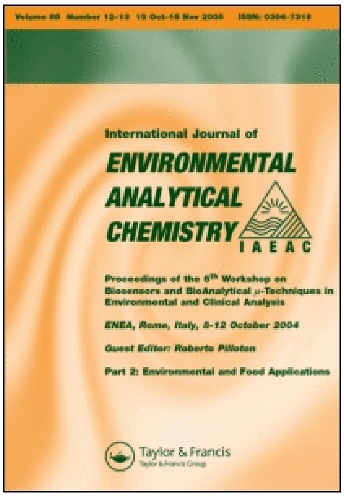


International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713640455

How Do We Prepare For Environmental Measurements Beyond 'EUROPE 1993'?

P. De Bièvreyg[a]
[a] Institute for Reference Materials and Measurements (formerly Central Bureau for Nuclear Measurements), Commission of the European Communities-JRC, GEEL, Belgium

To cite this Article De Bièvreyg, P.(1993) 'How Do We Prepare For Environmental Measurements Beyond 'EUROPE 1993'?', International Journal of Environmental Analytical Chemistry, 52: 1, 1 — 15
To link to this Article: DOI: 10.1080/03067319308042843
URL: http://dx.doi.org/10.1080/03067319308042843

Intern. J. Environ. Anal. Chem., Vol. 52, pp. 1-15
Reprints available directly from the publisher
Photocopying permitted by license only

HOW DO WE PREPARE FOR ENVIRONMENTAL MEASUREMENTS BEYOND "EUROPE 1993"?

P. DE BIÈVRE

Institute for Reference Materials and Measurements (formerly Central Bureau for Nuclear Measurements), Commission of the European Communities—JRC, B-2440 GEEL, Belgium

(Received, 20 October 1992; in final form, 23 February 1993)

Chemical measurements play a rapidly increasing role in modern society and more and more form the basis of important decisions and regulations (legal, medical, environmental, etc.). Therefore, their reliability is of the utmost importance. Environmental measurements are becoming a particular case in point with the public's and governments' growing awareness about the environment. These measurements will also have to be comparable across borders. Chemists must worry more about this reliability in order not to lose credibility and not to generate a loss of faith in chemistry.

In order to achieve both credibility and comparability, the fulfilment of the following requirements seems to be needed:

a) express all results of (environmental) measurements in the SI unit for amount of substance, the mole;
b) ensure their traceability to the mole, our SI unit for amount of substance;
c) assess their real performance through regular participation in external (border-crossing) quality assessment programmes using undisclosed, real-life samples.

Examples of highly unsatisfactory measurement performances will be discussed as well as ways which may contribute to improve the situation.

The need for an ethical approach by (environmental) scientists will be briefly highlighted.

KEY WORDS: Comparability, traceability, SI units, mole, quality assessment.

INTRODUCTION

In March 1991, the journal "New Scientist" reported an annoying story[1] for analytical chemists—environmental or other—about a shipment of grapes from Chile, South America, which had arrived in the docks at Philadelphia and which were banned on the basis of a FDA claimed cyanide content. This analysis was proven to be wrong by a University of California scientist and Chilean fruit growers filed a $400 million legal claim against the US Government.

If confirmed, this story would have the following ingredients:

— faulty analysis, and a blame for analytical chemists, *or*

— suspicion of less than impartial verification by authorities, *but certainly*

— rendering more difficult the flow from goods from southern countries in need of hard currency, to northern countries able to pay in currency which can be used to buy essential know-how and technology for self-development of the southern countries.

In the first round of CBNM's International Measurement Evaluation Programme (IMEP-1), six analytical laboratories of prominent psychiatric institutes were asked to determine Li in serum[2]. Li appears to be an excellent drug to cure mild forms of manic depressions. Since one does want to help the patient, yet not intoxicate him, some control over the Li dose is desirable and hence measurement capability is needed. The pictures developing were less than thrilling as far as agreement or comparability of measurements is concerned (Figure 1). Taking the average of the results produced, would have worsened the case: the certified

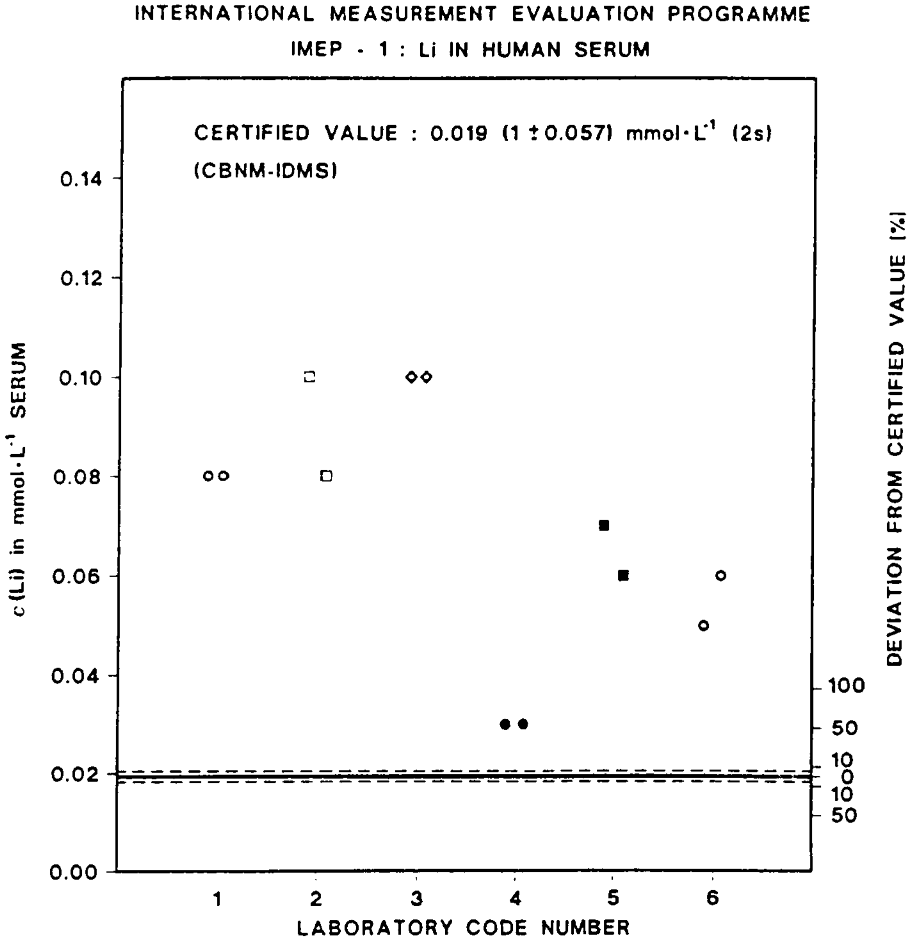


Figure 1 International Measurement Evaluation Programme IMEP-1: Li in Human Serum.

value provided in the IMEP-1 programme a posteriori showed *all* values to be wrong (Figure 1). Since all laboratories made up their own aqueous calibration solutions, an undisclosed aqueous solution was also provided—with equally disastrous results (Figure 2). One of the laboratories (No. 3) could not even distinguish between therapeutic (~ 1.0 mmol·L-^{1}) and toxic concentrations (~ 1.5 mmol·L-^{1}) (Figure 2). IMEP-2 on the determination of Cd in polyethylene (Figure 3)[3] revealed—once more—the fallacy of removing "outliers" in interlaboratory comparisons: if the outlier laboratory (No. 4) were an official or regulatory or customs laboratory, how can one possibly remove it as an outlier? Since analysts must also respect the rule of law which designated such a laboratory, must we now question how the law was made?

There is clearly something wrong in the world of measurements and therefore—of necessity—in decisions which are based on such measurements. The worrying thing is the impression of unreliability which might result from this and diffuse to the general public as well as to government responsibles in cases where important decisions can/must be made

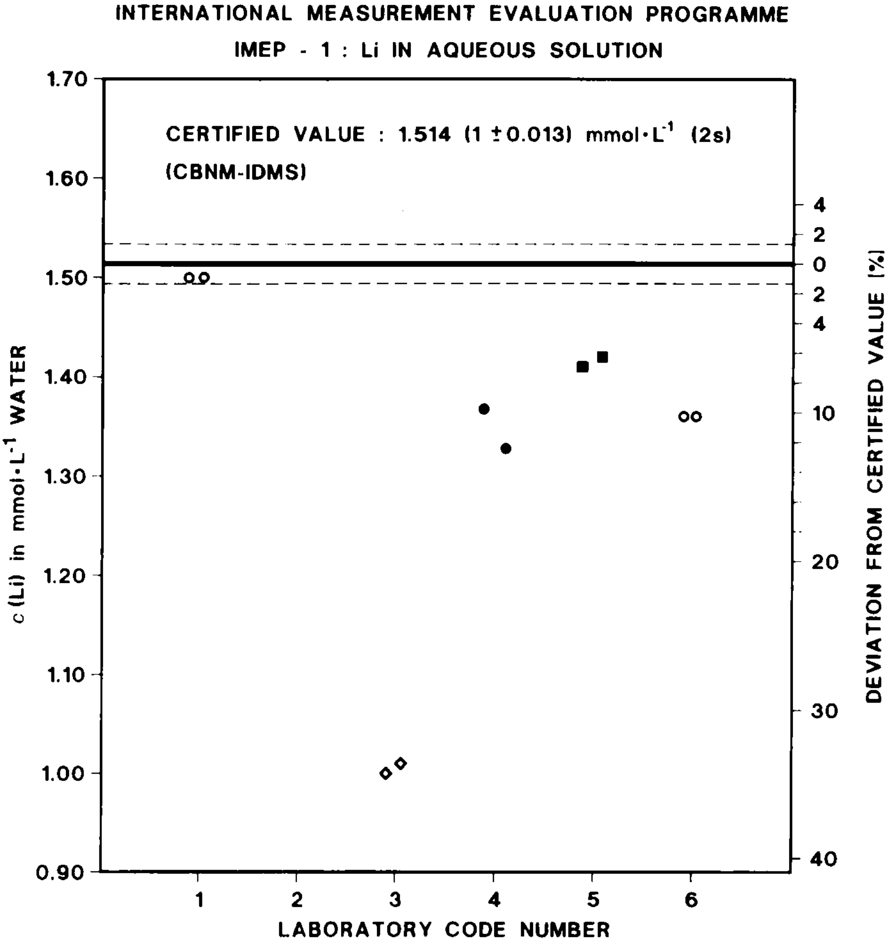


Figure 2 International Measurement Evaluation Programme IMEP-1: Li in Aqueous Solution.

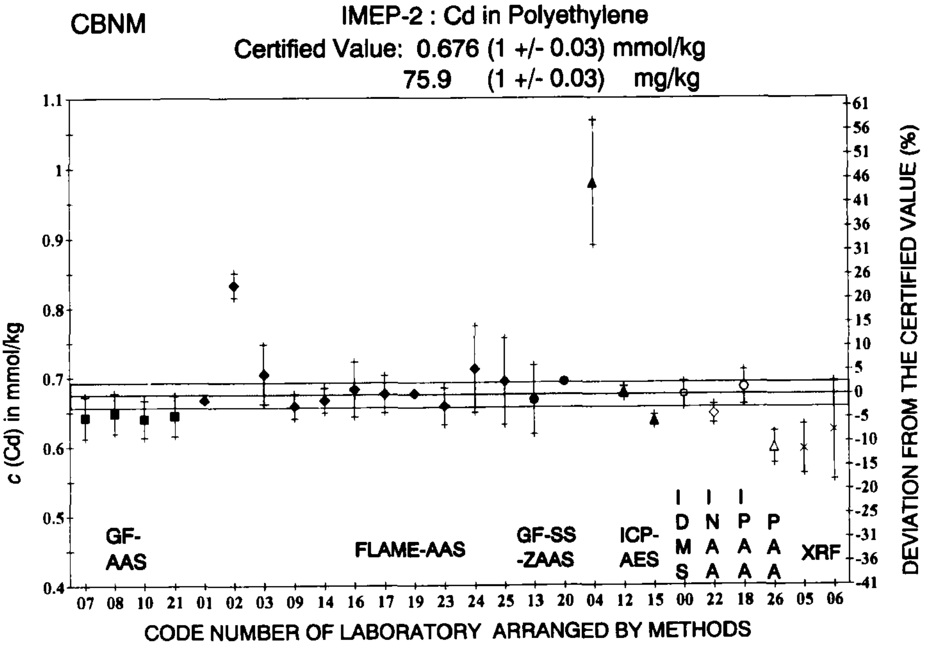


Figure 3 International Measurement Evaluation Programme IMEP-2: Cd in Polyethylene.

on the basis of chemical measurements. Unless chemists are ready with satisfying answers to pointed—and justified—questions about this (un)reliability in the future, chemistry might well suffer further damage to its image in the 1990's.

The area of environmental chemistry should be of particular interest and concern for us as chemists. Therefore, with an eye on the future—if we like more specifically on "Europe 1993"—I propose to examine how we could tackle a number of problems which need to be sorted out. At the same time we want to make sure that the push of "Europe 1993" does not get us into pitfalls.

Are our measurements of amount traceable?

If we look at physical measurements of length, time and mass, we observe that they have been organised in a consistent international system which ensures that they are "traceable" to the appropriate international SI unit. This traceability is ensured via the National Standards Institutes (Figure 4) to the Bureau International des Poids et Mesures (BIPM) which overlooks the system under the guidance of the International Committee on Weights and Measures (CIPM). There is nothing similar in the world of chemical measurements. Traceability of measurements of amount to the SI unit for amount of substance, the mole, must still be built up. Of all seven basic quantities with their units (Figure 5) the quantity "amount of substance" and its SI unit "mole" are the only ones for which traceability has

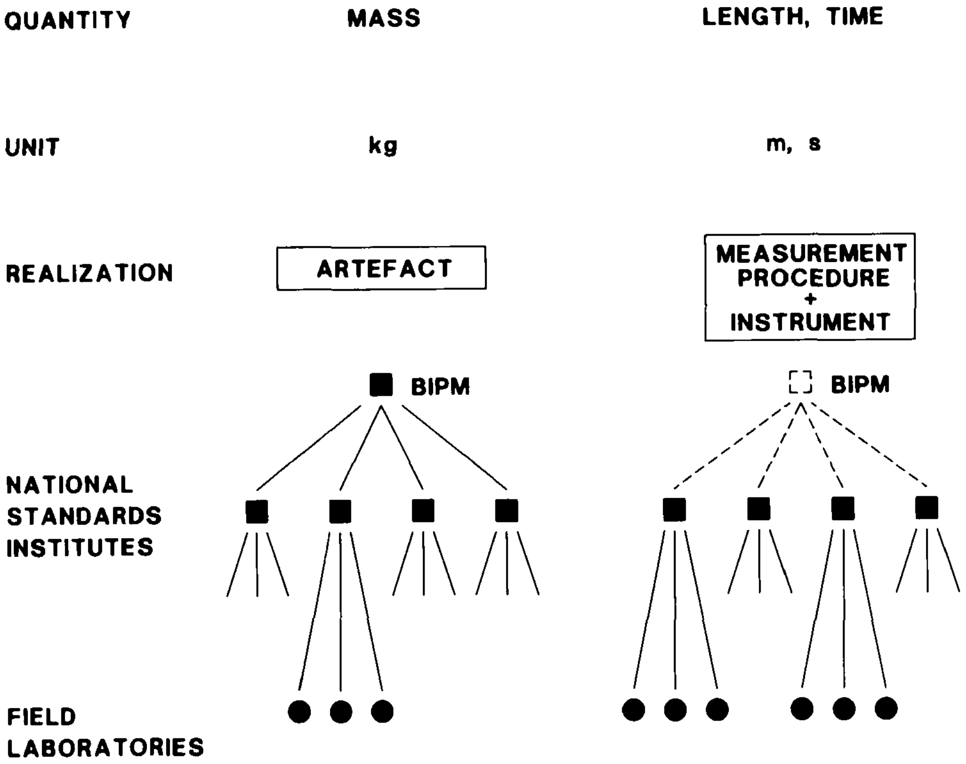
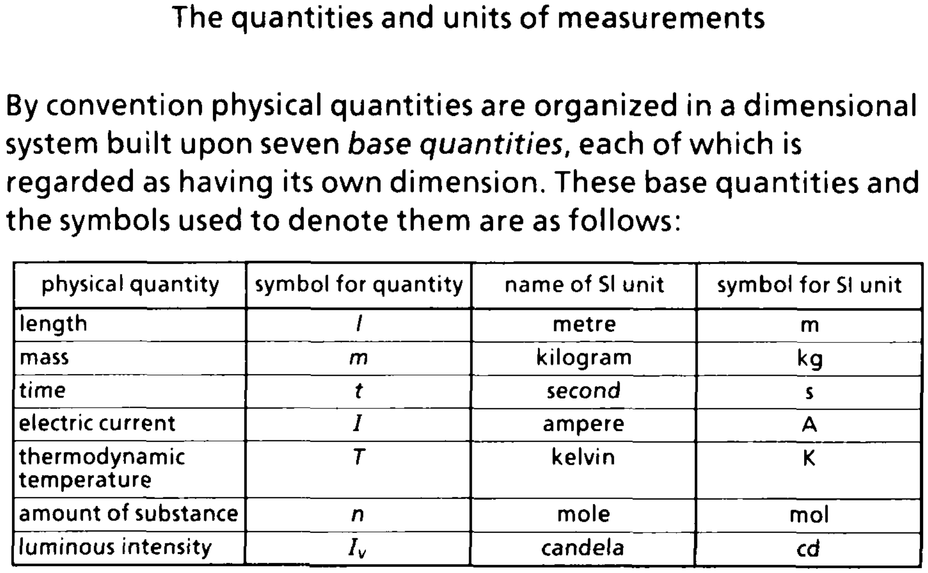


Figure 4 Physical measurements such as of mass, length and time are traceable to the appropriate SI units 'kg', 'm' and 's'.



The quantities and units of measurements

By convention physical quantities are organized in a dimensional system built upon seven *base quantities*, each of which is regarded as having its own dimension. These base quantities and the symbols used to denote them are as follows:

physical quantity	symbol for quantity	name of SI unit	symbol for SI unit
length	l	metre	m
mass	m	kilogram	kg
time	t	second	s
electric current	I	ampere	A
thermodynamic temperature	T	kelvin	K
amount of substance	n	mole	mol
luminous intensity	I_v	candela	cd

Figure 5 The SI quantities and units of measurements.

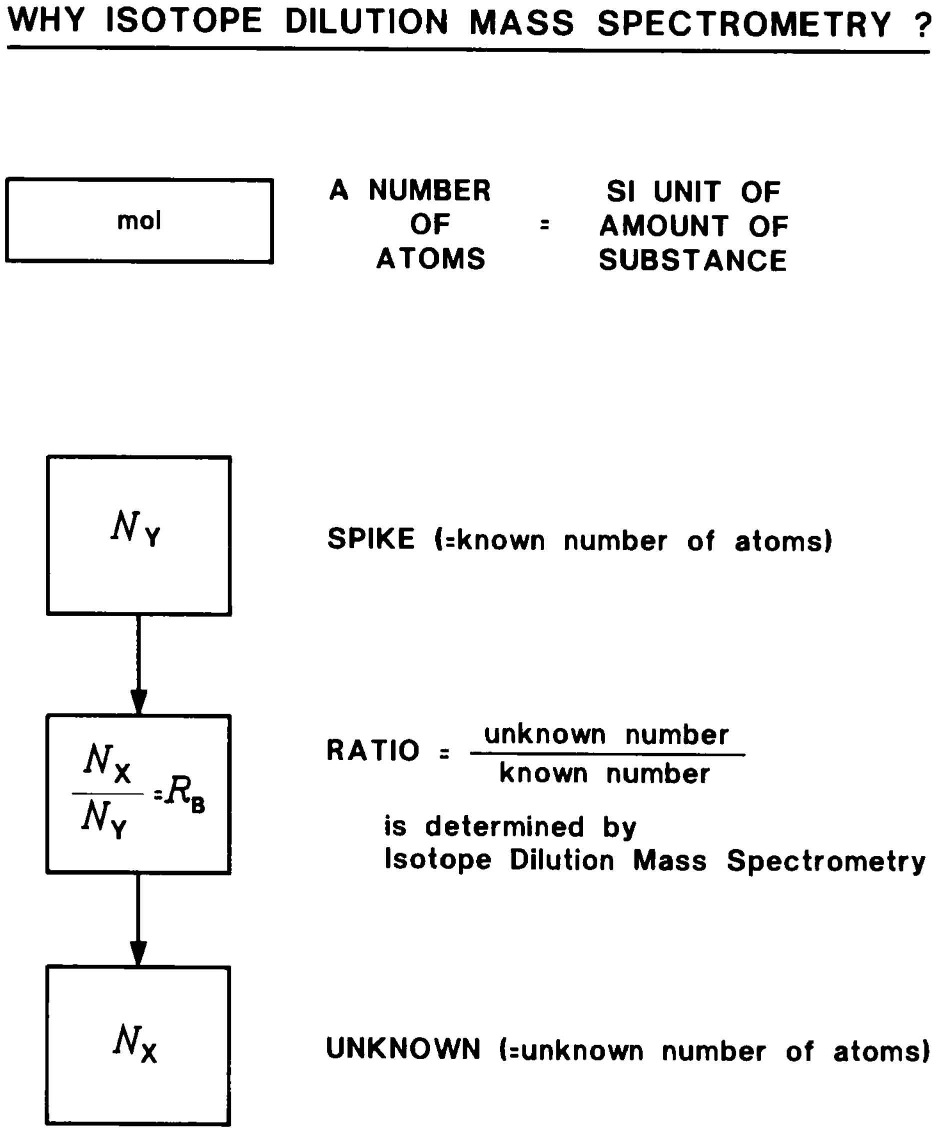


Figure 6 Why isotope dilution mass spectrometry?

not been organised properly. Under the pressing need for comparability of chemical measurements for 'Europe 1993', the CIPM has now set up a Working Group to start thinking on this matter, of course aiming at the worldwide scene. It looks as if a few methods have already been identified by the WG as primary chemical methods: isotope dilution mass spectrometry (IDMS), gravimetry, titrimetry and coulometry, from which IDMS seems to be the most relevant so far for environmental measurements. It has indeed some extraordinary features for high accuracy analysis of trace concentrations (Figures 6 and 7)[4].

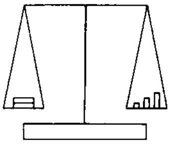
QUANTITY	SYMBOL OF QUANTITY	SI UNIT	REALISATION OF SI UNIT IN TERMS OF	MEASUREMENT IS A COMPARISON BETWEEN THE QUANTITY IN AN UNKNOWN TO THAT QUANTITY IN A REFERENCE MATERIAL	MEASURED AMOUNT IN UNKNOWN
MASS	m	kg	THE INTERNATIONAL PROTOTYPE OF THE KILOGRAM AT SÈVRES	 $(m - m_s)$ kg	$m_s + (m - m_s)$ kg
AMOUNT OF SUBSTANCE	n	mol	N_A THE NUMBER OF ATOMS IN 0.012 kg OF ^{12}C	 n/n_s	$n_s(\frac{n}{n_s})$ mol
				-PHYSICAL MEASUREMENTS -SMALL, KNOWN CORRECTIONS ONLY	

Figure 7 The (isotope) mass spectrometer (IMS): the modern chemist's balance.

Are we using the proper unit?

In 1971 the CIPM defined the 7th quantity in our international measurement system as "amount of substance" (symbol "*n*") and gave it the 7th SI unit: the "mole" (symbol "mol")^5.

The mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kg of ^{12}C. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles (14th CGPM 1971)^5.

Ever since, all measurements of amount, otherwise called "chemical measurements" should have been expressing their measurement results in "mol". That is not yet the case. It is useful to point out that the introduction of the SI unit mole (symbol: mol) is the reflection of the simple fact that it is numbers of atoms and molecules which are chemically active, not masses or weights of chemical substances.

The physical quantity "amount of substance" is of special importance to chemists. Amount of substance is proportional to the number of specified elementary entities of that substance, the proportionality factor being the same for all substances; its reciprocal is the Avogadro Constant. The physical quantity 'amount of substance' should no longer be called 'number of moles', just as the physical quantity 'mass' should not be called 'number of kilograms'. The name 'amount of substance' may often be usefully abbreviated to the single word 'amount', particularly in such phrases as 'amount concentration' and 'amount of N_2'.

It is a basic requirement for comparability that measurements are expressed in the same unit. For measurements of amount or "chemical measurements", there is no question that this must be the "mole".

There is even an EC Directive "applicable in all member states" that fixes the mole as the compulsory unit for measurements of amount[6]. Legally this problem is solved. And, even scientists must apply the law.

Are we accurately documenting reality?

Interlaboratory comparisons have been seen and used as a way to compare measurement results. As long as they satisfied our academic curiosity, could be treated statistically and gave us an idea about interlaboratory spreads, that was allright. Time has come to evaluate which specific criteria should be met by such comparisons in the case of measurements and concentrations that form the basis for important decisions.

Consider, for example, the case of an interlaboratory programme on Pb in cabbage with results as shown in Figure 8[7]. It has the merit of indicating that there is a problem, but the combination of the results with the knowledge that there could be something as a regulatory upper limit of 0.5 mg·kg^{-1} (wrong unit!) does not really help in deciding whether the food (cabbage) is fit for consumption.

The Swiss judge who would need measurements before ruling whether or not to order the withdrawal of batches of palm nuts from commerce, is really not helped by analysts who

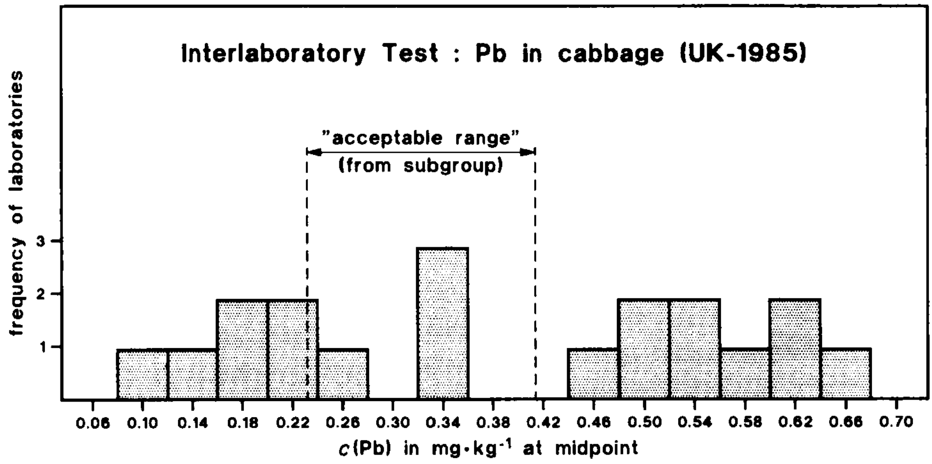


Figure 8 Is the cabbage fit for consumption if the regulatory upper limit for Pb in cabbage is 0.5 mg·kg^{-1}?

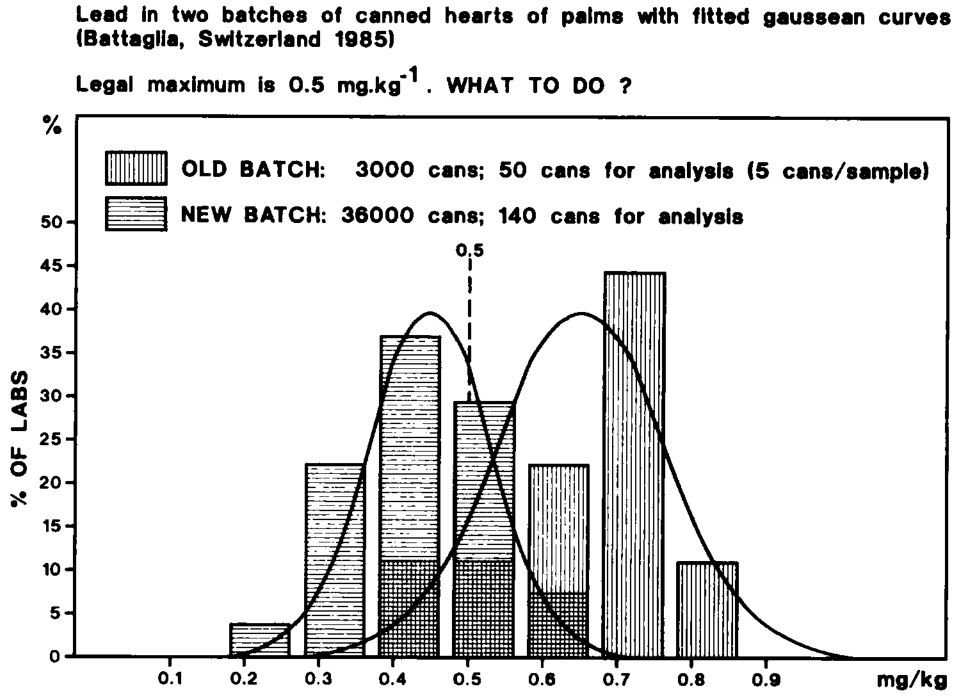


Figure 9 A judge must decide on the withdrawal from commerce of batches of canned palm hearts on the basis of their Pb content. What must the judge decide?

report measurement results which are nicely (and normally?) distributed (Figure 9)[8] around the legal maximum he must use in his ruling!

What is more troubling is that the self-assessment of laboratories is rather independent from reality as Figure 10 displays: the self-declaration of participants in this comparative programme did not bear any correlation with reality[9].

In many comparison programmes, the mean of participating laboratories or simply a "consensus" value is taken as "closest to the true value" and performance of laboratories are evaluated against it. This practice lacks demonstrated scientific grounds and is in fact arbitrary. The EUROANALYSIS VIII Conference in Edinburgh (1993) will have a section on"Round Robins: the wrong way to do it"! When important decisions must be made on the basis of measurements, we want these measurements to be anchored as much as possible in the physical and chemical reality of the substances we are interested in. This might not always be possible and especially in environmental measurements and samples, problems such as homogeneity may form additional and big barriers to this approach, but as far as measurements are concerned, these requirements must be fulfilled whenever possible. Specifically this means that for the amount under investigation (the "measurand") in the test samples for International Measurement Evaluation Programmes, sufficiently accurate values for the measurand must be obtained by methods giving results which are:

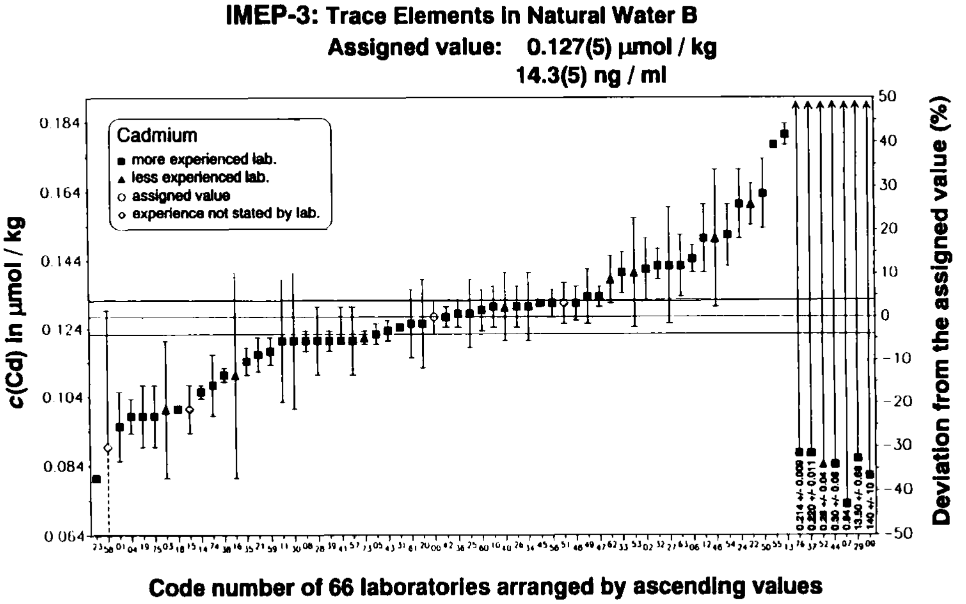


Figure 10 Do laboratories accurately know the reliability of their own measurements? Self-declaration of laboratories in IRMM (CBNM)'s International Measurement Evaluation Programme IMEP.

— expressed in SI units
— independent of sample matrix
— proven to be free from systematic errors within measurement reproducibility
— three to ten times less uncertain than needed for the purpose.

I am afraid that we must also rather drastically change the aim of interlaboratory comparisons. The main reason for participation of many laboratories in a proficiency test programme is NOT to have many results for "statistical evaluation" (of the laboratories, of the method, . . .) BUT that any of them might be involved in a problem caused by the discrepancy in their measurement results for the same measurand (shipper vs receiver, inspector vs inspected).

Has not the elimination of outliers so far been an artificial means to reduce the total spread, to embellish the picture and to be a concession to statistics which require more homogeneous sets of data for their methods to be applicable? In the former case we might be fooling ourselves. In the latter case, maybe we should require from statistics and statisticians that they develop methods which suit the reality of chemical measurements rather than that chemists have to adapt the reality of their measurements to the statistical models in existence?

An attempt to develop a statistical method without removing outliers has been published but largely ignored[10]. This approach yields a realistic standard deviation for the interlaboratory spread without being influenced by non-rejected outliers.

Reference Materials or Reference Measurements?

Many of us argue: we need Reference Materials, more and better RMs. In fact as many RMs as there are matrices in which we must measure in order to calibrate our measurements.

We must slowly realise that it will be inherently impossible to prepare that many RMs: it would require many thousands of them to be made in addition to those in existence and all the production facilities of the world could not achieve that.

But, apart from that, Reference Materials are not always successful in curing problems in measurements for the following reasons:

a) their "best value" is known (within a given uncertainty range), so users sometimes give special care to obtain this value; they tend to work "towards" this value;

b) the material is often "idealised" compared to "real-life" samples: it is (very) pure or does not correspond in concentration level and/or unknown element and/or matrix to the "real-life" case.

Therefore, future development of methods should be directed towards those methods which are matrix-independent and would have the potential of acting as "Reference Method" and enabling us to perform "Reference Measurements". The values obtained by these Reference Measurements on real-life samples could then be used to calibrate other, quicker and cheaper, methods in the field. The concept of Reference Measurements may well become as important in the future as the concept of Reference Materials.

Imposing methods and procedures? or
Imposing the obligation to produce good results?

It is generally known that some authorities have ventured into imposing certain methods or instruments for obtaining measurement values needed for regulatory purposes. This practice should be abandoned since it is built on several fallacies.

It is a fiction that a given method to measure an unknown amount in a given material has a *constant* repeatability and accuracy. The fact from observation is rather that any method has a *potential* for an *optimal* repeatability and accuracy (which may differ from method to method), but, when used in practice, it is subject to variable systematic errors and has, therefore, *non-optimal, not constant* repeatability and accuracy.

What is still more important: imposing everything relieves the person responsible for the measurement from responsibility for the end result, as this person is entitled to decline this responsibility on the basis of having followed exactly the prescriptions. In other words: one has followed the rules, therefore one's results must be right, reliable and . . . accepted. I am afraid that the harsh reality is that reliable measurement results cannot be "ordered", in the same way that a "true value"—or—"closest to the truth"—cannot be designated by decree.

The corollary of these two observations is, of course, that the analytical world itself must come up with a consistent and credible measurement system which makes its measurement results acceptable to the outside world. And the answer to that challenge is: use the same unit, make measurements traceable to the mole, indicate that the result may not be the "true value" nor even "closest to the truth", but deliver uncertainty ranges around the results which claim to contain the true value and which one is capable of defending with that meaning.

In this time of more and more automated analysis, we must call for a renewed responsibilisation in measurements and therefore for a massive increase of training and education in chemical measurements in our schools, laboratories and universities.

Are regulatory or decision concentration levels well founded?

More and more analytical chemists in general and environmental chemists in particular, are faced with regulations fixing maximum contents of given substances in given materials. The spread of toxic materials in the environment is a very specific case in point where such regulations are needed and where upper limits are necessary. It seems to me that this will force analytical environmental chemists to look with particular attention to the reliability of their measurements at and around these levels. But it also seems very indicated that regulatory authorities should:

— look for well-documented pictures of the State-of-the-Practice (SOP) and feasibility of meaningful measurements at and around decision concentration levels

— indicate an uncertainty on the decision concentration level, or, alternatively

— accept the principle that a measurement has a stated uncertainty and that e.g. the measured value should be lower than the decision concentration level, even if the uncertainty of the measured value would encompass this level.

It may be useful to note that the IUPAC will start to examine this particular problem.

It seems as if measurements of amount or "chemical measurements" have not conclusively demonstrated that they are reliable and do not seem to have a consistent international system to prove that. They do not come very close to what may be expected from "exact" measurements produced by "exact" sciences. It looks indeed as if one must rather talk for some time about "so-called" exact measurements. Other authors agree: "Considerable evidence exists in the literature that few analytical chemists pay attention to the question of the reliability of the analytical results they produce"[11].

Environmental measurements—as measurements in general—are becoming overwhelmingly important in our world. The quantification of amount of substance will determine more and more the use of these materials. The amount of toxic or useful substances will determine what we can do—or not do—with air, water and soil. At the same time, chemistry is facing one of the most serious challenges—if not *the* most serious—in its history: being treated by the public and their governments as having the ability to pollute the environment.

CONCLUSIONS

There is a need to put order in our measurements so as to make them comparable between laboratories and across borders. It seems that at least the following points must be seriously—and quickly—addressed by the chemical measurement community:

1. All measurements of amount (of substance)—otherwise called "chemical measurements"—should deliver their results in the agreed and decreed international SI unit

for amount: the mole (or fractions thereof). "Europe 1993" should not add just another contribution to the existing international confusion.

2. All measurements but especially those around important "decision concentrations" such as maximum contents, must be traceable to the mole through an uninterrupted chain of calibrations, symbolically pictured in Figure 11.
3. Interlaboratory comparisons should present an assigned value, traceable to the mole, the uncertainty of which claims to contain the true value on solid, proveable and transparent grounds. They should essentially abandon the practice of taking the average of a number of results as the "closest approximation to the true value", either before or after elimination of outliers, except where it is—temporarily—the only way to proceed. And even then the only scientific way is to present a range of observed values rather than just one value.

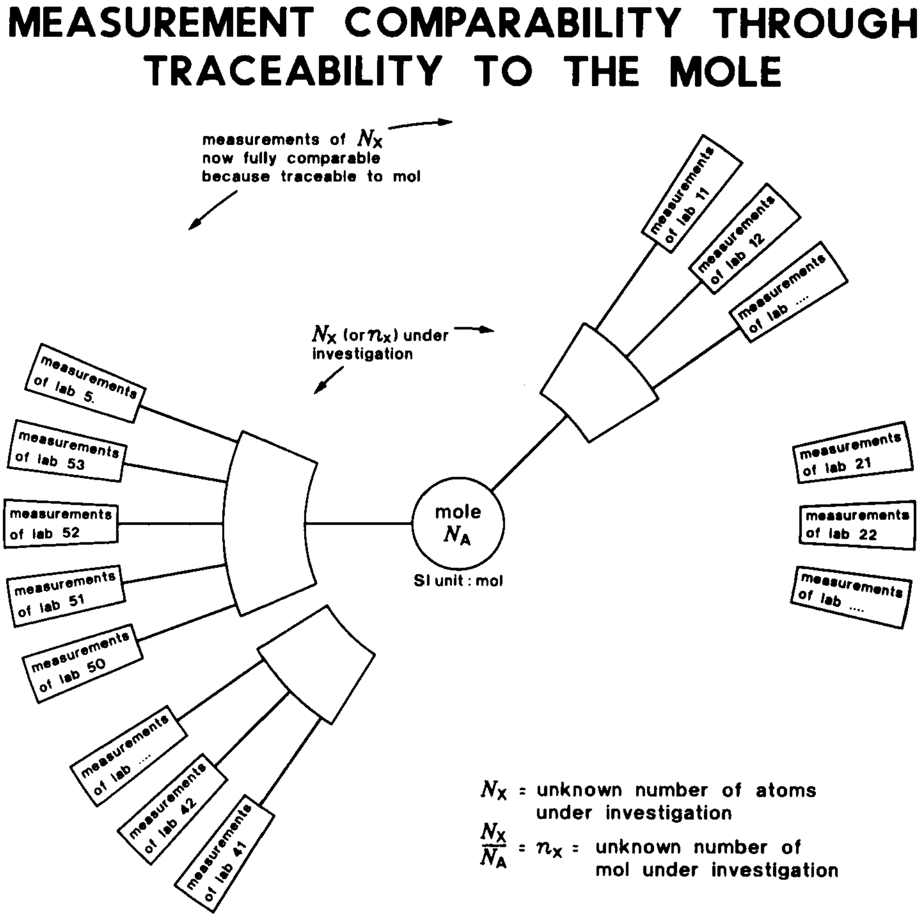


Figure 11 One of the possible ways measurement comparability could be established.

4. It is the authorities' role to decide *WHICH* measurements must be done. It is not for them to impose how they must be done because that would take away the intrinsic responsibility for the results from the measurement laboratory. It is, however, imperative for them to demand the set-up of mechanisms for impartial evaluation of the results.
5. It is the analysts' role and responsibility to determine *HOW* they make the measurements. They are responsible for the choice of method, instrument, procedure and . . . for the results.
6. External impartial quality assurance programmes seem to be the best approach so far to judge measurement performance. With these in hand, authorities can set the minimum criteria for each given problem where measurements are needed if they feel this is needed. In other words, it is for them to *decide* on a numerical Target Value for a (total) uncertainty of a measurement (see Figure 12). These Target Values can then serve as a useful guide for analysts to decide whether or not performance meets the "fit-for-the-purpose" requirement or whether further development is needed.
7. Regulatory rules, EC Directives, legal maxima should, whenever possible, be decided upon with impartial and properly documented pictures of the real state of the practice at hand.

The sole and only intention in this paper has been to attempt to localise several responsibilities correctly. Without such a clear identification, any attempts to remedy the chemical measurement situation in general and the environmental measurement situation in particular, could be ill-fated, because wrong in the concept and therefore disastrous in the results.

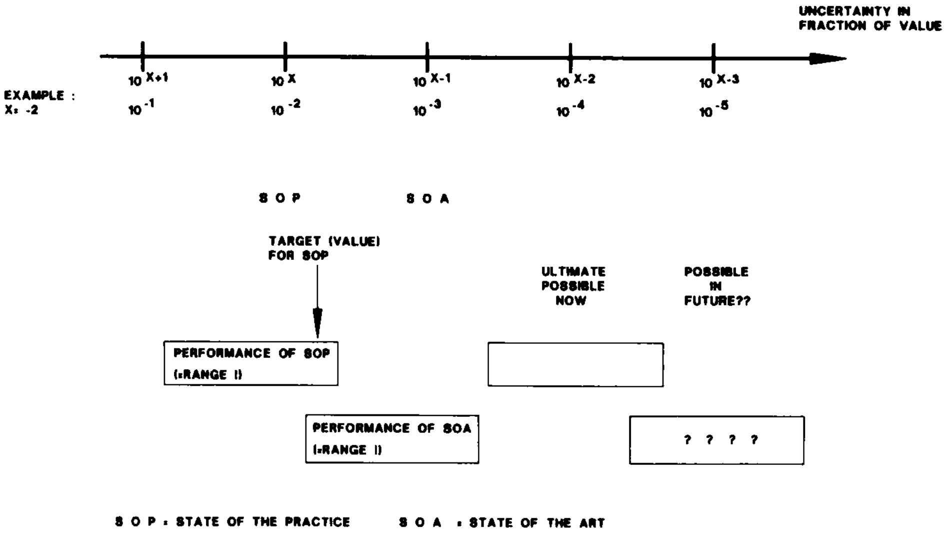


Figure 12 A Target Value for uncertainty is what is decided a priori. Performance is the experimental result of what has been achieved a posteriori.

We cannot afford that. It would violate elementary ethics which impose on us as on any other responsible human being, *a drastically increased care for our environment* and more so *for the environment the next generations will inherit from us.*

It would also contribute to a drastic loss in confidence in chemical measurements and chemical technology. We cannot have that either, since both chemical measurements and chemical technology are absolutely needed in the challenging task of stopping the deterioration of our environment and—do I dare say it?—the still more challenging task of starting to reconstruct it. It is a task where environmental chemistry has a formidable mission.

References

1. *New Scientist*, **March**, 15 (1991).
2. P. De Bièvre, J. Savory, A. Lamberty and G. Savory, *Fresenius J. Anal. Chem.*, **332**, 718–721 (1988).
3. A. Lamberty, P. De Bièvre and A. Götz, *Fresenius J. Anal. Chem.*, **345**, 310–313 (1993).
4. P. De Bièvre, *Fresenius J. Anal. Chem.*, **337**, 766–771 (1990).
5. *Quantities, Units and Symbols in Physical Chemistry*, International Union of Pure and Applied Chemistry, (Blackwell Scientific Publications, Oxford, 1988).
6. *Off. J. Europ. Comm.* No. L 243 (1971-10-29), No. L 39 (1980-02-15), No. L 109 (1983-0-26), No. C 185(1991-07-17).
7. *Chem. in Britain* No. 11 1019–1021 (1985).
8. R. Battaglia, *The Legal Limit: Nirvana for Food Analysts?* (Proc. 3rd Conf. Food Chemistry EUROFOOD CHEM III De Sikkel B–2150 Malle, Belgium 1985) Vol I pp. 1–18.
9. W. Beyrich, W. Golly, G. Spannagel, P. De Bièvre, W. H. Wolters, and W. Lycke, *Nucl. Technol.*, **75**, 73 (1986).
10. W. Beyrich, W. Golly, N. Peter and R. Seifert, *KfK 4721—EUR 11398 EN* (1990).
11. W. Horwitz, *J. of AOAC International*, **75**, 368–371 (1992).