MedicalBiostatistics.com

Indicators, Indexes and Scores

The terms indicators, indexes and scores seem to be widely used *interchangeably* for tools of measurement. Technically though, these terms have distinct meaning. Broadly speaking, indicator is a quantitative univariate measurement for specific aspect of health. Index combines two or more indicators, and score may incorporate even the qualitative characteristics such as signs and symptoms. There are exceptions. Intensity of pain on visual analogue scale is called score although this is univariate.

Indicators

Factor is a characteristic and indicator is a tool to measure its level. This assumes that the characteristic is graded and it has a level. This is not always true. Site of cancer and gender of a person are characteristics that cannot be assigned a grade at individual level. For groups, though, percentage of patients with cancer of different sites, and sex ratio in a group of arthritis cases are quantities. These are indicators at group level.

The term indicator is generally used when the focus is on a *specific* aspect of a characteristic. Severity of disease in a patient can be assessed by quality and quantity of pain, by inability to perform essential functions of life, by prognostic implications, by the chance of death within a week, etc. Each of these is an indicator since each concerns with a particular aspect of severity. Indicator is a univariate assessment and perhaps the most direct measurement of the particular aspect of characteristics of interest. Indicators provide the exactitude that you may like to have in your assessment, or you may like to have in the literature you consult to update knowledge.

Despite all the advances that medical science has made, there are factors that defy any kind of measurement. Psychological feelings such as compassion, love and affection come under this category. Metric scale is just not available for these factors and even ordinal assessment as none, mild, moderate and intense is difficult. Psychological feelings are important considerations in psychiatric illnesses and they also affect recovery rate of some somatic disorders. In the absence of any worthwhile indicator, such factors tend to be ignored in clinical practice or medical research, or at best tentatively assessed in an approximate manner.

Merits and Demerits of Indicators

Critiques may be concerned with limited utility of indicators because they focus on one particular aspect of health and ignore the other aspects, howsoever related. Case-fatality rate is an indicator of the chance of death but is oblivious to the pain and suffering it brings to the patient and the family. Weight of a person is an indicator that by itself looses importance unless related to height. Glomerular filtrations rate (GFR) is better interpreted in the context of creatinine excretion. At the same time, though, focus of an indicator on a particular aspect is its greatest strength. Case-fatality does measure the most important aspect of prognosis that is important for both patient and the doctor; weight does help when monitored in the same person over a period of time; and GFR level does indicate about functioning of kidneys. They are good indicators as stand alone, although putting them in the context of other related parameters does help in providing a better interpretation.

Choice of Indicators

Quite often multiple indicators are available for apparently the same characteristic and you may have to make a choice. Triglyceride, high-density lipoprotein (HDL) and low-density lipoprotein (LDL) are indicators of different lipids. Many times these are considered together but if one has to choose, the selection would depend on the relevance for the condition of the patient and the specific aspect of lipid of interest. For example, triglyceride level can be a good indicator of lipoprotein—called natural fats, and LDL level can be a good indicator of risk of developing atherosclerosis. Conversely, in a research set up, you may like to find which of the three is a better marker of say, coronary events, and why others are not.

Nutrition level of a child can be assessed anthropometrically by weight, height and skinfold thickness. They are assessed in relation to age yet are univariate, and thus are indicators in the relaxed sense of the term. Weight-for-height requires two measurements—thus is bivariate, and technically transgresses to the fold of index as discussed in the next section. The choice between weight, height and skinfold thickness as anthropometric indicators of nutrition level in children would depend on whether one is looking at short term or long-term nutrition, or at the adiposity.

Blood hemoglobin (Hb) concentration and hematocrit (Hct) (also called PCV) values assist in evaluating plasma dilution, blood viscosity and oxygen carrying capacity. Hb is generally easily affected whereas Hct is a long-term measure. A related measure is red blood cells (RBC) (also called erythrocyte) count. Each is used in specific context and you should be able to decide which one or which ones would serve your purpose.

At the community level, events such as child mortality can be measured by neonatal mortality, postneonatal mortality and infant mortality rate. Neonatal period can be divided into early (<7 days) and late (7 - <28 days) periods. Although all of them could be used together to provide a holistic picture but a professional may like to concentrate on one indicator that best measures the specific aspect of interest. If the focus is on antenatal care, maternal nutrition and skilled attendance at birth, perhaps early neonatal mortality rate is the best indicator. If the focus is on breastfeeding, infections and child nutrition, postneonatal mortality may be better. If the interest is in inability to thrive in the face of repeated infections and mid-term sequel of low birth weight, mortality between one and four years may be better. Indeed, almost all around the world, the interest has shifted to neonatal mortality as the mortality beyond this period has relented.

In summary, different indicators may seem like similar but each indicator has specific application. The focus of the application will help you decide which particular indicator to use.

Indexes

Technically, index is a combination of two or more indicators. A popular example is body mass index, which is mathematically derived from weight and height. In that sense, mean corpuscular hemoglobin concentration (MCHC) is an index, which is calculated from Hb and

Hct values. Similarly, mean corpuscular hemoglobin (MCH) is an index based on Hb and erythrocyte (RBC) count, and mean corpuscular volume (MCV) is an index based on Hct and erythrocyte count. A large number of such indexes are used in medicine and health.

Some Commonly Used Indexes

Body mass index (BMI) is a classical example that illustrates how an incompletely looking stand-alone indicator such as weight attains very useful meaning when considered in conjunction with another indicator, namely, the height in this case. An indicator value in the context of another can provide a very different perspective. **Ankle-brachial (pressure) index** is the ratio of systolic blood pressure (BP) in the ankle to that of in the arm. Normally, BP at ankle is the same or slightly higher so that normal ratio is 1.0 or 1.1. A ratio much less than 0.95 indicates peripheral arterial disease. A ratio between 0.8 and 0.5 indicates moderate disease and less than 0.5 a severe disease. A value greater than 1.3 is also abnormal, suggesting peripheral vascular disease instead of peripheral arterial disease.

Another useful index is **bispectral index** that has many applications but is mostly used to measure the effect of an anesthetic agent on the hypnotic state of patient's brain during surgery. Thus this measures the depth of anesthesia. This index is based on analysis of electro-encephalogram (EEG) signals. The value ranges from 0 for complete EEG silence to 100 in fully awake adults. A value between 40 and 60 is generally recommended for general anesthesia. Sometimes anesthetic agents are titrated accordingly. Although called an index, among intermediary in the calculation of bispectral index is a scoring system for grade of hypnotic level. This score however is not for a qualitative characteristic. The name derives from a complex multivariate statistical procedure called bispectral analysis, which is used to select 'best' variables for computing the index.

Bispectral index is an example that emphasizes that no index should be indiscriminately used in varying conditions. It may be a good predictor of movement and hemodynamic response to incision but not good for assessing movement after administration of common drugs such as opioids.

Glycemic index measures the blood glucose increasing potential of a food after human consumption. For this, 2-hour area under the curve of blood glucose level is measured for that food. This is compared with the same area for the reference food, namely, the glucose sugar in the same quantity. Foods with low glycemic index are considered to contain slowly digested carbohydrate. An index less than 55 is considered low, between 55 and 69 is considered middling, and 70 or more is considered high.

In a simple format, **craniofacial index** is the ratio of the breadth of the cranium to the breadth of the face. This is used to assess facial deformities, especially in the newborn. One of its useful articulations is the **craniofacial variability index**. A series of anthropometric measurements of head and face is taken and each measurement is converted to a Z-score relative to the standard. A high variance of these scores occurs in pronounced differences in direction and size of individual anthropometry, and indicates craniofacial anomalies.

Occupation stress index measures burden of job stresses, particularly in relation to cardiovascular risk. This index is mostly based on cognitive ergonomics and brain research rather than on social constructs. For example, an employee may feel handicapped due to limited options for decision-making or may find that the information required for decision-making is either incomplete or contradictory. The job may involve high risk for a momentary lapse. All this go into making the occupation stress index.

Example 1 Occupational risk index among female physicians

Novi Sad in Yugoslavia is perceived as a region with high prevalence of lifestyle-related cancer factors such as smoking, obesity, sedantriness and alcohol intake. Belkic and Medic [1] studied these risk factors among 112 female physicians in Novi Sad in relation to physician-specific occupational risk index (OSI). This index adds 0 to 2 scores for 'not present' to 'strongly present' for items such as underload (e.g., simple work, lack of recognition) and high demand (e.g., essential communication, rapid task performance).

Side Note: The study found significant association of lifestyle related cancer risk factors with the total OSI and some of its components. The findings suggest diminishing work stress burden while developing strategies to reduce the incidence of these risk factors.

As mentioned earlier, the term index is widely used in generic sense for any measurement, which is not necessarily combination of two or more indicators. It is also used for marker of disease or condition such as detection of tissue inhibitors of metalloproteinase-1 and -2 (TIMP-1 and TIMP-2) as a diagnostic *index* for hepatic fibrosis, prevalence of hydrocele as a rapid diagnostic *index* for community burden of lymphatic filariasis, and serum angiotensin converting enzyme (ACE) and C4 protein as a combined diagnostic *index* for alcohol liver disease.

Advantages and Limitations of Indexes

Our purpose of describing some commonly used indexes is two-fold. (a) Illustrate a cross-section of situations where various kinds of indexes are used in health and medicine. (b) Whereas some indexes are simple, others could be quite complex. These two features also express the advantages and limitations of indexes. They allow two or more indicators to be considered together and enhance their utility—sometimes generate new information by placing them in proper context.

Index is quantitative and may involve calculation. This is nemesis for some clinicians. Some indexing instruments come ready with a software to do the calculations and directly provide the result. Bispectral index is automatically calculated by Aspect Medical Systems. HPLC calculates peak area of intensity of signals corresponding to concentrates of drugs – evoked potential. Thus, for some indexes, calculations are not much of a problem. Perhaps greater problem is their validity and reliability. A large number of indexes are available and many are being devised every year but studies that provide evidence of their reliability and validity are rare. For example, occupational stress index in Example 1 is face-valid and its content validity can be assumed but its construct validity needs to be assessed. Most widely used index is body mass index but its utility too is sometimes questioned in comparison with waist-hip ratio (WHR), which is seen as a better correlate of coronary events. Utility of WHR in other setups such as for cancer and lung disease has not been fully evaluated.

Bispectral index is calculated using sophisticated algorithm to assess the hypnotic/sedation level. One can wonder about its need at all since the sedation level is easily directly assessed by 5-point system that gives a score of 5 to 'ready response', 4 to lethargic response, 3 to response only on shout, 2 to response only after mild prodding, 1 to no response to mild prodding, and 0 to no response to noxious stimulus. Many clinicians still prefer this simple assessment that appeals to the conscience. Perhaps the advantage of bispectral index is that it is automatically calculated rapidly in varying situations and removes some of the subjectivity that might creep in assessing the sedation level by the response to a stimulus. Also the index can help in finding exact titration of anesthetic agent that a subjective criterion may not be able to do.

Scores

Medical professionals can be statistically dichotomized as those preferring qualities and subjective assessments using clinical acumen, and those who prefer quantities that bring in exactitude. Some of those belonging to the latter group may like to turn even signs-symptoms into some kind of quantities. How to convert qualities into quantities at individual level? Through scoring system! APACHE and Apgar scores are everyday examples. Pain on visual analogue scale is an indicator of pain intensity but McGill pain *score* comprises qualities like throbbing pain, shooting pain, and stabbing pain. Scoring systems also reduce multivariate entity into a univariate quantity—thus increasing the comprehension and utility. They also help in reducing some of the epistemic uncertainties that can arise from inadequate realization of how much weight to be given to various pieces of information.

Scoring systems are gradually gaining importance similar to laboratory and radiological investigations. There was a time when such investigations were considered below the 'dignity', of a clinician but now hardly any medical decision is taken without recourse to such investigations. A large number of scoring systems on various aspects are being developed: the list already runs into hundreds. These systems have found useful applications in gradation of severity as they consider several individual assessments together. Scoring system has also been applied for establishing a diagnosis, particularly for differentiating one condition from other similar looking condition. Both types—for diagnosis and for severity—can add to the intelligence with which a medical condition is managed.

Scoring System for Diagnosis

Attempts have been made from time to time to quantify medicine and developing scoring systems that can help in diagnosis.

Hypothyroidism poses a challenge to physicians at the time of diagnosis. It is not easy to distinguish between hyper-, eu- and hypothyroidism on the basis of clinical signs and symptoms. Tests such as T_3, T_4 and TSH help in reaching a diagnosis but these tests are slightly expensive, and some hospitals in developing countries may not have facility for such testing. Clinically clear cases of hyper- and hypothyroidism can be treated without worrying about laboratory results, and the tests ordered only in doubtful cases. This can reduce the load also on laboratory services. See Example 2 on a scoring system that can help.

Table 1 A scoring system for thyroidism

Symptoms/sign	Score	
	Present	Absent
Physical tiredness	0	+2
Slow cerebration	–3	+2

Diminished sweating	+6	–2
Dry skin	+3	–6
Cold intolerance	+4	–5
Dry hair	–2	+2
Wight increase	+1	–1
Constipation	+2	–1
Hoarseness of voice	+5	–6
Parasthesiae	+5	–4
Deafness	+11	0
Slow movements	+4	–3
Coarse skin	+7	–7
Cold skin	+3	–2
Periorbital puffiness	+4	–6
Pulse rate <75 /min	+4	–4
Ankle jerk	+15	–6
Total score ≤ –30	Euthyroid	
–29 to +24	Doubtful	
≥ +25	Hypothyroid	

Example 2 A hypothyroid diagnostic index

A hypothyroid diagnostic index was devised by Billewicz et al. [2] considering the frequency of various symptoms and signs. They developed the scoring system given in Table 1 excluding four features with zero score.

The scores are based on the logarithm of the ratio of frequency of presence and absence of symptoms in established hypothyroid and euthyroid cases. According to this scoring system, a clear diagnosis can be made on the basis of signs and symptoms alone if the score is 25 or more (hypothyroid) or –30 or less (euthyroid). Laboratory help is required only when the total score is between –29 and +24. This can reduce the load on the laboratory by more than 50%.

Side note:
1. This scoring system was later simplified by Zulewski et al. [3]
2. Although this is a score but the authors liked to call it an index. Many such examples of mixed use of the terms exist in the literature.

Similar scoring system is available for diagnosis and prognosis of ischemic heart disease [4], to differentiate between supratentorial intracerebral hemorrhage and cerebral infarction [5], for diagnosis of left ventricular hypertrophy based on ECG [6], and for diagnosis of respiratory failure [7]. See Example 3 for scoring for diagnosis of catheter related infections. These are just some examples of various scoring systems available for diagnosis purposes but many others are available.

Almost all such scoring systems are based on data from developed countries. Because of nutrition and environmental factors, they may not directly apply to the subjects in developing countries. Consider modifications before using them for patients in such countries.

Such systems should be used only when a valid diagnosis is difficult to establish otherwise. Or when the diagnosis depends upon the physician's preferences, his expertise, or on results of laboratory or radiological investigations that lack reliability. Such inadequacies in the diagnostic process are more common than are otherwise apparent. A useful strategy is to use these scores as just one more evidence in addition to the clinical and laboratory evidence, and take a decision in the holistic manner.

Example 3 Scores for diagnosis of catheter related infections

Clinical diagnosis of catheter-related infections is difficult. Lugauer et al. [8] show that it can be done with relative ease with the help of a scoring system comprising: rate of rise of body temperature, attendant shivering, identification of pathogens in blood and/or catheter tip cultures, improvement in the clinical course after catheter removal, signs of catheter exit site inflammation, and results of diagnostic tests for other possible sources of infection. These criteria were graded using points and weighted according to their specificity. The patients were also diagnosed using the existing but complex clinical criteria. The agreement of scoring system with the clinical diagnosis was 85 percent in a group of 65 cases. There was no false negative and 10 were false positive. It turned out that 9 of these 10 were not false when additional findings are considered (clinical diagnosis criteria was expanded). Thus the scoring system appears to be more sensitive than existing diagnostic criteria without loss of specificity.

Scoring for Gradation of Severity

Prognostic assessment and the management of a patient depend to a large extent on the severity of the disease. There is considerable epistemic uncertainty about how to assess this severity. Different professionals use different methods. For uniformity and exactitude, at times scoring is considered desirable.

The Glasgow Coma Scale [9] is used to grade coma patients by using numeric scale for eye, motor and verbal response. Acute Physiology and Chronic Health Evaluation (APACHE) is used to assess critically ill hospitalized adults [10]. Various variations of this score are available. Mortality Prediction Modeling [11] uses systolic blood pressure (BP), level of consciousness, prior cardiopulmonary resuscitation (CPR), age, presence of cancer, and presence of infection to assess chances of survival in patients admitted to intensive care unit. Yale's Observation Scale [12] is used to identify serious illness in febrile children. Indrayan's smoking index measures the life-long burden of smoking. Apgar score is used to assess the prognosis in a neonate. These are just few examples. In case you are stuck with a problem of grading the severity of patients, see if a valid scoring system is available. If not, you may like to devise a scoring system yourself.

As already stated, all scoring systems try to convert multiple measurements into a unified single but meaningful index. They transform multivariate data into a univariate score. Sometimes several scoring systems are available for the same condition and it would be difficult to choose the right system. For example, severity in peritonitis cases can be assessed by APACHE score, Peritonitis Severity Score (PSS), Mannheim Peritonitis Index (MPI), Hacettepe Score, American Society of Anesthesiologists (ASA) Score, etc. Choose a scoring system that looks more appropriate for your patients.

Table 2 Scoring system for death in Chagas heart disease

Risk factor	Points
New York Heart Association Class III or IV	5
Evidence of cardiomegaly on radiography	5
Left ventricular systolic dysfunction on echocardiography	3
Nonsustained ventricular tachycardia on 24-hour Holter monitoring	3

Low QRS voltage on electrocardiography	2
Male sex	2

Example 4 Risk score for predicting death in Chagas heart disease
Chagas disease is an important health problem in Latin America and cardiac involvement in this disease increases the severity and risk of death. Rassi et al. [13] developed a risk score based on evaluation of 424 outpatients from a regional Brazilian cohort as given in Table 2. Score is obtained as sum of these points. They divided score into three groups: low risk, 0-6 points; intermediate risk, 7-11 points; high risk, 12-20 points. The risk pertains to the risk of death. The 10-year mortality rates for these three groups were 10 percent, 44 percent, and 84 percent, respectively.

REFERENCES

1. Belkic K, Nedic O. Workplace stressors and lifestyle-related cancer risk factors among female physicians: assessment using the occupational stress index. *J Occup Health* 2007; 49:61-71.
2. Billewicz WZ, Chapman RS, Crooks J, et al. Statistical methods applied to the diagnosis of hypothyroidism. *Q J Med* 1969; 38:255-266.
3. Zulewski HK, Muller B, Exer P, Miserez AR, Staub J. Estimation of tissue thyroidism by a new clinical evaluation of patients with various grades of hypothyroidism and controls. *J Clin Endocrinol Metab* 1997; 82:771-776.
4. Rautaharju PM, Warren JW, Jain U, Wolf HK, Neilson CL. Cardiac Infarction Injury Score: an electrocardiographic coding scheme for ischaemic heart disease. *Circulation* 1981; 64:249-256.
5. Poungvarin B, Viriyavejakul A, Komontri C. Siriraj Stroke Score and validation study to distinguish supratentorial intracerebral haemorrhage from infarction. *BMJ* 1991; 302:1565-1567.
6. Romhilt DW, Estes EH Jr. A point score system for ECG diagnosis of left ventricular hypertrophy. *Am Heart J* 1968; 75:752-758.
7. Wood DW, Downes JJ, Lecks HI. A clinical scoring system for the respiratory failure: preliminary report on childhood status asthmaticus. *Am J Dis Child* 1972; 123: 227-228.
8. Lugauer S, Reganfus A, Boswald M, et al. A new scoring system for the clinical diagnosis of catheter-related infection. *Infection* 1999; 27 (Suppl 1):549-553.
9. Jennett B, Teasdale G, Braakman R, Minderhoud J, Heiden J, Kurze T. Prognosis of patients with severe head injury. *Neurosurgery* 1979; 4:283-289.
10. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100:1619-1636.
11. Lemeshow S, Teres D, Avrunin JS, Gage RW. Refining intensive care unit outcome predictions by using changing probabilities of mortality. *Crit Care Med* 1988; 16:470-477.
12. McCarthy PL, Sharpe MR, Spiesel SZ, et al. Observation scales to identify serious illness in febrile children. *Pediatrics* 1982; 70:802-809.

13. Rassi A Jr, Rassi A, Little, WC, et al. Development and validation of a risk score for predicting death in Chagas heart disease. *N Engl J Med* 2006; 355:799-808.

Adapted from Medical Biostatistics, Third Edition

by A. Indrayan (http://indrayan.weebly.com)

Chapman & Hall/CRC Press, 2012 (http://www.crcpress.com/product/isbn/9781439884140)

Details of the book and reviews available at http://MedicalBiostatistics.synthasite.com

---