Ann. Ist. Super. Sanità
Vol. 27, N. 3 (1991), pp. 427-436

STANDARDIZATION OF IMMUNOASSAYS

R.F. MASSEYEFF

Labs, Nice, France

Summary. - *The aim of standardization is to ensure that assays of the same analyte in the same samples, done at different places or at different times or both, can be readily compared. Standardization is especially desiderable for immunoassays because all external quality assessment surveys have shown that this type of method involves a much greater variability than traditional assays. A major problem in immunoassays is that the recognition of the analyte is determined by the reagent (i.e. an antibody or conversely, if antibodies are to be determined, an antigen) using a limited portion of the molecule, the epitope or the paratope, respectively. Therefore, the specificity of whole process is questionable, even if the specificity of the epitope/paratope reaction is fair. Biological variability is such that many molecules possessing the same recognition site coexist in biological fluids. A first example is the different secretion forms of hormones, their fragments, and other related hormones which occur simultaneously in the blood. A second example is the high variety of antibodies possessing the same paratope. Many other examples can be cited. Thus, a precise definition of the analyte should be given, dependent on the aim of the assay. This is a major conceptual difficulty because several different molecules having the same recognition site may cooperate in a physiological function. In this context, it may be useful to assay them simultaneously, but it is impossible to express their concentrations in common mass units. Another major difficulty in immunoassay standardization is due to the fact that such methods operate at very low concentrations and are sensitive to the environmental conditions created by ambient molecules existing in much higher concentrations. These "non-specific" effects create problems of non-comparability due to the prevalent conditions in different samples. From the different causes mentioned above, it follows that the major difficulty is to ensure perfect identity of the conditions in the calibrator and in the sample (analyte and matrix). The current procedures of assay standardization are reviewed and their deficiences are stressed. Recent proposals for rationalization of standardization of immunoassays are described.*

KEY WORDS: standardization, immunoassay, antigen and antibody heterogeneity, matrix effects, reference standard, test of validity.

Riassunto (La standardizzazione del dosaggio immunologico). - *Lo scopo della standardizzazione è quello di garantire che i dosaggi dello stesso analita negli stessi campioni, eseguiti in posti diversi o in tempi diversi (o in entrambe le situazioni), siano direttamente confrontabili. Una standardizzazione, in particolare, è auspicabile nel dosaggio immunologico, che i programmi di valutazione esterna di qualità hanno dimostrato comportare una variabilità assai più grande di quella dei metodi chimico-clinici. Uno dei maggiori problemi del dosaggio immunologico deriva dal fatto che il riconoscimento dell'analita da parte del reattivo (cioè l'anticorpo, o, se oggetto della misura sono gli anticorpi, l'antigene) interessa una porzione limitata della molecola, rispettivamente l'epitopo o paratopo: la specificità dell'intero processo resta pertanto dubbia per quanto elevata sia la specificità della reazione epitopo-paratopo. La variabilità biologica è tale che nei fluidi corporei coesistono molte molecole con lo stesso sito di riconoscimento: un esempio è dato dalle differenti vie di secrezione degli ormoni, dai loro frammenti e dai molti altri ormoni correlati presenti nel sangue; un altro, dalla grande quantità di anticorpi che possiedono lo stesso paratopo; molti altri esempi potrebbero essere ricordati. Così, una definizione precisa dell'analita dovrebbe essere data in dipendenza dello scopo del dosaggio. E questo comporta una grande difficoltà concettuale, dato che molte molecole differenti, che hanno lo stesso sito di riconoscimento, potrebbero cooperare nello stabilire una certa funzione fisiologica; in tale contesto, potrebbe essere utile determinarle assieme, ma sarebbe impossibile esprimere la loro concentrazione in una comune unità di massa. Un'altra grande difficoltà che si incontra nella standardizzazione dei dosaggi immunologici è dovuta al fatto che tali metodi operano a concentrazioni estremamente basse e sono sensibili alle condizioni ambientali determinate da molecole presenti nei campioni a concentrazioni molto più elevate: questi effetti "aspecifici" creano problemi di non confrontabilità dovuti alle condizioni che prevalgono nei diversi campioni. Da quanto detto consegue che la difficoltà più grande è quella di assicurare la perfetta identità delle condizioni del calibratore e del campione (analita e matrice). Sono passati in rassegna i procedimenti correnti di standardizzazione, sottolinean-*

done la fragilità; vengono anche descritte recenti proposte per razionalizzare la standardizzazione del dosaggio immunologico.

PAROLE CHIAVE: standardizzazione, dosaggi immunologici, eterogeneità antigenica e anticorpale, effetti matrice, standard di riferimento, prove di convalida.

Rationale of assay standardization

Aim

The aim of assay standardization is to ensure that the assay of the same analyte in the same samples, made at different places or at different times or both, can readily be compared [1]. In laboratory medicine, standardization is essential not only for diagnosis and monitoring of the individual, but for physiological or epidemiological studies or simply for the most economical approach to health care. Standardization is especially desirable for immunoassays because all external quality assessement surveys have shown that this type of assay implies a much greater variability than others based upon chemical or biochemical principles.

Assay definition

According to metrologists, an *assay* (or measurement) consists in the *quantitative description of a property of the analyte* [2]. In practice, the assay consists in submitting the analyte in the sample to some physical or chemical interaction whereby the analyte induces an *effect*, generally the production of some signal directly perceived by human sense or most often by a non-human sensor. This effect is generally measured in a physical device, but it may be expressed in a much more complex integrated system such as a living cell or whole organism. This latter condition defines bioassays.

Classification of assays

Recently, Ekins [3] has expressed differently the fact that there are two categories of assays with different aims. The former, called *comparative assays*, are oriented towards the assessment of *effects* (e.g. biological effects), irrespective the structure of the substance assayed. A good example would be the assessment of the effect of a drug on diuresis, which can be obtained with drugs differing widely in their chemical structure. Obviously, such assay cannot lead to the expression of common weight quantities. However, *ad hoc* "activity" units can be defined *within a certain method*. The latter category is represented by *analytical assays*, which are aimed at measuring the *amount* of one particular substance. Therefore, they can measure only one structurally defined substance at a time, since the ultimate aim is the expression in mass or molar units. The measurement of several related substances in a single operation is not possible, in theory, because no common unit can express the quantities of different substances. As it will be discussed later, this represents a major conceptual difficulty in immunoassay standardization, because immunological recognition most often is not able to ensure the identification of a single component in a mixture, but reacts also with related material. It can be considered that closely related molecules may be treated as a single entity. The degree of resemblance of the different molecular forms of the analyte is a matter of discussion depending on the particular case. The purpose of the assay should be taken into account in order to consider only the members of the family of molecules related to this purpose. Ekins' paper points out that it is impossible to express activity units and quantities with the same unitage, and that it is impossible to express activity units with the same unitage when different methods are used. Clearly, immunoassays have analytical objectives. However, the distinction between comparative and analytical assays should be constantly kept in mind, because there are many ambiguous situations where guidelines are needed.

True and consensus values

Ideally, perfect comparability is achieved when the *true* value of the analyte in the sample can be measured, within known limits of confidence. This ideal goal can be achieved in a very small number of cases, when extremely accurate methods (e.g. methods involving mass spectrometry) and when fully reliable reference materials (e.g. internal standards labelled with stable isotopes) are available. In most cases this goal can be only approached. This is the reason why, in most cases, the accepted value is the one reached by *consensus*.

Methods of standardization

Theoretically, comparability can be achieved in two ways:

1) By the universal use of a single carefully described method using the same reagents. Although great efforts have been pursued in this direction, it is generally not feasible to describe such methods, which would be called definitive (in rare circumstances) or, at least, reference methods.

2) By applying acceptable methods to the same material containing the analyte, *i.e.* different samples and calibrators. At least one well studied common calibrator must be distributed. If this material has the required features, later it may be called "reference material". The results are evaluated by the currently recognized statistical methods.

Since in most instances (especially in immunoassays) there are no definitive nor reference methods, the latter approach is the sole possibility. Therefore, over time standardization has been primarily based on the selection of analyte containing material, which could serve as common basis for result comparison (reference preparation) [4]. The image of "common currency" has often been employed, although it is open to criticism.

Relative potency

The assay of different samples is aimed at measuring their *relative potencies*: the term "potency" is supposed to include not only biological activity, but also immunoreactivity, within *in vivo* and *in vitro* systems. The ambiguity of this concept should be noted, bearing in mind the Ekins' remarks. The relative potency of a preparation W is measured by reference to preparation R (stable candidate reference preparation), which has been defined to contain x units of active substance per unit volume or mass. Therefore the potency (P) of W is xP_{WR} units of activity per unit volume (or mass). Similarly, if P_{ZW} is an estimate for the relative potency of a third preparation Z in terms of W, then Z can be calibrated *via* W in terms of the units defined with reference to R, *i.e.* Z is estimated to contain $xP_{ZW}P_{WR}$ units of activity per unit volume (or mass). If a stable standard R of defined unitage is available, a calibration system in which all measurements of potency can be expressed in common units can be proposed. Almost all assays work by comparison of the signal obtained with the analyte in the sample to that of a reference preparation containing the same substance in pure form. The construction of a dose/effect curve (calibration curve) with the reference material, or with a secondary standard, provides the basis for a quantitative assay, the result of the sample being interpolated from the calibration curve. The shape of the dose/effect curve is important. With increasing analyte doses, the first observed effect indicates the sensitivity threshold of the method; then, usually, a concentration zone occurs where the dose/effect relationship can be mathematically modelled; this is the usable part of the curve. Finally, there is a second threshold when the slope of the dose/effect curve is such that it cannot be used for quantitation. In many instances its slope may be reversed. The best way to determine the usable part of the dose/effect curve is to generate the precision profile curve which was proposed by Ekins for immunoassays [5].

Processing the results

The results obtained using the candidate reference material by the chosen method(s) in different laboratories should be analyzed and compared with the results obtained by other methods. All these comparisons should be made using the statistical procedures of external quality assessment. If different methods are used with the same calibrators, it is essential to check that the results from one method are *commutable* to another one. Commutability can be defined as the existence of an identical and constant relationship, within the limits of experimental error, between results by two assay methods for all samples, including both reference materials and patient samples [6].

Drawbacks in immunoassay standardization

All assays are subject to two causes of error: *imprecision* and *bias*. The high sensitivity of many immunoassays (which are often used in a zone close to their detection limit) increases imprecision. On the other hand, many different causes of bias exist in immunoassays that are avoided with traditional assays. For example, immunological reagents are far less well characterized than conventional reagents, and the same applies to the interferences that occur. Special attention should be given in this definition to the term *analyte* as it is used in immunochemistry.

The definition of the analyte may be imprecise in immunochemistry

The *analyte* is the substance assayed in the sample. A precise definition of the analyte could avoid many subsequent problems. In principle, the analyte is defined as a set of molecules possessing an *identical chemical structure*. In certain cases (never in the field of immunochemistry), the definition of the analyte is straighforward. Take the example of ^{27}Al assayed in a sample: this substance has unambiguous physico-chemical characteristics and the assay by the correct spectrographic methods can be enterely specific. However, when considering much more complex substances, such as biological molecules and especially macromolecules, the definition may not be so clear, because of their potential intrinsic or acquired heterogeneity. This is already well known in biochemistry. For example an enzyme, in most instances, is assayed by determining its catalytic activity: however, it may be present in different isoenzymatic forms which may or may not have the same quantitative catalytic activity. From the user's point of view (*i.e.* that of the clinical biologist or the clinician), these different isoenzymes may or may be not considered as equivalent. In immunoassays, a wide variety of molecules with analogous immunoreactivity may coexist in samples, but they may have a very different physiological or even diagnostic significance. The question of the heterogeneity of molecules recognized by immunoreagents will be discussed in detail below. It can be extreme when a single monoclonal antibody is used for the recognition of carbohydrate or glycolipid residues. Then, the analyte is an epitope rather than a molecule. Thus, the kind of molecules which are assayed by immunological methods exhibits an especially broad heterogeneity, to such an extent that *the definition of the analyte itself may become ambiguous*. The definition of certain analytes reveals another form of ambiguity: for obvious historical reasons, the molecule has been first described considering its *physiological effects*. This has been true for most hormones. Quoting, Ekins [3] has commented on the fact that reference preparations which have been approved for the definition of international units (IU), are presented with such strange statement as "the IU for TSH is the activity contained in 36.6 mg of the International Reference Preparation". (In this text, activity stands for biological activity). Lets citate Ekins: "this misconception has led in turn to the mistaken belief that the IU represents a unit of *amount* of the active substance(s) present in the preparation, an erroneous concept reflected in the current practice of expressing the *concentrations* of many heterogenous

substances (e.g. the glycoprotein hormones) in terms of IU/ml. This is evidently a metrological nonsense, since the International Reference Preparation *defines* but does not *contain* the unit of activity".

Limits for standardization

Four important conditions should be observed in order to obtain a valid assay: (1) the reactivity of the immunoreagent shoud be known, (2) the analyte should be unambiguously defined, (3) the sample (and the calibrator) should not contain substances capable of generating, either positively or negatively, the same effect as the analyte, (4) the calibrator should be identical to the analyte in the sample. These conditions are not always fulfilled. Such failures create limits for assay standardization.

Reactivity of the immunoreagent

A specific antibody can be considered as a recognition unit which can identify a part of the surface of the antigen molecule, the antigenic site or epitope. The structural conditions that create an epitope are still controversial. It is likely that the potential number of epitopes on the surface of a protein is very high [7]. The counterpart on the immunoglobulin molecule, the antigen binding site or paratope, has a variable area, which according to present knowledge, can extend to 6 nm^2. It contains a set of aminoacids which may bind the antigenic site through a set of weak interactions assembled in a unique stereospecific shape. It may contain subsites which can explain certain cases of multispecificity [8]. An intrinsic weakness of immunochemical recognition is that it is not global: only a limited part of the analyte is identified. However, the selectivity of the paratope for the immunizing epitope is not absolute and resembling epitopes can be accomodated. For hapten molecules, the epitope and the paratope are in the same range of magnitude, but for large macromolecular antigens, they are very different in size. Because of the absence of global recognition, there may be errors in the analyte identification, especially when only one antibody is used. Altered non-biologically active molecules may be recognized on the basis of epitope-paratope reactivity. In general, there is no compulsory parallelism between immunoreactivity and biological activity. The errors in immunological recognition are due to the existence of *cross-reactions*. These fall into two groups. The former are due to the existence of *shared epitopes* in different biological molecules (e.g. shared epitopes in the group of pituitary hormones such as TSH, LH, FSH, hCG). In this case, the antigen-antibody reaction works as expected, but the epitope is not properly selected for ensuring specificity. Unfortunately, these cross-reactions cannot always be foreseen: for example, recent genetic studies have revealed the existence of gene families which code for proteins which previously were not supposed to be related (e.g. immunoglobulins and carcino-embryonic antigen). The latter group are related to the *true cross-reactivity*, which occurs when the paratope accomodates structures different from that of the initial epitope. They may be related and therefore expected or, in certain cases, completely unexpected. Even if it is weak, cross-reactivity may lead to major problems. An antibody against T3 cross-reacting with T4 at the level of 1% cannot be used since T4 is present in plasma at a concentration 100 times higher than T3. An example of unexpected cross-reaction was recently observed when it has been shown that digoxin assays can be disturbed by an unknown natural substances occurring in the blood of pregnant women, infants and dialyzed patients. Many immunochemical methods still employ polyclonal reagents (antisera or purified preparations from antisera). In the majority of cases, the antibodies obtained by artificial immunization with purified antigens recognize unknown epitopes. The nature of the complementary structures of antigen and antibody have been extensively studied only in a very few cases, (but these are of great importance as they provide us with models). Furthermore, large molecules simultaneously express numerous epitopes, which stimulate the formation of multiple, different, corresponding antibodies in hyperimmunized animals. The antiserum contains a mixture of antibodies in variable and undetermined proportions. Therefore, the specific reagents used in immunoassays, *i.e.* the antisera, depend on the recognition of antigen epitopes unknown in their number and structure by specific antibodies, the specificities and the concentrations of which are themselves unknown. Furthermore, each batch of a specific antiserum, even if it originates from a single animal, may differ in its antibody composition and reactivity. In such conditions, the most remarkable feature of immunoassays is that, nevertheless, they work!

The increasing availability of monoclonal antibodies directed against a wide variety of epitopes bring a certain degree of simplification into this complexity. Although the precise antigen structures which are recognized remain unknown in most cases, monoclonal antibodies have the advantage of being homogeneous and permanently reproducible (*i.e.* they are not liable to batch-to-batch variation).

Similar problems, due to the insufficient knowledge of the reactivity of the immunoreagent, arise in the determination of antibodies, since it is always difficult to ensure that the antigen preparation does not contain epitopes common to completely different structures (e.g. another infectious agent).

Insufficient scientific knowledge of the analyte structure

Despite active progress in science, there are still many antigens that are insufficiently biochemically characterized and it is therefore difficult to secure standardization. Many examples can be found in allergic diseases, mycoses, parasitic diseases, etc.

Intrinsic and acquired heterogeneity of the molecules recognized by immunoreagents

This question relates to the definition of the analyte, as discussed above. Molecular heterogeneity is a widespread phenomenon in biochemistry. There are multiple causes of

heterogeneity such as genetic allotypy, substances secreted in different forms by different organs (e.g. glucagon in pancreas and gut), substances secreted in a precursor form (preproprotein and preprotein), substances easily degraded (fibrinogen, parathyrin), occurrence as single chain, the original protein being multichain (light chain of immunoglobulin, β-hCG); quite often the isolated chain coexists with the parent protein and therapeutic products coexist with the physiological substance (e.g. porcine and human native insulin). Last but not least, the analyte may be a pathological "abnormal" product (such a myeloma protein, often deleted, or an abnormal ACTH in Cushing's disease). The extreme example of intrinsic heterogeneity is observed with antibodies: a wide variety of antibodies react with the same antigen molecule or even with the same epitope (with different intrinsic affinity constants), and each of these antibodies may be associated with an array of different isotypes. It is therefore impossible to give a scientifically valid expression of an antibody assay in mass terms, because the molecules are different. However, compromises may be proposed in order to define groups of antibodies [9]. Another form of heterogeneity is commonly observed in drug assays, *i.e.* the occurrence of various metabolites. These many different forms may or may not possess the biological activity attached to some of these substances, and if they have it, it may be quantitatively different. Therefore, it is not surprising that frequent discrepancies between the results obtained by immunoassays and bioassays are observed. For these many reasons, immunoassay results are not always expressed in moles or g/l. Cooperative studies and external quality assessment of results usually show quite large scatter and, even for relatively well described antigens such as α-fetoprotein, there was a recommendation to express results in IU. In fact, most laboratories use ng/ml, but it is likely that the results obtained using different kits are not always comparable. Although it is desirable to use molar units in the future, for many antigens there is insufficient control of the assay variables, and the use of common arbitrary units is justified at least provisionnally. Therefore, in any attempt of standardization, the goal which is pursued should be precise, since it could strongly influence the selection of the reagents and the planning of the survey.

Carbohydrate or glycolipids grafts on macromolecules

One special but frequently observed case is that of the use of a monoclonal antibody for the assay of substances still poorly characterized, just because this monoclonal antibody itself has been the tool which has permitted the discovery of the antigen. Many so called "tumour markers" have been introduced by this process. For substances such as CA 19-9, it turned out that the antibody recognizes a carbohydrate component (2,3,-disialo-lacto-N-fucopentaose II) which is grafted on a protein molecule, which may be not entirely specific for this unique protein structure. Then the analyte cannot be defined as a single molecular entity. It is just an epitope, defined by the reactivity of the antibody. It cannot be expressed in mass units.

Influence of the method on the quality of immunological recognition

Certain drawbacks can be avoided, in some cases, by proper selection of the reagents, provided that the problems can be identified. Immunoassays in which the analyte is first extracted by a reagent and then recognized by a second reagent (e.g. sandwich), have greater possibilities than the competition design, which relies on the recognition of the epitope by a single antibody. The sandwich method, especially when using monoclonals with well studied specificity, allows selection of the nature of the molecules that will be assayed to some extent. When recognition is obtained by a single reagent (typically the case in the assay of small hormones and drugs), the result is strongly influenced by the presence of cross-reacting material. However, the nature of the substances capable of cross-reacting is itself highly dependent on two factors: the design of the immunogen which has been used to obtain the antibodies, and the design of the labelled antigen. They should be coordinated in order to obtain the best possible response. If the bridges linking the carrier on one hand and the tracer on the other hand are located in too dissimilar zones of the molecule, a complete loss of recognition may be observed [10].

Interference in immunoassays (matrix effects)

Neither sample nor calibrator should contain any factor that could interfere positively or negatively, changing the final signal. Any substance in the sample (or in the calibrator) other than the analyte and the immunoreagent that will alter the reaction is classified as a *matrix effect*.

Influence of the method on the sensitivity to interferences. - Methods operating in reagent excess largely escape the problems created by the various factors inhibiting the antigen-antibody reaction (see below). The methods most sensitive to these factors are those in which only a limited amount of reagent is available (competition methods). If an interfering factor is present in the sample, the reaction proceeds as if less reagent is available. Therefore, the calibration curve is invalid. Since these curves are strongly influenced by the quantity of reagent, large errors (through excess) may be observed.

Diluent. - The diluent used may influence the development of the antigen-antibody reaction. In certain cases, additives are intentionally added in order to accelerate the reaction. For example, polyethylene glycol additives are frequently employed in immunonephelometry. It is common experience to notice that a calibration curve constructed from experiments using a protein-rich mother solution of an antigen and its various dilutions in buffer

does not parallel one made in a protein enriched buffer. There is a shift towards lower signals at high protein concentration. This is due to some type of interference with the antigen-antibody reaction, which is usually slowed down. The intensity of matrix effect depends on the quality of the environmental molecules (proteins?). Matrix effects are usually more important in techniques with analyte excess than in techniques with reagent excess. A commonly used method to equalize the matrix effect in all tubes is to make all dilutions in a reasonably high protein diluent (bovine albumin solution, diluted serum, casein, skimmed milk, gelatin, various buffers, etc.)

Zero calibrator. - A related problem is that of the *zero calibrator*, which may be difficult to prepare. For example, it is difficult to obtain a preparation resembling serum, but devoid of thyroid hormones T3 and/or T4. In some cases, the subtraction procedure creates interferences in the assay.

Anti-globulin factors, heterophilic antibody, immune complexes. - The antigen-antibody reaction may be disturbed by antibodies reacting against the antibody reagent. They may be auto-antibodies in the patient plasma (rheumatoid factor, cross-reacting with animal immunoglobulins, which produces erroneous results both in antigen and antibody assay) or other varieties frequently found such as anti-mouse antibodies [12]. Anti-globulin antibodies directed against the Fab fragment may interfere considerably in competition assays. The occurrence of circulating immune complexes in certain patients give artifacts in some assay methods (e.g. the assay of thyroglobulin in the presence of spontaneously occurring anti-thyroglobulin antibodies). Antibodies against peroxidase epitopes may also creates problems in immunoenzymatic methods (author's unpublished results). The best method to avoid the problems created by all varieties of anti-globulin antibodies is to add an excess of immunoglobulins originating from the same animal species from which the antibody reagent comes.

Interference with the signal. - The samples may contain substances interfering with the final production of the signal. This does not occur frequently with assays involving a separation step, because the potentially interfering factors are eliminated. However, it may occur in all non separation assays (nephelometric assays, so called homogenous assays). For example, a drug may be present in the sample which could interfere with a fluorometric assay.

Non-identity of the assay conditions in the calibrator(s) and in the samples. - From the above discussion, it is clear that the basic requirement for validity of an immunoassay, *i.e.* the identity of assay conditions in the calibrator and in the sample may be difficult to reach and almost impossible to garantee in many cases.

The critical points are: (1) the diluent (in assays where the sample is not much diluted, it is difficult to ensure that the same "matrix" exists in the sample and in the different dilutions of the calibrator); (2) the other different "matrix" effects, which cannot always be controlled, because the origin of some is still unknown, and some other are occasional; (3) the non identical chemical structure of the reacting material in the sample and in the calibrator.

These major problems may have a very different importance according to the considered test. It is easier to ensure the identity of the calibrator and the sample with small hapten molecules (steroid hormones, drugs). However, the sample may contain cross-reacting material (other hormones, metabolites).

The immunoassay of cyclosporin has provided a most illustrative model of the difficulties caused by the occurrence of cross-reacting metabolites. These metabolites differ in relative quantity according to the duration of the treatment, they differ in their pharmacological activity since some are still active, and they may be recognized by certain antibodies and not by others, according to different patterns. The various immunoassay (kits or in-house methods) may give very different results to the HPLC assay (this difference changing with time, as the treatment progresses) and between themselves. This assay offers an almost complete set of the theoretical and practical problems of the immunoassays. What must be assayed? Only cyclosporin? Cyclosporin plus the active metabolites? What is the best antibody to reach the chosen goal? Polyclonal, monoclonal, a mixture of monoclonals? How should the results be expressed? It is clear that no answer is satisfactory, and the decisions are always the result of compromises. With protein antigens, there are many causes of heterogeneity, with respect to both sample and calibrator. There is no general answer to the problems that are encountered.

Another illustrative example is that of parathyrin, a hormone liable to spontaneous degradation. According to the method used and to the selection of antibodies, very different results may be obtained. However, the use of a sandwich assay with selected monoclonal antibodies permits the assay of only the "native" molecule. This solution is preferred, not because it is intrinsically the best, but only because the results show a better correlation with the clinical findings.

Each of the different protein hormones gives specific problems; there is no general answer, and each case should be considered separately according to the aim of the assay, taking into account the available scientific knowledge. These considerations should deeply influence the criteria for the selection of any candidate reference preparation. The different International Reference Preparations (IRP), which were initially based on bioactivity and expressed in activity units, are not suited for immunoassays and there are difficulties in ensuring the continuity of standards when an IRP is exhausted. When a new candidate preparation for replacing IRP 78/549 for FSH was tested *in vivo* and *in vitro*, discrepancies appeared in the average ratios. A wide scatter was observed. The new preparation was assigned activity units, but this unitage was unable to ensure the continuity of the standardization for FSH immunoassays.

Improper separation of the free reagent

Many assays require observation of the distribution of a labelled reagent into two compartments: bound to the immunoreagent or free. In the past, one frequent cause of error was the lack of reliability of the process of separation of these two forms. The situation has greatly improved with the popular use of efficient solid phases. However, there are still causes of variability, such as the occurrence of immune complexes or mechanical causes. One potential cause is non-specific absorption on the solid phase.

Solid phase non analyte-specific absorption

In most separation immunoassays, the final reaction is developed from the solid phase. Thus, any kind of non-specific absorption may artificially increase the readings. This non antigen-specific absorption is especially important with use of the technology of coated tubes or beads. In this case, the plastic *solid phase* used in many methods must be considered as a reagent. Its perfect standardization is a formidable problem for all reagent firms. Non analyte specific absorption is especially important at high protein concentration, and therefore is common in antibody determination. In this case, the amount of specific antibody is usually small with regard to the quantity of non antigen specific immunoglobulins. Since the specific antibody is most often identified *via* a second layer of labelled anti-Ig antibody, this undesired absorption causes high background noise and may lead to falsely positive results, especially when the final response is qualitative (positive/negative, such as in testing hepatitis or HIV antibody). The quality of the solid phase surface is a determinant factor of this Ig non antigen-specific absorption. IgG absorption is associated with positively charged hydrophobic surfaces. Methods for the control of this side effect have been recently been suggested [11]. Such methods are needed for the improvement of immunoassays.

The high-dose "hook" effect

The sandwich method using two different monoclonal antibodies in a one step assay (Tandem® method) is now frequently used for protein assay. A major drawback may occur when the sample contains a very large amount of the analyte. The readings may be very low corresponding to minute amounts of antigen, even approching zero in a few instances. When an extended calibration curve, with large antigen excess is constructed, it may be observed that the curve is reversed after a plateau, and may reach negligible readings in extreme antigen excess. In fact, this phenomenon is not an example of the "hook effect" as was described in 1974 by Miles *et al.* [12]. In this latter case, what is observed is similar to the "prozone effect", well known in reverse agglutination techniques. All the antibody sites, both in the solid phase and in the labelled reagent, are saturated by excess antigen and the antigen bridge cannot be established. But, if the same test is made with an intermediate washing, the high-dose effect is not anymore observed. The true "hook effect" as described by Miles *et al.* is not limited to on-step methods and is more rarely observed. It is usually less pronounced and the mechanism is still controversial. These drawbacks are observed with analytes showing a wide pathophysiological range and are exceptionally encountered. It is possible to limit the problem by displacing the high-dose effect towards greater amounts of antigen simply by increasing the amount of solid phase antibody; however, it is not suppressed. The problem is that the high-dose effect is impossible to predict. The only way to prevent its occurrence is to routinely perform the test at two dilutions, and look for discrepancy, but this measure doubles the analytical cost.

Criteria of validity

The validation of an immunoassay method is made with different tests. None of them can truly validate the assay; therefore, they may rather be called "invalidity tests". However, the accumulation of negative invalidity tests is generally accepted as a positive global test for likely validity. The following test can be recommended: trials with possible cross-reacting substances, trials with potential interferring substances (analyte free samples, chemicals, antiglobulins, etc.), general test performance (between-assay and day-to-day variability, imprecision profile), assay of reference samples (if available), comparison with reference methods (if available), comparison with different antisera (or antigen for the determination of antibodies), pathophysiological validation, recovery test (addition of known amounts of standard), dilution test (serial dilution of the sample and standard), stability test (trends of the daily mean and median).

Ekins [3] has recently suggested a new test for detecting the ultimate lack of identity of the analyte in the sample and in the calibrator. The test refers to the fractional occupancy of the antibody by the antigen when the antibody concentration varies widely. In a competitive assay, the vacant antibody sites are measured using a labelled anti-idiotypic antibody, permitting the evaluation of fractional occupancy (ratio of occupied sites to total number of sites). Another different label on antibodies permits the estimation of the total number of sites. In a labelled antibody method, the second antibody is labelled with a tag different from that of the capture antibody. In this case, calculation shows that the relative potency of a preparation (as compared to the standard) tends to K_S/K_U when the free antibody concentration $[fAb] \rightarrow 0$ (K_S and K_U standing for the affinity constants of the standard and the unknown, respectively). Therefore, by varying the antibody concentration and plotting the fractional occupancy, it is possible to demonstrate the existence of difference in affinity constants, which may reflect the existence of molecular differences of the reacting material in the unknown sample and in the standard. This test is said to be far

more sensitive than the usual dilution test. However it requires mastery of this new technique using two different labels.

Procedures for establishment of standard reference material (SRM)

Despite the numerous theoretical and practical problems which are still not resolved, standardization of immunoassays remains an obvious need and has been in operation for more than two decades [14-16]. The following summarizes the practical process of standardization. Standardization may be initiated at different possible levels, rarely regional, often national, and in many instances international (group of countries, e.g. BCR "Bureau Communautaire de Référence" for EC, or worldwide with IFCC or WHO). The commercial companies have their own internal problems of standardization and usually actively participate in national and international efforts. The current number of immunological test systems is very large and the first step is to determine priorities. Once the decision has been taken, the first task of the expert panel is to define the specifications of the candidate SRM. The following criteria should be met by this material:

- It must be adequately characterized. There may be differences between different batches of the same preparation according to their origin and/or the method of purification. Synthetic preparations may also differ from natural ones.
- It must be readily available.
- It must have the same property in the assay system as the substance to be measured. This point is especially important for the highly sensitive immunoassays in which matrix effects may be important. For example, a solution of the purified substance (e.g. digoxin) in the appropriate solvent is usually not a good working standard, because digoxin is currently not measured in a solvent but in serum or plasma, the components of which interfere in the antigen-antibody reaction. Therefore, standards are usually prepared in the matrix which is most commonly used in assays.
- It must be stable. This is, of course, an essential point which is self explanatory. Quite frequently, purified antigens are much less stable than the same antigen in a more natural protein environment. This is a further reason why standards are usually not made of purified antigen.
- It must be physically homogeneous.
- It must be free from bacterial contamination.
- It must be compatible with accurate division into aliquots.
- It should be compatible with freeze-drying with minimal denaturation.
- It should have a uniform moisture content of less than 1%.
- It must give a clear solution after reconstitution.
- Its antigenic content must be high enough to permit optimal dilution in the appropriate "matrix".

In addition of these general specifications, the expert committee should prepare a report on the specifications related to the particular analyte. It can be predicted that this aspect will become more important in the future. The next step concerns the source of the material and the way it is procured, treated, transported and prepared. In some cases, this step may be dependent on the intended use of the standard. For example, the addition of azide should be avoided for a preparation which will be used with a peroxidase-linked immunoreagent. Usually scientific and/or commercial organizations are requested to present candidate preparations, which will be examined in a first phase study in order to select the best one.

Collaborative study

When the material is in its final condition, its uniformity must be tested as well as its stability, usually by conventional accelerated degradation tests at 56 °C. There is no strict rule for the acceptable degree of stability, but obviously it should at least be of several years for a practically useful standard.

The candidate preparations are then tested by several expert laboratories in a collaborative study. The design of this study should be carefully prepared with the help of a biostatistician. The constitution of the panel samples that will be tested needs special consideration. The selection of expert laboratories is another important point. However, the major point is the selection of a commonly agreed method, in the absence of a standard method. The optimal requirements for such a method are of analytical, epidemiological, economic and practical nature. Its overall imprecision should be determined as well as its overall day-to-day reproducibility. It is desirable that the diagnostic performance of the method is known; *i.e.* specificity (the proportion of negative specimens that are correctly classified as negative by the method) and sensitivity (the proportion of positive specimens that are correctly classified as positive by the method). Other important parameters are simplicity, availability of reagents and cost of equipment and reagents. It is also desirable to compare data obtained by other methods in use in the different laboratories.

The results are collected for a statistical study. Whenever possible, dose/response curves are constructed and compared. They should run parallel over the selected assay range. A unitage should be proposed and finally an instruction sheet is written indicating the intended use and limits of the material and the way in which it is to be used.

SRM are precious reagents; they should never be used as working reagents, nor for experimental purposes. They are primarily intended for the calibration of working (secondary) standards. Tertiary standards are usually made available by commercial companies.

Approval and distribution

The standard reference material has to be approved by official authorities. They may suggest further studies or changes. There are many different initiatives in this do

main, and there is a need for coordination. In fact, many projects are done in cooperation between the different agencies. An important aspect of standardization is the need for continuity. Once the stock of an SRM is approaching exhaustion or end of shelf-life, it should be renewed. A special part of the new collaborative study should be devoted to the calibration of the new SRM in terms of the former one.

Role of quality assurance in standardization

The methods of quality assurance are in constant use during the process of standardization to detect the possible bias, to monitor the reproducibility of the results and, finally, to validate the method. When a method reveals poor quality assurance controls, this is incentive for undertaking the steps for standardization. Finally, external assurance control may reveal the problems occurring with certain standards.

Conclusion and perspectives

The different facets discussed above show that, by nature, standardization cannot be the work of isolated scientists. It is a cooperative task, involving individual scientists, scientific organizations, commercial firms and international organizations. Therefore, the process is submitted to all the hazards and problems of human cooperation, including divergent views of experts, national or commercial pressures, etc. Indeed, standardization implies a long, complex, systematic and expansive collaborative effort, involving unavoidably considerable administration, even bureaucracy. Finally, a *consensus* should be reached between the different experts about the best available procedure and material allowing an approach to the initial goal. According to their degree of reliability, they may or may not be accepted as reference method and preparation. Therefore, as stated above, it is evident that standardization is a long and evolutionary process, whose progress parallels that of scientific knowledge on the analyte and on analytical methods. It is also clear that it has a economic impact, due to the growing importance of commercialized reagents, especially in laboratory medicine.

The present standards are not always satisfactory for a number of analytes (especially those showing polymorphism) for the reasons emphasized above. The theoretical constraints limiting the feasibility of such standards are such that the existence of a fully satisfactory solution is questionable. However, the need for comparing results remains obvious. For certain assays presenting specific problems, a wild kind of standardization exists where commercial companies tend to adjust their assay results on those of the market leader. Although this non-scientific practice cannot be recommended, it demonstrates both the real need for standardization and the difficulty in finding simple scientific solutions.

Here, two examples of proposals for the future are given, which evolved from the same type of caution about the present methods of standardization. Ekins [3] observed that, on one hand, it is incorrect to use the same units for defining biological activity and amount of a preparation (as determined by immunoassay), and that, on the other hand, it is "futile to assign units value to the international reference preparations of many heterogenous analytes". Therefore, the unitages for *in vivo* and *in vitro* assays should be completely separated. He proposes an empirical approach where the international references preparation would be distributed, with no assigned unitage, but with international agreement that the "normal" analyte concentration range in human serum would extend between arbitrarily-specified numerical limits. Individual laboratories would subsequently assign a "local" unitage to the international standard, the (dimensionless) units so defined reflecting the relative potency of standard and analyte (in normal samples) in the particular assay system used. The author's idea is that this pragmatic and less ambitious approach is still superior to the present system, which is judged as scientifically misleading. The basic problem of the different analyte heterogeneity in pathological samples could be solved in the future by multiple analysis of the different isoforms, and the author has a project allowing this challenge to be taken up.

Delaage (personal comunication) recently suggested a renewed approach to the standardization of the assay of macromolecules whose sequence is known. After having selected a source of material for purification (organ, biological fluid or cell cultures), a small quantity of antigenic material will be purified. Then, the different existing monoclonal antibodies will be tested and classified into clusters, by mutual exclusion. This preliminary operation is simpler than epitope mapping, and should be sufficient for the purpose. One antibody per cluster will be selected. The basic idea is to study the reactivity of the different antibodies in each relevant pathophysiological situation. At the end of the study of a particular antigen, it would be possible to have information about the relevance of the different isoforms or fragments that may exist. The decision could be made on the validity of a single assay for all kinds of pathophysiological situations, or if it is advisable to set up other assay methods aimed at the assessment of particular isoforms and adapted to specific diseases.

Standardization is one amongst many other problems raised by the adaptation of science and techniques to the needs of our society. This adaptation may be difficult as shown by these controversies. Improvements will follow due both to the advancement of knowledge and to consensus on reasonable compromises.

Review submitted on invitation by the Editorial Board of the *Annali*.
Accepted for publication: 1 February, 1991.

REFERENCES

1. BANGHAM, R.D. 1983. XI. Assays and standards. In: *Hormones in blood*. C.H. Gray & V.H.T. James (Eds). Academic Press, London. 3rd Ed. Vol. 5. pp. 256-299.
2. ZENDER, R. 1989. Measurements in biological systems. Metrological principles and terminology. *Scand. J. Clin. Lab. Invest.* **49**(Suppl. 193): 3-10.
3. EKINS, R. 1991. Immunoassay standardization. *Scand. J. Clin. Lab. Invest.* **51**(Suppl. 205): 33-46.
4. MASSEYEFF, R. & MALVANO, R. 1986. Problems in standardization of enzyme-immunoassays for antigens, In: *Methods Enzym. Anal.* 3rd Ed. H.U. Bergmeyer (Ed.) VHC, Weinheim. Vol. 9. pp. 38-53.
5. EKINS, R.P. 1981. The precision profile: its use in RIA assessment and design. *Ligand Q.* **4**: 33-43.
6. MOSS, D. 1989. Measurement of enzyme activity. *Scand. J. Clin. Lab. Invest.* **49**(Suppl. 193): 20-25.
7. BENJAMIN, D.C. & BERZOFSKY, J.A. 1984. The antigenic structure of proteins: a reappraisal. *Ann. Rev. Immunol.* **2**: 67-120.
8. DAY, E.D. 1990. *Advanced immunochemistry*. Wiley-Liss, New-York.
9. MASSEYEFF, R. 1991. Antibody quantitation. *Scand. J. Clin. Lab. Invest.* **51**(Suppl. 205): 63-69.
10. MALVANO, R. 1982. Error in RIA: sources and evaluation. *J. Nucl. Med. Allied Sci.* **26**: 205-226.
11. GRAVES, H. 1988. Noise control in solid-phase immunoassays by use of a matrix coat. *J. Immunol. Methods* **111**: 167-178.
12. MILES, L.E.M., LIPSCHITZ, D.A., BIEBER, C.P. & COOK, J.D. 1974. Measurement of serum ferritin by a two site immunoradiometric assay. *Anal. Biochem.* **61**: 209-224.
13. NAHM, M.H. & HOFFMANN, J.W. 1990. Heteroantibody: phantom of the immunoassay. *Clin. Chem.* **36**: 829.
14. WORLD HEALTH ORGANIZATION. 1978. Expert committee on biological standardization, 1978. Guidelines for the preparation and establishment of reference materials and reference reagents for biological substances, 29th Report. Geneva. *WHO Tech. Rep. Ser.* n. 626.
15. WORLD HEALTH ORGANIZATION. 1984. Expert committee on biological standardization. *WHO Tech. Rep. Ser.* n. 700.
16. WORLD HEALTH ORGANIZATION. 1984. *Biological standards. Reference preparations and reference reagents*. Geneva.