hers.com, for detailed information on obstetrics & gynaecology books, visit our website www.jaypeebrothers.com, for detailed information on obstetrics & gynaecology books



Textbook for MRCOG-1 Basic Sciences in Obstetrics & Gynaecology

Includes more than 1,200 SBAs & Image-based Questions:
As per the latest RCOG Modules

Richa Saxena

3rd Edition

JAYPEE

Contents

1. Introduction to MRCOG Part 1 1

- Training in Obstetrics and Gynaecology in the UK *1*
- Pattern of MRCOG Examination *2*
- Eligibility *2*
- Venue and Timings *2*
- Examination Format *2*
- Syllabus of MRCOG Part 1 Examination *2*
- Tips for Studying *3*
- Scoring System *3*
- Single Best Answer *4*
- Single Best Answer versus Multiple Choice Question *4*
- Cover Test *4*

2. Ethics, Law, and Professionalism 5

- Levels of Healthcare in the National Health Service *5*
- Antenatal Care (Midwifery System) in the UK *6*

Clinical Governance *6*

- Audit *6*
- Risk Management *7*

Consent *9*

- Informed Consent *9*
- Types of Consent *10*
- Components of Consent *10*
- Gillick's Competence *10*
- Fraser Guidelines *11*
- Axon *11*
- Bolam *11*
- The Mental Capacity Act 2005 *11*
- Rights of the Unborn and Newborn Children *12*
- Jehovah's Witnesses *12*

3. Anatomy..... 14

Anatomy of Thorax *14*

- Diaphragm *14*
- Anatomy of the Female Breast *15*

Anatomy of the Abdominal Wall *17*

- Muscles of the Anterior Abdominal Wall *17*
- Blood Supply to the Anterior Abdominal Wall *19*
- Lymphatic Drainage of the Upper Abdominal Wall *20*
- Nerve Supply of the Anterior Abdominal Wall *20*
- Rectus Sheath *21*

Abdominal Cavity and Its Contents *22*

- Abdominal Aorta *22*
- Peritoneal Reflections *23*

Inguinal Region/Groin *26*

- Inguinal Canal *26*

Anatomy of the Female Pelvis *29*

- Pelvic Inlet *30*

Pelvic Organs *34*

- Female External Genitalia *34*
- Female Internal Genitalia *36*
- Male Internal Genitalia *43*
- Epithelium Lining the Reproductive Tract *43*
- Pelvic Organs: Part of the Gastrointestinal Tract *44*
- Blood Supply to the Pelvis *47*
- Nerve Supply to the Pelvis *48*
- Lymphatic Drainage of Lower Limb and Pelvis *51*

Anatomy of the Urinary Tract *56*

- Urinary Bladder *56*
- Ureter *57*
- Female Urethra *60*
- Adrenal Glands *60*

Anatomy of the Thigh *61*

- Femoral Triangle *61*
- Femoral Sheath *63*
- Femoral Canal *63*
- Adductor Canal *63*
- Femoral Artery *64*
- Great Saphenous Vein *64*
- Nerve Supply of the Lower Limbs *65*

Perineum *67*

- Anatomy of the Urogenital Region *67*

Anatomy of the Foetus *68*

Miscellaneous *71*

- Blood Supply to the Brain *72*

4. Physiology 73

Haematology *73*

- Blood *73*
- Blood Coagulation *76*

Homeostasis and the Fluid Balance *81*

- Body Fluids *81*
- Transport Mechanisms *82*
- Acid-base Balance *83*
- Abnormalities of Acid-base Balance *84*
- Arterial Blood Gas Analysis *85*
- Electrolyte Imbalance *86*

Physiology of Respiratory System *89*

- Lung Volumes *89*
- Lung Capacities *89*
- Respiratory Distress *90*
- Oxygen-haemoglobin Dissociation Curve *93*
- Disturbances in Respiration *95*

Physiology of Cardiovascular System ***96***
- Blood Flow Through the Heart *96*
- Cardiac Cycle *96*
- Lymphatic Vessels *101*
- Veins *102*

Physiology of Urinary System ***104***
- Microanatomy of Kidney *104*

Physiology of Gastrointestinal System ***108***
- Digestive System *108*
- Physiology of Liver *112*

Neuromuscular System ***116***
- Muscle Structure and Function *116*

Physiology of Sleep ***119***
- Sleep Cycle *119*

Physiology of Normal Pregnancy ***120***
- Placenta *120*
- Physiological Changes in Pregnancy *122*

5. Biochemistry and Nutrition 132

Biochemistry ***132***
- Structure and Function of Normal Cell *132*
- Carriers of Genetic Information *132*
- Carbohydrate Metabolism *135*
- Metabolism of Glucose *135*
- Other Fuels: Fructose and Galactose *146*
- Regulation of Blood Glucose *147*
- Metabolism of Fats *148*
- Metabolism of Nucleotides *152*
- Proteins, Peptides, and Amino Acids *154*
- Cell Signalling and Second Messengers *158*
- Cell Adhesion Molecules *161*

Nutritional Physiology in Health and Disease ***166***
- Biochemistry of Vitamins *166*

6. Pathology 171

General Pathology ***171***
- Inflammation *171*
- Acute Inflammation *171*
- Chronic Inflammation *172*
- Wound Healing *174*
- Adaptive Disorders *177*
- Cellular Response to Injury *178*
- Tumourogenesis *179*
- Diseases of the Blood Vessels *182*
- Shock *184*
- Skin Lesions *186*
- Diseases of Lymph Nodes *186*

Malignant Tumours of the Genital Tract ***190***
- Vulva *190*
- Vagina *191*
- Cervix *191*
- Uterus *191*
- Ovary *191*

7. Microbiology 195

Bacteria ***195***
- Streptococci *196*
- *Staphylococcus aureus* *199*
- *Listeria monocytogenes* *201*
- *Neisseria* *201*
- *Corynebacterium* *202*
- *Clostridia* *203*
- *Chlamydia* *204*
- *Mycoplasma* *205*
- Enterobacteriaceae *205*
- Spirochaetes *206*
- *Actinomyces* *207*
- *Pseudomonas* *207*
- *Campylobacter* *207*
- *Bacteroides* *208*

Principles of Infection Control and Outbreak Management ***211***
- Infection *212*
- Sterilisation and Disinfection *213*
- Specific Infections *214*

Non-bacterial Infections ***217***
- Viruses *217*
- Herpes Group of Viruses *217*
- Human Papillomavirus *220*
- Hepatitis Viruses *221*
- Parvovirus *224*
- Human Immunodeficiency Virus *225*
- Zika Virus *228*
- Mycology *228*
- Protozoal Infections *229*

8. Immunology 234

Organisation of the Immune System ***234***
- Introduction *234*
- Immune Response *234*
- Innate and Acquired Immunity *234*
- Cells and Humoral Elements of Acquired Immunity *235*
- Immunogenetics and Principles of Antigen Recognition *238*
- Immunodeficiency *239*
- Autoimmunity *240*
- Immunology of Transplantation *240*

Immunological Problems in Pregnancy ***242***
- Antiphospholipid Syndrome *242*
- Pregnancy and the Immune System *243*

Hypersensitivity and Complement System ***245***
- Hypersensitivity *245*
- The Complement System *246*
- Vaccines *248*
- Immunisation for *Haemophilus influenzae* *248*
- MMR Vaccination *249*

9. Embryology 251

Formation of Gametes ***251***
- Cell Division *251*
- Gametogenesis *253*
- Spermatogenesis *253*
- Oogenesis *254*
- Ovulation *256*

Embryogenesis ***257***
- Development of Human Embryo *257*
- Development of Human Placenta *260*

Organogenesis ***264***
- Development of Genitourinary System *265*
- Pharyngeal Arches *270*
- Development of the Central Nervous System *271*
- Development of Foetal Heart *272*
- Foetal Circulation *272*

10. Genetics 279

Basics of Genetic Disorders ***279***
- Deoxyribonucleic Acid *279*
- Ribonucleic Acid *280*
- Process of Protein Synthesis *281*
- Gene *281*
- Histones *282*
- Mitochondrial DNA *283*
- Human Chromosomes *283*
- Human Sex Chromosome *284*
- Principles of Inheritance *285*

Karyotypic Abnormalities ***289***
- Microdeletions *290*
- Genetic Carriers *290*
- Down Syndrome *292*
- Edwards Syndrome *293*
- Patau Syndrome *294*
- Klinefelter Syndrome *294*
- Testicular Feminisation Syndrome *295*
- Müllerian Agenesis (Mayer–Rokitansky–Küster–Hauser Syndrome) *296*
- 5-α Reductase Deficiency *296*
- Turner Syndrome *297*
- Noonan's Syndrome *299*
- Triple X Syndrome *299*

Chromosomal Abnormalities ***300***
- Haemophilia A *300*
- Von Willebrand's Disease *300*
- Duchenne Muscular Dystrophy *300*
- McCune-Albright Syndrome *301*
- Glucose-6-Phosphate Dehydrogenase Deficiency *301*
- Thalassaemia *302*
- Cystic Fibrosis *302*
- Tay–Sachs Disease *304*
- Fragile X Syndrome *304*
- Mutations of *BRCA1* and *BRCA2* Genes *305*
- Ambiguous External Genitalia *305*
- Trans-sexual Patients *306*
- Hermaphroditism *306*
- Pure Gonadal Dysgenesis *306*
- Short Stature *306*
- Sweating *306*

11. Biophysics 309

Radioactivity ***309***
- Ionising Radiation *309*
- Ionisation and Excitation *309*
- Units of Radioactivity *310*

Imaging Modalities ***311***
- Principles of Ultrasound *311*
- Doppler Ultrasound *314*
- Magnetic Resonance Imaging *314*
- Dual Energy X-ray Absorptiometry *315*

Electrosurgery ***317***
- Surgical Diathermy *317*

Lasers ***319***
- Classes of Lasers *320*
- Safety Precautions at the Time of Using Lasers *320*

12. Epidemiology/Statistics 321

Disease Epidemiology ***321***
- Prevalence *321*
- Incidence *321*
- Evaluation of a Clinical Test *321*

Statistics ***327***
- Distribution *327*
- Indices of Central Tendency *329*
- Variation *329*
- Correlation *330*

Obstetric Epidemiology ***333***
- Maternal Mortality *333*
- Confidential Enquiry into Maternal Deaths *333*
- Perinatal Mortality *334*

Clinical Trials ***335***
- Steps for a Research Study *335*
- Inferential Methods *336*
- Different Types of Clinical Trials *336*
- Different Types of Epidemiological Studies *338*

13. Endocrinology 342

General Endocrinology ***342***
- Introduction *342*
- Hormones Produced by Maternal-Foetal-Placental Unit *342*
- Half-life *343*
- Diurnal Variation *343*
- Chemical Structure of Hormones *344*
- Mechanism of Action of Hormones *345*
- Hypothalamus *346*
- Pituitary Gland *347*
- Pineal Gland *348*

- Growth Hormone *348*
- Prolactin *349*
- Hormones of Posterior Pituitary *350*

Disorders of Thyroid Gland *354*
- Thyroid Gland *354*
- Hyperthyroidism *355*
- Hypothyroidism *356*

Understanding of Pancreas *358*
- The Pancreas *358*
- Diabetes Mellitus *359*

Understanding the Sex Hormones *362*
- Gonadotropin-releasing Hormone *362*
- Testicular Hormones *362*
- Androgens in Women *363*

Female Reproductive System *364*
- Hormones of Female Reproductive System *364*
- Normal Menstrual Cycle *366*

Puberty and Adolescence *371*
- Puberty *371*
- Anovulation *376*
- Menopause *376*
- Treatment for Menopausal Symptoms *379*

Principles of Parathyroids *382*
- Parathyroid Glands *382*
- Hyperparathyroidism *383*
- Hypoparathyroidism *384*

Understanding of Adrenal Structure and Function *385*
- Adrenal Glands *385*
- Adrenocorticotrophic Hormone *386*
- Adrenal Hormones *387*
- Miscellaneous Hormones *390*

14. Clinical Pharmacology and Therapeutics ... 393

General Pharmacy *393*
- Drug Interactions *393*
- Adrenergic Agents *394*

Drugs Acting on Central Nervous System *395*
- Hypnotics, Sedatives, and Anxiolytics *395*
- Antipsychotic Drugs *396*
- Antidepressant Drugs *396*
- Analgesics *397*
- Antiepileptic Drugs *398*

Drugs for Inducing and Inhibiting Vomiting *401*
- Antiemetic Agents *401*
- Emetic Drugs *401*

Drugs Used in Obstetrics *402*
- Tocolytic Agents *402*
- Oxytocin *402*
- Methergine *404*
- Drugs for Treating Galactorrhoea *404*

Antimicrobial Drugs *406*
- Antibiotics *406*
- Beta-lactam Antibiotics *406*
- Antiamoebic Drug *410*
- Antiviral Drugs *410*
- Antifungal Drugs *411*

Miscellaneous Drugs *413*
- Anticancer Agents *413*
- Anticoagulants *414*
- Diuretics *416*

Drugs Acting on Cardiovascular System *418*
- Antihypertensive Agents *418*

Anaesthetic Agents *421*
- Local Anaesthesia *421*
- General Anaesthesia *422*
- Regional Anaesthesia *423*

Drugs Acting on Hormonal Imbalance *425*
- Drugs Interfering with Glucose Metabolism *425*
- Drugs for Thyroid Abnormalities *425*

Teratogenesis *427*
- Teratogenic Drugs in Pregnancy *427*
- Safe Drugs During Pregnancy *428*
- Cocaine and Pregnancy *428*
- Alcohol Consumption and Pregnancy *428*
- Foetal Alcohol Spectrum Disorders *428*
- Drugs Secreted into the Breast Milk *430*

Drugs in Gynaecology *432*
- Clomiphene Citrate *432*
- Luteinising Hormone-releasing Hormone Analogues *433*
- Danazol *434*
- Mifepristone *435*
- Misoprostol *436*
- Cyproterone Acetate *437*
- Ulipristal Acetate *438*
- Letrozole *438*
- Progestins *439*
- Oxybutinin *439*

15. Clinical Management: Obstetric Disorders 441

Early Pregnancy Care *441*
- Pregnancy Dating *441*
- Early Pregnancy Loss *441*

Early Pregnancy Complications *444*
- Gestational Trophoblastic Diseases *444*
- Ectopic Pregnancy *448*

Medical Disorders During Pregnancy *451*
- Iron Deficiency Anaemia *451*
- Intrauterine Growth Restriction *454*
- Epilepsy and Pregnancy *454*
- Cardiac Disease During Pregnancy *455*
- Diabetes in Pregnancy *455*
- Thyroid Disorders During Pregnancy *458*
- Urinary Tract Infection During Pregnancy *460*

Conditions Specific to Pregnancy *462*
- Pregnancy Dermatosis *462*
- Multiple Pregnancy *462*
- Hydramnios *465*

- Preeclampsia *467*
- Eclampsia *469*
- HELLP Syndrome *470*
- Hyperemesis Gravidarum *470*
- Preterm Labour *471*
- Screening for Foetal Anomalies *471*
- Anaesthesia in Pregnancy *473*
- Normal Labour and Delivery *475*
- Elective Episiotomy *478*
- Foetal Distress *478*
- Breech Presentation *479*
- Caesarean Section *480*

Post-partum Complications *486*
- Venous Thromboembolism *486*
- Post-partum Haemorrhage *489*
- Disorders of Mood *489*
- Endocrinology of Lactation *489*

Structural Changes in the Newborn *490*
- Normal Neonate *490*
- Congenital Anomalies *491*
- Jaundice in the Newborn *492*

16. Clinical Management: Gynaecological Disorders 495
- Surgical Skills (Preoperative) *495*
- Post-operative Changes *496*

Contraception *499*
- Barrier Contraception *500*
- Progestogen-only Contraceptive Methods *501*
- Intrauterine Contraceptives *502*
- Combined Oral Contraceptives *503*
- Post-coital Emergency Contraception *504*

Gynaecological Abnormalities *506*
- Amenorrhoea *506*
- Infertility *507*
- Male Infertility *507*
- Female Infertility *509*
- Genital Tract Fistulae *511*
- Pelvic Prolapse *512*
- Incontinence *514*
- Hirsutism *516*
- Ovarian Masses *517*
- Uterine Retroversion *518*
- Fibromyomata *518*
- Primary Dysmenorrhoea *521*
- Uterine Malformations *521*
- Endometriosis *522*
- Vaginal Discharge *522*
- Sexually Transmitted Infections *524*
- Urinary Tract Infections *526*

17. Data Interpretation 532
- Introduction *532*

Interpretation of CTG *532*
- Electronic Foetal Monitoring *532*
- Features of the Heart Rate Observed on the Cardiotocograph *532*
- Categorisation of Heart Rates *535*
- Management *537*

Interpretation of ECG *544*
- Hypokalaemia *544*
- Hyperkalaemia *544*
- Hypocalcaemia *545*
- Hypercalcaemia *545*
- Acid–base Imbalance *549*

Interpretation of X-rays *549*

Interpretation of a Urogram *551*
- Normal Findings in Filling Cystometry *551*
- Normal Findings in Voiding Cystometry *551*

Risk of Malignancy Index *554*

Female Genital Mutilation *554*

18. Image-based Questions 556

19. High-Yield Topics and Mock Examinations 573

Index 575

CHAPTER 12

Epidemiology/Statistics

DISEASE EPIDEMIOLOGY

PREVALENCE

The prevalence is the proportion of people in the entire population who are found to be with a disease at a certain point in time. Prevalence illustrates how often a situation occurs and is expressed as a proportion. It depends on the number of individuals who contract the disease in a particular time period.

Since the prevalence looks at the number of individuals with a disease at a given point in time, or within a defined time interval, if a patient has recovered from the illness in that duration then they would not be included in the prevalence rate. Prevalence is one measure, which can assess the health requirements of a community. The prevalence rate of a disease has the following features:

- It can be estimated from a cross-sectional study.
- It can be used to determine the health needs of a community.
- It is dependent on the duration of illness.
- It is dependent on the incidence of the disease.
- It measures all the current cases in the community.

INCIDENCE

The incidence of the condition is the number of newly diagnosed cases of the condition over a specific time period. The incidence tells how common a situation is.

EVALUATION OF A CLINICAL TEST

A clinical diagnostic test can normally be evaluated using four indices: Sensitivity, specificity, positive predictive value (PPV), and the negative predictive value. Sensitivity and specificity are most easily understood from a 2 by 2 table **(Table 1),** which gives the number of subjects showing positive and negative tests, and with and without disease in a cross-sectional random sample from a target population. It is only for such a sample that all four indices can be calculated from the same data.

Sensitivity

The sensitivity of a test is the probability of the test being positive in somebody with the condition. Sensitivity indicates the proportion of individuals with positive screening tests that actually have the disease. Sensitivity is equal to true positives/(true positives + false negatives).

TABLE 1: 2 by 2 table for calculating the sensitivity, specificity, positive predictive value, and the negative predictive value.

	Disease present	*Disease absent*	
Test positive	a (true positive)	b (false positive)	a + b
Test negative	c (false negative)	d (true negative)	c + d
Total	a + c	b + d	a + b + c + d

a = true positives (Persons who have disease and test positive on the screening test)
b = false positives (Persons who do not have disease and test positive on the screening test)
c = false negatives (Person with disease but with negative test)
d = true negatives (Persons who do not have disease and are test negative)
Sensitivity = a/a + c (Ability of the test to be correctly positive among those who are known to have the disease)
Specificity = d/b + d (Ability of the test to be correctly negative among those who are known to be without disease)
Positive predictive value = a/a + b (Ability of the test to correctly predict presence of the disease)
Negative predictive value = d/c + d (Ability of the test to correctly predict the absence of disease)

Specificity

The specificity of a test is the probability of test being negative in the absence of the condition. Specificity is equal to true negatives/(true negatives + false positives).

Positive Predictive Value

Positive predictive value can be defined as the probability of having a condition if the test is positive. It is the probability of an event in the active group divided by the probability of the event in the control group. It is calculated as follows:

$$\text{Positive predictive value} = \frac{\text{True positive}}{(\text{True positive} + \text{False positive})}$$

Negative Predictive Value

Negative predictive value can be defined as the proportion of individuals with negative tests that are free of the

JAYPEE BROTHERS

condition. In other words, it is the probability of the condition being absent in those where the test is negative. It is calculated as follows:

$$\text{Negative predictive value} = \frac{\text{True negative}}{(\text{True negative} + \text{False negative})}$$

Number Needed to Treat

Number needed to treat (NNT) can be described as the number of individuals in a general population, which need to be treated in order to prevent one case. NNT is the inverse of attributable risk [absolute risk reduction (ARR)]

$$\text{NNT} = \frac{1}{\text{ARR}}$$

ARR = CER – EER (event rate in the control group minus event rate in the experimental group)

The ideal NNT is 1, where everyone improves with treatment and no one improves with control.

The higher the NNT, less effective would be the treatment.

Likelihood Ratio

Likelihood ratio can be described as the likelihood that a given test result would be expected in a patient with the target disorder compared to the likelihood that the same result would be expected in a patient without the target disorder. Two versions of the likelihood ratio exist, positive likelihood ratio (LR^+) for positive test results and negative likelihood ratio (LR^-) for negative test results. LR^+ is a ratio of true positive rate (TPR) and false positive rate (FPR), whereas LR^- is the ratio of false negative rate (FNR) and true negative rate (TNR). LR^+ gives the odds of having the disease in relation to not having the disease when the test is found to be positive. Higher the LR^+, higher the likelihood of having the disease. LR^- is the reverse of LR^+. It gives the odds of having the disease in relation to not having the disease when the test is found to be negative.

$$LR^+ = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \text{ or } \frac{\text{TPR}}{\text{FPR}}$$

$$LR^- = \frac{1 - \text{Sensitivity}}{\text{Specificity}} \text{ or } \frac{\text{FNR}}{\text{TNR}}$$

Receiver Operating Characteristic Curve

The Receiver Operating Characteristic (ROC) curve is a fundamental tool for evaluation of the diagnostic test. It shows pairs of sensitivity and specificity corresponding to various cut-off points for the continuous test result **(Fig. 1)**. In this curve, the TPR (Sensitivity) is plotted as a function of the FPR (1–Specificity) for different cut-off points of a parameter. The area under the ROC curve (AUC) is a representative of how well a parameter is

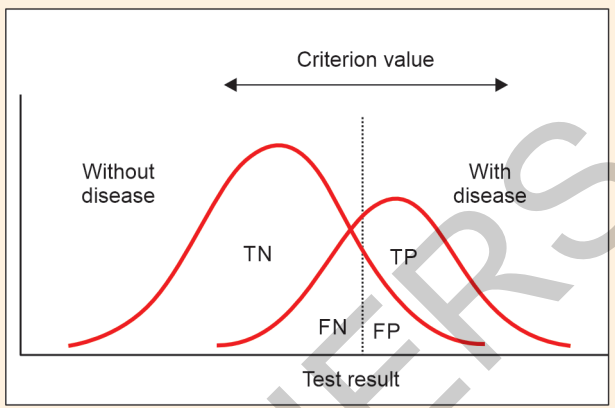

Fig. 1: Overlapping datasets generating false positives and negatives. (FN: false negative; FP: false positive; TN: true negative; TP: true positive)

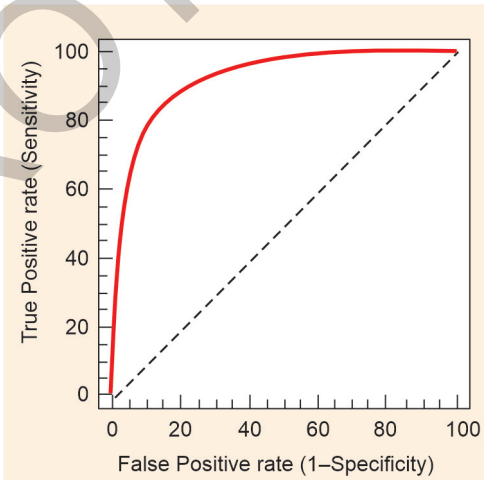

Fig. 2: Receiver Operating Characteristic curve passing through the upper left corner.

capable of distinguishing between two diagnostic groups (diseased or normal). A test that has a higher AUC is better. A test with the AUC of 1 is perfect. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner. This is indicative of 100% sensitivity and 100% specificity. The closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test **(Fig. 2)**. ROC curve is used for assessing the accuracy of a test to discriminate diseased cases from normal cases. It is also used for comparing the diagnostic performance of two or more laboratory or diagnostic tests.

Disease Causation

Disease is a dynamic process and denotes disharmony and deviation from the normal functioning of the various

JAYPEE BROTHERS

bodily functions. Besides the causative agent, there are many factors related to the man and environment, which may be responsible for disease causation. Criteria for assessing causation include:

- *Strength:* Strong associations are more likely to result in a disease in comparison to the weak ones.
- *Consistency:* Multiple studies finding the same thing are more likely to be causal.
- *Specificity:* If a variable is associated with a single outcome, and the outcome is associated with only a single possible cause, then the relationship is more likely to be causal.
- *Temporality:* Causes must definitely precede effects to suggest causation.
- *Biological gradient:* If an increased exposure is associated with an increased rate or severity of disease then causality is more likely.
- *Plausibility:* Hypotheses should sound reasonable (apparently valid).
- *Coherence:* Causal association is strengthened if epidemiological data fit in with pathology.
- *Experiment:* If the cause is removed and disease frequency declines, the likelihood of a causal link is strengthened.
- *Analogy:* If a similar association has been shown to be causal then the association under investigation is more likely to be causal.

Screening Program

The following criteria must be met before a screening program can be successfully implemented for disease detection and screening:

- The disease entity needs to be well characterised.
- Appropriate treatment or interventional strategies must be in place to provide a positive outcome due to early detection and for the test to be well validated and acceptable. It does not necessarily have to be painless **[4B]**.
- Natural history of the disease must be clearly defined **[4A]**.
- There should be an effective treatment available for the condition **[4C]**.
- There should be evidence that the screening test is effective at reducing morbidity or mortality **[4D]**.
- Benefits of the screening test should outweigh potential harm of diagnosing and treating an asymptomatic individual picked up through screening **[4E]**.

Rare Disease

A rare disease (RD) is a medical condition that occurs much less frequently than common diseases. There is no universally agreed-upon definition for RD. However, the 2021 Rare Disease Framework in the United Kingdom defines RD as one that affects fewer than 1 in 2,000 individuals.*

TABLE 2: Methodology for data measurement.

Continuous data	***Categorical data***
Single group:	*Single group:*
• *Normally distributed:* Mean, standard deviation, and 95% confidence interval • *Non-normal:* Median, range, and interquartile range	• Rate (risk) • 95% confidence interval
Two groups:	*Two groups:*
• Mean difference • 95% confidence interval	• Risk difference, 95% confidence interval • Relative risk, 95% confidence interval • Odds ratio, 95% confidence interval

Data Measurement

Methods for medical data measurement are mentioned in **Table 2**. Measurements may be performed for either continuous or categorical data, which has been classified into single or two groups.

Odds Ratio

The odds ratio (OR) can be defined as the measure of association between the exposure and outcome. It is the ratio of two odds under two conditions, i.e. odds of having the disease or not having the disease under exposure and non-exposure. The OR is generally used for assessing risk in case–control studies. It is calculated using the following formula:

$$OR = (a/b)/(c/d) \text{ or}$$

$$OR = (a \times d)/(b \times c) = ad/bc \text{ or}$$

$$\frac{\text{odds of an event occurring in the exposed group}}{\text{odds of an event occurring in the non-exposed group}}$$

	Cases	***Controls***
Exposed	a	b
Not exposed	c	d

Relative Risk

Relative risk (RR) is the ratio of two probabilities and is calculated when we study the outcomes for two groups.

$$\text{Relative risk} = \frac{\text{Incidence rate of exposed group}}{\text{Incidence rate of the unexposed group}}$$

The OR may or may not be equal to RR, depending upon the probabilities of outcome in the two groups:

- OR is equal to RR if probabilities of outcome are equal in the two groups (exposure and non-exposure)

*Lord BRS, Vaughan G, Mairi G. (2021). The UK rare diseases framework. [online] Available from https://www.gov.uk/government/publications/uk-rare-diseases-framework [Last accessed October, 2024]

JAYPEE BROTHERS

- OR is greater than RR if the probability of outcome in the exposed group is higher than that of the non-exposed group.
- OR is lower than RR if the probability of outcome in the exposed group is lower than that of the non-exposed group.

Attributable Risk (Absolute Risk Reduction)

Attributable risk (AR) refers to how many more cases are there in one group which is exposed to a risk factor in comparison to another group which is not exposed to a risk factor.

AR = Incidence rate of exposed group – Incidence rate of the unexposed group

Percentile

It is a measurement used in statistics indicating the value below which a given percentage of observations in a group would fall. This results when the frequency is divided into 100. For example, 10th percentile is the value below which 10% of the population lies and above which 90% of the population lies.

Confidence Interval

A confidence interval (CI) is a statistical range with a specified probability that a given parameter lies within the range. Thus, the CI refers to the target population, because 95% CI would imply that 95% of the population distribution would be contained within the CI **[6C]**. While interpreting a result from the sample, it is useful to express it within a range and enlist possible values it might have taken if other samples of the same size have been selected. This range of values is known as CI and is usually set as a percentage. Commonly used CI in the research studies is 95% CI, because this corresponds to the *p* value of 0.05 **[6A]**. CIs allow a true estimate of the breadth of a difference, reducing the likelihood of a type II error. The CIs decrease as the population sample increases **[6B]**. CIs can be both parametric and non-parametric **[7A]**. It is not a test of the null hypothesis **[7B]**. Rather, the confidence level is equivalent to 1 – the alpha level (*p* value). Therefore, if the significance level is 0.05, the corresponding CI would be 95%.

Ninety-five per cent CIs can be used for both distributional and distribution-free data. A 95% CI points towards the range of values within which the researcher can be 95% confident that the true population parameter lies. CIs increase the accuracy when comparing means with another population by looking at the spread of differences **[7D]**. A wide CI indicates that the estimate is imprecise. If the 95% CI crosses zero (the value of no difference), it indicates that the treatment has no effect **[7C]**. CI is calculated using the following formula:

$CI = \bar{x} \pm Z \times SEM$

CI = Confidence interval

$\bar{X}$ = Mean

TABLE 3: Z value of corresponding confidence interval.

Confidence level	*Z value*
90%	1.65
95%	1.96
98%	2.33
99%	2.58

Z = Confidence level value (**Table 3**; Z value varies based on the selected CI. The commonest CI used in clinical studies is 95% for which the Z value is 1.96, usually rounded off to 2.)

$$SEM = \frac{\sigma}{\sqrt{n}}$$

SEM = Standard error of mean

σ = Standard deviation

n = Sample size

Thus, 95% CI may be calculated as the mean ±1.96 times the SEM for population greater than 30 **[6D, 7E]**.

Probability

Probability (*p*) can be defined as the frequency with which an event would occur at random, given a total frequency of 1. A *p* value of 0.05 indicates that an event would occur randomly 5 out of 100 times or to put it simply, 1 out of 20 times. *p* value of 0.05 is traditionally considered as the cut-off point for statistical significance. *p* value of 0.05 or below is considered non-random or statistically significant **[6E]**. Unfortunately, statistical significance does not connect with clinical significance. For example, a doubling of risk from 0.0001% to 0.0002% may be completely insignificant, even though it is statistically highly significant.

Statistical significance is usually accepted at the *p* value <0.05. This implies that in <5% of occasions could such an observation be seen. A *p* value of 0.01 implies that the observation would be seen to occur by chance in 1 out of 100 occasions.

Power

Statistical power of the study is the ability of the study to detect a statistically significant difference between the two groups (i.e. if the difference really existed). Power of a study increases with the increasing sample size. This means that a large sample has a greater ability to detect a clinically important effect, if it exists.

Variables

Variable can be defined as a value or a characteristic that can be changed. It is generally useful for the purpose of research studies. Two types of variables have been described **(Fig. 3)**: numeric (variables whose values are in numbers) and categorical variables (those variables which can be put into categories). Numeric variables can be discrete or continuous. Two other types of numeric variables include interval and ratio variables. Categorical variables can be nominal or ordinal variables.

JAYPEE BROTHERS

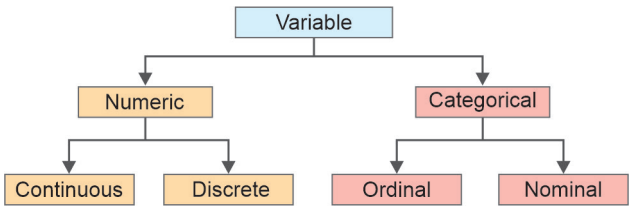

Fig. 3: Different types of variables.

Numeric Variables

Discrete variable: A discrete variable is a variable, which can take on only a certain number of values, i.e. they cannot have an infinite number of values. These variables do not change with time and circumstance, e.g. counts, per cent, or numbers.

Continuous variable: Continuous variables are variables, which can take up an infinite number of values. These variables change with time and circumstances. Continuous variables include blood pressure, height, weight, plasma concentrations, etc.

Interval variables: Variables are graded in equal increments (e.g. measurement of temperature in Celsius scale, time, and gestational age **[8C]**). There is no true zero point of these variables.

Ratio scales: It orders things and contains equal intervals similar to interval variables, but it also has a true zero point (e.g. height and weight).

Categorical Variables

Nominal variables: Variables where each category is described by a name and has no natural order. It is used for labelling variables, without any quantitative value, e.g. gender (male or female), hair colour (black, brown or white), ethnicity (Asian, African, Caucasians, British, etc.), control group versus the drug group, and blood group **[9B]**.

Ordinal variables: Categories can be placed in a specific order, e.g. parity (0, 1, 2). It can also measure non-numeric concept, e.g. pain sensation; 'Hot, hotter, hottest', or 'Agree, strongly agree, disagree'.

SINGLE BEST ANSWER

Q1. An experimental new MRI machine is being investigated to diagnose adnexal abnormalities in a population. The results are given below. What is the most likely sensitivity of the new MRI machine in diagnosing adnexal abnormalities?

MRI result	*Abnormalities present*	*Abnormalities not present*
Positive	100	25
Negative	10	150

A. 0.9%
B. 92%
C. 91%
D. 90%
E. 9.0%

Ans. C

Sensitivity = true positives / (true positives + false negatives)
100/(100 + 10) = (100/110) × 100 = 91%

Q2. A diagnostic test has a sensitivity of 90% and a specificity of 95%. Which of the following statements is likely to be true in this case?

A. 10% of positive tests are false positive.
B. 5% of negative tests are false negative.
C. Sensitivity is equal to true positives/(true positives + false negatives).
D. Sensitivity indicates the test will be negative in disease.
E. None of the above

Ans. C

Sensitivity implies probability of the test to be positive in those with the disease. Sensitivity of 90% implies that out of 100 cases with the disease, test is likely to be positive in 90 cases **[2A]**. On the other hand, specificity implies probability of the test to be negative in those without the disease. Specificity of 95% implies that out of 100 cases without the disease, test is likely to be negative in 95 cases **[2B]**. Sensitivity indicates the proportion of individuals with positive screening tests that actually have the disease **[2D]**. Sensitivity is equal to true positives/(true positives + false negatives) **[2C]**.

Q3. An experimental new kit is being investigated to detect ovulation in the infertile women of reproductive age group. The results are given below. What is the most likely specificity of the new kit in diagnosing ovulation?

Kit result	*Ovulation present*	*Ovulation not present*
Positive	800	50
Negative	200	950

A. 95%
B. 80%
C. 85%
D. 90%
E. 84.2%

Ans. A

Specificity = true negatives / (true negatives + false positives)
950 / (950+50) = (950/1,000) × 100 = 95%

Q4. The National Screening Committee criteria for establishing a population screening programme do not include which of the following?

A. Clearly defined natural history of the disease.
B. Should be entirely painless.
C. There should be an effective treatment available for the condition.
D. There should be evidence that the screening test is effective at reducing morbidity or mortality.
E. Benefits of the screening test should outweigh potential harm of diagnosing and treating an asymptomatic individual picked up through screening.

Ans. B

Q5. A press release by a pharmaceutical company regarding their new drug includes the following statement, "with the widespread use of this new medication, risk of pre-eclampsia will be cut in half as demonstrated by the pre-eclampsia OR of 0.48 [95% CI, (0.29, and 1.8)]. What is the most accurate statement that can be made using the reported OR and CI?

A. Pre-eclampsia increased in the control group, but this is not statistically significant.
B. Pre-eclampsia decreased in cases, but this is not statistically significant.
C. Pre-eclampsia increased by 52% in the control group versus cases with statistical significance.

JAYPEE BROTHERS

D. Pre-eclampsia decreased by 52% in the cases versus control with statistical significance.
E. There was no effect of this drug in cases.

Ans. B

Odds ratio is a comparative measure used to evaluate the effectiveness of an intervention against a control or placebo group.

- *OR > 1:* The control is more effective than the intervention.
- *OR < 1:* The intervention is more effective than the control.
- *OR = 1:* There is no difference in effectiveness between the two groups.
- *Confidence interval crossing 1 in this case* (e.g. 95% CI: 0.29–1.8): Indicates no statistically significant difference between the groups.

Q6. In comparing CIs with *p* values, which of the following statements is correct?

A. 95% CIs are equivalent to a *p* value of 0.95.
B. As the sample size increases, the CI will increase.
C. Confidence intervals refer to the target population.
D. Confidence interval is approximately equal to the standard deviation.
E. The *p* value >0.05 implies that the result is statistically significant.

Ans. C

Q7. Which of the following is true regarding 95% CI?

A. Can only be used in parametric data.
B. It is a test of the null hypothesis.
C. If zero difference lies within the 95% when comparing two groups to a treatment, it indicates the treatment is very effective.
D. Is not useful when comparing data with another population.
E. It is calculated at ±1.96 times the SEM.

Ans. E

Q8. Gestational age is which type of data?

A. Nominal
B. Ordinal
C. Interval
D. Regression
E. Ratio

Ans. C

Q9. What type of variable is ABO blood group?

A. Ordinal
B. Nominal
C. Continuous
D. Ratio
E. Interval

Ans. B

Q10. 700 women were enrolled in the study for the assessment of PPV for testing preterm labour. 320 women were test positive. Out of them, 80 had preterm labour. 380 were test negative and out of them 20 had preterm labour. What is the PPV of the test?

A. 25%
B. 50%
C. 70%
D. 10%
E. 100%

Ans. A

In this case, total number of women who tested positive (i.e. true positive + false positive) = 320.
Out of the 320 women who tested positive, only 80 had preterm labour. Therefore, 80 were true positive.

$$\text{Positive predictive value} = \frac{\text{True positive}}{(\text{True positive} + \text{false positive})}$$

Positive predictive value = 80/320 × 100 = 25%

Q11. What is the negative predicative value in the following data?

	Preterm labour	*Normal labour*
Fibronectin positive	5	20
Fibronectin negative	15	60

A. 75%
B. 80%
C. 60%
D. 70%
E. 90%

Ans. B

Negative predictive value is defined as the probability of the condition being absent in those where the test is negative. True negative in this case is 60. These 60 individuals were fibronectin negative and also did not have preterm labour. False negative were 15. These individuals tested negative for fibronectin, but still had preterm labour.

$$\text{Negative predictive value} = \frac{\text{True negative}}{(\text{True negative} + \text{false negative})}$$

Negative predictive value = 60/60 + 15 = 60/75 × 100 = 80%

Q12. A publication reports the outcome of a new therapy in a placebo-controlled primary prevention of venous thromboembolism (VTE). 200 patients were randomised to receive the new therapy and 200 allocated to placebo. The study was completed over a 2-year period. In the placebo group, 20 had VTE and in the group treated with the new therapy, there were 10 who had VTE. What is the NNT to prevent one VTE over the course of this study?

A. 2
B. 4
C. 10
D. 20
E. 40

Ans. D

Number needed to treat (NNT) is the inverse of absolute risk reduction (ARR).

ARR = CER (control event rate) – EER (experimental event rate).

In this case, in the placebo group, out of 200 patients, 20 had VTE, which is equal to the event rate of 20/200 in the control group. In the treatment group out of 200 patients, 10 had VTE, which is equal to the event rate of 10/200 in the experimental group.

ARR = CER – EER
= 20/200 – 10/200
= 10/200
NNT = 1/ARR
= 200/10
= 20

Q13. A prospective randomised trial was carried out to determine if the incidence of incisional hernia is higher after vertical incision in comparison to the Pfannenstiel incision following a caesarean delivery. The study enrolled patients who had undergone caesarean delivery between January 2021 and December 2023. They were split into two groups: Vertical and Pfannenstiel, with all patients being operated on by the same surgeon. All other pre- and post-operative parameters were identical. Patients were monitored for 2 years after surgery. All patients were thoroughly evaluated

JAYPEE BROTHERS

for incisional hernias and the results were documented. Given below are the results. Calculate the OR.

Incision group	*Incisional hernia*	*No incisional hernia*	*Total*
Vertical group	90	450	540
Pfannenstiel group	100	1,000	1,100
Total	190	1450	330

A. 2.0 B. 0.2
C. 0.02 D. 1.2
E. 1.5

Ans. A

Odds for developing incisional hernia in the vertical group = 90/450 = 0.2
Odds for developing incisional hernia in the Pfannenstiel group = 100/1,000 = 0.1
Odds ratio = 0.2/0.1 = 2
The OR > 1 indicates that the rate of incisional hernia is more in vertical group. (Odds ratio = 1 indicates there is no difference in the rate of event)

Q14. A research study showed that an allergic event occurred in 100 individuals given a new medication, and the event did not occur in 200 individuals given the new medication. It also found that the event occurred in 400 individuals not given the medication, and the event did not occur in 800 individuals not given the new medication. What is the OR of the event occurring?

A. 0.25 B. 0.50
C. 0.75 D. 1.0
E. 1.5

Ans. D

Odds of the event occurring in the exposed group = (number with an event in the exposed group)/(number without an event in the exposed group) = 100/200
Odds of an event occurring in the non-exposed group = (number with an event in the non-exposed group)/(number without an event in the non-exposed group) = 400/800

$$\text{The odds ratio} = \frac{\text{Odds of an event occurring in the exposed group}}{\text{Odds of an event occurring in the non-exposed group}}$$

$$\text{The odds ratio} = \frac{100/200}{400/800}$$

$$= (1/2)/(1/2) = 1$$

Q15. Ten patients with APGAR score <7 developed symmetrical and asymmetrical IUGR each. Ten patients with APGAR score of 7 developed asymmetric IUGR, while 40 developed symmetric IUGR. Calculate the OR of asymmetrical intrauterine growth restriction (IUGR) and symmetrical IUGR for APGAR <7.

A. 4 B. 2.5
C. 16 D. 1.5
E. 1

Ans. A

In this question for calculation of odds of asymmetrical IUGR and symmetrical IUGR for APGAR <7, the grid for calculation of OR would be as follows:

	Asymmetrical IUGR	*Symmetrical IUGR*
APGAR <7	10 (a)	10 (b)
APGAR = 7	10 (c)	40 (d)

= (ad)/(bc)
= (10 × 40)/(10 × 10) = 4

Q16. A survey conducted by Third National Health and Nutrition Examination evaluates relationship between depression and obesity amongst both males and females. In concern to the given below data, which of the following statements correctly explains the overweight value of 0.86?

BMI category, kg/m²	*Unadjusted Odds ratio*	*95% confidence interval*
Normal weight (BMI 18.5–24.9)	1.00	
Underweight (BMI <18.5)	1.17	0.49–2.80
Overweight (BMI 25.0–29.9)	0.86	0.53–1.41
Obese (BMI ≥30)	1.88	1.02–3.46
Class 1 (BMI 30–34.9)	1.28	0.64–2.56
Class 2 (BMI 35–39.9)	1.76	0.78–3.95
Class 3 (BMI ≥40)	4.98	2.07–11.99

A. The odds of having major depression among overweight individuals' are 14% lower than the odds of having major depression for normal weight individuals.
B. The odds of having major depression among overweight individuals' are 14% higher than the odds of having major depression for normal weight individuals.
C. The probability of major depression in an overweight individual is 0.86.
D. The odds of major depression in an overweight individual is 0.86.
E. The given data is not enough to interpret the overweight value.

Ans. A

Odds ratio helps in estimation of effects of other variables [different body mass index (BMI) values] with depression. Higher trends of odds ratios in relation to depression have been observed with increasing levels of obesity. The odds of overweight people suffering from depression is 14% less, i.e., 1.00 – 0.86 = 0.14, i.e., 14% more than normal weight people.

Q17. In a sample of 72 diastolic blood pressure reports amongst the women with preeclampsia, the SEM was 0.53 mmHg and the mean value was 84 mmHg. What would be the 95% confidence limits?

A. 86.96 to 89.04 mmHg B. 76.96 to 89.04 mmHg
C. 86.96 to 99.04 mmHg D. 56.96 to 79.04 mmHg
E. 82.96 to 85.04 mmHg

Ans. E

$CI = \bar{x} \pm Z \times SEM$
The sample mean plus or minus 1.96 times its standard error gives the following two figures:
84 + (1.96 × 0.53) = 85.04 mmHg
84 – (1.96 × 0.53) = 82.96 mmHg.

STATISTICS

DISTRIBUTION

The most commonly used type of distribution in medical statistics is the normal or Gaussian distribution. It forms the basis for many aspects of medical statistics. It is

JAYPEE BROTHERS

commonly used for large samples of observation (>20). Another type of distribution is the t distribution, which forms the basis for the Student's t-test.

Normal (Gaussian) Distribution

This is one of the most important and widely used methods for probability distribution in medical statistics. It generates a bell-shaped curve if it is plotted in terms of probability **(Fig. 4)**:

- Sixty-eight per cent of observations lie between 1 standard deviation (SD) on the either side of the mean value.
- Ninety-five per cent of the area under the curve lies between the points that fall 1.96 SD on the either side of the mean value **[2A]**. In this case, 2.5% of individuals will be above **[5E]** and 2.5% below 1.96 SD on either side of the mean **[3A, 4D]**.
- Ninety-nine per cent of the values respectively fall between 2.58 SDs on the either side of the mean value.

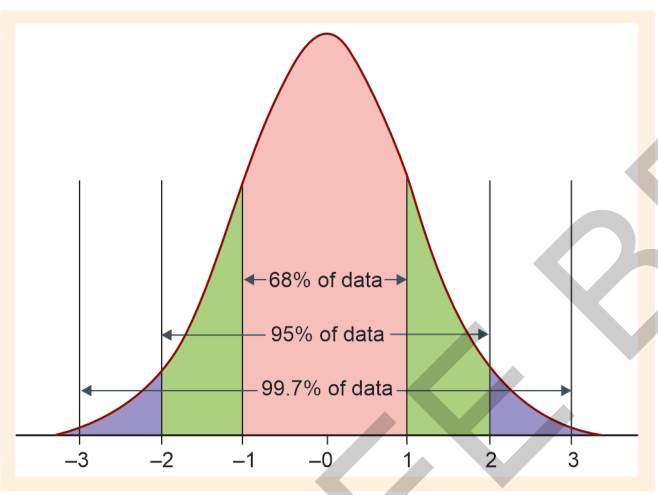

Fig. 4: Bell-shaped curve demonstrating normal/Gaussian distribution.

In a normally distributed data, the probability of attaining a value higher than 2 SDs above the mean is approximately 1 in 40 (p = approximately 0.025) **[4C]**. This implies that most of the population lies between 2 SDs on either side of the mean **[5D]**.

If a characteristic is normally distributed in a population, the mean, mode, and median are all equal and therefore they coincide **[3B, C, D, 5A, B]**. In case a characteristic is normally distributed, number of individuals below and above the mean are equal **[3E, 5C]**. Data from a normal distribution are suitable for parametric tests without prior transformation.

Different tests for null hypothesis have been developed for the data that are hypothesised to follow normal distribution. Some of these tests are Z-test for normal distribution, t-test for student's t-distribution, and f-test for f-distribution. Pearson's chi-squared test can be used for data, which does not follow normal distribution. Chi-squared test is the standard method for comparing distributions, e.g. between the observed and expected frequency of a given event **[6D]**. Chi-squared test is derived by squaring the normal distribution **[7D]**.

Student's t-test is a standard test for comparing the difference between the two means. On the other hand, f-test is the standard method for comparing the size of variance.

Skewed Distribution

In case of normal distribution, the numbers of individuals above and below the mean will be equal; if they are not equal then the distribution is said to be skewed. For positively skewed data, the mean is larger than the median which is always greater than the mode, and the mean usually lies to the right of the mode **[8A]**.

On the other hand, for negatively skewed data, the median is larger than the mean **(Fig. 5)**. When summarising skewed data, it is better to quote the median and the interquartile range, whereas while summarising

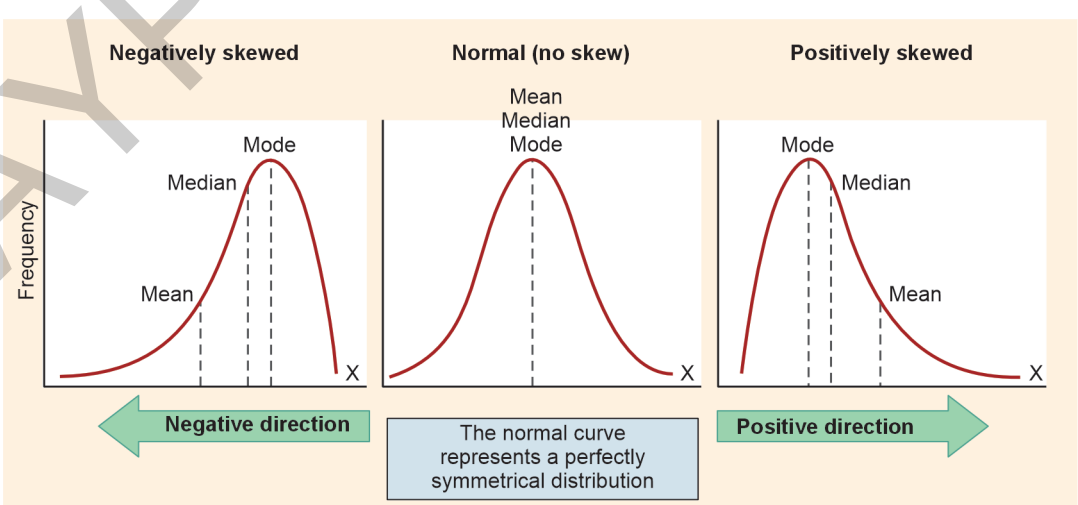

Fig. 5: Examples of normal and skewed distribution.

JAYPEE BROTHERS

the normally distributed data, the parameters most commonly used include the mean and the SD **[7B]**. In skewed data, the geometric mean is the most appropriate measure **[8B]**.

INDICES OF CENTRAL TENDENCY

Mean

The mean (denoted by the Greek letter mu, μ) is the average of the data. In case of normally distributed data (giving a bell-shaped Gaussian curve), the mean is same as the median and mode. However, in case of data, which is not normally distributed, the mean, mode, and the median may be different values.

The arithmetic mean is calculated by using sum of the values [(a1+ ...+aN)/N]. On the other hand, geometric mean is defined as the nth root of the product of n numbers [n√(a1.a2.a3....aN)]. It indicates the central tendency for a set of numbers by using the product of their values. Hence, the geometric mean will always be less than the arithmetic mean **[9B]**. The arithmetic mean is preferred in normal statistics as it generally represents the average. Also, it is preferred in the data which is normally distributed **[9A]**.

Median

The median is the middle observation value **[9C]**. Therefore, half the observations lie above this value **[6C]** and the other half lies below it **[4B]**. Sometimes, it may be more appropriate to quote the median rather than the mean, because it is less sensitive to the large outlying values. Also, in case of skewed data, it is better to quote the median rather than the mean.

Mode

The mode is the observation that occurs with the greatest frequency **[4A, 6A, 9E]**. Mode refers to the most frequently encountered value and in a normally distributed data coincides with the mean and median values.

VARIATION

Range

Range denotes the total collection of values between the largest and the smallest observation. This indicates how widely varied the data is. It is calculated by subtracting the lowest value from the highest value.

Interquartile Range

This is similar to the median although the lower value (Q1) indicates that 25% of the observations lie below it and the upper value (Q3) indicates that 25% of the observations lie above it. Thus, the interquartile range represents the central 50% range of values (Q3 – Q1). For the calculation of interquartile range, firstly the data is arranged in an ascending or descending order. The list is then cut into four equal parts by making three cuts, Q1, Q2, and Q3

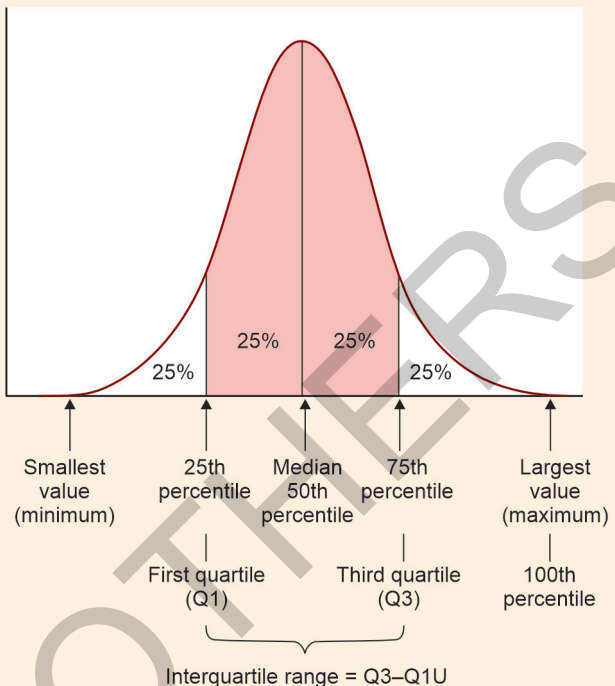

Fig. 6: Diagrammatic representation of interquartile range.

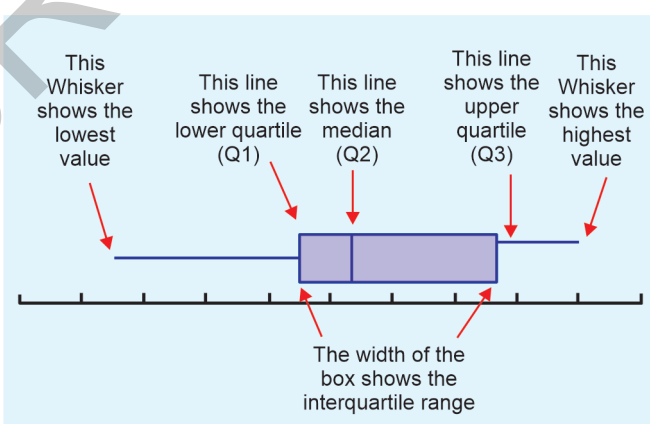

Fig. 7: Box and Whisker chart.

(where quartiles are the cut). As previously described, interquartile range = Q3 – Q1 **(Fig. 6)**.

Box and Whisker Chart

The interquartile range can also be demonstrated with help of a box and Whisker chart. A box and whisker plot is used for depicting the groups of numerical data through their quartiles **(Fig. 7)**. It uses medians (or quartiles) to divide the population into four groups. The median is known as Q2, because 50% of values lie above it and 50% of the values lie below it. This is further divided into four groups with help of upper quarter or Q3 and lower quarter or Q1. Vertical lines extend perpendicularly from the boxes and are known as whiskers. These lines indicate variability outside the upper and lower quartiles.

JAYPEE BROTHERS

Variance

The variance can be defined as the spread of the observations. This is an indication of variability of the observations. To calculate the variance, each observation is subtracted from the mean, squared, added-up, and divided by the number of observations minus one.

Standard Deviation

Standard deviation is represented by the Greek letter sigma (σ). It is known as the root mean square deviation and is the square root of the variance **[7A]**. Therefore variance is equal to the square of SD **[2E]** and is a measure of observation variability **[11C]**. The SD provides a good indication of distribution about the mean **[7E, 9D]**. It is a measure of observation variability and is greater than the SEM **[2B, 7C]**.

Standard Error of Mean

Standard error of mean can be defined as the SD of the sampling distribution of a statistical measurement, most commonly mean **[11D]**. This can be defined as SD divided by √n or √n—1 **[11A]**.

$$\text{SEM} = \frac{\sigma}{\sqrt{n}} \text{ Or } \frac{\sigma}{\sqrt{n-1}}$$

where SEM is standard error of mean
σ is SD
n is the size (number of observations)

This gives an estimate of the limits to the mean of the total population from which the sample observations were drawn. In other words, SEM quantifies the accuracy with which the researcher knows about the true mean of the population. By definition, SEM is always smaller than the SD **[11B, E]**. Also, SEM gets smaller as the samples become larger since mean of the large sample is likely to be closer to the true population than is the mean of a small sample. If testing to see if there is a difference between two population means (for example, t-test), then t = mean/SEM.

CORRELATION

This is the relationship between the two variables (between a dependent variable and one or more independent variables).

Linear Regression

This is the standard method for determining whether two variables are correlated or not. Linear regression can be of two types: Simple and multiple linear regression. Correlation between one independent variable and one dependent variable is known as simple linear regression. Correlation between more than one independent variable and one dependent variable is known as multiple linear regression.

Coefficient of Correlation

This is indicative of the degree of correlation between the two variables, X and Y. The symbol denoting coefficient of correlation is r. Correlation coefficients vary between –1 and +1. A value of 1 is indicative of perfect correlation, whereas the value of 0 indicates no correlation at all.

SINGLE BEST ANSWER

Q1. Identify the correct statement about normally distributed data.

A. Approximately 97.5% of observations fall within 2 SDs of the mean.
B. The Student's t-test is not applicable for comparing two means.
C. A 2 × 2 table analysis results in one degree of freedom.
D. Non-parametric tests are not suitable for comparing data sets.
E. None of the above

Ans. C

About 68% of data is within one SD, 95% within two SDs, and 99.73% within three SDs **[1A]**. The Student's t-test can compare two means **[1B]**. Degrees of freedom for a 2 × 2 table is calculated as (number of columns – 1) × (number of rows – 1), which equals 1 **[1C]**. Non-parametric tests can be used for comparing data sets **[1D]**.

Q2. Which of the following statements related to statistics is not correct?

A. In a normally distributed population, 95% of the values fall in the range of the mean plus or minus two SD.
B. Standard error and SD are synonymous.
C. The chi-squared test can be applied to non-parametric data.
D. The statement '*p* is <0.01' means that there is <1% likelihood of an event having occurred by chance.
E. Variance is equal to the square of the SD.

Ans. B

Chi-square test is a non-parametric test **[2C]**. A *p* value of less than 0.01 implies that the observation would be seen to occur by chance in less than 1 out of 100 occasions **[2D]**.

Q3. Which of the following statements is correct if a characteristic is normally distributed in a population?

A. 5% of individuals will be more than two SDs from the mean.
B. The mean will be greater than the median.
C. The median will be less than the mean.
D. The mode will be greater than the median.
E. The numbers of individuals above and below the mean will be equal.

Ans. E

Q4. Which of the following statements is correct regarding statistical distribution?

A. The mode is the value that occurs the most frequently.
B. The median is that point on the scale of measurement above which lie exactly half the values and below which lie the other half.
C. Of a normally distributed variable, the probability of attaining a value higher than two SDs above the mean is approximately 1 in 40 (*p* = approximately 0.025).
D. Having a normal distribution, approximately 95% of the values will lie within the range between (mean +2 SDs) and (mean –2 SDs).
E. All of the above

Ans. E

JAYPEE BROTHERS

Q5. Which of the following is true if a characteristic is normally distributed in a population?

A. The median will be less than the mean.
B. The mean will be of lesser value than the mode.
C. The numbers of individuals above and below the mean will be equal.
D. This implies that most of the population comprises normal individuals.
E. 20% of individuals will be more than two SDs from the mean.

Ans. C

Q6. In statistics, which of the following statements is not correct?

A. The mode of a distribution is the most frequently occurring value.
B. Wilcoxon's rank sum test is a non-parametric test.
C. In any set of observations, half of the observations are greater than the median.
D. The chi-square test compares the observed and expected frequencies of an event.
E. Infant mortality rate (IMR) is the number of infants dying during the first year per 10,000 live births.

Ans. E

Wilcoxon Signed-Rank test is an example of a non-parametric test, which helps in comparing two sets of observations on a single sample **[6B]**. IMR can be defined as number of infant death (deaths during first year of life) per 1,000 deliveries **[6E]**.

Q7. Which of the following is true regarding the SD of a group?

A. Is the square of the variance.
B. Is a valid statistical parameter for observations that do not have a normal distribution.
C. Is numerically equal to the SEM.
D. Can be used for the calculation of chi-squared test.
E. Is a measure of the scatter of observations about the mean.

Ans. E

Q8. Which of the following statements regarding statistical terms is true?

A. In a positively skewed distribution, the mean always lies to the left of the mode.
B. In distributions which are markedly skewed, the arithmetic mean is a more appropriate measure than the geometric mean.
C. In non-parametric data, the mode is usually different in value from the mean.
D. All of the above
E. None of the above

Ans. C

Non-parametric tests are those in which no assumptions can be made regarding the defining characteristics of the population distribution from which one's data is drawn. The population on which the non-parametric tests are performed are unlikely to have a normal distribution. Therefore, in such a population, the mean, mode, and median, all are likely to be different **[8C]**.

Q9. Which of the following statements regarding statistical evaluation is not correct?

A. The arithmetic mean is the measure to be preferred in data which are symmetrically distributed.
B. The geometric mean is always less in value than the arithmetic mean.
C. The median is also called the measure of central value.
D. The SD is a good indication of distribution about the mean.
E. The value of the variable which occurs with the least frequency is the mode.

Ans. E

Q10. In a random sample population of 9 women, the average resting pulse rate is 80 beats/min, while the sample SD is s = 5. Calculate the standard error of the sample mean.

A. 1.667
B. 0.557
C. 0.745
D. 2.778
E. 4.035

Ans. A

The standard error of the sample mean depends on both the SD and the sample size, as shown by the simple formula: SE = SD/√(sample size).

So, SE = 5/√9
= 1.667

Q11. Which of the following is true regarding the SD?

A. SD = SEM/(square root of n)
B. The SD equals the SEM in non-parametric tests.
C. The SD is a measure of observation variability.
D. The SEM determines the accuracy of measurement of the observations.
E. The SD is equal to the SEM.

Ans. C

Q12. Which of the following statements concerning statistical tests is true?

A. Wilcoxon's rank test needs equal sample sizes.
B. Student's t-test is a non-parametric test.
C. Correlation coefficients vary between −10 and +10.
D. Sigma (σ) is the symbol denoting coefficient of correlation.
E. y = a + bx is a regression equation.

Ans. E

Wilcoxon's rank test is a non-parametric test, compares two sets of observations on a single sample **[12A]**. The line of regression of Y on X is given by Y = a + bx, where a and b are unknown constants. This equation is used to predict the unknown value of variable Y when value of variable X is known **[12E]**. Parametric tests are preferred over the non-parametric tests, because fewer observations are required to provide evidence in favour of the hypothesis if it is true, e.g. student's t-test is a typical parametric test **[12B]** for comparing the mean values of a continuous variable between two groups. The symbol denoting coefficient of correlation is r **[12D]**. Correlation coefficients varies between −1 and +1 **[12C]**.

Q13. For the data series: 2, 1, 6, 4, 2, which of the following statements is true?

A. The mean is 6.
B. The mean is always identical to the median.
C. The median is 3.
D. The mode is 4.
E. The SD is 2.

Ans. E

Mean = addition of number of observation divided by total number of observations

Therefore, Mean = (2+1+6+4+2)/5 = 15/5 = 3

The median is the middle observation value, which in this case is 2. Mode is the most frequently occurring value which again in this case is 2.

JAYPEE BROTHERS

Variance = Each observation is subtracted from the mean, squared, added-up, and divided by the number of observations minus one.

$$\text{Variance} = \frac{(3-2)^2 + (3-1)^2 + (3-6)^2 + (3-4)^2 + (3-2)^2}{4}$$

$$= \frac{1+4+9+1+1}{4} = \frac{16}{4} = 4$$

Variance = 4
Standard deviation = √variance
Standard deviation = √4 = 2

Q14. In a trial of a new therapeutic agent, the required sample size varies with which of the following?

A. Experience of the investigator
B. Level of acceptance of failing to discover a true effect
C. Level of statistical significance required
D. Type of statistical test to be employed
E. None of the above

Ans. C

Sample size can be considered as an important parameter for determining statistical significance. As the sample size increases, the statistical significance increases.

Q15. A study is conducted to investigate whether there is a relationship between soy milk and improvement of menopause symptoms. The null hypothesis represents which of the following statements?

A. A relationship between soy milk and improve menopausal symptoms
B. A positive relationship between soy milk and improvement of menopausal symptoms
C. A negative relationship between soy milk and improvement of menopausal symptoms
D. A difference between soy milk and improvement of menopausal symptoms
E. No relationship between soy milk and improvement of menopausal symptoms

Ans. E

Null hypothesis is a hypothesis which states that there is no difference between the two sets of data. Therefore, the null hypothesis regarding relationship between soy milk and improvement of menopausal symptoms states that there is no difference regarding improvement of menopausal symptoms between the two groups: One which is taking soy milk and the other group which is not taking soy milk. In other words, this implies that there is no relationship between intake of soy milk and improvement of menopausal symptoms.

Q16. The number of hours worked per week for a sample of students is shown below.

Hours	1	2	3	4	5	6	7	8
Students	12	9	29	28	15	10	18	15

What is the mode of the above data?

A. 10
B. 4
C. 3
D. 29
E. 28

Ans. C

From the analysis of above-mentioned data, it can be observed that the maximum number of students (29 in this case) studied for 3 hours. Therefore, the most frequently occurring value or the mode is 3.

Q17. What is the interquartile range (IQR) for the given data?

5, 2, 6, 3, 9, 4, 6, 6, 7, 8, 7

A. 6
B. 5
C. 3
D. 4
E. 7

Ans. C

The IQR for this set of data can be calculated through the following steps:

- The values of the data set are arranged on the ascending order: 2, 3, 4, 5, 6, 6, 6, 7, 7, 8, 9
- The median as calculated is 6.
- The medians of the upper and lower portions of the data are then calculated:
  Median of upper portion – 4
  Median of lower portion – 7
- The difference of the upper and lower medians is: 7 – 4 = 3

This value gives the interquartile range, which is 3 **(Fig. 6)**.

Q18. In a set of given data, mean BMI is 25% and SD is 5, what is the percentage of obese class 2 and 3?

A. 2.5%
B. 5%
C. 16%
D. 32%
E. 34%

Ans. A

According to the World Health Organization (WHO), the distribution of normal weight and different classes of obese individuals according to their BMI values are as follows:

BMI value	*Classification*
<18.5	Underweight
18.5–24.9	Normal weight
30.0–34.9	Class I obese
35.5–39.9	Class II obese
>40.0	Class III obese

Here, the given mean BMI is 25%, hence μ = 25. SD, σ = 5.

Hence, for class II obese individuals, it would be μ + 2σ = 25 + 2 × 5 = 35.
Similarly, for class III obese individuals, it would be μ + 3σ = 25 + 3 × 5 = 40.
In this question, we need to find out the number of individuals that lie between (μ + 2σ) and (μ + 3σ).
According to the normal Gaussian distribution, 99.74% of people would lie between (μ – 3σ) and (μ + 3σ). Also 95% of people would lie between (μ – 2σ) and (μ + 2σ). Hence subtracting the two values (99.74–95%), we get 4.74%, which represents the number of people lying between (μ + 2σ) and (μ + 3σ) and (μ – 2σ) and (μ – 3σ). Now, dividing it by 2 we get a value of 2.37% which is near to 2.5% and represents the number of people lying between (μ + 2σ) and (μ + 3σ). Hence, the correct answer is 2.5%.

Q19. A study compares the birth weight of the 1,500 babies delivered by vacuum with the 2,000 babies delivered by caesarean section in a year on a labour ward. What would be the most suitable statistical test to use in this comparison?

A. Analysis of variance
B. Friedman test
C. Paired T-test
D. Pearson correlation coefficient
E. Unpaired T-test

Ans. E

JAYPEE BROTHERS

for detailed information on obstetrics & gynaecology books, visit our website www.jaypeebrothers.com, for detailed information on obstetrics & gynaecology books, visit our w

Textbook for
MRCOG-1
Basic Sciences in Obstetrics & Gynaecology

Textbook for MRCOG-1: Basic Sciences in Obstetrics and Gynaecology is an essential resource for candidates preparing for the MRCOG Part 1 examination. This comprehensive guide covers the fundamental sciences that underpin obstetric and gynaecological practice, with content meticulously updated in four colours to reflect the latest guidelines and best practices. The third edition introduces a dynamic blend of learning tools, including two online mock papers and nearly 1000 interactive flash cards, to facilitate thorough revision and self-assessment. Updated with the latest Single Best Answer (SBA) and image-based questions, this edition offers a complete and engaging approach for mastering the MRCOG Part 1 syllabus, making it an indispensable companion for examination preparation.



Richa Saxena is an obstetrician gynaecologist, an RCOG Associate with more than 1.5 decades of academic experience and an esteemed author of many best-selling books for undergraduate and postgraduate medical students as well as MRCOG aspirants (inclusive of the three parts of the MRCOG examinations). She is Co-Author of the book, *Jeffcoate's Principles of Gynaecology, 9th edition*, considered as the bible of gynaecology for the postgraduates. She is also the Mentor and Director of MRCOG online course modules on an E-learning platform, DigiNerve, for all the parts of the MRCOG examination. She did both her graduation in medicine as well as postgraduation (MD obstetrics and gynaecology) from New Delhi, India. She has been outstanding in academics throughout the medical course and subsequently only followed the upside of growth curve through hard-work and determination. Besides being a prolific author to more than a dozen books in obstetrics and gynaecology, she is an excellent teacher, revolutionary thinker, exceptional doctor, and an ardent promoter of women's health. Let's talk woman's health is her initiative dedicated towards ***E***mpowering and ***E***nlightening women through ***E***ducation and ***E***-consultation. As a step towards patient education and empowerment, she has taken many courses on Udemy related to the topics of importance from the perspective of women's health.

She is the recipient of the Rotary Vocational Excellence Awards (2017–2018) for her exemplary service towards medical education in the speciality of obstetrics and gynaecology. She has also received the Women of Substance Award (category healthcare) for the year 2022 and Dr APJ Abdul Kalam Health Awards (2024) for excellence in obstetrics and gynaecology.

To know more about her, kindly visit her website, www.drrichasaxena.com.

Printed in India

Available at all medical bookstores or buy online at www.ejaypee.com

JAYPEE BROTHERS
Medical Publishers (P) Ltd.
EMCA House, 23/23-B, Ansari Road,
Daryaganj, New Delhi - 110 002, INDIA
www.jaypeebrothers.com

Join us on facebook.com/JaypeeMedicalPublishers
Follow us on instagram.com/JaypeeMedicalPublishers

Shelving Recommendation
OBSTETRICS AND GYNAECOLOGY

ISBN 978-93-6616-080-1