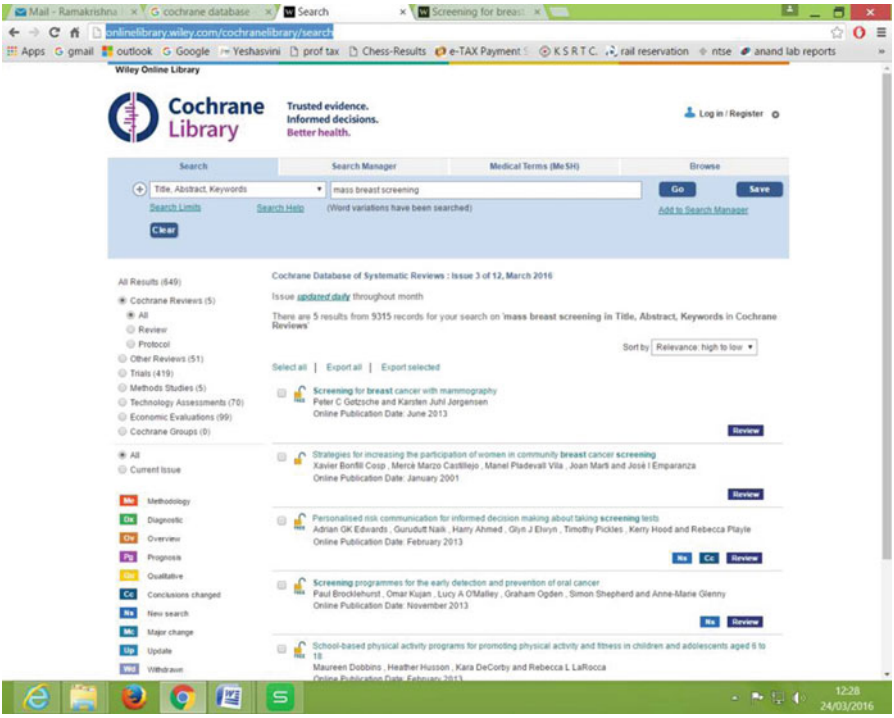
expenditure in the libraries of affiliated colleges and conceptualized the Health Science Library and Information Network (HELINET) for seamless access to the world-class health science literature and information resources, round the clock to the students, teachers, and researchers in all the affiliated colleges of RGUHS (http://www.rguhs.ac.in/HELINETHOSTCONSORTIUM/homehelinethost.htm).

Ovid Online Portal to Clinical and Educational Content, Plus Rich Multimedia Ancillaries for Teaching, Learning, and Practice

LWW Health Library is far more than electronic texts—providing highly intuitive, interactive access, and simple search capabilities to essentials texts, as well as rich multimedia ancillary content comprised of procedural videos, images, real-life case studies, and quiz banks specifically tailored for the specialty (http://www.ovid.com/site/index.jsp).

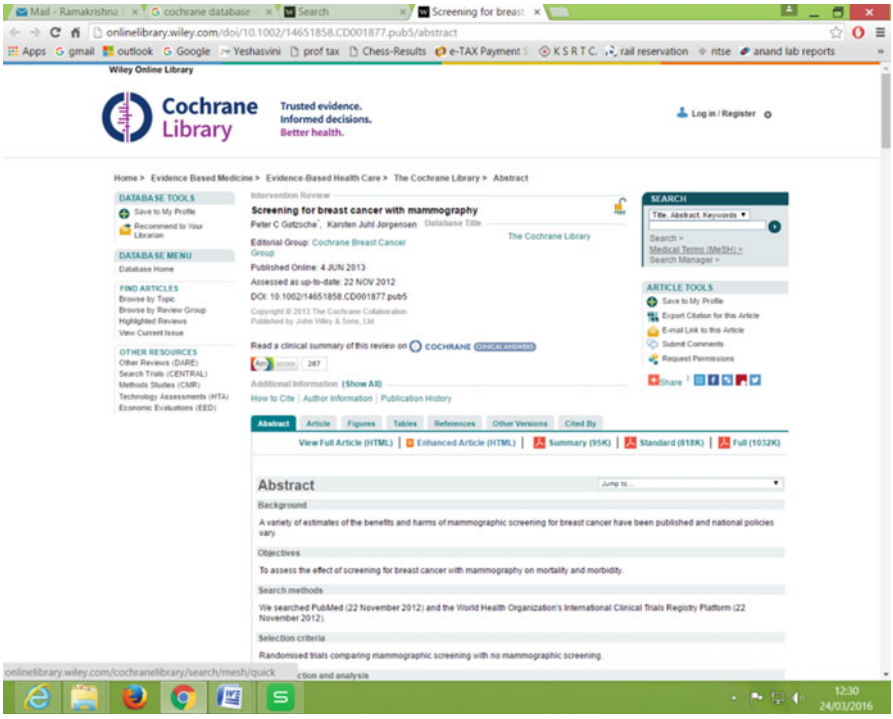
Medscape offers specialists, primary care physicians, and other health professionals the Web's most robust and integrated medical information and educational tools. After a simple, one-time, free registration, Medscape automatically delivers to you a personalized specialty site that best fits your registration profile (http://www.medscape.com/public/about).

Cochrane Library/Database http://onlinelibrary.wiley.com/cochranelibrary/search is the link to Cochrane library. It has a library of systematic reviews. If key words "mass breast screening" is entered, it shows five different articles.



Cochrane library web site: Searching with key words (http://onlinelibrary.wiley.com/cochranelibrary/search)

The first article is selected.



The first article selected: shows details (http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001877.pub5/full)

Many detailed full article can be accessed free of cost. It also shows where the article is cited.

Limitations and Issues

Everybody cannot access all information and articles. Full texts of the articles are not available on most of the websites. Each article can be accessed on pay-per-article/view basis.

Many sites require subscription/registration and substantial amount of payment. Institutions can subscribe and all its members can then access them.

The information on the Internet is so vast that it is difficult to get what we want. Sometimes it takes hours on the Internet to get the required information.

It takes a lot of patience and perseverance to read, understand, analyze, and write down an article.

Open-Access Journals

These journals are free to access for anybody who has an Internet connection. They do not have any financial or legal barrier. But they charge the authors to publish their articles. Some of them are sponsored by a society or an institution or government. They bear the cost of publishing and hence readers need not pay to access the articles. https://doaj.org/subjects is a directory of open-access journals. There are several thousands of such journals covering all fields. For example, *BMJ Case Report* is an award-winning journal that delivers a focused, peer-reviewed, valuable collection of cases in all disciplines so that healthcare professionals, researchers, and others can easily find clinically important information on common and rare conditions. This is the largest single collection of case reports online with more than 11,000 articles from over 70 countries (http://casereports.bmj.com/). These journals are useful both to submit articles and also to get information for writing articles.

Open-access medical journals are listed in Wikipedia (https://en.wikipedia.org/wiki/List_of_open_access_journals):

- *Annals of Saudi Medicine*
- *Bangladesh Journal of Pharmacology*
- *Biomedical Imaging and Intervention Journal*
- *BMC Health Services Research*
- *BMC Medicine*
- *British Medical Journal*
- *British Columbia Medical Journal*
- *Canadian Medical Association Journal*
- *International Journal of Medical Sciences*
- *Journal of Clinical Investigation*
- *Journal of Postgraduate Medicine*
- *The New England Journal of Medicine*
- *PeerJ*
- *PLOS Medicine*
- *PLOS Neglected Tropical Diseases*
- *PLOS Pathogens*

Publish or Perish?

This topic is debated in recent times very frequently. An estimate shows that each day more than 34,000 articles are added to the literature from more than 4000 journals! Each minute a new article is added to the literature!

Scientific commitment should be the prime driver for publication. However, one of the main reasons for publishing the articles is career development. Two publications under research paper or original article are required to be promoted as

associate professor from assistant professor or professor from associate professor under MCI India guidelines. So, each associate professor tries to produce at least two articles. Some authors publish only for pride, to put it in their CV. Papers written out of compulsion will result in **substandard papers**. It can also lead to **scientific fraud**. Many postgraduate dissertations are converted into paper: many of these papers are also substandard (not all, some are really good).

A paper written and published out of real interest without any of the greed will be useful. If written out of pressure, there will be more problems than benefits. Those who cannot write on their own pay a *ghost writer* to get the article written in their names and publish. In some articles there are *too many authors*. It is doubtful if all the authors have made any substantial contribution to the study. It leads to a **lack of quality and integrity. Rejection rate of the articles increases**. Publishing has become a **profit-oriented business**. Many journal publishers have made substantial wealth out of publishing by asking the authors to pay for their article to be published.

The pressure to publish **adds stress** on the authors.

In spite of all the ill effects of pressure to publish, many good articles with good study designs are published by genuinely interested researchers. They help in improving patients' care.

Problems of Compulsion to Publish
Substandard papers.
Scientific fraud.
*Ghost writer*s.
Too many authors in a single article.
Lack of quality and integrity.
Increase in rejection rates.
Publishing has become a profit-oriented business.
Adds stress to the authors.

Model Example of Writing an Article

Let us consider the example of an imaginary study on the Drug X helping in the passage of a ureteric calculus given in Chap. 7 on Design of a Study. Please read the example in that chapter again before continuing here. How to write an article using that data? *Please note that the inference*, *conclusions*, *data, etc. mentioned here are purely imaginary and are used here only for simple explanation and understanding purposes. It should not be used in the clinical decision-making*.

Title: Prospective Randomized Control Trial on Efficacy of "Drug X" in the Spontaneous Passage of Ureteric Calculus

(**Clear message**: **Since it is a new drug**, **the reader is interested in knowing if the drug is really effective. Small ureteric calculi are a common problem faced by clinicians in everyday practice. Hence the topic is relevant. It merits publishing.**)

(**Type of the study design**: **It is a prospective randomized double-blind control trial.**)

(**Types of journals suitable for submission of article**: **surgery journals**, **medicine journals**, **urology journals**, **pharmacology journals, etc.**)

Authors:
XXX, Professor of surgery, AA Medical College, Mumbai
YYY, professor of Medicine, AA Medical College, Mumbai

Corresponding author:
XXX, Professor of surgery, AA Medical College, Mumbai
e-mail: aaa@sss.com
Mob no: 88xxxxxx77
Address for correspondence: No 42, uu road, Mumbai Central, Mumbai.

Key Words: Ureteric calculus, Ureteric calculi, Drug X, Spontaneous passage of ureteric calculus, Randomized double-blind trial.

Abstract:

Ureteric calculi are fairly common in clinical practice seen by general practitioners, surgeons, physicians, urologists, etc. Larger calculi require some form of interventions. Small calculi less than 8–9 mm size are known to pass spontaneously via the *naturalis* over a period of time. Usually about 4–6 weeks of observation is advised, unless there are definite indications for intervention like repeated severe colicky pain without progressive descent, etc. It is estimated that about 70–75 % of small ureteric calculi pass out spontaneously during such observation. However, there are some drugs which hasten the passage with higher frequency. Drug X is one of them, which acts by blocking alpha receptors. We conducted a randomized prospective double-blind control trial to study the efficacy of the Drug X in spontaneous passage of ureteric calculus. A total of 268 patients diagnosed on ultrasound scanning (USG) as ureteric calculus is included in the study. They are divided randomly into study and control groups. Study group was given Drug X and control group was given a placebo. Both the patient and the investigator were blinded to intervention. Data was collected by an independent observer. Seven patients in control group and ten patients in study group were lost for F/U. In the study group, two patients were excluded from the study as they developed adverse effects to the drug. One patient in control group developed severe pain and was operated. Under observation, 76/131 in study group and 59/117 patients passed the calculus by 30 days. Chi-square test is applied to this data, and the difference in the results was not significant at P 0.05.

(**264 words are used. Observe how the reader is led to understanding the problem. The current management is explained. The data are presented very briefly. No abbreviation is used. Read the previous discussion on abstract and observe how this example follows the guidelines.**)

Introduction:

Patients presenting with typical loin-to-groin renal colic type of pain are common in clinical practice. Usually an ultrasound scanning (USG) is done. USG is a

very sensitive investigation for detecting small renal calculus and hydronephrosis [1]. Its sensitivity and specificity are 95 % and 96 %, respectively [2]. Renal colic is managed conservatively with analgesics. Further management depends upon the size of the calculus. Larger calculi require some form of intervention like extracorporeal shock wave lithotripsy (ESWL), ureterorenoscopy (URS), basketing, ureteroscopic lithotripsy, etc. [3–6]. For small ureteric calculi, usually observation is advised. Patients are observed for spontaneous passage of calculi via the *naturalis*. About 70–75 % of calculi passes out over a period of 4–6 weeks [1, 7]. There are claims that some drugs can increase the frequency with which calculi are passed. Drug X is claimed to help in the passage of the calculi by acting on alpha receptors, resulting in relaxation of smooth muscles of the ureter and sphincter. This trial studies the efficacy of the drug in expelling the small ureteric calculus less than 8 mm size as assessed on USG.

(***Observe how the reader is introduced to the problem. Note how references are quoted in Vancouver style. The first reference cited is given number 1 and the subsequent references were serially numbered. The first reference is quoted again along with number 7. So it is not given a new number but given the same original number*** [1]. ***For each statement reference from the literature of published article is cited. If the reader is facing this clinical situation in his day-to-day practice, he will be definitely interested in knowing the efficacy of this drug as this knowledge will be useful for him in his practice. More detailed explanation and more references can be cited.***)

We conducted a randomized prospective double-blind control trial. …. (**Continue writing**….)

Materials and Methods:

All cases with a diagnosis of ureteric calculi on ultrasound scanning at XYZ hospital during January 2015 to December 2015 period were studied. Patients presented with typical renal colic type of pain. Ultrasound scanning was done for all patients. Out of 578 patients, 268 patients satisfied inclusion criteria and recruited to the study.

Inclusion Criteria:

1. Should have ureteric calculi on ultrasound scanning
2. Calculus size as measured by ultrasound scanning 8 mm or less
3. Age between 18 to 60 years

Exclusion Criteria:

1. Age less than 18 and more than 60 years
2. Calculus size 9 mm or more
3. Patients allergic to the drug
4. Pregnant and lactating women

These 268 patients were randomly assigned to study group and control group. Study group was assigned of 143 patients and control group was assigned of 125

patients. All the patients were given a unique ID. Both the patient and investigator were blinded. To the study group of patients, the Drug X was given along with analgesics. Patients were asked to drink lots of water. To control group of patients, a placebo tablet (which looked similar to Drug X tablet but without active ingredient) was given along with the same analgesics. These patients were also asked to drink lots of water. All patients were instructed to watch urine (by means of a filter) for the passage of calculus for a period of 6 weeks time. Patients were followed up for a period of 6 weeks. When a patient passed a calculus, it was recorded by an independent observer.

(***Note how the study population was defined including the study period*, *setting, etc.*, *definition of the problem in question mentioned*, *strict criteria to include or exclude a patient in to the group*, *assigning the patients to groups done*, *intervention is described and follow-up protocol, etc.***)

Results:

Sex distribution

	Males	Females	Total
Study group	69	62	131
Control group	59	58	117

Chi test: $P = 0.82$: statistically not significant

Age distribution

Age in years	Study group	Control group
11–20	4	6
21–30	26	25
31–40	38	30
41–50	36	29
51–60	27	27
Total	131	117

T test: $P = 0.718$: statistically not significant

(**It is important to show that the two groups did not differ in age and sex distribution. Otherwise critics may argue that females have better expulsion rates or younger age group has better rates. This can undermine the significance of results and conclusions.**)

Out of 268 patients, two patients developed adverse effects to the drug in the study group and so were excluded from the study. In the study group, ten patients were lost for follow-up. Similarly, in the control group, seven patients were lost for follow-up and one patient developed severe pain and was operated. These patients were excluded from the study. So, 131 patients in study group and 117 patients in control group (a total of 248) were available for final analysis.

(**Check the numbers several times for accuracy. The total**, **group total**, **and individual numbers should be tallied. Suppose it is mentioned that study group**

has 132 patients (or some similar inaccuracies are mentioned), then the numbers will not be tallied. 143-2-10 = 131.)

Data are presented in Table 8.1.

Table 8.1 Frequency of passage of calculus

	Passed calculus by day 30	Didn't pass calculus by day 30	Total
Study group	76	55	131
Control group	59	58	117
Total	135	113	248

(**The data mentioned in the table or graph need not be and should not be repeated in the text. Again check the numbers and totals for accuracy.**)

Discussion:

In the study group, 58 % of patients passed calculus by day 30. In the control group, 50.4 % of patients passed calculus by day 30. These figures are less when compared to other reported series. ABC et al. have reported 72 % of spontaneous passage in 4 weeks' time (8). Other series report between 40 and 75 % (6, 7, 10). The reason may be less intake of water, higher temperature in our country or other factors like consumption of alcohol, etc.

(**This is only a model imaginary report; hence kept short. Discussion can be in more detail.**)

We applied chi test to test statistical significance between these proportions. P value was 0.2318. The difference in results is not statistically significant as $P > 0.05$.

Conclusion:

The Drug X does not have beneficial effect on passage of ureteric calculus of 8 mm or less in size in the age group of 18 to 60 years.

(**Since in the inclusion criteria, we fixed age group as 18–60 years and calculus size as 8 mm or less in size, study results cannot be generalized to all age populations or all sizes of calculi. A single conclusion backed by adequate data is stronger than too many conclusions with inadequate or no data in the main text. It is wrong to mention statements like Drug X decreases the severity of pain, does not affect the requirement of surgery for the calculus, etc., since these were not the objectives and there are no data in the article on these parameters.**)

Disclosures:

Authors do not have any interest in the pharmacological companies producing Drug X. No financial assistance was taken from any source.

(**The sponsorship from pharmaceutical companies can undermine the validity of the studies especially if the results show benefits.**)

Acknowledgments:

We are thankful to XYZ hospital for allowing us to utilize patients' data.

References:

1. Acb.(2012). Spontaneous passage of ureteric calculi: Journal of uuu.25:1:75–80.
2. …

3. …
4. …
.
.
10. …

(**Note that references are written in Vancouver style. Do not cite any article unless information from the article is utilized in the writing the present article**.)

As it is already mentioned, this model is only an imaginary general presentation and serves as an example. Papers have to be written with care avoiding all inaccuracies.

Other types of papers have different format and style of writing.

Case Reports

The main objectives of case reports are to highlight some learning points. Usually rare cases are reported. Rare presentation of a common case also merits reporting. Any single series cannot accumulate sufficient number of rare cases to present the data as observational study. So they are reported as case reports.

Guidelines for title, **authors**, **key words, and other heads remain the same**.

(Read the article HK Ramakrishna, UJ Vaidya: Post operative recurrent acute jejuno-jejunal intussusception: Indian J. Surg (May–June 2008) 70:147–148 to serve as a model.)

Title: Postoperative Recurrent Acute Jejuno-Jejunal Intussusception

Abstract: *A case of recurrent acute jejuno-jejunal intussusception presenting in the postoperative period of the surgery for acute ileocolic intussusception is presented. Postoperative intussusception is defined as intussusception occurring within 30 days of the primary surgery. This is a rare entity. Jejuno-jejunal intussusception is also rare. Recurrent intussusception is uncommon. The present case is a combination of all these rarities*.

(**Explain how rare the condition is. Explain what the special features of the present case are. Materials' and methods' heading is not applicable as it is not a trial. Instead**, **a case report is written.**)

Case Report

A 6-month-old female baby presented with vomiting of 1-day duration. In the night, i.e., about 12 h of the initial symptoms, the baby had one bout of minimal bleeding per rectum.

The next morning, the baby was feeding well but used to vomit 15–20 min after the feed. On examination, the baby was irritable and not dehydrated. Abdominal examination revealed no palpable mass….

(**Describe the case as a case record is written. The presenting complaints**, **examination findings**, **relevant investigation findings**, **clinical photos**, **operative**

photos, and imaging photos, if applicable, are all explained in detail. At the same time, it is important to avoid writing unnecessary details like [in this case] pulse rate was 92 per minute, general condition was satisfactory, moderately built, etc. It all depends on the case. Report should be complete. If it was operated, what happened after the surgery, whether there was complication, mortality, etc.)

Discussion:

Postoperative intussusception is defined as acute intussusception occurring within 30 days of primary surgery. This can follow any surgery. This is rare. Eke N and Adotey [1] found only two cases after a literature review on postoperative intussusception….

(**Cite other reference of similar articles reported in the literature. Literature search is discussed above helps here. Explain different types of presentations. How the present case is different from the cases reported in the literature.**)

Conclusions:

After a thorough search of literature, we failed to find a similar case. Though we could find recurrent, postoperative and jejuno-jejunal intussusception cases separately, a combination of all these was not found. Hence, we are reporting the case. High index of suspicion is the key to success as symptoms are less dramatic.

(**Do not give conclusions which cannot be drawn by reading the case report. Actually, the reader himself will be able to draw conclusions: The author's conclusions should also be similar. Highlight the learning points.**)

Meta-analysis and Review Articles

A **review article** may be written by collecting articles from various journals on a particular condition.

Writing **meta-analysis** and review articles differs from others in methods of collecting data. Other parts of writing remain the same.

(Read H. K. Ramakrishna and K. Lakshman. Intra Peritoneal Polypropylene Mesh and Newer Meshes in Ventral Hernia Repair: What EBM Says? Indian J Surg. 2013 Oct; 75(5): 346–351 as a model example.)

Materials and Methods*: We searched for electronic data on the Internet by Google search, Google scholar, MEDLINE, PubMed, Cochrane library, etc. for as much of information as possible. We also searched websites like www.who.int/hinari/en/, www.inasp.info/file/68/about-inasp.html, www.nlm.nih.gov/pubs/factsheets/medline, www.nlm.nih.gov/pubs/factsheets/pubmed, www.Intute.ac.uk, www.tripdatabase.com/, etc. The articles which did not provide the necessary information we wanted (on the incidence of complications of the intraperitoneal mesh) or meet the necessary criteria (intraperitoneal PPM or newer mesh) were discarded. We got a heterogeneous data, some in support of and some against the use of PPM intraperitoneally. It was difficult to compare the data as demographics, protocols, methods, techniques, and meshes used were different in different studies. We could not find a prospective randomized controlled double-blind study*

comparing intraperitoneal placement of PPM with newer meshes. So, we concentrated on the data on complications of intraperitoneally placed PPM and newer meshes by both open and laparoscopic techniques in the reported literature. Recurrence is not much of controversy. Arguments against the use of intraperitoneal PPM are because of incidence of complications like enterocutaneous fistula, bowel adhesions with consequent intestinal obstruction, chronic infection, sinus formation, erosion of the mesh into the viscera, etc. Based on these data, we drew our conclusions.

(**Observe the difference in the description of how data is collected. Mention all sources of information. Mention the inclusion criteria**: **how articles are selected. Mention the exclusion criteria**: **how articles are rejected and not considered for database creation.**)

This chapter can serve at the best as a guideline as to how to write an article. Writing an article is as much an art as a science. Some people are more blessed with the art of writing and presentation. For others, a lot of efforts, patience, and practice is required. By these qualities, everyone should be able to write a good article, if the material is good.

Conclusions

Writing a journal article is both an art and a science. Every clinician should aim at documenting his data and results. Use the data documented to write various types of articles and submit them to appropriate journals.

It is important to learn how to use computers and the Internet so that required information can be searched on various websites.

This indirectly helps one to keep up to date with knowledge: Thus, it helps in professional development also.

Wonderful thing about game of life is that
Winning and losing are only temporary,
.....unless you quit. Dr Fred Mills

Whether the article will be accepted or not is not important: **but the important thing is to keep writing**.

Evidence-Based Medicine

9

Learning Objectives
What, why, and how evidence-based medicine?
Pyramid of studies: increasing values
Levels of evidence
Benefits of evidence-based medicine
Limitations of evidence-based medicine

> People almost arrive at their beliefs not on the basis of proof,
> but on the basis of what they find attractive.
>
> Blaise Pascal

What Is EBM?

Definition: ***Evidence-based medicine (EBM) is finding evidence obtained from a well-designed and conducted research and using that evidence to make clinical decisions***.

It means our treatment decisions should be based on a combination of experience and a sound evidence from research results published in the literature and not simply on our experience or opinion of some "expert." Ultimately it is for the benefit of the patients. It aims at improving the quality of treatment and outcome. It provides some standardization of the treatment decision. It helps in forming protocols. The concept of EBM started around 1980s. Its popularity is rapidly increasing as a tool for recommendations for clinical decision.

Why Do We Need EBM?

Many times what we practice is what our teachers have taught us. They are "experts" in their field. We trust them so much that whatever they teach is the ultimate truth. However, another teacher with equal experience may have a different view and

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_9

teach altogether a different method. We get confused. Whom to follow? Both are experts. Suppose a complication happens and patient drags you to the court. The judge may call a different expert, who gives a different opinion. You cannot say "my teacher had taught me like this." Nobody will accept the claim or management decision if there is no evidence in the literature to support.

One expert says "In a particular problem situation, I managed the patient in this way. Patient did well." Under similar situations you manage the patient in the same way but still the result was not exactly the same. The "expert" might have just boasted and hid the failure. Or there may be other factors which you have not noticed. Or it may be simply because of biological variation.

Patient may question your decision to treat in some particular way. You should be in a position to justify your decision.

The medical knowledge and concepts of management also change with time. If you have to transfer the benefits of recent advances to your patients, you need to have updated knowledge. If you read two different journals on the same subject, you may find two, seemingly opposite, conclusions. How will you conclude which conclusion is currently acceptable?

For example, consider the conclusions in this article (Malik FI, Mirza TI. Intraperitoneal mesh plasty. *Professional Med J* Sep 2010; 17(3): 360–365) "Intraperitoneal Meshplasty with conventional polypropylene mesh is a safe, quick, convenient method of incisional hernia repair with minimum morbidity and mortality; the results are comparable to any other procedure being practiced today. The complications associated with intraperitoneal placement of the conventional polypropylene mesh were not seen in our experience."

Another article concludes (Keith W. Millikan et al., Intraperitoneal underlay ventral hernia repair utilizing bilayer expanded polytetrafluoroethylene and polypropylene mesh. The American surgeon (2003) Volume: 69, Issue: 4, Pages: 287–291). "Bilayer prosthetic mesh composed of ePTFE and polypropylene can be safely placed intraperitoneally without causing intestinal obstruction or enteric fistula."

Now, you cannot decide whether to use conventional polypropylene mesh or should you go for newer bilayer mesh. The problem is newer mesh is costlier by 15 times. Whether the extra cost is worth? How would you decide?

Suppose you have a problem in your current management line. You want to improve the results. You need some guidelines to change your current line of management. You want to know what are the results with the new line of management and how reliable the results.

Answer to all these problems is EBM.

How?

While reading the journals or analyzing a study conclusion, you must know what is the strength of the results. All types of articles or study do not have the same value or strength in its results. Some evidences are very strong so that you can trust it and use in your practice. It can also be quoted in the court of law as an evidence to support your management decisions. But conclusions of some of the studies have questionable value.

The value of different types of articles or studies can be represented as a pyramid. It is also referred to as *hierarchy* or *ladder of evidence*. If you Google the word "evidence pyramid," you get many types of pyramids. One such example is (Fig. 1):

Fig. 1 Pyramid of Evidence, Medicine

(http://www.slideshare.net/kpadron_libraries/evidence-based-practice-8412826).

They basically show in vitro studies have the least value for clinical application, whereas systematic reviews/meta-analysis of published randomized controlled double-blind trials have the highest value.

This pyramid should not be confused with levels of evidence. Different types of evidences carry different value or strengths. Depending upon the strength, evidence is classified into different levels. As the level increases, its value in clinical application decreases. There are many methods adopted by different countries/centers. But they all have a general agreement. They only differ in terminology, for example, the Oxford Centre for Evidence-Based Medicine – Levels of Evidence.

(http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009) uses a system (1a,1b,1c,2a,2b,2c, 3a,3b, 4 and 5) where level 1 has the highest value and level 5 has the lowest value of evidence. Each level (from 1 to 3) is subdivided again into a, b, c, etc.:

Level 1: Evidence shows evidence from systematic review of high-quality RCTs.
Level 2: Evidence from low-quality RCTs.
Level 3: Evidence from case control studies.
Level 4: Evidence from case series or observational studies.
Level 5: Evidence from expert opinion or anecdotal case reports.

As level increases, the strength or value of evidence decreases.

The UK-NHS system uses levels A to D:

- Level A: Consistent RCT or cohort study
- Level B: Consistent retrospective cohort, exploratory cohort, case control study, or extrapolations from Level A studies
- Level C: Case series or extrapolations from level B studies
- Level D: Expert opinion without critical appraisal, or based on physiology

Now, having understood the evidence pyramid and levels of evidence, we shall consider how to apply this knowledge to practice. It is a five-step approach.

First of all you should identify the problem. Turn it into a question. Taking the above example on mesh, our problem is to decide whether we must use newer bilayer mesh or can we use conventional polypropylene mesh for intraperitoneal placement in the repair of a ventral hernia. Theoretical problem of risk of conventional mesh is that it can form an intestinal fistula or produce intestinal obstruction from bowel adhesions. Problem with the newer mesh is it is costlier by almost 10–15 folds. Is this extra cost worth? Now turn this problem into a question: is there sufficient evidence to say that conventional polypropylene mesh (PPM) produces more complications than newer bilayer tissue-separating mesh?

The second step is to search the literature for the evidence to see if there are increased complications. We know that the highest value or the most reliable evidence is from a systematic review of RCTs or a large randomized controlled double-blind study comparing these two types of meshes. If such RCT is not available, then meta-analysis of literatures of this subject can also be used. If you consider only observational studies, the value is questionable.

So the third step is to analyze the results critically. Ask many questions to yourself. Read carefully in between the lines. Note if the two arms of the study are really comparable in all aspects except the type of mesh. If authorities review the available literature systematically and give conclusion, we can use it as a guideline. If you draw conclusion based on a single article, probability is that it may be wrong. There may be other experts who can quote another article to prove that you are wrong.

Once you are satisfied that you have found a satisfactory reliable answer, you can apply it in your practice. That is the fourth step. So let us say, you find a meta-analysis which concludes as (Intra Peritoneal Polypropylene Mesh and Newer Meshes in Ventral Hernia Repair: What EBM Says? HK Ramakrishna, K Lakshman - *Indian Journal of Surgery*, October 2013, Volume 75, Issue 5, pp 346–351) "Complications of intra peritoneal PPM (adhesions, infection, intestinal

fistulisation, sinus formation, seroma and recurrence) can occur with newer mesh also. There is no statistically significant difference in the incidence of these complications between these meshes." So you decide conventional mesh can be used. Start using the conventional mesh.

Now, the fifth step. Record your result. Record any complications you may face. Analyze your own results to see whether your conclusion to use conventional mesh was justified. You may get results to support its use or to recommend against its use. Now you can form a final conclusion and guideline/s.

The Five-Step Approach to EBM

1. Frame a question.
2. Search literature for the evidence.
3. Critical appraisal of the evidence.
4. Apply the best evidence.
5. Evaluate the outcome and develop guidelines.

Benefits of EBM

- Minimizes errors in clinical decisions
- Improves quality of patient care
- Promotes lifelong and self-directed learning
- Continuous professional development

When clinical decisions are made based on some evidence, the errors are minimized. The decisions should not be made just because some expert said or taught by a professor. We have to question and analyze critically whether the advice given is correct. There should be evidence in the published literature so that if somebody questions or if there is litigation in the court of law, the decision can be defended. For example, in stage IV cancer, if literature says surgery is not of any benefit, patient should be advised against surgery. If surgery is done on flimsy indications and a complication occurs, surgeon may be held responsible. The ultimate goal of evidence-based medicine is to improve the quality of care of the patients. It also sets uniform type of treatment for a medical condition. Based on evidence-based medicine, protocol of treatment or guidelines for treatment can be drawn. EBM encourages clinicians to learn new things to keep updated knowledge of recent advances in the management of patients. With all these factors, EBM helps in developing profession in a better way.

Limitations of EBM

- RCT unethical
- Expensive trials
- Funding and conflicts of interests
- Time-consuming
- Obsolescence of research findings
- Unavailability of evidence
- Publication bias/retrieval bias
- Ghost writers

There are problems with EBM also. To produce evidence, we need good randomized control trials. More and more trials have to be done which sometimes may be unethical. To test anticancer drugs, many cancer patients are to be treated with inevitable suffering and adverse drug reactions. Some of the patients who are in control groups do not receive any active drugs. If there is an option to treat the condition under study, withholding treatment can be unethical.

Also, good randomized control trials are very expensive. A lot of funding is required. If a pharmaceutical company (producing the drug under evaluation) is funding the project, there may be conflict of interest. Randomized control trials sometimes take a lot of time. By the time the trial concludes, the findings would have become obsolete and not of any clinical relevance. For many clinical decisions, evidence may not be available in the literature. In the above-quoted example of the use of a regular polypropylene mesh or one of the newer meshes (intraperitoneal) in the repair of ventral hernia, there was no prospective randomized double-blind study comparing these two types of meshes. When the evidence is lacking, decision is essentially based on personal preference and cost rather than EBM.

The problems of publication bias and ghost writers are discussed in an earlier chapter (*Writing a Journal Article*: Chap. 8).

Conclusions

In spite of limitations, evidence-based medicine helps the clinicians to take appropriate clinical decisions with a combination of experience and evidence available in the published literature.

Evidence-based medicine is a five-step approach as explained above.

Clinicians must know the value of different types of study and methods of interpreting the data.

Clinicians should critically analyze the claims in the literature/pharmaceutical companies to arrive at proper decisions.

As far as possible, clinicians should follow guidelines and protocols in the management.

Model Example 10

Objectives
To understand concepts studied so far with a model example of a study design and writing a paper

I assume by now the reader has all the basic knowledge of medical statistics, design of a study, the art of presentation, collection of information from the Internet whenever required, and converting the information into an article. Combining all the knowledge, let us try to design a model example trial and write an article to a journal.

As it is only a model example, I try to keep it very simple. Wherever there is a doubt about terms used, the reader is advised to refer to appropriate chapters for refreshing the knowledge and clear the doubt. I have used *italics* wherever thoughts and explanations interrupt the flow of writing the study and article.

****

A surgeon was facing the problem of patients complaining of moderate-to-severe headache in the postoperative period, who were operated under spinal anesthesia (subarachnoid block (SAB)). This we call postspinal headache. He discussed the problem with the anesthesia colleagues and considered one of the causes for spinal headache was the size of the spinal needle. The opinion was the thinner the needle, the lesser will be the incidence of headache. However, all would not agree.

So now we have a problem. I turn this problem into a question.

"Is the incidence of postspinal headache less when thinner-gauge spinal needle is used for spinal anesthesia?"

A study was planned. The aim and objective of the study is to evaluate if there is difference in the incidence of postspinal headache when different sizes (gauges) of spinal needles are used.

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4_10

Trial has to be planned well. The variables to be considered are to be planned.

1. *Two different needle sizes.*
2. *What is the outcome measurement? It is the assessment of headache. As pain is a subjective phenomenon, a method should be used to quantify the headache: otherwise we cannot use statistics and tests of significance. Visual analog scale is one such useful tool to quantify the severity of headache.*
3. *What type of study? We know that prospective randomized controlled double blinded study is the best. So we have to recruit the patients as and when they come for surgery. Note that the records of the past operated cases were taken for database creation.*
4. *Tests of significance and level of significance. P value less than 0.05 is fixed as the level of significance. Student T test was used for continuous variables, and chi-square test was used for categorical variables.*
5. *Subjects for the study: patients who are undergoing lower abdominal surgery under spinal anesthesia.*

An ethical committee is formed with a group of anesthesiologists and surgeons. A written consent from all the patients entering the study is taken.

Initially a pilot study was started and continued for 2 months. Patients who are undergoing lower abdominal surgery are selected (***Prospective study***, as patients are recruited prospectively) and randomly assigned to two groups. For the first group of patients, 23 G spinal needle is used. For the second group of patients, 26 G spinal needle of the same type and company is used.

Randomization is done by tossing a coin by a nurse with prior fixation of group to heads or otherwise. Tossing the coin is done only once. Under no circumstances the group assigned by the toss is changed. (***Random***, as we know with each flip of coin probability of heads is 50 %: so each patient has 50 % chance to enter either group.)

Anesthesiologist would write the name of the patient, ID number and 23 G (or 26 G), put it in an envelope, and seal it (*23 G group acts as the* ***control group***, *which is compared with 26 G group*). The surgeon would not know which size needle was used (***double blind***, neither the patient nor the evaluator knows to which group the patient belongs). The envelope is not opened until when documentation is completed and database creation is started. Anesthesiologist did not visit the patients postoperatively. The surgeon looked after pain management in the postoperative period and also followed up the patients.

So now the study design is (*see each bold letters in brackets above*) ***prospective*, *random*, *controlled*, and *double blind***.

Postoperative Management Protocol to Be Followed

1. Foot-end elevation for all patients in the immediate postoperative period and continued for 24 h. Patients should be advised rest in bed for 24 h.

2. Three liters of IV fluids should be given in the postoperative period.
3. All patients should receive IV infusion of diclofenac 75 mg diluted in 100 ml of normal saline twice daily, IV tramadol 1 ml every 8 h, and, when oral feeds were resumed, paracetamol tablet 650 mg three times daily. This regimen should be given for the first 2 days (operated day and the first postoperative day). From the second postoperative day, tablets of piroxicam (20 mg two times a day for 2 days and continued with once-daily dosage) be given along with paracetamol tablet 650 mg three times daily. Oral analgesics should be continued up to the fifth postoperative day. The same protocol of pain management was followed in all patients.
4. Patients should be followed up for 1 month.
5. In the postoperative period, if any patient complained of headache, the severity of headache should be assessed by using visual analog scale (VAS), and the VAS score is recorded by the surgeon.

(*Observe how the protocol is written explaining in detail how patients are managed and clarifying all parameters. If different postoperative pain management is used in different patients, it is difficult to compare VAS scores, as analgesics can affect VAS score.*)

During the pilot study, it was noted that some patients had migraine and complained of exaggeration of headache. It caused some confusion. Also, children and older age group patients were more difficult to assess with VAS. So inclusion and exclusion criteria were considered:

Inclusion criteria: age group 15–60 years
Exclusion criteria: prior history of migraine

(*Note how the pilot study helps in finding out flaws or difficulties we may face during the study. Also note how inclusion and exclusion criteria are defined. After ensuring that the trial is running smoothly for a 2-month period, the actual trial was started.*)

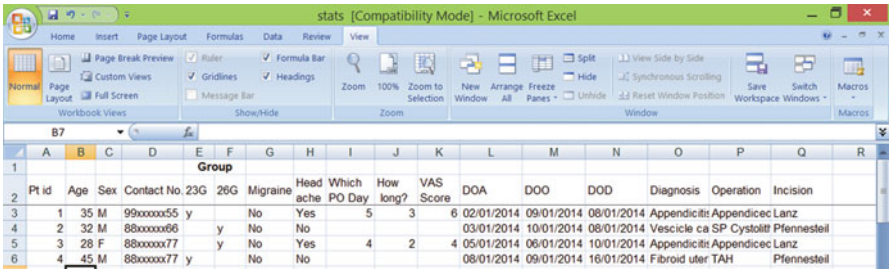
Study Period All patients satisfying the above criteria and undergoing lower abdominal surgery under spinal anesthesia during the period Jan 2014 to Dec 2015 are included in the study.

A data entry sheet was printed to enter each patient's data.
(A model) data entry sheet:

Data entry sheet		
ID number		DOA
Age		DOD
Sex		Date of operation
Contact number		

Data entry sheet		
Diagnosis		
Operation		
Incision		
H/O migraine	Yes	No
Follow-up		
Headache	Yes	No date
After how many days of post-op?		Days
VAS score		
Lasted for how many days?		Days

When the trial ended, a master chart is created on the Excel sheet to enter all patients' details using data entry sheet.



	Pt id	Age	Sex	Contact No.	Group 23G	Group 26G	Migraine	Headache	Which PO Day	How long?	VAS Score	DOA	DOO	DOD	Diagnosis	Operation	Incision
3	1	35	M	99xxxxxx55	y		No	Yes	5	3	6	02/01/2014	09/01/2014	08/01/2014	Appendicitis	Appendicec	Lanz
4	2	32	M	88xxxxxx66		y	No	No				03/01/2014	10/01/2014	08/01/2014	Vescicle ca	SP Cystolit	Pfennesteil
5	3	28	F	88xxxxxx77		y	No	Yes	4	2	4	05/01/2014	06/01/2014	10/01/2014	Appendicitis	Appendicec	Lanz
6	4	45	M	88xxxxxx77	y		No	No				08/01/2014	09/01/2014	16/01/2014	Fibroid uter	TAH	Pfennesteil

From the master chart, data are compiled into tables for various fields.

Results

A total of 212 patients who are undergoing lower abdominal surgery are randomly assigned to two groups. The first group consisted of 98 patients who were given subarachnoid block using 23 G needle. The second group consisted of 114 patients who were given subarachnoid block using 26 G needle.

Number of patients assigned to groups

23 G group	98
26 G group	114
N	212

(*Present data of sample size first in clear terms. All other data going to be presented subsequently are based on this table*).

Average age of patients

Age	
23 G group	38.51 ± 13.47
26 G group	39.29 ± 13.17

(The *age of the patients in each group has to be mentioned. Age is an important parameter in medical statistics. It is not possible to mention the ages of all patients individually, and hence mean age ± standard deviation is mentioned*).

Male to Female ratio

Males	Female	M:F
59	39	1.512821
62	52	1.192308

(*Sex ratio is one of the important basic data.*)

Incidence of headache			
	N	Headache	Percentage
23 G group	98	23	23.46
26 G group	114	13	11.4
	N = 212		

(*Our important data*: *the outcome recorded. This data is categorical*: *whether developed headache or not. The severity or magnitude of the problem cannot be made out of this data.*)

VAS score		
	N	Score
23 G group	23	4.52 ± 1.15
26 G group	13	3.46 ± 1.36
	N = 212	

(*The severity of headache is numerically documented. The data is continuous variable*, *depicted as mean ± standard deviation.*)

Severity of headache				
	N	Mild 1–3	Moderate 4–6	Severe 7–10
23 G group	23	4	18	1
26 G group	13	8	4	1
	N = 212			

(*The severity can also be categorized into groups*: *mild*, *moderate, and severe*.)

Paper

Title:
Impact of Size of the Spinal Needle on Postspinal Headache
(*Details should be sent in a separate page*, *indicating the corresponding author.*)

Authors:
Swarnalatha MC, MD (Anesth.), Ramakrishna HK* MS (Gen. Surg.), DNB, FMAS.

*Corresponding author
(*Details should be sent in a separate page*, *indicating the corresponding author.*)

Key words: Postlumbar puncture headache, Postdural puncture headache, Postspinal headache, 23 G spinal needle, 26 G spinal needle.

Aims and Objectives:
To study the effect of different sizes of spinal needles on:

1. The incidence of postspinal headache
2. Severity of postspinal headache

Synonyms:
Postlumbar puncture headache (PLPHA), Postdural puncture headache (PDPHA), Postspinal headache (PSHA)

Abstract:
Postspinal headache is a common problem after a dural puncture for either a diagnostic CSF tap or for subarachnoid block for anesthesia. Sometimes it is so severe as to cause suspicion of meningitis. It is sometimes so incapacitating that patient cannot get up. Many times it increases hospital stay. Various methods are advised for the prevention of headache after dural puncture. One of them is to use finer spinal needles. We undertook a randomized controlled double-blind study to test if finer spinal needles produce lesser incidence of postspinal headache.

(*The problem is very briefly touched. Of the many solutions to the problem, only one is picked to concentrate on one problem so that conclusion will be simple*, *easy, and strong.*)

Introduction:
Postspinal headache is a common problem after dural puncture for either diagnostic CSF tap or for subarachnoid block for anesthesia. It is said to be more common in younger age group [1] and pregnant with lower body mass index [2].

Typically spinal headache is bilateral and more occipital, aggravated by sitting up, and relieved by lying down. It occurs within 7 days of dural puncture and disappears by 14 days. Other causes of (which may be coincidental during postdural puncture period) headaches do not have these typical features of postspinal headache. If the headache lacks these features, clinician should be on alert to rule out other causes (sometimes serious like meningeal infection) of headache.

The overall incidence of postspinal headache is 0.1 to 36 % [4]. Factors affecting the incidence of postspinal headache are [5]:

1. Needle size
2. Direction of the bevel
3. Needle design
4. Replacement of the stylet
5. Number of lumbar puncture attempts

Contrary to the common belief, following factors do not affect the incidence of postspinal headache [5]:

1. Volume of CSF removed
2. Position during lumbar puncture: lateral or sitting up
3. Improving hydration: either oral or IV fluids before the procedure
4. Bed rest

Though many factors are there, we wanted to fix to only one factor that is the needle size. This will make conclusion easier. The other factors may be tested in another trial separately. Study design was such that all other factors were similar in all patients in the study. We undertook a randomized controlled double-blind study to test if finer spinal needles produce lesser incidence of postspinal headache, comparing 23 G needle with 26 G needle.

(*The reader is introduced to the existing problem. Information from the literature regarding the incidence of the headache*, *possible variables which can affect the incidence*, *and solutions are all given briefly. This creates interest in the reader to read further as the information would be useful in the clinical practice.*)

Materials and Methods:
All patients who are undergoing lower abdominal surgery under spinal anesthesia during the period January 2014 to December 2015 in Lakshmi Hospital, Bhadravathi, were included in the study.

Ethical committee permission was taken. A written consent from all the patients entering the study was taken.

Patients were randomly assigned to two groups. For the first group of patients, 23 G spinal needle is used. For the second group of patients, 26 G spinal needle of the same type and company is used.

Anesthesiologist would write the name of the patient and 23 G (or 26 G), put it in an envelope, and seals it. The surgeon would not know which size needle was used. The envelope was not opened until all cases are documented and database creation was started. Anesthesiologist did not visit the patients postoperatively. The surgeon looked after pain management in the postoperative period and also followed up the patients.

Postoperative management protocol was followed:

1. Foot-end elevation for all patients in the immediate postoperative period and continued for 24 h. Patients were advised to rest in bed for 24 h.
2. Three liters of IV fluids were given in the postoperative period.
3. All patients received IV infusion of diclofenac 75 mg diluted in 100 ml of normal saline and IV tramadol 1 ml every 8 h. When oral feeds were resumed, paracetamol tablet 650 mg three times daily was introduced. This regimen was given for the first 2 days (operated day and the first postoperative day). From the second postoperative day, tablets of piroxicam (20 mg two times a day for 2 days and continued with once-daily dosage) were given along with paracetamol tablet

650 mg three times daily. Oral analgesics were continued up to the fifth postoperative day. The same protocol of pain management was followed in all patients.
4. Patients were followed up for 1 month.
5. In the postoperative period, if any patient complained of headache, the severity of headache was assessed by using visual analog scale (VAS), and the VAS score was recorded by the surgeon.

Inclusion criteria: age group 15–60 years
Exclusion criteria: prior history of migraine

Student *T* test and chi-square tests were used to test the statistical significance in the difference between the groups and calculate the P value.

(*Methodology followed is described in detail and accurately leaving nothing of worth mentioning. Details about where the study was conducted, what is the study period, who were the patients, etc. are all described in detail. Strict definitions of inclusion and exclusion criteria are used leaving no room for confusion. The outcome measurement is also defined properly. What types of data are expected and what statistical test will be used are also mentioned.*)

Results:
A total of 212 patients were included in the study. The number of patients assigned to different groups is shown in the Table 10.1.

Table 10.1 Number of patients assigned to each group

23 G group	98
26 G group	114
N	212

Age distribution is shown in Table 10.2. There was no statistically significant difference in the age distribution between the groups using Student *T* test ($P = 0.668$).

Table 10.2 Age (Groups are comparable as $P > 0.05$)

Age	
23 G group	38.51 ± 13.47
26 G group	39.29 ± 13.17
T test *P* =	0.668

(*Please note that the data are continuous: hence, the Student T test is used.*)

(*It is important to show that the groups did not differ with respect to age of the patients (for that matter, they should not differ in any respect like sex, etc. except the study intervention). If one of the groups had younger patients, it may be argued that results could be due to the fact that the group had younger age patients, as age of the patient is known to affect perception of pain and threshold of pain.*)

Male to female ratio is shown in Table 10.3. The P value for the table is 0.53 (chi-square test), showing there is no statistically significant difference in the sex distribution between the groups.

Table 10.3 Male to female ratio (comparable as $P > 0.05$)

Males	Female	M:F
59	39	1.512821
62	52	1.192308
Chi test for M:F $P = 0.53$		

Headache incidence was 23.46 % in the patients who received spinal anesthesia with 23 G needle in comparison with 11.4 % in patients who received spinal anesthesia with 26 G needle. The data is shown in Table 10.4. Chi test results $P = 0.019$. The incidence of postspinal headache is significantly higher in the group who received spinal anesthesia with 23 G needle.

(*Please note that the data are categorical*: *Hence*, *the chi-square test is used.*)

Table 10.4 Incidence of headache

	N	Headache	Percentage
23 G group	98	23	23.46
26 G group	114	13	11.40
	$N = 212$		
Chi test $P =$	0.019		

(*Please note that the data are categorical*: *hence*, *the chi-square test is used.*)

Also, the severity of headache was significantly higher in the group who received spinal anesthesia with 23 G needle. Table 10.5 shows the comparison of the VAS scores of the two groups. Student *T* test was used to this data: $P = 0.005$.

Table 10.5 VAS score

	N	Score
23 G group	23	4.52 ± 1.15
26 G group	13	3.46 ± 1.36
	$N = 212$	
T test $P =$	0.019	

(*Please note that the data are continuous*: *Two possibilities*: *23 G group has lower incidence or higher incidence of headache*; *both types of results are important*: *So a two-tailed test is to be used. Hence*, *a two-tailed Student T test is used*).

When the severity of headache was classified into mild (VAS score 1–3), moderate (VAS score 4–6), and severe (VAS score 7–10), it was found that more number of patients had moderate headache in 23 G group in comparison with more number of patients who had mild headache in 26 G group (*P* value = 0.005, highly significant). The results are shown in Table 10.6.

Table 10.6 Severity of headache

	N	Mild 1–3	Moderate 4–6	Severe 7–10
23 G group	23	4	18	1
26 G group	13	8	4	1
	$N = 212$			

Chi test $P = 0.005$

(*Please note that the data are categorical*: *it is a 3x2 contingency table. Hence*, the *chi-square test is used.*)

Discussion:
The incidence of postspinal headache is related to the size of the needle used for dural puncture. It is postulated that the CSF leaks from the puncture site causing a low CSF pressure which is the cause of headache. The thicker the needle, the bigger will be the puncture. Hence, more CSF leaks, resulting in higher incidence of headache. The incidence of postspinal headache in different studies varies from 0.1 to 36 % [4]. Our overall incidence is 16.98 % (36/212). The incidence was significantly less when thinner-gauge needle was used. In our series, headache occurred in 11.4 % in patients when 26 G needle was used in comparison with 23.46 % when 23 G needle was used. The results were statistically significant ($P = 0.019$). Reported series also shows similar higher incidence when thicker-gauge needles were used [6, 7].

The severity of headache is also related to the needle size. When the severity was quantified with VAS score, the average score for 23 G group was 4.52 ± 1.15 and that of 26 G group was 3.46 ± 1.36. When Student *T* test was applied, the result was significant at significance level of $P < 0.05$. ($P = 0.019$). PPP also reported similar difference in their study [8].

When VAS scores were categorize as mild (VAS score 1–3), moderate (VAS score 4–6), and severe (VAS score 7–10), we found that in 26 G group, more cases fell into mild category, and in 23 G group, more cases fell into moderate category.

There are other factors which affect the incidence of postspinal headache after dural puncture. But our objective in the present study was only to study the effect of the size of the needle.

(*Detailed discussion of the results*, *comparison of the results with the reported results in the journals, etc. are written. Observe the Vancouver style of reference citing. Some references are imaginary. If necessary more details and more references can be added. If the results differ from the reported results*, *mention the possible explanation for the same.*)

Conclusions:
The use of thinner spinal needles is associated with significantly lesser incidence of postspinal headache. Headache when occurs will be of lesser severity if a thinner needle is used. We strongly recommend 26 G needles for spinal anesthesia whenever feasible.

(*It is a repetition of aims and objectives given as conclusion. Observe that only two factors are mentioned in the conclusions*: *incidence and severity when it occurs. There are data in the text to support the conclusions. No conclusion is given on different designs of needles, hydration, etc. which are affecting the incidence of headache. It cannot be overemphasized that a few conclusions based on strong data are better than a large number of conclusions without data in the study. Based on the conclusions*, *a recommendation for clinical practice may be given.*)

Disclosures: Authors have no association with spinal needle manufacturers. No conflict of interest to disclose.

References:

1. Leibold R A, Yealy D M, Coppola M. et al. Post-dural puncture headache: characteristics, management and prevention. Ann Emerg Med 1993221863–1870.1870 [PubMed]
2. Kuntz K M, Kokmen E, Stevens J C. *et al.* Post-lumbar puncture headaches: experience in 501 consecutive procedures. Neurology 1992421884–1887. 1887 [PubMed]
3. Olsen J, Bousser M-G, Diener H-C. *et al.* The International Classification of Headache Disorders: 2nd edition. Cephalalgia 2004249–160.160 [PubMed]
4. Kuntz KM, Kokmen E, Stevens JC, Offord KP, Ho MM. Post lumbar puncture headache: experience in 501 consecutive procedure. Neurology. 1992;42: 1884–7. [PubMed]
5. S V Ahmed, C Jayawarna, and E Jude. Post lumbar puncture headache: diagnosis and management. Postgrad Med J. 2006 Nov; 82(973): 713–716.
6. xxx. incidence of post spinal headache with different needle sizes: A comparative study. J of bbb:23:5:27–30.
7. aaa. post spinal headache and needle sizes: J.of ccc:26:8:57–60.
8. PPP. Severity of post spinal headache: Needle size matters.vvv Journal of anesthesia.2006.86.3.78-81.
9. ….

(*References are numbered serially in the order in which they appear in the article. The format of citing is also important*: *For details refer to Chap.* 8, Writing an Article for Journals.)

Index

A
Absolute risk reduction (ARR), 71, 81, 86
Analysis of variance (ANOVA), 51–56, 67–69, 103
ANCOVA, 80
Area chart, 115

B
Bar chart, 14, 107–112, 116, 117, 120
Bias, 92, 99–101, 103, 123, 145, 166
Biological sciences like dental, veterinary, agriculture, 2, 10
Biostatistics, 5, 6, 9–19, 45, 86
Blinding, 89, 100–101, 123
Boolean operators, 146–148

C
Calculator, 4, 30, 33, 38, 39, 42–47, 50, 56, 61, 65, 69
Case-control study, 36, 98, 138
Categorical variable, 36, 37, 56, 59, 103, 120, 168
Chance factor, 11
Chi square test, 1, 17, 35–39, 154, 168, 174–176
Clinical
  audit, 90, 120–127
  trials, 2, 4, 85–127
Cohort study, 99, 164
Component bar chart, 117
Conditional probability, 32, 34
Confidence
  interval, 31, 34
  level, 16, 31, 34
  limit, 13
Continuous variables, 36–37, 45, 56, 61, 103, 117, 120, 168, 171
Control group, 13, 46, 47, 55, 58, 59, 61, 63, 65, 68, 71, 79, 90, 98, 100, 102, 123, 124, 154–157, 166, 168
Control study, 65, 98
Correlation, 71–73, 77, 81–83
  coefficient (r), 73
Cox regression, 77
Crossover study, 90
Cross sectional studies, 99

D
Data
  analysis and inference, 90, 103–105
  entry sheet, 169, 170
Declaration of Helsinki, 87
Degree of freedom, 38, 57, 60
Designing a trial, 87–88
Dichotomous outcome, 76
Double blind, 89, 98, 101, 123, 137, 154, 155, 159, 160, 163, 164, 166, 168, 172, 173
Doughnut chart, 116–117

E
Editorial process, 142
Effect size, 32, 89, 93
Ethical committee, 87, 137, 168, 173
Evidence based medicine, 4, 161–166
Expected frequency, 12, 13, 37

F
Factor analysis, 80
False
  negative, 16, 31, 32, 126
  positive, 31, 32, 67, 68, 125
Fisher's exact test, 1, 17, 35, 40

© Springer Science+Business Media Singapore 2017

H.K. Ramakrishna, *Medical Statistics*, DOI 10.1007/978-981-10-1923-4

G
Gaussian distribution, 33–34
Ghost writer, 153, 166
Gosset's test, 45–48

H
Harvard system, 139
Hazard ratio (HR), 77, 78
Hierarchy (ladder of evidence), 163
Histogram, 117–119

I
Impact factor, 132–133
Incidence of a disease, 86
Independence, 21, 32–33, 36
Indexing, 132
Internet based calculators, 47, 50, 61, 65

K
Kaplan–Meier estimator or survival graph, 79

L
Levels of evidence, 161, 163, 164
Linear regression, 74–76
Line chart, 114
Literature review, 94, 146–148, 159
Logistic regression, 76, 80
Longitudinal studies, 99

M
Mann–Witney–Wilcoxon (MWW) test, 35, 54–56, 66, 103
MANOVA, 80
Manuscript submission, 133–142
Mass screening, 85, 125, 126
Mean, 10, 11, 13, 15–18, 22, 24, 28, 30, 46, 56, 71, 72, 101–103, 132, 171
  deviation, 26, 27, 34
Measures of central tendency, 21–27
Measures of dispersion, 21, 26–27
Median, 15, 17, 18, 22–25, 27, 30, 33
Meta analysis, 4, 17, 131, 138, 159, 163, 164
Mode, 15, 17, 18, 22–24, 27, 30, 33
Multiphasic screening, 126
Multiple bar charts, 116
Multiple regression, 76–77, 80
Multivariate analysis, 79–80

N
Natural variation, 11, 12, 35, 37
Negative
  correlation, 72–73
  predictive value, 32, 33
Nonparametric data, 19, 56, 69, 73, 103
Nonsampling errors, 102–103
Normal
  curve, 1, 13, 15, 69
  distribution curve, 13, 15, 17, 45, 54, 67, 73
Number needed to treat (NNT), 71, 81, 86

O
Observational study, 98, 158, 163, 164
Observed frequency, 37
Odds, 67, 71, 81, 83
  ratio, 71, 81, 83
One tailed test, 48–49
Open access medical journals, 152

P
Paired T test, 5, 35, 49–51, 56, 59, 65, 69, 100, 103
Parallel studies, 90
Parametric data, 13, 56, 73, 103
Pearson's correlation coefficient, 73
Peer review, 132, 143, 145
Phases of trials, 85, 89
Pie chart, 113
Pilot study, 90–94, 168, 169
Plagiarism
  ClinicalKey (previously MD Consult), 149
  Cochrane library/ database, 150
  Helinet, 149–150
  Hinari, 149
  Medline, 149
  Medscape, 150
  Ovid, 150
  Publishing houses, 149
  PubMed, 149
  ScienceDirect, 149
Poisson regression, 77
Positive
  correlation, 72, 73, 83
  predictive value, 32
Power of a study, 80–83, 92
Prevalence of a disease, 87
Probability, 9–19, 21, 32–38, 45, 65, 70, 76, 77, 164, 168
Prospective, 81, 133, 137, 153–155, 159, 166
  study, 98, 123, 168

P value, 9–11, 13, 16, 17, 35, 37, 38, 40, 42, 45, 46, 51, 56, 57, 59, 61, 68, 70, 157, 168, 174, 175
Pyramid of studies, 161

R
Randomization, 92, 102, 168
Range, 14, 16, 26, 30, 31, 33, 76, 140
Rank test, 54, 67, 69–70
Reference interval, 13
Regressions, 73–77, 80
  coefficient, 75
Rejection rate, 132, 153
Relative risk reduction (RRR), 71, 81, 86
Retrospective, 98, 137, 164
  study, 98
Review articles, 130, 131, 133, 138, 159
Risk, 16, 33, 46, 70, 71, 80–82, 86, 89, 92, 99, 126, 137, 164
  ratio, 70, 82
R module, 41, 42, 68

S
Sampling
  simple random sampling, 102
  stratified random sampling, 102
  systemic random sampling, 102
Sampling errors, 16, 102–103
Scholarly databases, 146, 149–151
Screening of at risk population, 126
Sensitivity rate of the test, 32
Single blind, 98, 137
Skewed distribution curve, 17–19, 54, 55
Spearman correlation coefficient, 73
Specificity rate of the test, 32
Standard deviation (SD), 15, 16, 26–31, 33, 34, 45, 46, 171
Statistical inference, 22
Stats calculator, 30, 33
Student's T tests, 1
Surgical audit, 4, 122
Survival analysis, 77, 78

T
Tables, 1, 10–12, 14, 16–18, 38, 105–106, 138, 170
Tails of the data, 16
Tests of significance, 2, 4, 5, 9–19, 35–66, 103, 104, 138, 168
Theory of probability, 10
Time to acceptance, 132
True negative, 32
Two tailed test, 1, 47–49, 175
Type1 error (alpha error), 31–32, 67
Type2 error (beta error), 31–32

U
Universal screening, 126

V
Vancouver system, 139
Variance, 27–30, 67, 69

W
Web calculator, 46
Which test to use? wizard, 51
Wilcoxon signed-rank test, 69–70
Write an article, 1–5, 129, 130, 153, 160, 167

Y
Yates' correction, 36