The Logic of Laboratory Medicine

Dennis A. Noe, M.D.

Second Edition

Wisdom is the principal thing;
therefore get wisdom
and with all thy getting get understanding.

Proverbs

ACKNOWLEDGMENTS

I am indebted to Dr. Louise Grochow for encouraging me in this project and for editing the entire manuscript. She has seen to it that the book is so much better than it otherwise would have been. I thank Dr. Sandra Wolin for giving me three months away from the laboratory to work on the book. And I thank my wife, Dr. Karen Kumor, for her understanding and patience.

CONTENTS

Chapter 1
LABORATORY-BASED MEDICAL PRACTICE

CLINICAL USE OF LABORATORY STUDIES

Physicians select laboratory studies and interpret the results. They integrate their laboratory interpretations into the clinical assessment through the synergistic interplay of quantitative analysis and clinical judgment. This process, which occurs repetitively throughout the care of a patient (Figure 1.1) constitutes the general pattern of use of laboratory studies.

If a laboratory study is to provide clinically useful information, it must be ordered with a specific clinical goal in mind (Table 1.1). The clinician relates each goal to a set of specific information needs based upon a pathophysiologic understanding of the disorder or disorders under consideration (Table 1.2).

For example, one medical goal in a patient with severe chest pain is to establish or exclude the diagnosis of acute myocardial infarction. Infarction means cell death so the foremost information need is to determine if myocyte death has occurred. In addition, though, the information needs might include an assessment of cardiac function and electrophysiology as infarcts typically cause regional contractile dysfunction and impaired myoelectric signal propagation in the ischemic myocardium.

Having identified the information needs for a patient, the clinician then orders clinical and laboratory studies to address the needs, here also depending upon pathophysiologic principles to select and interpret the studies. For instance, in pursuing the question of cardiac myocyte death in the patient with chest pain, the clinician will request that the plasma concentration of creatine kinase-MB be measured because there is a reliable pathophysiologic relationship between the plasma concentration of this enzyme and the presence of myocyte death.

The interpretations of the study results and of additional clinical observations provide the input for reevaluation of the clinical assessment and revision of the catalog of medical goals for the patient. And the cycle begins again.

This simple picture of study use does not, however, take into account a number of real-life considerations including (1) limitations in the availability of studies, (2) delays in the time it takes to receive study results, and (3) the physical, psychological, and financial costs of performing studies.

When planning the laboratory evaluation of a patient, the clinician must keep in mind the capabilities of the clinical laboratories. Not every study is available at every hospital. Fortunately, the standards of medical care in developed countries are so high that, usually, when a laboratory study is not available, either an adequate substitute study is, or a specimen can be sent to a reference laboratory, or the patient can even be transferred to a medical center that does offer the study. The availability of laboratory studies can also be constrained by limitations in the schedule for study performance—a study that cannot be done when needed may be no better

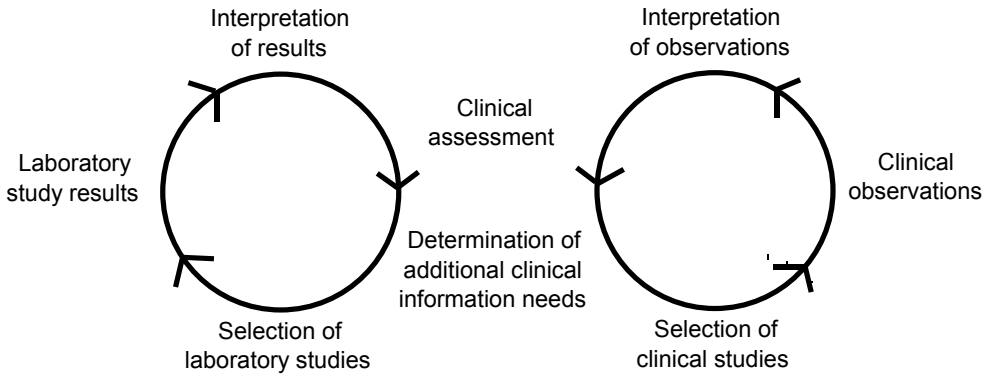

Figure 1.1 Cyclic structure of the use of medical studies.

Table 1.1
Goals of Medical Care

1. Detect and quantify risk of future disease
2. Detect subclinical disease
3. Establish and exclude diagnoses
4. Assess disease severity and establish prognoses
5. Select appropriate therapy
6. Monitor disease progress and treatment effect

Table 1.2
Clinical Information Needs

1. Assess organ function
2. Assess metabolic activity
3. Assess macro- and micronutritional status
4. Detect and monitor neoplasia
5. Detect and quantify tissue injury
6. Detect and identify genetic disorders
7. Detect and identify immunologic disorders
8. Detect and identify infectious agents
9. Detect and identify intoxicants and poisons
10. Monitor therapeutic agents

than a study that is not offered at all. Alternative studies that can be obtained in a timely fashion must then be used instead.

Another and more frequent consideration in ordering studies is the time that elapses between requesting a study and receiving the result. If the wait is short, a study can be ordered and the result received and interpreted prior to requesting the next study, if additional testing is indicated. This is sequential test ordering. It is efficient because each study ordered contributes to clinical care and is cost-effective because the number of studies ordered is minimized. If the turnaround time is long, however, the patient's care is compromised by the cumulative delay in obtaining study results occasioned by sequential testing. In that case most or all of the potentially useful laboratory studies can be ordered together and the results interpreted *en masse*. This is concurrent ordering. It is not efficient but, when properly used, is cost-effective since the costs of delayed care are minimized. Concurrent ordering must not be confused with the indiscriminate ordering of laboratory studies by those who have the misconception that the greater the number of studies ordered, the greater the amount of information that will be available for use in the care of the patient. Remember, data is not always information! Informative study results contribute to the care of the patient. Superfluous study results, at best contribute nothing to the patient's care, often obscure informative results, and sometimes even misinform.

Lastly, clinicians must keep in mind that laboratory studies are costly. The financial burden of medical bills, especially outpatient bills, is in no small part due to laboratory charges. In addition, much of the physical and psychological discomfort experienced by patients as part of their medical care is attributable to the invasive nature of most laboratory studies. Although it is not always possible, it is important to include cost as a consideration in study selection.

EXPRESSING LABORATORY RESULTS

Most laboratory studies are simply measurements. The information requested is of the type "How much, or many, of some analyte is present in this specimen?" As such, these studies quantify the analyte of interest. The level of quantification achieved varies depending upon clinical needs and the sophistication of the method of measurement. Qualitative studies are characterized by binary quantification. The analyte is reported as either "present" or "absent." Semiquantitative studies arrange study results into grades or categories. Results may, for example, be reported as "absent / trace / moderate / marked" or "zone I / zone II / zone III." Quantitative studies use a scale of measurement. The scale is graduated according to a reference measurement, called the unit of measurement. The value of a quantitative measurement indicates how many multiples of the reference measurement, or unit, are contained in the specimen.

SI Units

There exists an international system of units, le Systčme International d'Unités (abbreviated SI), advanced by the International Committee of Weights and Measures as the system of units to be adopted by all signatories of the Diplomatic Convention of the Meter, 1875 (Lehmann 1979). From this system there has evolved a recommended system of units (Recommendations 1978 and 1984) to be used in medicine (Dybkaer 1978a; Dybkaer 1978b; Siggard-Anderson *et al.* 1987). These recommendations are the product of the Clinical Chemistry Committee of the International Union of Pure and Applied Chemistry (IUPAC) and the International Federation of Clinical Chemistry. They are supported by the International Committee for Standardization in

Table 1.3
SI Quantities and Units Used in Laboratory Medicine

Quantities	Derivation	Unit	Symbol
Base quantities			
time		second	s
length		meter	m
number		1	[none]
mass		kilogram	kg
amount of substance		mole	mol
Derived quantities			
volume	length^3	liter	L
flux			
volume	volume/time		L/s
number	number/time		/s
mass	mass/time		kg/s
substance	amount/time		mol/s
concentration			
number	number/volume		/L
mass	mass/volume		kg/L
substance	amount/volume		mol/L
fraction			
volume	volume/volume		[none]
number	number/volume		[none]
catalytic activity	substance flux/volume	katal per liter	kat/L
pressure	force/area	Pascal	Pa

Hematology and the World Association of (Anatomic and Clinical) Pathology Societies. Although the medical community in the United States has not generally been supportive of the Recommendations, laboratories and medical journals here may someday accede to its implementation.

The system of units advanced in the IUPAC recommendations is based upon the SI, the use of substance quantities (such as mole) rather than mass quantities (such as grams), and the use of the liter as the preferred unit of volume (see Table 1.3). The medical arguments in favor of using molar units rest upon their physiologic appropriateness. The electrochemical activity and osmolarity of solutes are determined by their molar concentration as is the bioactivity of hormones and the binding capacity of ligand-binding proteins. In addition, the use of molar units preserves the quantitative relationships between metabolic precursors and products.

A special consideration in the use of substance quantities is the expression of the concentration of catalytic activity. Rather than defining the molar concentration of an enzyme present in a specimen, the molar flux of substrate acted upon by the enzyme is measured and reported. The reporting unit, the katal, is equal to one mole substrate transformed per second per liter. This approach is appropriate whenever the actual catalytic activity of an enzyme or set of enzymes with overlapping substrate specificities is the physiologic entity of interest.

The flexibility of SI units is increased by the use of magnitude prefixes. Rather than being restricted to using a scale graduated in unit divisions, measurements can be based upon divisions that are powers-of-ten multiples of the standard unit. For example, a substance that is present in a concentration of 1.6×10^{-9} moles per liter can be described as having a concentration of 1.6 nanomoles per liter. The prefix "nano-" takes the place of the factor 10^{-9}. Similarly, large numbers can be avoided by using unit prefixes. A partial pressure of 6×10^3 Pascal becomes 60 kiloPascals. Table 1.4 lists approved magnitudes prefixes. Notice that the prefixes used with SI units are only for magnitude changes that are third powers-of-ten. Prefixes are not generally used to describe multiples of time greater than one second because the common units of time, such as minute and day, are ingrained in scientific as well as everyday usage. Non-SI units of time and their preferred symbols are listed in Table 1.5.

Table 1.4
Magnitude Prefixes for Units

Factor	Prefix	Symbol
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a
10^{-21}	zepto	z

Table 1.5
Non-SI Units of Time

Unit	Symbol
minute	min
hour	h
day	d
week	wk
month	mo
year	y

Unit conversion

Until Recommendations 1978 and 1984 have been fully integrated into clinical laboratory practice for some years, physicians will have to deal with two unit systems, SI units and so-called common units. Inevitably some results expressed in common units will need to be converted to SI units and *vice-versa*. Measurements are converted from one scale of measurement to another by substituting a numerically equivalent number of units from the new scale of measurement in the place of the original unit. For example, to convert the measurement 1.2 mg creatinine/dl to its SI expression in μmol creatinine/L, dl must be converted to its L equivalent and mg creatinine must be converted to its μmol creatinine equivalent. There are 10 dl per L,

$$1.2\ \frac{mg}{dl} \times 10\ \frac{dl}{L} = 12\ \frac{mg}{L}$$

There are 8.85 μmol creatinine in 1 mg creatinine,

$$12\ \frac{mg}{L} \times 8.85\ \frac{\mu mol}{mg} = 106\ \frac{\mu mol}{L}$$

So the SI equivalent is 106 μmol creatinine/L.

The number of units of one scale of measurement contained in 1 unit of a comparable scale of measurement is called the conversion factor between the units. It is the number by which the value of a measurement is multiplied to re-express the measurement as multiples of the alternate unit. Here, mg creatinine/dl are converted to μmol creatinine/L using the conversion factor 88.5. Care must be taken when using conversion factors to make certain that the factor used is appropriate for the direction of the conversion. Extensive tables for unit conversion can be found in Lippert and Lehmann (1979).

Calculated values

When direct measurement of a quantity is impractical or impossible, its magnitude may be estimated by calculation from related measurements. The measurements that serve as quantitative input for these calculations may possess an exact theoretical relationship to the unmeasured quantity, such as that between bicarbonate concentration and the ratio of the partial pressure of carbon dioxide to the hydrogen ion concentration (Kassirer and Bleich 1965),

$$[HCO_3^-] \text{ in mmol/L} = 180 \times \frac{pCO_2 \text{ in kPa}}{[H^+] \text{ in nmol/L}}$$

Alternatively, a calculation may be based upon an empirical relationship between the measured and unmeasured quantities. The calculation of body surface area using body weight and weight is an example.

Calculated values can be obtained in two ways. If a mathematical formula is available, the value can be computed. This has become particularly simple since the advent of inexpensive, powerful hand-held calculators. Calculated values can also be found without performing computations by using tables, which are usually too large to be convenient, or graphical representations of mathematical equations, called nomograms. For instance, body surface area in m^2 can be calculated from body weight and height. Using the formula proposed by Gehan and George (1970),

$$\text{surface area in } m^2 = 0.0235 \times (\text{weight in kg})^{0.51456} \times (\text{height in cm})^{0.42246}$$

Using this formula, a person who weighs 64 kg and is 145 cm tall has a surface area of 1.64 m^2. The authors also provide a table in their paper. The table entry for 64 kg and 145 cm is 1.64 m^2. A nomogram of their formula is also offered by the authors for those readers who prefer not to perform calculations or to look up table entries. Because that nomogram is somewhat difficult to use, another

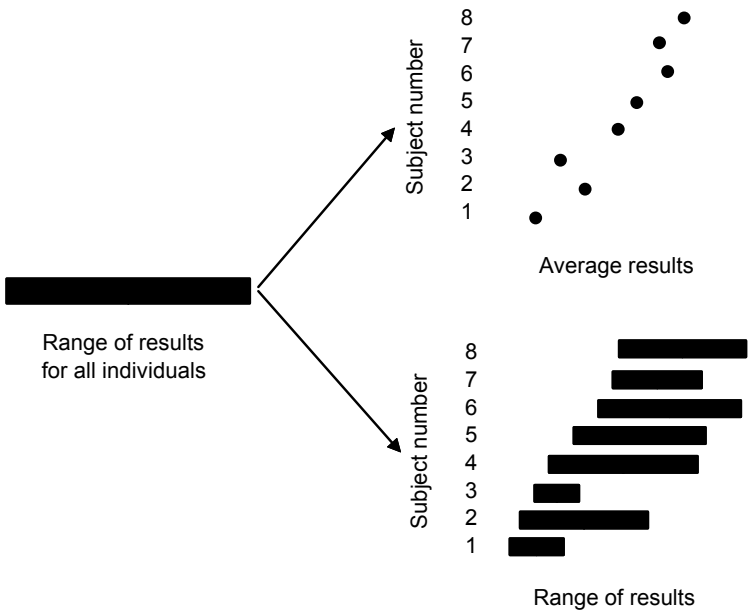

Figure 1.2 Separating inter- from intraindividual measurement variability

nomogram of the formula has been proposed (Noe 1991). Using that nomogram, the body surface area is determined to be 1.64 m^2. As expected, the same value is derived using each of the different approaches for its calculation.

VARIABILITY IN LABORATORY RESULTS

A laboratory measurement performed upon many different individuals or upon a single individual many times will show differences in the magnitude of the entity measured, that is, there will be measurement variability. This variability comes from a number of sources: biologic variability, preanalytic variability, analytic variability, and postanalytic variability.

Sources of variability

Biologic variability is due to the heterogeneity of physiologic influences among individuals and in individuals over time. It is distinguished from other sources of variability in that it cannot be controlled to reduce its effect. The two components of biologic variability are interindividual variability and intraindividual variability. Interindividual variability alludes to differences in the magnitude of a measurement among individuals. Important sources of interindividual variability include race, sex, and age. Intraindividual variability refers to differences in the results of a study in one individual when determinations are made at different times. Typically, intraindividual variability is smaller than interindividual variability. Consequently, interindividual variability largely determines the total biologic variability. This finding is illustrated in Figure 1.2. For eight individuals, here numbered one through eight, the range of a set of replicate laboratory studies is shown. This range is separated into its component parts: the interindividual component, i.e., the range of average study results among the individuals, and the intraindividual component, i.e., the range of study results for each individual. Clearly here the interindividual variability contributes more to the total variability. Indeed, even if there were no intraindividual variability at all, the total range of study results would not be lessened very much.

Preanalytic variability is due to physiologic influences that can be controlled in the individual upon whom the measurement is made. It also results from the effects of specimen collection and handling, factors that can be controlled by the laboratory. Important sources of physiologic preanalytic variability and, therefore, important considerations in patient preparation include: time of day, food intake (including caffeine and ethanol-containing beverages), physical exercise, and drug therapy (including self prescribed drugs). As a rule, it is recommended that laboratory studies be performed in the morning following an overnight fast.

Strenuous or stressful physical or emotional activity should be avoided. If drug therapy cannot be suspended, the laboratory results must be interpreted with consideration of the influence of the therapy upon the measurement. (Drugs may interfere with analytic methods as well as alter the physiologic state of the patient.)

Analytic variability is the variability in laboratory measurement that can be attributed to the analytic method generating the measurement. The analytic method includes materials, equipment, procedures, and personnel. Variable performance of each of these components contributes to the total analytic variability. Analytic variability is kept within acceptable limits by both rigorous analytic method assessment and ongoing surveillance of the method using a quality assurance program.

Postanalytic variability in laboratory measurement arises between the completion of the analytic method and the assimilation of the measurement by the clinician. A major source of postanalytic variability is transcription error. Such errors may be made by laboratory personnel, laboratory clerks, ward clerks, medical students, resident physicians, or attending physicians. The opportunity for such error increases with increasing numbers of transcriptions so the original laboratory report form is the most reliable source for the measurement.

Laboratory error

Despite the efforts to eliminate mistakes in the performance of laboratory studies, inevitably some study results that reach the clinician will harbor an error. The clinician must be vigilant for the evidence of such errors. At the same time, he or she must also remain open to the possibility that a study with a suspicious result was performed correctly and that the results, though surprising, are valid for the specimen received.

Laboratory error should be considered when (1) the result is unreasonable, unphysiologic, or impossible; (2) the result is inconsistent with previous results from the same patient or is incompatible with the results of other studies performed upon the same specimen; or (3) the result differs from that expected on the grounds of the clinical impression. In the third situation, the consideration of a laboratory error is appropriate, but reevaluation of the clinical impression is equally necessary. It may even be advisable to confirm that the result really is inconsistent with the impression.

When a laboratory error is suspected, the clinician must act to confirm or refute the suspicion. It is not enough simply to ignore the result. If the result truly is in error, the laboratory must be made aware of the problem so that steps may be taken to prevent its recurrence. If the result is valid, the clinician must confront the unpleasant fact that either his or her interpretation of the study result was faulty or that his or her clinical impression may be incomplete or even frankly incorrect. The clinician should evaluate the possible influences of known sources of biologic and preanalytic variability upon the laboratory study. Special attention should be paid to the effects of drug therapy upon both the physiologic state of the patient and the reliability of the laboratory study. If laboratory error is still suspected, he or she should request that the laboratory repeat the study upon the original specimen and, if possible, upon a new specimen. These actions, taken in the stated order, will detect the site of the error in almost all cases in which an error exists. Of equal importance, however, is that this regimen will also reveal the explanation for a puzzling but valid result and thereby facilitate patient care.

Variability and monitoring

When following, or monitoring, a patient by repeatedly performing a laboratory study upon him or her, it is essential that the clinician be able to tell if a change in the results indicates a change in the status of the patient or if it merely reflects variability in the study's measurement. In order to make this decision, the clinician must know the pattern of systematic intraindividual variability of the analyte (such as the diurnal oscillations in plasma iron concentration) and the magnitude of the random intraindividual variability of the analyte in the patient. Techniques for characterizing systematic variability, for quantifying random variability in individual patients, and for the quantitative analysis of serial laboratory data are discussed in later chapters.

Variability and clinical classification

The central problem in the use of laboratory studies for clinical classification is that of interindividual variability. Clinicians must be able to decide if a given study result is better explained by the variability in the measurement among persons having a disorder or by the variability among individuals who do not have the disorder. For

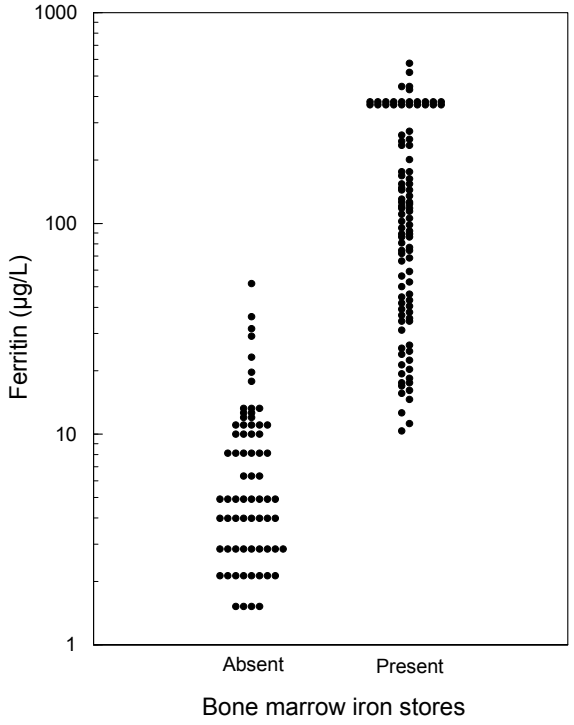

Figure 1.3 Relationship between bone marrow iron stores and ferritin concentration in the plasma.

example, Figure 1.3 (data taken from Ali *et al.* 1978) shows values of the concentration of ferritin in plasma in patients with iron deficiency and in patients who are iron replete. There is obviously considerable measurement variability in both groups. This variability results in a substantial overlap of values: the ferritin concentrations in patients with iron deficiency overlap those in patients with normal iron stores. Because ferritin concentrations in the region of overlap can be found in individuals in either clinical class, for such study results there is uncertainty as to the correct classification. As will be discussed in later chapters, to deal with this kind of classification uncertainty, clinicians must have access to quantitative descriptions of the distribution of study results in persons with and without the disorder. These descriptions are called frequency distributions. A frequency distribution that arises from the performance of a defined laboratory study upon a sample of subjects from a defined, or reference, clinical population is called a reference frequency distribution.

Establishing reference frequency distributions

The determination of a reference frequency distribution for a laboratory study proceeds in the order listed in Table 1.6 (Solberg 1987a). A

Table 1.6
Steps in the Determination of a Reference Frequency Distribution

1. Definition of the analytic procedure, equipment, and reagents that generate the measurement
2. Definition of the conditions under which specimens are obtained and the procedure for specimen collection, handling, and storage
3. Definition of the reference population by specification of the criteria for subject inclusion and exclusion and the criteria for partitioning subsets of the population
4. Solicitation of subjects and performance of the study
5. Calculation of the study result frequency distribution

satisfactory statement of the inclusion criteria for a population of individuals afflicted by a disorder is of paramount importance. The inclusion criteria, which amount to the basis for the diagnosis of the disorder, must rely upon a universally accepted method for identifying the disease. Such methods are called reference methods or "gold standards." A related concern is that the stage or severity of the disorder be considered when constructing the reference criteria (Ransohoff and Feinstein 1978). The inclusion and exclusion criteria applied to populations free of a disorder will determine the clinical settings in which these reference distributions will be useful. The criteria should define persons similar to those upon whom the laboratory study will be performed in practice. For instance, if the study is to be used to screen for a disorder in the general population, a sample of the general population should be used. On the other hand, if the study is to be used exclusively to identify the disorder in a select subset of patients, the sample should consist of members of that subset. At the same time, the sample subjects should represent as broad a spectrum of biologic variability as is possible within the confines of the stipulated criteria (Ransohoff and Feinstein 1978). In particular, sex and age should be considered when defining the desired subject composition of the sample because these characteristics contribute so much to the biologic variability of most laboratory studies. Alternatively, separate reference distributions can be constructed for males and females or for certain intervals of age. Partitioning of the clinical population into subsets should be considered when the subsets show significant differences in the location of their frequency distributions (Harris 1975, Harris and Boyd 1990).

The calculation of the frequency distribution of study results involves, at a minimum, the

computation of the central range of values. Patterned on conventional statistical practice for determining significance, this range is defined as the central 95% of the results. Thus, this representation of the frequency distribution consists of a statement of the 2.5 and 97.5 percentile values for the cumulative frequency distribution of the study results. A more informative description of the frequency distribution involves the complete characterization of the cumulative frequency distribution. Additionally, the utility of the study in clinical classification is increased if the frequency density distribution of results is also characterized. The frequency density distribution relates the frequency of occurrence of a study result to the value of the result while the cumulative frequency distribution relates a study result to the frequency of all study results equal to or less than it.

Frequency distributions may be described empirically (the nonparametric approach) in which case the frequencies assigned to a study result are those observed among the subjects studied or the data may be mathematically modeled (the parametric approach) in which case the assigned frequencies are those predicted by the fitted model. The nonparametric approach has the advantage of not depending upon the appropriateness of the model chosen to describe the relationship between the study results and their frequencies. The major disadvantage of the approach is the large number of study results that must be collected in order to describe the distribution precisely. This shortcoming is particularly troublesome when the distribution is defined only in terms the 2.5th and 97.5th percentiles of the distribution because the precision of the approach is worst in the distribution tails. In general, the nonparametric approach requires the enrollment of one and one-half to two times as many subjects as the modeling approach in order to derive precise estimates of the 2.5th and 97.5th percentiles values (Linnet 1987). This disadvantage can be partly overcome by mathematical smoothing of the distribution. (Smoothing differs from modeling in that no global formula for describing the distribution is assumed; instead, the smoothed line is computed from the data within successive intervals of the distribution). Methods are available for smoothing frequency density distributions (e.g., Willard and Connelly 1992 and Strike 1996) and cumulative frequency distributions (e.g., Shultz *et al.* 1985). The modeling of frequency distributions attempts to reveal the form of the underlying "true" relationship between the result values and their frequencies and thereby to describe the frequency distribution more accurately than is possible by empirical means. The advantages of modeling include not only the potential for greater accuracy in the description of the frequency distribution but also the ability to predict frequencies for all possible study values, not just those appearing among the study subjects, and the ability to describe the full distribution using simply the values of the parameters defining the model. This permits considerable ease in the manipulation of the frequency data for the purpose of generating quantitative estimates of diagnostic and prognostic probabilities. The major disadvantage of modeling is the potential for a less accurate description of the frequency distribution owing to the use of an inappropriate model. The application of statistical tests of the goodness-of-fit of a model substantially reduces the chances that an incorrect model will be used to describe a distribution (Solberg, 1987b).

Figure 1.4 illustrates the calculation of frequency distributions using nonparametric and modeling approaches. Panel A shows a set of 200 hypothetical study results generated from a normal distribution with a mean of 100 and a standard deviation of 10. The corresponding frequency density distributions are shown in panel B. The nonparametric distribution (symbols) has been plotted without data grouping to demonstrate the appreciable, and typical, degree of data irregularity. In general, the grouping of frequencies into result intervals (histogram bins) lessens the irregularity of nonparametric distributions (Scott 1979). Knowing that the data arose from a normal distribution, a normal distribution has been used for the model distribution (line). The mean and standard deviation derived from the data set are 100.35 and 10.37, respectively. The model distribution appears to correspond fairly closely to the nonparametric distribution but the scatter in the data makes it difficult to judge the quality of the fit. Panel C presents the frequency data as cumulative frequency distributions. The nonparametric distribution (symbols) is quite well-behaved; although some data irregularity persists, the scatter seen in the nonparametric frequency density distribution is largely eliminated when the data are plotted as cumulative frequencies. The normal distribution model (line) clearly fits the empirical data very well. The 2.5 and 97.5 percentiles calculated for these data are 82.98 and 121.03,

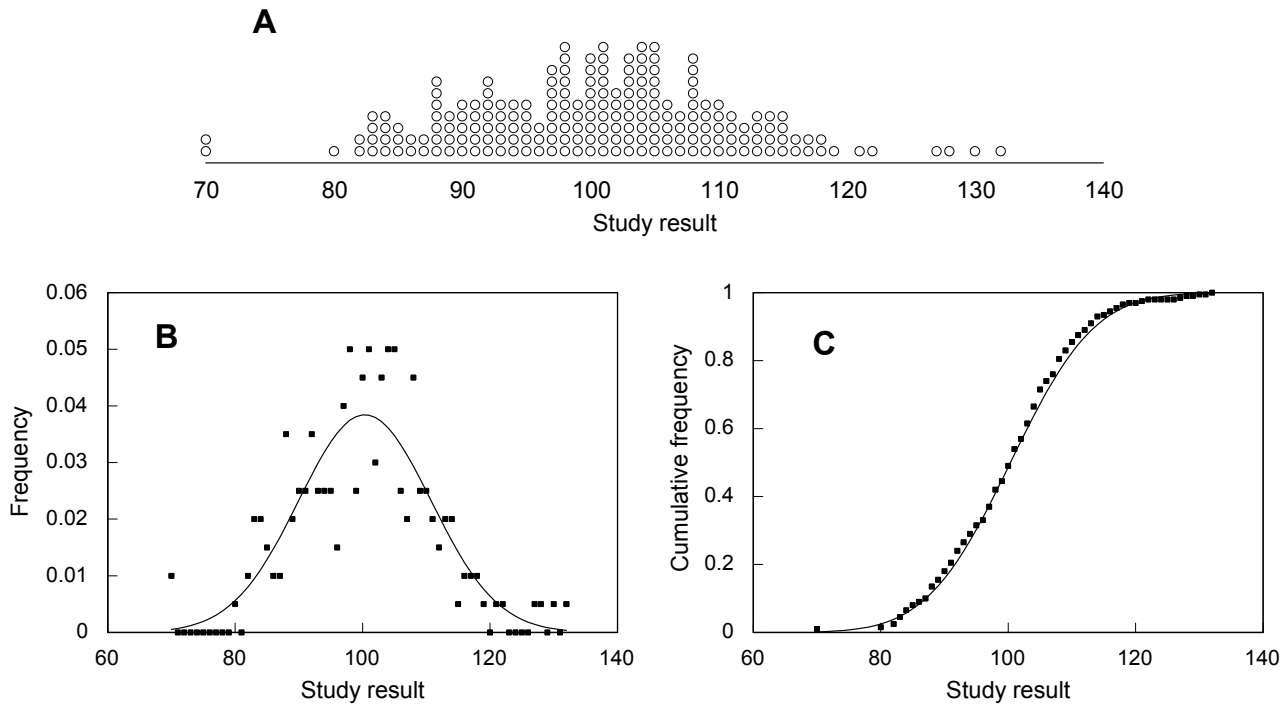

Figure 1.4 Graphical presentations of the distribution of a set of hypothetical study results. **A.** Individual values. **B.** Empirical values (squares) and the normal, i.e., Gaussian, model (line) of the frequency density distribution. **C.** Empirical values (squares) and the normal model (line) of the cumulative frequency distribution.

respectively (nonparametric) and 80.02 and 120.68, respectively (normal model)

The choice of the model is obviously crucial when modeling frequency data. In the example, a normal distribution was selected as the model and, both by visual inspection and by statistical criteria proved to be satisfactory. Often, though, a normal distribution will not serve well. The frequency distributions of most laboratory studies show appreciable skew. That is, more than half of the results occur at values to one side of the median (the value with the maximum frequency). Such data can often be well modeled using the lognormal distribution or the even more generally applicable three-parameter lognormal distribution (Royston 1992).

Reporting results

When a study result is reported to a physician it is necessary that he or she understand what the particular result means in terms of the medical reason for which the study was performed. When a study is ordered to aid in the classification of a patient the physician will need to know how the study result compares to the values found among individuals in the pertinent clinical classification category and also among individuals in complementary classification categories. It makes sense, then, to report not only the study result itself but also its relationship to the frequency distributions associated with the classification categories under consideration. This can be done in a number of ways (Dybkaer and Solberg 1987). The quantitative approach is to indicate where exactly the result falls within the frequency distributions of study results in the appropriate clinical reference groups. At present, though, it is still beyond the capability of clinical laboratories to provide this service because of the vast number of reference groups that are of interest to physicians ordering laboratory studies.

It is possible, however, and a practical standard, to take the much less quantitative approach of reporting the usual (95%) range of study results as found in normal individuals. This range is called the normal range by most clinicians and the reference range or reference interval by most laboratorians.

The identification of the members of this reference population is problematic (Solberg 1987a, Gräsbeck 1990). What, for instance, is meant by the inclusion criterion "normal"? And what are the exclusion criteria for normality? Some laboratories

construct a reference range from values that arise from the performance of the study upon a random sample of specimens submitted to the laboratory for other determinations (Solberg 1994). Some results are discarded on the basis of statistical rules, but no specimen, and therefore no sample subject, is excluded prior to the performance of the study. Because of these and other conceptual difficulties inherent in identifying a population of normal individuals, it is essential that laboratories in some fashion annotate their reference intervals to indicate the composition of the reference population.

REFERENCES

Ali MAM, Luxton AW, and Walker WHC. 1978. Serum ferritin concentration and bone marrow iron stores: a prospective study. Can Med Assoc J 118:945.

Dybkaer R. 1978a. Approved recommendation (1978). Quantities and units in clinical chemistry. (IUPAC and IFCC Recommendation 1978). J Clin Chem Clin Biochem 17:807.

Dybkaer R. 1978b. Approved recommendation (1978). List of quantities in clinical chemistry. (IUPAC and IFCC Recommendation 1978). J Clin Chem Clin Biochem 17:822.

Dybkaer R and Solberg HE. 1987. Approved recommendation (1987) on the theory of reference values. (IFCC and ICSH Recommendation 1987). Part 6. Presentation of observed values related to reference values. J Clin Chem Clin Biochem 25:657.

Gehan EA and George SL. 1970. Estimation of human body surface area from height and weight. Cancer Chemother Rep 54:225.

Gräsbeck R. 1990. Reference values, why and how. Scand J Clin Lab Invest 50(Suppl 201):45.

Harris EK. 1975. Some theory of reference values. 1. Stratified (categorized) normal ranges and a method for following an individual's clinical laboratory values. Clin Chem 21:1457.

Harris EK and Boyd JC. 1990. On dividing reference data into subgroups to produce separate reference ranges. Clin Chem 36:265.

Kassirer JP and Bleich HL. 1965. Rapid estimation of plasma carbon dioxide from pH and total carbon dioxide content. N Engl J Med 272:1067.

Lehmann HP. 1979. SI Units. CRC Crit Rev Clin Lab Sci 10:147.

Linnet K. 1987. Two-stage transformation systems for normalization of reference distributions evaluated. Clin Chem 33:381.

Lippert H and Lehmann HP. 1979. SI units in Medicine. Urban & Schwarzenberg, Baltimore.

Noe DA. 1991. A body surface nomogram based on the formula of Gehan and George. J Pharm Sci 80:501.

Ransahoff DF and Feinstein AR. 1978. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 299:926.

Royston P. 1992. Estimation, reference ranges and goodness of fit for the three-parameter log-normal distribution. Stat Med 11:897.

Schultz EK, Willard KE, Rich SS, Connelly DP, and Critchfield GC. 1985. Improved reference-interval estimation. Clin Chem 31:1974.

Scott DW. 1979. On optimal and data-based histogram. Biometrika 66:605.

Siggard-Anderson O, Durst RA, and Maas AHJ. 1987. Approved recommendation (1984) on physio-chemical quantities and units in clinical chemistry with special emphasis on activities and activity coefficients (IUPAC and IFCC Recommendation 1984). J Clin Chem Clin Biochem 25:369.

Solberg HE. 1987a. Approved recommendation (1986) on the theory of reference values. (IFCC and ICSH Recommendation 1986). Part 1. The concept of reference values. J Clin Chem Clin Biochem 25:337.

Solberg HE. 1987b. Approved recommendation (1987) on the theory of reference values. (IFCC and ICSH Recommendation 1987). Part 5. Statistical treatment of collected reference values. Determination of reference limits. J Clin Chem Clin Biochem 25:645.

Solberg HE. 1994. Using a hospitalized population to establish reference intervals: pros and cons. Clin Chem 40:2205.

Strike PW. 1996. *Measurement in Laboratory Medicine: A Primer on Control and Interpretation*. Buttersworth-Heinemann, Oxford.

Willard KE and Connelly DP. 1992. Nonparametric probability density estimation: Improvements to the histogram for laboratory data. Comp Biomed Res 25:17.