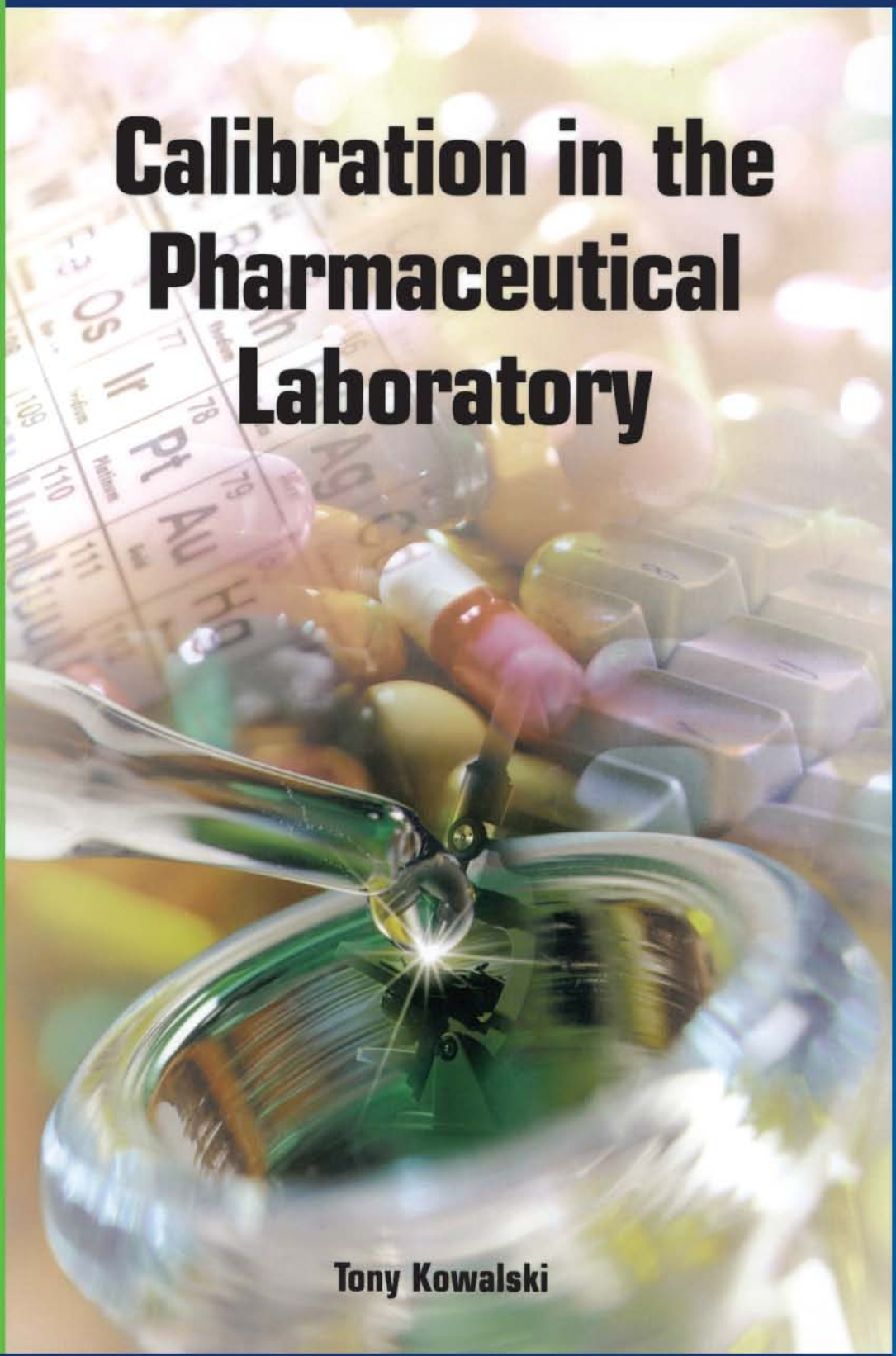
Calibration in the Pharmaceutical Laboratory

Kowalski

CRC

Taylor & Francis

Calibration in the Pharmaceutical Laboratory



Tony Kowalski

Calibration in the Pharmaceutical Laboratory

Tony Kowalski

Editor

Taylor & Francis
Taylor & Francis Group
Boca Raton London New York

A CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa plc.

Published in 2001 by
CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2

International Standard Book Number-10: 1-57491-092-2 (Hardcover)
International Standard Book Number-13: 978-1-57491-092-6 (Hardcover)
Library of Congress catalog number: 00-054188

Library of Congress Cataloging-in-Publication Data

Catalog record is available from the Library of Congress

informa
Taylor & Francis Group
is the Academic Division of Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

CONTENTS

To my wife, Pauline Kowalski,
and my daughters,
Katherine Louise Kowalski and
Rachel Caroline Kowalski.

PREFACE

As with many product developments and discoveries, this book was written as a result of a series of coincidences and circumstances. First, I was asked to make a presentation to a major pharmaceutical company on calibration weights—their use, storage and which class should be used. From this and subsequent similar presentations, I was able to deduce that, although such companies were expert in their field of drug discovery, manufacturing and quality control, there was generally a huge gap in their knowledge of metrology and weighing technology.

As I made more presentations, mostly on demand from pharmaceutical companies, I gradually expanded the content to include GLP/GMP, understanding and determination of uncertainty, choosing the correct balance and aspects of in-house maintenance and external service contracts. It was after completing a series of calibration seminars during the London Laboratory

Show, I began to develop this book—as a series of relevant and related chapters by recognised international experts in their respective fields. This was not only very interesting, exciting and informative but also incredibly frustrating since all of the authors hold demanding positions and required varying amounts of time to produce their chapters. Nevertheless, with a lot of effort and help, this book has been completed and will, I hope, offer answers to many questions as well as insight into better mass measurements, increasing the reader's understanding and appreciation of the associated errors.

As with all modern technology, technical and scientific books are almost certainly guaranteed to be out of date or suffer from changes in opinion and shifts in emphasis due to the ever-changing world in the pharmaceutical industry. A good example of this was the U.S. Food and Drug Administration's introduction of a directive whereby all mass measurements must be made with a maximum uncertainty of 0.1 percent. This directive, section 41 of the U.S. Pharmacopeia regulations, lay virtually undetected by pharmaceutical scientists for some three years. Today it not only has been discovered but is highly topical due to misunderstanding and misinterpretation of the document. As a result, most calls from the pharmaceutical industry to balance manufacturers are for help in determining the smallest mass that can be measured on

particular balances and still comply with this directive. A short section, written by Dirk Ahlbrect, has therefore been added to Chapter 1 as an appendix, which outlines the requirements and details the procedures involved in determining the minimum weight that can be measured on individual balances. In my experience, uncertainty of measurement has not previously been understood or taken into consideration by any balance operators outside mass calibration even though it can be a significant hidden influence on the accuracy of results. A full description and outline of the factors affecting the uncertainty of measurement is given in Chapter 8.

Next year, undoubtedly, the flavor of enquiry will change as the FDA not only moves the goal posts farther apart but also may introduce new regulations. I am quite sure that one of the next hot topics will be validation of the software embedded within the microprocessor of measuring equipment such as balances. So far, manufacturers have not needed to provide documentation since validation of the measuring accuracy (described in Chapter 1 under hardware validation) and repeatability has satisfied the FDA.

During my 16 years at Sartorius, I have been grateful for knowledge imparted to me by my colleagues within product management, including Horst Nagel, product manager for moisture analysis and pipette calibration, and Dirk Mueller and Thomas Pertsch, product managers

of precision laboratory balances. In addition, I am especially appreciative for the commitment from my co-authors who contributed chapters for this book.

Tony Kowalski
February 2001

CONTRIBUTORS

Tony Kowalski, Editor

Tony Kowalski is currently the product manager for laboratory balances and head of marketing for weighing equipment at Sartorius Limited at Epsom, Surrey. He was educated at the College of Science and Technology Newcastle upon Tyne and trained as a laboratory technician. Before joining Sartorius, he worked in quality control and process development for Harcross, a bulk chemical manufacturer, and for the Cancer Research Unit at Newcastle upon Tyne.

M. J. Buckley

Mike Buckley joined the trading standards service of Leicestershire Council and Rutland County Council in 1972 and later became standards officer and general manager with the South Yorkshire County Council in 1976.

Mr. Buckley now heads the South Yorkshire Trading Standards Unit, which is a UKAS accredited calibration laboratory that provides services to industry and trading standards services to the South Yorkshire Metropolitan District Council. Mr. Buckley has presented papers on mass calibration to conferences in Europe, the United States and the Far East and was chairman of the Weighing 2000 Conference.

John P. Clark

John P. Clark has over 20 years of experience at the Westinghouse Savannah River Company. A metrological engineer, he has technical oversight for the mass and pressure calibration laboratories. Prior to working at the Savannah River site, he set up and managed the chemical standards laboratory at the Barnwell Nuclear Fuels Plant and worked in the analytical laboratories at the Idaho National Engineering Laboratory. Mr. Clark is currently implementing a site-wide scale and balance calibration program and measurement control programs. He is a frequent presenter and session developer at national and international professional meetings.

Chris Jenkins

Trained in mathematics and statistics, Chris Jenkins became a trading standards professional and has held legal metrology enforcement posts

in both England and Scotland. Since his appointment as Calibration Services Manager at Kent Scientific Services, he has lead the development of a modern metrological laboratory, which has become the Trading Standards calibration house in South East England. Mr. Jenkins is a member of the Institute of Trading Standards Administration and a licentiate member of the Institute of Measurement and Control. He is also a an advisor on legal metrology.

Juergen Ober

Trained in electrical and control engineering, Juergen Ober joined Sartorius AG in 1972 as a scientist in the development department for weighing technology, primarily responsible for all linear circuits inside the balances. Since 1993, he has been head of the DKD (German Calibration Services) accredited calibration laboratory for electronic balances and is product manager for mass comparators.

Ted Scorer

Ted Scorer joined Glaxo Wellcome in 1969 and since then has worked in both quality assurance and quality compliance. In 1982, he joined the Central Analytical Services Department (now the Laboratory Support Group) based at Barnard Castle. This group is a multidisciplinary analytical science group providing worldwide support

for Glaxo Wellcome in laboratories and the introduction of measurement science to production lines. He has developed manual and automated check-weighing applications and in-house software for controlling these applications. Mr. Scorer is also a founding member of the National Physical Mass, Weighing and Density Club.

A. Harper Shull

A. Harper Shull was a statistical consultant for Eastman Kodak for 17 years before joining the Westinghouse Savannah River Company in 1990. He is currently a principal scientist in the Advanced Planning and Process Support Section of the Analytical Laboratories Department. He provides technical support and statistical oversight to the laboratory in the areas of standards, control systems, calibration, measurement uncertainty estimation, variance propagation and experimental design.

Stephan Weyhe

Trained in physics, Dr. Stephan Weyhe was a product manager for medical devices in the therapy of cancer. He joined Sartorius AG as product manager for micro and analytical balances in 1989. Dr. Weyhe subsequently became the head of product management for laboratory and industrial products.

Brian Wichmann

Dr. Brian Wichmann has over 30 years of experience on software quality issues. He was responsible for the PASCAL validation suite used to check compilers internationally. He also was part of the team that designed Ada and was project editor for the ISO Technical Report on using Ada in high integrity system. Dr. Wichmann is a visiting professor at the Open University.

Chapter 1

INTRODUCTION

Tony Kowalski
Sartorius Limited
Epsom, Surrey, United Kingdom

As the regulation of test and measuring equipment used in pharmaceutical manufacturing and quality assurance is ever increasingly becoming an administrative nightmare with continually moving goalposts and where new regulations could be introduced from either government, the Food and Drug Administration (FDA) or the Medicines Inspectorate, external help is often required from publications that specifically detail plans to incorporate equipment into quality systems.

Today's scientist is not only required to be a specialist in his or her own field but needs to become familiar with legislation that governs the

use of equipment in the pharmaceutical industry. Within Europe, a new directive was introduced in 1993 (89/336/EEC) that radically changed the type of equipment used in quality control laboratories. This directive becomes final in January 2003 following a 10-year derogation period and is fully covered in the chapter covering the verification of balances (chapter 2). Having decided on the correct balance, which should be governed by the determination of required and actual accuracy of the weighing machine, the end user must then follow a rigid process in order to validate the equipment prior to bringing it into use, which is also covered later in this chapter.

Scientists need to understand a great deal of physics in order to make decisions on pass and fail limits when testing laboratory balances with traceable mass standards and must also make informed decisions on which quality, ultimately connected to accuracy, of calibration mass should be used to test each type of balance. Taking into account the errors associated in measuring the mass of an object is also crucial in order that the uncertainty of measurement can be calculated. Without this knowledge, a scientist cannot be sure that all measurements are made with the intended accuracy.

As the requirements of the FDA are constantly tightening, with the sole aim of reducing margins for error, managers responsible for quality control and quality assurance require the best possible tools for the job and up-to-date

accurate information in order to set out Standard Operating Procedures (SOPs) for balances and scales used throughout the manufacturing and testing process. Having been introduced some years earlier, section 41 of the USP (U.S. Pharmacopeia) covering the use of balances for quality control analysis of drugs destined for the U.S. market is currently the topic of discussion and in particular the determination of the minimum weight that can be measured on each balance. Details of how to comply with this directive can be found in Appendix 1.1 at the end of this chapter. The goal of this book is to provide some general background information on laboratory balances and precision scales, together with concise detailed information on key quality issues associated with weighing.

The Laboratory Balance

The modern laboratory balance is taken for granted by many end users; however, it is the result of many years of development to both the mechanical weighing cell and the sophistication of the most recent microprocessor technology. Yet this is not the full picture, since every balance requires complex software in order to produce even a simple weighing result because the weight readout is a dynamic average value of measurements integrated over a short time period. This readout is constantly updated at a rate

of some 100–200 milliseconds, which means that the balance can react to the smallest changes in mass quickly. Depending on how well the digital filter algorithm is written, the balance should stabilise quickly and be reasonably resistant to external influences.

The first commercial microprocessor-controlled electronic balances were introduced by Sartorius in 1973, instigating a new revolution in weighing technology as other manufacturers followed suit. These first generation electronic balances gave the operator the ability to tare containers and weigh in components without having to make subtractions for the container weight. The weighing process was shortened even further: Since there is no requirement for preweighing with a mechanical system on partial release prior to the final measurement, some estimates show that the total measuring time has been reduced by a factor of five. A further advantage of digital balances is the possibility of connection to a data printer for hard copies of all measurements, which drastically reduces the risk of transcription errors, increases integrity, and provides an easy means of storing measured values.

Printed Results

Printouts from balances can offer far more than just simply the measured value. This is particularly useful since compliance to quality systems

requires records that show the date, time, equipment used and operator's signature. Most balances today have the means to print this information, and the top of the range models have the facility for additional alphanumeric header text that allows entry of the company and location of the balance. This type of printout is usually referred to as a GLP (Good Laboratory Practice) or ISO (International Organisation for Standardisation) compliant printout.

A typical example of a GLP/ISO printout is as follows:

28.04.1998 16:00:00	*Date/Time*
SARTORIUS	
Model LP4200S	*Balance Model*
Ser. no. 030319914	*Unique serial number*
Vers. no. 01–30-01	*Software Version*
ID 123456789ABCDE	*Alpha ID (user definable)*
L-ID 123456789ABCDE	*Lot or Location ID*
N + 520.00 g	*Weighed Value*
N + 1340.00 g	*Weighed Value*
S-ID 123456789ABCDE	*Sample ID*
N + 1530.00 g	*Weighed Value*
28.04.1998 16:01:00	*Date/Time*
Name:	*Operator Signature*

Information concerning the unique serial number should preferably be read directly from the main microprocessor in the balance rather than relying on a text entry in the memory of the data printer. This prevents the obvious from happening should the printer be connected to another balance. Time and date stamps are important and in most cases are taken from the clock in the data printer. In cases where balances are connected to other devices, the top of the range models from some manufacturers have internal real-time clocks which allow date and time stamping of the results.

Balances with internal adjustment weights also generate printouts just after the adjustment routine has finished to give records of all changes in the calibration of the balance. (Please refer to the terms for calibration and adjustment for detailed definitions.) Many manufacturers offer balances that can automatically activate the internal adjustment weight, where a change in temperature is significant enough to cause a change in the calibration; if the manufacturer has taken into account that air density may change whilst the temperature remains constant (air conditioning), the balance will automatically adjust every four hours irrespective of temperature. As always, the top of the line models offer the feature of storing up to 50 adjustment routines for printing at a later or more convenient time. This helps to prevent loss of the smaller individual printouts from single routines. In order for the end user to be confident that there is no

excessive inherent drift and that the automatic adjustment routine is only there masking this, there should be some indication of drift expressed on the calibration routine printout.

The following is a typical example of an adjustment routine printout:

28.04.1998 16:00:00	
SARTORIUS	
Model LP4200S	
Ser. no. 030319914	
Vers. no. 01–30-01	
ID 123456789ABCDE	
Internal calibration	*Calibration Function*
Start: isoCAL/temp.	*Reason for isoCAL (automatic adjustment)*
Diff. + 0.01 g	*Calibration result (drift since last adjustment)*
Internal adjustment completed	*Adjustment function*
Diff. 0.00 g	*Adjustment result*
28.04.1998 16:01:00	*Date/time*
Name:	*Operator signature*

A question often asked in working to a quality system is, "Can I have a certificate for the internal adjustment weight and is it traceable to a national standard?" In short the answer is a resounding "No" to both parts of the question; however, this does not render the internal weight obsolete. The internal weight is there only to

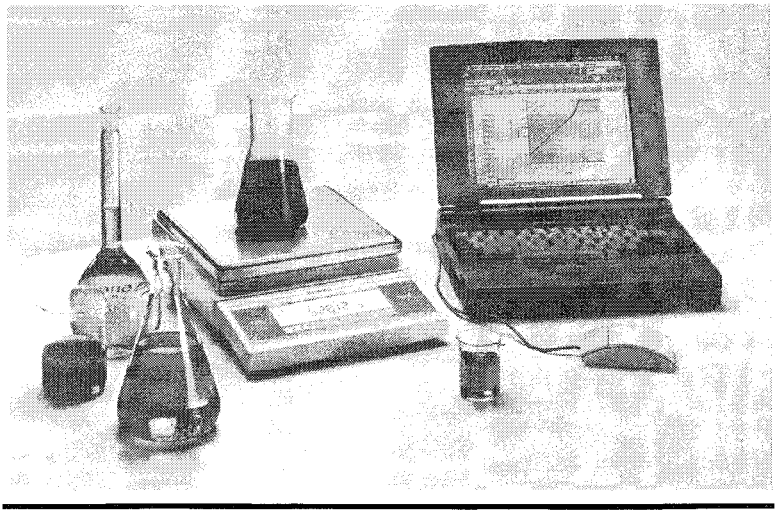
adjust the balance if during a calibration test using an external calibration weight the operator notices that the indicated value on the balance display deviates from the apparent mass value of the calibration weight by more than the tolerance set out in the SOP. This means that the balance is only adjusted using the internal weight and is then calibrated using an external certified and traceable calibration weight. Almost every balance with internal adjustment weights uses conveniently shaped pieces of stainless steel that can be easily raised and lowered and stored in a fixed position close to the weighing system; the weights are manufactured to a value close to the target value. It is therefore quite feasible that a weight with a target value of say 200.0000 g actually measures only 198.6321 g. It is because of the non-OIML Organisation Internationale Metrology Legale compliant shape and odd mass value that the internal weight cannot be given a calibration certificate. However, because its actual mass has been accurately determined, a factor can be stored in the microprocessor so that the balance is always adjusted correctly.

Data Communications Via a Personal Computer

Serial communications via RS232 have been the standard format for interface connections on laboratory balances since the earliest personal

computers (PCs) adopted this standard. Because of the incredible processing power of the commercially available software packages, the option to download data from the balance directly into a PC seems highly attractive. However, anyone simply connecting a balance via the RS232 link and hoping to directly input data into a spreadsheet will be in for a rather rude awakening since software of this nature expects data entry via the keyboard and has no option for accepting data via the serial communications port. Help, as always, is at hand since most manufacturers can supply a solution in the form of a software wedge and connection cable. A software wedge is the description given to the utility written to allow the user to obtain full two-way serial communications between a Windows® application and an external instrument, in this case a laboratory balance, although the device could be anything from a pH meter to a spectrophotometer (Figure 1.1). The software wedge operates on two levels: (1) by converting incoming serial data to keystrokes, which means that the applications receive data as though they had been typed in or (2) the data can be downloaded using dynamic data exchange (DDE). DDE is vastly more powerful than the keystrokes method, however, as one would expect this superiority to require additional complexity in the set-up. These software packages are relatively low cost and as such are not available as fully validated for use in the quality control of tablets or medicines.

Figure 1.1. Serial Communications from a Balance to a PC



Moreover, there is no restriction preventing data from being entered manually or previous data being changed.

Software Integrated into Balances

Even the lowest cost laboratory balances are equipped with some applications software to perform calculations or routines that provide an instant solution, saving valuable time and effort on behalf of the operator. With the advent of lower costing memory and faster microprocessors, it is now common to have a large suite of applications software available in every balance, although in most cases the balance would have

been purchased with the intention to use only one of the programs available. When choosing a balance to solve problems, be aware that (paradoxically) a lower cost balance with less sophisticated software is often more difficult to use than the top of the range model with more complex programs. The reason for this is that the more expensive top of the range models have more keys and usually have one key for each function; in addition, the better display on the more expensive balances will provide easy-to-follow prompts throughout setup and execution of programs. In particular, look for a balance that utilises soft keys in combination with a dot matrix display and giving plain English prompts for easy operator instruction. The next few sections detail some of the most commonly used applications.

Counting

The counting program has two distinct areas of use for precision laboratory balances: (1) high-cost, small precision electronic components and (2) tablet counting in a pharmaceutical dispensary. Simply count out a small number of tablets, usually five or ten, place them on the balance and push a key to store the total weight. The balance automatically calculates the average tablet weight. The display of the balance now shows how many tablets are on the pan; all you need to do is add more tablets until the required total is reached.

Percentage Weighing

Where small recipes are written using a percentage rather than weight to define the amount of each component, then this program offers an easy solution, relieving the operator from the burden of calculating values in grams. Also be aware that for the printout generated, it is usually possible to print the tare weight of the container as well as individual net weights and a gross or total weight. If the balance has a numeric keypad, then it may be possible to define the weight of the first component as its percentage value in the recipe. This is particularly useful when one of the components is difficult to measure due to its nature. Simply dispense a convenient amount and define its percentage value in the recipe and weigh out the rest of the components in relation to this amount.

Dynamic Weighing

Dynamic weighing is used to give a close estimation of the weight of an object that causes extreme fluctuations in the displayed result. It is often used to weigh live animals. This program will take a series of measurements over a short time frame and give the value as an average of these measurements. Often the facility of an automatic start helps, as all the operator needs to do is load and unload the balance. In order to improve accuracy, a range window can be set so

that the balance will only start to record measurements when they fall within a predefined limit. For example, when animals are initially placed on a balance, their movements may cause fluctuations in the display of up to 75 percent. The window is set to start measurements when the fluctuations are no more than 25 percent, therefore, the balance waits until the animal calms down, leading to a final result that is much closer to the true value—a simple but effective method! A further use for dynamic weighing is where the location prevents reasonable stability from being achieved on the balance, for example, in fume cupboards or on oil rigs!

Statistics

Usually used in statistical process control (SPC) applications rather than tablet weighing since pharmacopoeial limits are set as limits of uniformity rather than as statistical limits fixed around a set target value, a specific pharmacopoeial program exists in its own right and can be integrated into a standard balance for some manufacturers or is available in a PC program. However, where statistics are required, then these programs are usually configurable to print out and display the important parameters and set-up via the menu and will allow the user to select automatic data transfer and taring between additions of samples to speed up the process.

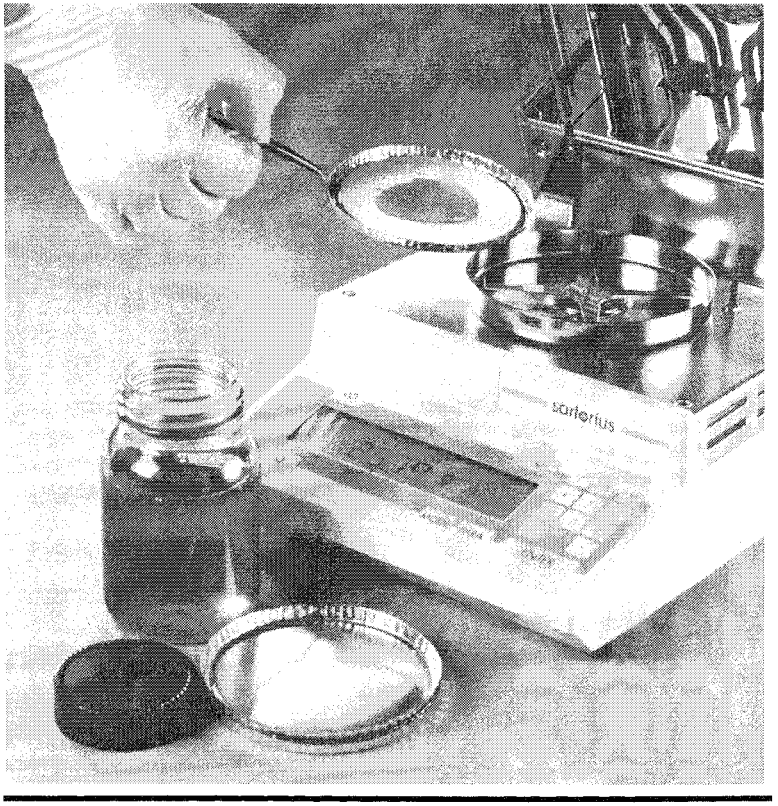
Other typical programs available include density, checkweighing, and the conversion of mass units, none of which have typical applications that spring to mind within the pharmaceutical industry.

Moisture Analysers

There are many different moisture analysers (Figure 1.2) available on the market, all making different claims for temperature ranges, maximum sample size and novel heating sources. All of the different types work on broadly the same principle: An integrated balance measures the starting weight of the sample, and infrared energy emitted from a source causes the volatiles in the sample to evaporate. After an elapsed time or pre-set end-point determination, the energy source is turned off, and the results are displayed either as a percentage of moisture lost or solids remaining. Since the moisture unit simply measures all volatiles lost as moisture, this value cannot be considered as an absolute value for water, and care must be taken to avoid denaturing the sample due to excessive bombardment with infrared energy, as this will cause weight loss that will erroneously be calculated as moisture.

This method is ideal for providing rapid results for single samples, where drying times can be as short as 5 min, but typically 10 to 20 min based on a typical sample size of 3 to 6 g. Such moisture

Figure 1.2. Samples Being Loaded into a Moisture Balance



analysers are used widely throughout the pharmaceutical industry, often in production as a means of measuring the moisture of in-process material prior to compression into tablet form.

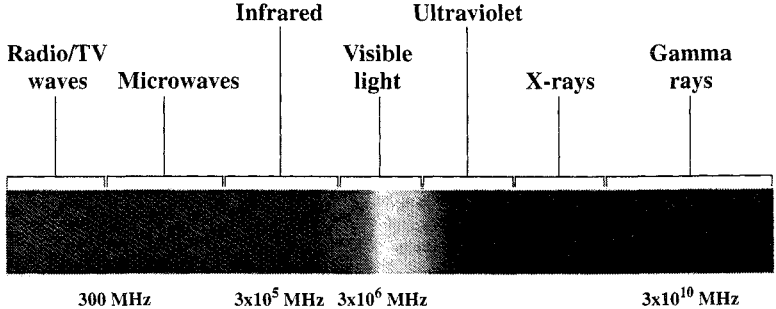
There are a few misunderstood concepts of infrared moisture analysers that are well worth discussing at this point so as to dispel some of the myths and untruths that may be causing concern

when comparing one moisture analyser to another or to a different, perhaps absolute, moisture determination process. First, all of these analysers dry the sample by means of infrared energy, although there are many different energy sources, including ceramic surface radiators, metal rod radiators, quartz rod radiators, or halogen elements.

Although infrared energy is invisible to the human eye, this type of energy is classified as light energy and as such is governed by the laws of optical physics. Any visible red light emitted is only a by-product of when the radiator converts electrical energy into infrared energy. This red visible light is seen only on certain types of infrared radiators.

The next topic of contention is that of the drying temperature that the analyser will raise the sample to during the drying process. One manufacturer quotes 40 to 160°C, another 50 to 200°C;

Wave Spectrum



one unit quotes temperatures as high as 400°C—so which will dry the sample the quickest? Since the sample is dried by absorbing infrared energy, and energy is measured in watts or joules per second, how could it be possible to quote temperatures in degrees Celsius? The answer to this comes in two parts, namely how the temperature of the sample is affected by its own character and how the output from the infrared source is measured and calibrated. Consider two similar samples: The first sample is a light-coloured powder, and the second sample is almost black, although both have the same temperature content. Once again, the laws of optical physics once again play their part in the drying process. The dark-coloured sample readily absorbs infrared energy, and the sample temperature will ultimately be higher than the light-coloured powder which in turn reflects more of the infrared energy and reaches a lower temperature. The drying process is also likely to take a little longer for the lighter sample. Common sense will tell you that if a temperature probe is located closer to the infrared radiator, then it will register a different value than if it were placed at a distance. Furthermore, the type of probe used will also cause slight differences in the measured temperature within the chamber to be recognised. A thermocouple in conjunction with a digital readout is the most common and reliable method for measuring the reproducibility of the infrared radiator within a moisture balance. If in doubt, it is

always best to consult the manufacturer and inquire how they would check and adjust their own moisture balances. If you are then able to utilise the same method, this will give you the ability to calibrate the heating source of the moisture balances throughout your company to the same standard. Sartorius can supply as an accessory the same type of thermocouple and digital readout as used in manufacture and servicing together with an SOP for testing the reproducibility of the heating source.

In order to test the reproducibility of the instrument, there is a method detailed in the Sartorius SOP for moisture balances that utilises a saturated saline solution with a specified criterion for the settings of the heat source and the anticipated results. Since all moisture balances calculate the moisture content of a sample as being the total weight of all components lost by evaporation during a rapid heating process, the method cannot be considered as capable of providing an absolute value as would be given in the case of, say, a Karl Fischer analyser; therefore, moisture standards that are used to calibrate such instruments cannot be used to calibrate infrared moisture balances.

Multireweighing or Backweighing Software for Moisture Determination

Where the volume of samples is too great to use a moisture balance (which can analyse only one

sample at a time), most manufacturers have integrated software into their balances that will allow the storage of container weights and sample weights in a non-volatile memory. The samples can then be processed and reweighed to determine the moisture content as a percentage. As the name *multireweigh* suggests, samples can be processed more than once in order to calculate the loss after different processes, for example, drying at 105°C and then after ashing in a muffle furnace at 800°C. Flexibility in the software should allow the operator to weigh the sample in any order, since it would be unlikely that samples will be removed from a dessicator in the same sequence that they were originally weighed. Such a program can, of course, be used to calculate changes other than loss of moisture, for example, measuring the change in mass after time of inhalers or measurement of repeated single doses.

Calibration of Micropipettes in a Quality System using a Balance and PC

Since according to any quality management system pipettes are classified as test and measuring equipment, they too must have an SOP that includes a documented calibration log. Most companies have a contract with the manufacturer or a qualified service organisation to test, calibrate

and adjust their pipettes on a six-monthly basis. This usually involves sending them away, although some companies offer a mobile calibration service. Both options offer a calibration and repair service, although the latter may seem at first the more attractive option since your pipettes will be out of action for less time. Consider, however, that in order to provide this service, the agent has to transport the balance all over the country, set it up and start calibration within a short time, which is usually not enough time for the balance to fully equilibrate temperaturewise. Furthermore, it is quite unlikely that in today's well utilised (crowded) laboratory that the visiting technician will be given a perfect location with a solid stone weighing table and air-conditioning.

Whichever route you decide upon to take care of six-monthly calibration testing, there is still the question of how best to calibrate the pipettes in the interim in order to meet the requirements of the quality management system. Naturally, a balance will feature somewhere in the plan but which balance, what conditions and which international standard apply? How can I be sure that I am reducing external influences such as evaporation, all of which contribute to the accuracy and repeatability values for my pipette?

The main objective set out in any quality system must be to ensure that any measurements are made with the intended accuracy. Pipettes are judged by the following criteria: (in)accuracy and (im)precision, both of which can be

measured using an appropriate balance. According to ISO 10012, this means that the balance should have an accuracy of one-tenth of the permissible error of the pipette to be tested. Additionally, a documented system must be employed covering the measuring equipment and any measuring standards used in the calibration process, together with all significant uncertainties identified, including those contributed by personnel and the environment.

For the gravimetric determination of the accuracy of a micropipette, the volume of distilled water aspirated from the micropipette is calculated from the measured mass using the following equation:

$$\text{volume} = \frac{\text{measured mass}}{\text{density of water}}$$

The corrected density of distilled water can be taken from tables published in reference books—The density will be corrected for changes in both temperature and air pressure. The factor to correct for different combinations of air pressure and temperature is known as the Z factor.

The first software packages appeared on the market in 1994, and some incorporated a database facility to allow the entry of individual test plans for each type of pipette as well as long-term storage of results. The choice of testing in accordance with the current British, American or European standard was given to the operator,

and all measured mass values are processed to give the correct result expressed as a volume, correcting for the effects of temperature and air pressure.

There are two schools of thought as regards to the optimum solution for counteracting the effect of evaporation—potentially the largest factor influencing the quality of your results. One approach is to measure the rate of evaporation of liquid from the vial on the balance pan and then to incorporate a factor in the calculation to correct for this phenomenon. At first, this solution seems plausible, however, I would ask you to consider that if liquid has been aspirated down the side of the container, the rate of evaporation will be different to that of a vial containing liquid in the bottom, and the settling time of a balance can easily double from one measurement to another, therefore invalidating any correction factors applied.

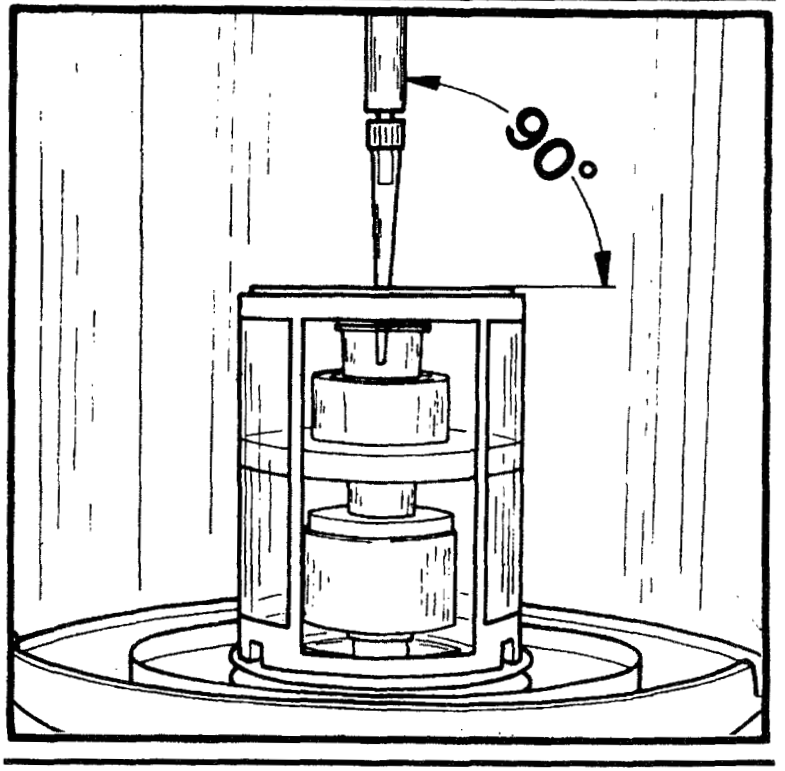
A better solution utilises a vapor or humidity trap where the vial is housed in a chamber where the humidity is between 60 percent and 90 percent, thus removing the gradient between the vial and current ambient conditions, therefore reducing evaporation to a negligible amount compared to the reproducibility of the pipette. As a further endorsement, the latest international standard for the calibration of pipettes, ISO 8655, recommends the use of such a vapor trap.

The Humidity Trap

The humidity trap (Figure 1.3) is easily assembled unit that replaces the standard pan on the balance and allows quick and easy aspiration of liquid from the pipette into the container within the housing. Water located in the trough provides the required humidity to prevent evaporation.

Balances with an automated door can be controlled from the PC, which means that the process of calibrating pipettes is totally automated, releasing the operator to concentrate on his or

Figure 1.3. The Humidity Trap



her technique of accurate and reproducible liquid aspiration and speeding up what in the past has been a laborious but essential procedure. Furthermore, the requirement to reduce systematic errors (errors due to uncontrollable factors in the measurement process) means that all measurements should be of the same time order so that systemic errors are of the same magnitude and hence of lower influence to reproducibility (imprecision). This is better achieved when all the operator has to do is dispense liquid and press a data transfer key.

Intrinsically Safe Weighing Equipment for Hazardous Areas

From country to country, the regulations and regulating bodies are of course different, therefore, the content of this section is general rather than in-depth or country specific. There are occasions in pharmaceutical manufacturing where an area is designated as a zoned area, since there is a risk of fire or explosion from a spark or naked flame. As such, any electrical equipment used in this location must be approved and certified by the relevant body. The hazard is due to the presence in the atmosphere of

- dust—flour, dust, fine powders;
- liquids—solvents, diluents, coating materials; or
- gases—methane, propane, and so on.

As a manufacturer, you are governed by legislation to provide a safe working environment and, in this case, to provide suitable equipment for the hazardous area. Depending on the material causing the hazard and the length of time it is present in the atmosphere, the risk can be graded from extremely likely down to a mere possibility of fire or explosion, and equipment is manufactured and certified with the relevant coding according to this hazard rating.

Laboratory balances and scales fall into two categories. The first category is described as intrinsically safe, which means that the components and printed circuit boards used in the equipment operate on very low voltage (less than 30 V and 50 mA current). Therefore, there is no possibility of sparking, and the operating temperatures of the equipment are very low. Operating temperature is, of course, an important factor, since some gases have very low flashpoints. Furthermore, the equipment must remain safe even if a fault develops (a higher rating is awarded if the equipment remains safe when two faults develop). The second category is described as flame-proof because the components that are likely to cause sparking have either been sealed in an enclosure or the equipment has been filled with granular quartz that will suppress any sparks generated. Flame-proof equipment is not considered as totally safe for use in zones with extreme hazard ratings. Depending on the type of equipment, there are different notations used to describe the degree of protection inherent in a

balance or scale. As part of the regulation, these markings must be clearly visible to the user.

In the United Kingdom equipment may be labeled as follows:

EEX ib IIB T6	*where the markings indicate that the equipment is*
EEX	*approved under European directive*
ib	*intrinsically safe by design*
IIB	*the types of gases present in the atmosphere for which the balance or scale is safe*
T6	*maximum surface temperature does not exceed 85°C*

In countries other than Europe, there are similar schemes for grading equipment deemed safe for use in a hazardous area, although the markings and symbols will vary from country to country.

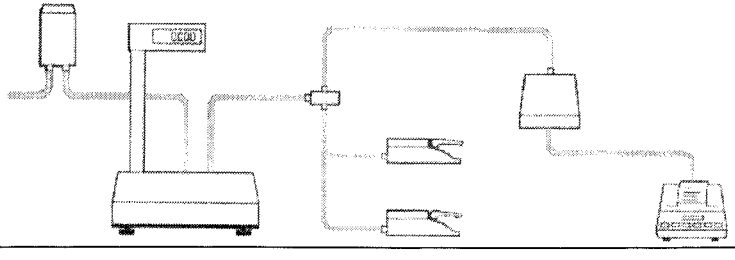
Symbols are used in Europe, the United States and Canada to indicate compliance to the required standard for those countries. These symbols must be clearly visible on the product to which they relate.

Data acquisition from scales or balances in hazardous areas is by no means exempt from the requirement to download directly to a PC or

networked system; therefore, the design of equipment has needed to keep pace with customers' expectations, and digital communication via a safe area network is possible. The most simple option is to equip the intrinsically safe scales with a 10 mV communication port. The data output can be transmitted to a converter in the safe area where conversion to RS232 takes place in a black box and a PC link established (Figure 1.4). Data generated by the weighing machine is recorded at the PC; however, communications are only possible in a simple format, which means that the operator has no feedback from the PC terminal. The overall effect is to reduce the amount of data transfer to a few simple weighing results.

Where the requirements are more demanding, for example, in a dispensary where a full recipe

Figure 1.4. Data Acquisition Processes



(A) Scale connected in hazardous area; RS232 including footswitches for print/tare.

(B) Signal converted to the safe area.

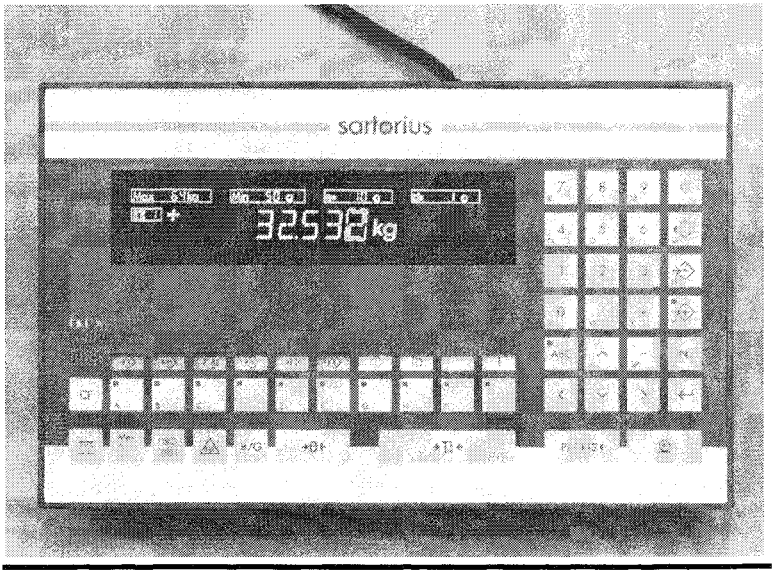
system may be installed, PC terminals approved for use in hazardous areas are now available, and these may well be integrated into the weighing scales. Communication will be on a network approved for use in the hazardous area, for example, fibre optic or in the past ArcNeT Novel would have been used. Ultimately, all base data and software are stored in the safe area on a conventional PC and fileserver and downloaded to the terminals in the hazardous area when operators call up particular recipes for dispensing (Figure 1.5).

Due to the potential consequence of installing non-appropriate equipment in a hazardous area, it is imperative to involve not only the expert from the scale manufacturer but also the health and safety officer from your own company in order to properly categorise the zoned area and subsequently to specify the appropriate explosion-proof rating for the weighing scales.

Commissioning of Balances

When manufacturing drugs and medicines for export or sale in the United States, the FDA inspector will be a well-known person to you. In a constant effort to reduce as far as possible the risk of errors in manufacturing or quality control, the FDA develops and improves protocols for qualifying the performance of new equipment being commissioned and for regular test-

Figure 1.5. Intelligent Terminal That May Be Connected to a Scale in the Hazardous Area



ing when in use. No matter whether the new balance or scale has a readability of 1 μg or 1 kg, there would be little value in any of the measurements made with such an instrument unless the equipment had been properly installed, tested and passed as fit for its intended use. This has become known in the pharmaceutical industry as equipment qualification or EQ for short. As a function of GLP and GMP (Good Manufacturing Practice) quality systems, which are covered later in greater detail, equipment purchased must be appropriate for the purpose and function for which it will be used. This criterion

is, of course, doubly crucial in circumstances where the quality of the product being manufactured or tested depends on the accuracy of measurements made on this equipment. There are three stages of (EQ), namely, design qualification (DQ), installation qualification (IQ) and performance qualification (PQ).

Design Qualification

In terms of weighing equipment, DQ is least likely to require much input from the scientist, since the fundamental principle of the modern electronic balance has changed little since 1975 and neither have the weighing techniques employed by chemists and technicians; furthermore, most of the main manufacturers comply to the ISO 9000 quality system covering manufacturing quality. DQ would become more relevant should a manufacturer introduce a totally new instrument or technique to make measurements, for example, when a microwave moisture analyser is purchased to replace a conventional convection oven, qualification would be required to prove that the drying method used by this equipment is properly documented and that the manufacturer's specification is appropriate to the application.

Other important factors in DQ are the responsibility of the manufacturer to provide and include some or all of the following:

- Clear and well-documented development records, which may often impinge on a manufacturer's right to secrecy in order to maintain market superiority
- Certificates of conformance to specifications or a mutually agreed performance criterion
- Certificates of conformance to international or European legislation covering, for example, safety or conformity to electrical regulations
- Clear and concise manuals of operation including advice on how to get started and some troubleshooting hints. This should include a version number or date of issue to avoid confusion when revisions are made.
- In the case where a weighing system has been supplied that includes PCs and complex control software, an agreement should be made with a third party to store source codes usually at a bank under the terms of an escrow agreement. Embedded software in microprocessors so far has not become subject to such agreements; however, it is becoming more of an issue in terms of validation. Chapter 4 covers this topic.

Installation Qualification

IQ is the responsibility of the end user (meaning the head of the laboratory) and may be performed by either the purchaser, the manufacturer or a competent service agent. IQ covers the unpacking of the balance and completion of the first page in the EQ manual, including

- a completeness check for all items
- the serial number
- the model number
- the asset or internal ID number
- the software version number
- its location and any environmental influences (drafts, vibrations, direct sunlight, etc.), and
- leveling and checking that functions are okay after power up.

Operational Qualification

In the case of laboratory balances, OQ may be performed by a competent person within your own company, although it is usually deferred to the manufacturer or a third-party service organisation. In practice, most companies use the manufacturer, as OQ involves routine quality testing procedures very similar to regular service protocols. A qualified service technician is usually better placed than an in-house operative. There

is a small overlap between what may be considered part of IQ or OQ; leveling, powering up and checking the basic functionality of the balance could become part of OQ instead.

OQ is essentially confirmation of the performance of the new balance against a predetermined criterion and is concluded by the technician signing a ticked list of acceptable test results, recorded modifications and calibration results. (There is a crucial difference between calibration and adjustment; discussions of these two different but easily confused terms are included in this book.)

There are two schools of thought governing the setting of the criterion for acceptance. You could simply take the manufacturer's specification as the pass/fail limits for repeatability, linearity and so on. While there is nothing fundamentally wrong with this approach, there are other factors that may influence you toward choosing the second and perhaps more practical approach. The specifications set by manufacturers should be achievable in most laboratories and production areas where care has been taken to properly site the unit; however, what is the ultimate in performance that you are likely to expect from your balance or scale? Why not determine the pass/fail criterion after establishing the maximum permissible error that you can tolerate in the measurement process? State the error as a percentage and then according to ISO 10012 set your limits to ideally one-tenth of this value or at worst one-third in the case where there are

severe cost implications in achieving the former. To further illustrate this scenario, consider the case of using a microbalance with a resolution of 1 μg and the same reproducibility which provides an uncertainty of measurement equivalent to one-tenth of the maximum error. Using a semi-microbalance costing only **half the price,** however, meets the requirement by having a total uncertainty within one-third of the limit. In this instance the semi-microbalance would seem to be an acceptable alternative.

Having decided on which method to determine your pass/fail limits, then the process of qualification begins. As an example of the components in the testing procedure, consider including the following (taken from the balance manual):

Capacity	*xxx g*
Readability	*x.xxx g*
Linearity	*±x.xxx g*
Reproducibility	*±x.xxx g*

Details of the weights used for the test:

UKAS or equivalent certificate details

_____

Serial numbers _____

Date calibrated _____

Recalibration Date _____

Before any testing commences, ensure that the balance has been "warmed up" by connecting to a power supply for the minimum time indicated in the manual and that a calibration (adjustment) routine has been performed where there is in internal adjustment weight inside the balance; otherwise, external adjustment will be required. (Adjustment corrects for differences in regional gravity that will affect the accuracy of the balance.) Of course, the balance must be leveled prior to adjustment and calibration testing!

Reproducibility Test

A reproducibility test can be performed using a single calibration weight that is closest to the maximum capacity of the balance under test, for example, 200 g for a capacity of 210 g or 100 g for a capacity of 160 g. Never use two weights in order to be able to test closer to the maximum capacity of the balance, as this will introduce eccentric loading errors, whereas a single weight can be placed close to the centre of the pan for each measurement. Tare the balance, then make a minimum of 6 and a maximum of 10 repeat measurements, in each case removing the weight and replacing it as close as possible to the centre of the pan.

Value 1 *xxx.xxx g*

to

Value 6 (or 10) *xxx.xxx g*

Calculate the standard deviation using the formula

$$S = \sqrt{\frac{\sum_{i=1}^{n} (\bar{v} - vi)^2}{n-1}}$$

where S is the standard deviation, vi is the individual value, $\bar{v}$ is the mean value, and n is the number of measurements

Test the linearity of the balance after taring at a minimum of 4 and a maximum of 10 (quite often 6) equally spaced points throughout the weighing range

25% Load Measured Value ____ *xx.xxx g* _____
Accept/fail

50% Load Measured Value ____ *xx.xxx g* _____
Accept/fail

75% Load Measured Value ____ *xx.xxx g* _____
Accept/fail

100% Load Measured Value ____ *xx.xxx g* _____
Accept/fail

Adjustments Required ____ *Yes* ____ *No*

Observations or Comments:

_____

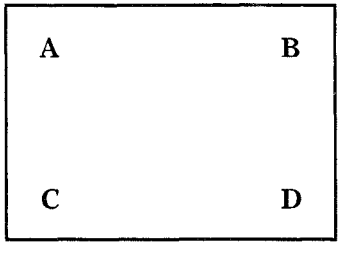
_____

_____

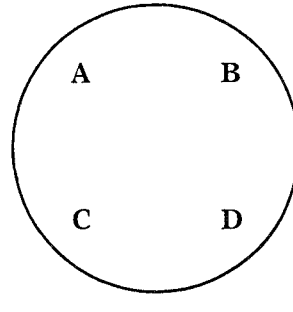
_____

Any adjustments, repairs or routine servicing should be recorded in an appropriate section at the back of the balance manual.

Eccentricity errors, sometimes referred to as corner load errors, must also be checked because these are subject to deterioration due to shocks and vibration that may well have occurred in transport from the manufacturer. Generally, there are no values for this error given in the normal specifications tables in the balance manual, although if you contact the service department of the manufacturer, they should be able to offer you some assistance. As a general rule of thumb, one should expect between two to three digits error at a maximum of half the capacity of the balance. Therefore, a balance with a capacity of 4,000 g and a readability of 0.01 g would have an acceptable error of up to ±0.03 g at 2,000 g when loaded in the positions indicated below:



Rectangular Pan



Round Pan

Indicate the tolerance. Record measurement values at positions A, B, C, and D in a similar manner as with the linearity. Record any observations or comments. If an adjustment is required, ensure that details are recorded in the log.

Appendix 1.1
Determination of the Minimum Sample Weight According to the USP

Background

The United States Pharmacopeia is a quality management system from the USA and applies to the pharmaceutical sector. It is the result of the U.S. ASTM and ANSI quality standards.

Because of the international ramifications in the pharmaceutical industry, more and more German pharmaceutical companies are introducing the USP. It is mandatory to apply the USP particularly in the pharmaceutical sectors that are audited by the FDA (Food and Drug Administration).

Section 41 of the USP specifies the use of weights and balances. It requires that the measurement uncertainty must be known for all balances that are used for pharmacopeial tests and assays. According to the USP, the measurement uncertainty is calculated as three times the standard deviation. As the USP only states the measurement uncertainty for a minimum quantity, this calculation is a good approximation.

Moreover, Section 41 stipulates that the minimum amounts weighed may not be less than

---

Source: Dick Albrecht. Marketing Sales Communications News of Sartorius Weighing Technology, No. 6, September 2000, pp. 7–9.

1,000 times the measurement uncertainty (in other words, the measurement uncertainty may not exceed 0.1% of the reading of the minimum sample weight). Tare weights, such as vessels used for weighing samples, may not be added to the minimum sample weight.

Original text quoted from the USP 23, Section 41 "*Weights and Balances*"

> . . . Pharmacopeial tests and assays require balances that vary in capacity, sensitivity, and reproducibility. Unless otherwise specified, when substances are to be "accurately weighed" for Assay the weighing is to be performed with a weighing device whose measurement uncertainty (random plus systematic error) does not exceed 0.1% of the reading. For example, a quantity of 50 mg is to weighed so that the absolute error does not exceed 50 μg. Measurement uncertainty is satisfactory if three times the standard deviation of not less than ten replicate weighings divided by the amount weighed, does not exceed 0.001. . . .

Given that a balance is most inaccurate **in the relative sense** as a load approaches the balance's zero point, the USP requires that all amounts be weighed with a minimum accuracy of 0.1% and a confidence interval of 99.9% (three times the standard deviation).

Procedure for Buying a New Balance for Use According to the USP

It is not possible for the manufacturer to explicitly specify the minimum sample weight for a balance. The reason is that the standard deviation, as the significant factor, depends on the location. This means that the more a balance's place of installation approaches ideal conditions, the better the standard deviation, and the lower the minimum sample weight will be.

There are various approaches for specifying the minimum sample weight:

Specification of the Minimum Sample Weight Based on the Standard Deviation Specified by the Manufacturer

The standard deviation (repeatability) describes the ability of an instrument to obtain corresponding results under constant testing conditions. Therefore, this can be designated as the most important metrological feature of a modern electronic weighing instrument. For this reason, the "repeatability" (also called "reproducibility") is always given in the specifications for a balance.

The simple standard deviation (1s) is mostly used to quantify the repeatability. With this specification, the minimum sample weight can be calculated using the following equation:

Repeatability · 3000 =
minimum sample weight

The factor 3,000 is calculated based on the USP requirement that **three times the standard deviation** (= measurement uncertainty according to the USP) must be used and that the measurement uncertainty may not exceed 0.1% of the reading of the minimum sample weight (0.1% corresponds to the factor 1,000).

Smallest Possible Weight Below the USP Minimum Sample Weight

The actual minimum sample weight can be determined only at the place where a balance is installed because, as explained above, the standard deviation depends on the ambient conditions of the place of installation. Therefore, a balance can be used for weights below the USP minimum sample weight, because the USP definitely limits the minimum sample weight.

Example:

If nine out of ten prescribed measurements yield the same result, and if the tenth weight measured differs by only 1 d (digit) from the others, you obtain a standard deviation of s = 0.32 d. This yields an uncertainty of 0.96 d, hence approximately one digit (scale interval or division), and the minimum sample weight cannot be less than 1,000 digits. This applies to many analytical balances with a scale interval (scale division) of 0.1 mg.

If all ten measurements yielded the same result (s = 0), the minimum sample weight would be

calculated as 0 d. However, from a metrological standpoint, this is incorrect (although some beg to differ) because one should not specify an uncertainty of less than 0.5 d due to the rounding error of measuring equipment with a digital display. As a result, a minimum sample quantity below 1000 d should not be used.

If several measurements differ from the others, this will always yield a higher standard deviation and thus a higher minimum sample weight. Usually, you can expect a minimum sample weight of 1,500 d or 2,000 d. This applies to many analytical balances with a scale interval (division) of 0.01 mg and to micro- and ultra-microbalances.

Hence, the following minimum sample weights are yielded under excellent ambient conditions and for optimal settings on the balance, based on the metrological specifications:

Type of Balance	*Minimum Sample Quantity*
LA 230 S	approx. 150 – 300 mg
ME 215 S	approx. 15 – 30 mg
MC 5	approx. 1.5 – 3 mg
SC 2	approx. 150 – 300 µg

Chapter 2

CONTROL OF TEST AND MEASURING EQUIPMENT IN A QUALITY SYSTEM

Balances and Scales in GLP/GMP/ISO9000

Stephan Weyhe
Sartorius AG
Göttingen, Germany

An important element of all quality systems is the control of inspection, measuring and test equipment. The quality element *control of inspection, measuring and test equipment* requires that the supplier of a product or service

develop and maintain Standard Operating Procedures (SOPs) for inspecting, calibrating and servicing test and measuring equipment. The purpose of these SOPs is to ensure that the supplier's products conform to defined quality standards. When referring to the control of inspection, measuring and test equipment, we mean an orderly sequence that ensures that the equipment is inspected in a timely fashion and, if necessary, appropriate measures are taken so that the equipment corresponds to the given requirements.

Using the laboratory balance as an example, this chapter explains how one can establish an acceptable level of confidence in the test and measuring equipment being used. Suitability of the equipment is the initial requirement for obtaining reliable results.

Motivation

Recently, extreme ranges of resolution have been attained in the field of analytical weighing technology. Reaching these new limits, however, has opened up discussion about the competence of individual laboratories. For this reason, most laboratories currently keep certificates of their credentials on file. These certificates provide objective evidence of the laboratory's performance and assure those using the laboratory that analytical questions will be answered by an expert.

In addition, the flood of analytical data has confronted laboratory employees with a problem. Namely, they must test and validate many measured values for plausibility and accuracy. Here again, quality assurance measures are essential for correct, comparable and verifiable results and are fundamental to long-term success.

Good Laboratory Practice (GLP), Good Manufacturing Practice (GMP), the ISO 9000 series and the EN 45000 series are regulations and standards of the most prominent quality systems that relate to the control of inspection, measuring and test equipment. They have been generalized to cover a large number of devices and procedures and, therefore, must be interpreted accordingly. ISO 10012 provides a more extensive and concrete explanation of the requirements for test and measuring equipment. Accordingly, a series of measures for using test and measuring equipment can be summarized into a few general, basic requirements.

Selection of Suitable Test and Measuring Equipment

Data obtained with the test and measuring equipment are simply observed values of a quality feature that, for example, is created in a laboratory and exhibits a tolerance defined as acceptable. During measurement, systematic and random

influences from the weighing procedure, as well as from the test and measuring equipment itself, are superimposed on the quality feature—in this case the weight. The recorded result is subject to uncertainty, which we refer to as uncertainty of measurement. This uncertainty essentially must be indicated for every weighing procedure. Specific factors that should be taken into consideration when weighing will be discussed in the following sections.

When selecting a suitable piece of equipment, one must ask what degree of uncertainty of measurement the test equipment can exhibit in order that a reliable judgement of the quality feature is ensured (adherence to a required tolerance).

A guiding principle for selecting the proper equipment is the "golden rule" of measurement technology, which states that the uncertainty of measurement of a piece of test equipment should only amount to 1/10 of the tolerance of the feature being tested. An example will explain this principle: A 10 mg sample is to be weighed with a degree of accuracy of 1 percent, which corresponds to 0.1 mg. According to the "golden rule," the total uncertainty of the balance may not exceed 0.01 mg.

In this regard, however, it is also important that this requirement be met in an economically sensible manner, especially when performing costly procedures. More than ever, when it comes to the type of equipment used, cost-effectiveness is gaining acceptance as an

important aspect of quality-conscious business practices. Therefore, under certain conditions, the uncertainty of measurement may deviate from the rule mentioned above. A ratio of 1/3 is still acceptable if the suitability of the test equipment is ensured through proper supporting measures (such as the frequency of inspection).

Manufacturer specifications, which are the basis for selecting test and measuring equipment, are briefly explained and interpreted in the following sections.

Weighing Range

The limits of a measurement, i.e., the range within which the defined certainty of measurement is maintained, determine the **weighing range.**

Repeatability

Also called reproducibility, **repeatability** describes the ability to display corresponding results under constant testing conditions when the same load is repeatedly placed on the weighing pan in the same manner. Either the standard deviation or the difference between the largest and the smallest result for a defined number of measurements is used to specify this quantity. For rating the quality of an instrument on the basis of the specifications, both values are essentially comparable if the minimum and maximum capacity specifications are compared with the

standard deviation times three. Ninety-nine point nine percent (99.9%) of all representative values of a series of measurements are found within the range of the standard deviation times three. In other words, they fall within the range that is also defined as the difference between the maximum value and the minimum value.

In accordance with known statistical data, the standard deviation times two defines the range in which 95 percent of the measured values are likely to occur. In practice, it has become acceptable to use this interval as a percentage for the statistical error when calculating the uncertainty of measurement.

Repeatability is essentially independent from the load on the balance or scale. It is regarded as an important technical feature because its influence on the uncertainty of measurement—especially with smaller loads—becomes a dominant factor. Therefore, in most cases it is sensible to run a short series of repetitive measurements before purchasing a high resolution balance or scale to be convinced that this technical feature can be maintained.

Linearity

The **linearity error** (usually referred to as **linearity**) indicates how much a balance or scale deviates from the (theoretically) linear slope of the interdependent load and display. Over the entire characteristic curve, the linearity error can

typically be described as a monotone curve of the second or third order. Therefore, its influence on the accuracy of small loads or differences is negligible. With heavier loads (>10 percent of the weighing capacity), this influence can have an effect on the precision of the measurement. Therefore, when developing an SOP (see below), it may be necessary to indicate the uncertainty of measurement, depending on how the equipment is used.

Temperature and Sensitivity

The **temperature coefficient** indicates the change of a value when the temperature changes divided by the degree of the temperature change. Therefore, it plays a decisive role for rating the stability of a balance or scale when the ambient temperature in the laboratory changes. In general, a distinction is made between the temperature coefficient of the zero point and the **sensitivity.**

For practical purposes, both values can easily be separated. If a light load is left on the balance, over time one will recognize the drift of the zero point in the display. With heavier loads that fall within the overall weighing range, the drift visible in the display is the sum of both effects. The difference, therefore, is the sensitivity drift.

The sensitivity drift is almost always an important factor because the zero point component can be zeroed by taring before the weighing

procedure (exception: long-term analyses, such as thermogravimetric experiments). For this reason, the manufacturer only indicates the sensitivity drift.

If, for example, the temperature coefficient is 2×10^{-6}/K, the load 10 g and the temperature change is 5 K, the system error due to the temperature coefficient can be calculated as follows:

$$2 \times 10^{-6}/\text{K} \times 10\ \text{g} \times 5\ \text{K} = 0.1\ \text{mg}$$

Off-Center Load Error

The **off-center load error** is not specified in the literature accompanying the balance or scale because it is highly dependent on the test methods used. This error is understood as a change in the readout when the same load is placed in different positions on the weighing pan.

To verify the error, a weight is placed exactly in the middle of the weighing pan and the balance is tared. Then the weight is placed in 3 to 4 different locations on the edges of the weighing pan; if the pan is rectangular, the weight is placed in the **corners.** The off-center load error can then be directly read in the display. This value can be negative or positive and usually ranges from 1 to 10 digits. Therefore, especially when using balances with high resolution, the sample to be weighed should always be placed exactly in the middle of the weighing pan.

Determination of the Uncertainty of Measurement

Chapter 3 is devoted to the use of clearly defined mathematical algorithms for determining the uncertainty of measurement based on the balance's specifications. In addition, other factors that can substantially influence the weighing results must be taken into account. Since these factors can only be roughly quantified, it is recommended that they be avoided if at all possible.

Employee Qualifications

Today, leading manufacturers offer balances with a readability of up to 0.1 μg and a resolution of up to 21 million digits. It almost goes without saying that the user must receive proper training in order to capitalize on the accuracy and precision of these instruments. For instance, the user must be aware of basic rules, such as placing the sample in the middle of the weighing pan (to avoid off-center load errors). The user must also attempt to work as consistently as possible (to maintain the specified repeatability) and make sure that the balance is set up on a level surface (to prevent a system sensitivity error).

A simple calculation example underscores the importance of setting up the balance on a level surface. For instance, if the angle of inclination between the balance and the surface is 0.3°, the

mass of a 20 g sample will register 0.25 mg lighter than it should. This value corresponds to 2.5 times the repeatability of an analytical balance!

Weighing Location

Important influence quantities at the weighing location include gravitational acceleration, mechanical disturbances, temperature, humidity and barometric pressure as well as electromagnetic radiation that may or may not be caused by the electrical connection.

Because of the earth's rotation and geographical features, the **gravitational acceleration** varies depending on where the balance is set up. We therefore recommend that the balance be adjusted each time it is set up in a new location and before initial start-up. Another effect that often goes unnoticed is a change in altitude and how it can influence the gravitational acceleration when, for example, the balance is moved to a higher location. For instance, a change in altitude of only 4 m can alter the weight of a 200 g sample by 0.26 mg.

As a result of the moment of inertia, **mechanical disturbances** (e.g., caused by pumps, laboratory vibrators, turbulence under laboratory fume hoods and so forth) register on the balance as periodic or stochastic "weight changes" depending on their attributes. A digital filtering feature on the balance, which can be activated by selecting a suitable integration time, can counter

these disturbances. Low-frequency interference, however, is less likely to be filtered out because the filter can no longer differentiate between interference and a slowly changing weight result (in the case of a filling procedure, for example). We generally recommend that specially designed weighing tables be used with balances that have extremely high resolutions. If vibrations in the building cause the disturbances, we recommend that the balance be set up on a lower floor. If this is not possible, the balance should be used with a specially designed wall console.

We touched on the influence of **temperature** previously. Special care must be taken when the ambient temperature changes very rapidly, for example, when a room is aired out. A weighing system is made up of many individual, different-shaped components of various materials. These components have varying heat capacities and, therefore, do not adapt to external changes at the same rate. This leads to mechanical stresses, which show up as changes in the weighing result over time.

One manufacturer recently reported the development of a new **dynamic temperature compensation** feature integrated within the balance. With this feature, individual temperatures are no longer corrected for semi-stationary conditions; rather, each gradient of temperature change is evaluated.

Humidity is a significant influence quantity, especially in balances and scales that are

equipped with force-compensating systems. Since the introduction of encapsulated compensation coils (especially in balances and scales with high resolutions and those verified for legal metrology), this factor has not had a great influence on weighing accuracy. However, when using an older model balance or scale, changes in humidity must be kept as minimal as possible.

For standard weighing procedures, **barometric pressure** is a negligible source of error. It is only necessary to allow for additional buoyancy, which lifts an object during weighing, when conducting weighing procedures that require a high degree of precision. The same holds true, of course, for climatic influences, such as temperature and humidity.

All interference that travels from the main power connection to the electronic components of the balance or scale and could possibly influence the computation of the weighing result is referred to as **disturbances caused by the electrical connection.** The spectrum of disturbances ranges from short-term power outages to pulsed power surges. In most cases, filters installed in the area of the balance's power supply input can alleviate this problem. When an especially high degree of weighing accuracy is required, the balance can also be used with a rechargeable battery. Accessories such as this are available for almost all laboratory balances today.

Disturbances not caused by the electrical connection consist mainly of electromagnetic

radiation in the range of a few kHz up to several GHz, which is frequently used for wireless communication. The balance's external power cord is often the site where an induced signal interferes with the effective signal. Countermeasures for this type of interference must usually be determined empirically.

The Sample

In the majority of cases, the properties of the sample itself are the cause of inadmissible results. The most important factors that influence weighing accuracy are electrostatic charges, magnetic or magnetizable materials, hygroscopic materials and sample temperatures that deviate too much from the ambient conditions in the laboratory.

Electrostatic charges, which are particularly noticeable when the humidity is low, are characterized by a weight readout that drifts considerably and poor repeatability. This effect is based on the interaction of electrical charges that have built up on the sample weighed and on the fixed parts of the balance that are not connected to the weighing pan. This phenomenon primarily affects substances that have a low electrical conductivity and can therefore pass on charges (caused by friction with air, internal friction or direct transfer) to their environment only slowly. Examples of these substances are plastics, glass and filter materials, as well as powders and

liquids. Depending on the polarity of the charged particles involved, this force either attracts or repels, so a weighing result may deviate in either direction. The existence of electrostatic charges shows up in the drift of the weighing result because the charges slowly dissipate above the weighing pan, and the interference resulting from these charges recedes at a corresponding rate. This problem can be eliminated by shielding the sample (using a metal container), increasing the surface conductivity of the sample by raising the level of humidity inside the draft shield of an analytical balance or directly neutralizing the surface charges using so-called static eliminators.

If a sample is **magnetic** or **magnetizable,** i.e., contains a percentage of iron, nickel or cobalt, forces of a different origin are generated, which also have a significant influence on the weighing result. Two interactive mechanisms can be differentiated: attractive and repulsive.

Hygroscopic samples cannot be precisely analyzed because they absorb moisture, which causes a constant increase in weight. If appropriate steps cannot be taken to keep the humidity to a minimum at the weighing location, we recommend that the sample be weighed in an enclosed container that is suitable for its size.

The **sample temperature** is an influence quantity that is often underestimated. Especially when conducting very precise weighing procedures, it is imperative that the sample be adapted

to the ambient temperature. Otherwise, convection currents on the surface of the sample can lead to major errors in measurement. Research has shown that when beakers with a large surface area are used during weighing, temperature differences of a few degrees can lead to deviations in the gram range.

Test Methods

All manufacturer specifications are based on "idealized" weighing conditions. Otherwise, comparisons could not be made between different instruments. But the methods actually used in the field almost always differ from those used by the manufacturer. Variations in the methods used should be documented appropriately in the SOP, and allowances should be made for deviations in the weighing accuracy that may result. For example, if a hanger for below-balance weighing is used to weigh a magnetic sample, the manufacturer specifications, which were determined under the best weighing conditions, cannot be maintained. In this case, preliminary tests must be run using reference samples to verify the attainable degree of accuracy.

If the sample is magnetized, as is the stirring bar of a magnetic stirrer, the forces of attraction that this magnet exerts on the magnetizable parts of the balance will override the weight of the sample. Vice versa, the influence that the residual magnetic field of the electromagnetic-

compensating weighing system has on a sample cannot be ruled out.

Magnetic forces manifest themselves as a loss of repeatability of the weighing result because they depend on the orientation of the sample within the field of interference. Unlike electrostatic interference, magnetic interference is stable over time.

Among the many measures to counter this effect, the method of choice is to increase the distance between the sample and the weighing pan. If increasing this distance is strictly limited by the available space, shields made of soft magnetic materials are good alternative solutions. Special weighing pans provide an optimal combination of increasing the distance and shielding and are offered as accessories for analytical balances.

Calibration and Adjustment

The previous sections covered a series of influence quantities that can adversely affect the accuracy of test and measuring equipment in a variety of ways. Therefore, it is hardly surprising that the test and measuring equipment standards used in all quality systems require that errors in measurement be quantified. In addition, measures for eliminating such errors must be specified. This is done through calibration and adjustment.

Calibration checks the deviation between the weight readout on the balance and a reference weight (in the field of weighing technology, this is a weight whose value is indicated on an accompanying certificate). Calibration is the most important source of information for checking the balance's uncertainty of measurement under actual installation and operating conditions. Therefore, it plays a central role in controlling inspection, measuring and test equipment.

Adjustment always entails corrective intervention in the balance to eliminate the existing error as far as possible. During adjustment, the weight readout is compared to the "correct" value of the calibration weight, and the resulting correction factor is stored in the balance's processor until the next adjustment.

A variety of instruments and methods exist for performing both of these procedures. In general, a distinction is made between internal and external calibration and adjustment.

The external procedure is used mainly on older-model balances and scales or those with high capacities. Comparison and correction are accomplished using one or more weights whose value and uncertainty must be known and documented. National testing laboratories, calibration laboratories and qualified manufacturers provide appropriate certificates for this purpose.

For internal calibration and adjustment, a reference weight that is built into the balance or scale is used. The exact value of this weight was

previously determined during manufacture and stored as a fixed value in the electronically programmable read-only memory (EPROM) of the balance processor. On the simplest models, the user places a weight on the balance's weighing system with the help of a mechanical device. The motorized calibration weight feature, which is operated at the touch of a button, has recently become the standard. The most advanced balances and scales are equipped with a fully automatic calibration and adjustment device that initiates calibration after a preprogrammed or user-defined amount of time has elapsed. In addition, an internal sensor continuously monitors the balance temperature (as a parameter for determining accuracy) and triggers automatic calibration once a certain temperature difference has been exceeded.

Besides the advantages offered by this convenience feature, internal calibration is generally considered preferable over external calibration. The internal weights are better protected from dirt and damage and are always at the same temperature as the balance or scale, per se. Moreover, the motorized calibration feature ensures that the weight is placed on the balance or scale in the most reproducible manner possible. And the fully automatic mode ultimately ensures that one of the most important requirements of the test and measuring equipment is fulfilled.

Traceability of a Measurement

To enable the comparison of results obtained with various balances and scales, we must be able to trace these results to a defined standard. The balance's weighing results are traced and monitored by comparing them to a standard that represents the value of the measurand (quantity subject to measurement) that is required to be correct. This standard is also traced to the national prototype through an uninterrupted chain of such standards for comparison.

Mass is one of seven base units of the International System of Units (SI). Each unit is considered independent of another concerning the unit definitions. Among these units, the kilogram is the only one whose representation and definition are determined by a specific object. The necessity of tracing other units to the kilogram by mass comparison has given rise to the hierarchical structure of mass standards. In this hierarchy, the uncertainty of measurement at a certain level depends on the number of previous mass comparisons.

Weights of various accuracy classes are available to the user. The minimum values and corresponding uncertainties of these weights are specified in R20 of the OIML (Organisation Internationale de Métrologie Légale). Today, certain manufacturers also provide comprehensive information about which accuracy classes are

suitable for the particular application and resolution of the balance or scale being used.

Whenever possible, certified weights should be used. Calibration weights must be unmistakably labeled as such, and their handling (procurement, storage, inspection, transport) must be regulated and documented.

In this context, one often asks the question of how the traceability of the internal weight of a balance can be ensured. This is accomplished by tracing the internal weight, whose materials and surface properties must fulfill all the requirements of a classified weight, to a highly accurate set of reference weights from the manufacturer.

As is the case with all external weights, this internal weight must also be tested at defined intervals to determine whether its tolerances are maintained. This is usually done when the balance or scale is serviced.

Documentation

A characteristic element of all quality systems is the requirement of documentation. A documented system must be set up and maintained for managing, verifying and using test and measuring equipment. Requirements as to the extent and depth of the documentation vary significantly depending on the system being used.

In any case, it is helpful to use the rule of five Ws as a guide when developing a set of

instructions that must be followed. This rule states that procedures must be documented in such a way as to answer the question.

"Who did What, with What, When and Why?"

In the area of management of test and measuring equipment, experience has shown that this requirement is best met by introducing and maintaining an SOP and a logbook for the weighing instrument. While all aspects of operation are laid out in the SOP, the logbook contains entries about the maintenance, service and repair procedures for the particular balance or scale.

SOPs should be written in a way that enables a qualified employee to perform the specified operations without special training. When laboratory balances are to be used, it is imperative that the SOPs include references to the most important basic functions, calibration and maintenance schedules, limitations of use and reporting requirements when implausible deviations are detected in the weighing results. The following information is to be used only as a guideline for developing documentation records; practical examples of SOPs and logbooks are available from leading manufacturers of laboratory balances.

In particular, the following must be recorded:

- *Description and identification of the test and measuring equipment:* This includes general information about the type of

weighing instrument (e.g., analytical balance with motorized draft shield); the most important manufacturer specifications; and the model, serial number or inventory number at the weighing location.

- *Calibration equipment and results:* These two factors are decisive for maintaining the desired degree of weighing accuracy. Depending on the resolution of the balance or scale and its construction features (motorized placement of the weight on the weighing pan, fully automatic calibration function), determinations must be made about the nominal value, the maximum permissible errors and how the weights are to be used. The weights or sets of weights used are also considered test and measuring equipment and must be labeled and identified accordingly. Intervals for recalibration of the weights must also be defined. Especially when there are large deviations in the calibration results, control limits must be defined, and a procedure must be developed for reporting such deviations.

- *Defined control limits:* The overall uncertainty of measurement, which was determined using the test and measuring equipment described above, must be traceable. On the basis of this value, the user can determine whether the balance or scale is suitable for the tolerance indicated in the SOP (e.g., the analysis).

- *Ambient conditions and corresponding adjustments:* The specifications which characterize the balance or scale are determined by the manufacturer under well-defined standard conditions. In reality, however, certain usually unfavorable conditions often cannot be avoided. For example, if the balance or scale is located under a fume hood in the laboratory or in a place where there are great fluctuations in temperature, the analysis can be adversely affected. Modern balances and scales can be adapted to the ambient conditions at the weighing location by varying the set of parameters in the operating system so that the balance or scale may be used in that location. However, this almost always results in the accuracy being reduced. For example, if the "stability range" parameter is increased, the balance or scale can deliver accurate results even when it is subjected to a field of interference of a great amplitude. The attainable repeatability, however, is sacrificed in the process. In this case, the change in the parameters of the balance or scale operating system and the influence on the uncertainty of measurement must be documented.

- *Maintenance procedures:* Determinations must be made about when the balance or scale should be cleaned, who should service it and at what intervals and how to proceed if a repair is necessary. The

results of regularly performed maintenance procedures can also be useful for analyzing the trend of certain deviations. This facilitates appropriate definition of the interval of confirmation.

- *Modification of the weighing instrument:* A variety of technical applications require that the standardly equipped balance or scale be modified. For example, a hanger for below-balance weighing might be used if either the size of the sample or special ambient conditions (such as magnetic fields, temperature, humidity and so forth) dictate the manner in which the analysis should be conducted. Weighing pans of modified shapes and sizes and analytical balances with specially designed draft shields are also often used. Today, leading manufacturers are in a position to offer their customers application-specific solutions with respect to digital filters or other weighing parameters. Dynamic weighing procedures constitute one of the main application areas for which this type of modification is necessary.

  It almost goes without saying that this type of balance or scale will have specifications that differ from those of a standard model. When selecting test and measuring equipment, these deviations must be determined and documented in the SOP.

- *Restrictions on the suitability of test and measuring equipment:* If a confirmation or

calibration procedure determines that the test and measuring equipment can no longer operate within the defined maximum permissible errors, even if corrective intervention is taken, the balance or scale should no longer be used for the intended purpose. Of course, it is possible to use the balance or scale for analyses that do not require such a high level of accuracy. In this case, the limited application range must be clearly denoted on the instrument and indicated in the SOP.

- *Identification of responsible personnel:* The laboratory manager appoints a person to oversee the test and measuring equipment. This person is responsible for the appropriate use of the balances or scales.

Defining the Interval of Confirmation

We use the term *confirmation* to summarize all activities which ensure that the predefined properties of the test and measuring equipment are maintained. Therefore, the interval of confirmation corresponds to the time interval or number of analyses performed with the test and measuring equipment between two successive inspections.

From an economic standpoint, testing should be optimized so that it is performed before a balance or scale exceeds the maximum permissible

errors. This is also closely connected to the previously mentioned rule which states that the uncertainty of measurement of the test and measuring equipment should be much lower than that required by a particular weight measurement application.

The following must be taken into account when first defining the interval of confirmation:

- The extent of possible adverse effects on the analysis due to nonconforming test and measuring equipment. When should data obtained with a nonconforming instrument be rejected?

- What additional expenditures can result from overfilling expensive substances? Can the customer assert product liability claims in such cases?

- Laboratory balance manufacturers—if they provide service and maintenance for their products—have an extensive amount of data at their disposal with respect to all important features of the balance. This is especially true given their various areas of use and ranges of application.

- *Tendency toward component wear and drift:* Modern laboratory balances are designed and constructed to keep component wear to a minimum when they are operated according to the manufacturer's instructions. The readout might drift in individual cases and after prolonged use of

the balance or scale due to the electronic components.

- *Environmental influences:* The range of uses for balances and scales is specified according to temperature and humidity classes. If a weighing instrument is mainly or constantly subjected to temperatures or levels of humidity that border on the allowable limits of these classes, the specifications will likely be affected and must be taken into account accordingly.

- *Demands of customers, standards or laws:* If the equipment is to be used in sensitive areas with very high security standards (e.g., in the aerospace industry, for medical technology, for pharmaceutical production and so forth), the customer will place high demands on the supplier's quality system. These demands can go far beyond the standard requirements and, therefore, can also have an influence on the control of inspection, test and measuring equipment.

- *Experience with similar test and measuring equipment:* Because of the multitude of factors that must be considered when defining the interval of confirmation, a general recommendation on how to do so cannot be made. It makes more sense to follow your technical "intuition" and consider the relevant factors to determine a suitable interval. Statistical data from the current inspection can be used to check

and optimize the interval that is initially selected. For example, the interval of confirmation can be gradually adjusted by cutting the test interval in half, if the maximum permissible load errors are exceeded, or doubling it if the requirements are met satisfactorily.

From an economical standpoint and to ensure the traceability of test results, it may be useful to combine extensive inspections at longer intervals with additional short-term tests or calibration procedures using suitable working standards.

Summary

The control of inspection, measuring and test equipment is an element of functional quality management. It is a prerequisite for objectively demonstrating the performance of a laboratory as well as for introducing and maintaining processes that can be controlled. This starts with the selection of a suitable test or measuring device based on the tolerances to be tested, which, for instance, are indicated in the laboratory's SOPs. Measuring equipment suitable for this purpose has an overall uncertainty of measurement that is much lower than the sample with respect to the specifications of the equipment and all factors that have an influence on the measurement. Suitable SOPs should be indicated in writing to

ensure that the test requirements are always met, and all related data should be documented.

BIBLIOGRAPHY

(German titles have been translated into English in parentheses for convenience.)

Christ, G. A., S. J. Harston, H.-W. Hembeck, and K.-A. Opfer. 1998. *GLP-Handbuch für Praktiker* (*GLP Handbook for Experienced Professionals*). Darmstadt, Germany: Gt Verlag GMBH.

Deutsche Gesellschaft für Qualität e. V. (German Society for Quality). 1998. *Prüfmittelmanagement* (*Management of Inspection, Test, and Measuring Equipment*). Frankfurt, Germany.

Deutsches Institut für Normung e. V. (German Institute for Standardization). 1995. *Leitfaden zur Angabe der Unsicherheit beim Messen* (*Guidelines for Indicating the Uncertainty during Measurement*). Berlin, Germany.

DIN ISO 10012. 1996. *Forderungen an die qualitätssicherung für Meβmittel, Meβunsicherheit und Fähigkeit, Qualität und Zuverlässigkeit* (*Quality Assurance Requirements for Measuring Equipment, Uncertainty of Measurement, and Capability, Quality and Reliability*). Geneva, Switzerland: International Organisation for Standardisation.

Verein Deutscher Ingenieure (Association of German Engineers). 1998. *Prüfmittelmanagement und Prüfmittelüberwachung* (*Management and Control of Inspection, Test, and Measuring Equipment*). Düsseldorf, Germany.

Weyhe, S. 1997. Wägetechnik im Labor (*Weighing Technology in the Laboratory*). Landsberg/Lech, Germany: Verlag Moderne Industrie.

Chapter 3

EXPRESSION OF UNCERTAINTY FOR DISPLAYED NET VALUES

Juergen Ober
Sartorius AG
Goettingen, Germany

Everyone who uses a measuring instrument should be thoroughly acquainted with the characteristics of the instrument in order to estimate the accuracy of the results produced by it. This is why knowledge of equipment characteristics is required by the standards established for quality management systems. One common method of

determining these characteristics is to calibrate the measuring equipment.

There are two basic regulations that may be used: (1) the OIML Recommendation R 76 for Non-Automatic Weighing Instruments (used in legal metrology and explained in detail in Chapter 7) and (2) the ISO (International Organisation for Standardisation) Guide to the Expression of Uncertainty in Measurements. In this chapter, methods are shown that comply with the ISO guide.

In some countries, publications from the accreditation bodies recommend different calibration procedures. Most of them lead to estimated uncertainties for gross load values. Taking into consideration that almost all modern electronic balances allow easy tare balancing by pressing the tare button, the user very often will read net values in the display. The procedures described here will lead to estimates for the uncertainty of both the displayed net values and gross values.

Quantities Influencing Uncertainty of Measurement

The uncertainty of a measurement result should cover all possible influences. For weighing results, uncertainty depends on influences from the

- weighing instrument itself,
- weighing procedure,

- ambient conditions and
- user.

The influences from the weighing instrument can be determined during calibration. The result depends on the calibration procedure, which should cover all weighing parameters, and on the calibration weights used. The weighing parameters are

- repeatability;
- rounding error;
- deviations in the characteristic curve, causing performance/linearity errors; and
- errors with eccentric loads.

Relevant characteristics of the calibration weight (weights recommended for calibration meet the requirements of the OIML R 111 International Recommendation) are

- conventional mass value and
- uncertainty of the conventional mass value or
- maximum permissible error of the weight.

Almost all weighings in a laboratory will be performed only once. Therefore, influences from the weighing process normally are covered by the calibration procedure; no additional increase or decrease factor is necessary.

Influences from ambient conditions are taken into consideration by performing the calibration at the place of installation of the weighing instrument under typical ambient operating conditions, especially considering vibration in the supporting surface and drafts and by estimating the possible influence of variations in temperature.

If the calibration is performed by the user of the instrument, person-related influences are also included in the result. If the instrument is used by several persons with different qualifications, an increase factor may be chosen.

Normally, time-dependent variations are not considered in calibration procedures. Of course, the characteristics of high-resolution weighing instruments will change with time. These variations can be covered by an increase factor, depending on facts such as the following:

- How many different persons are using the instrument?
- How often is the instrument used?
- How often is the span adjusted with the internal weight or a separate weight?
- How often is maintenance service performed?

DETERMINATION OF THE COMPONENTS

The aim of the calibration is to obtain an estimate of the uncertainty value that is representative for the normal weighing process. Therefore, the calibration procedure should be oriented on the normal use of balances, which includes tare balancing in many weighing processes. The calculated uncertainty values must be valid for displayed net values as well as for displayed gross values.

Single-Range Balances

For single-range balances, the following procedures are recommended.

Repeatability

The standard deviation s is determined from at least 6 measurements. The test weight should be at least 20 percent of the maximum capacity of the weighing instrument. Before each loading, the pan must be empty, and the display must show zero. The variance is the square of the standard deviation, s.

$$s^2 = \left(\frac{1}{n-1}\right)\left[\sum_{i=1}^{n}\left(I_i - \bar{I}\right)^2\right] \tag{1}$$

$$\bar{I} = \left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} I_i\right) \tag{2}$$

- s^2 = variance related to repeatability
- n = number of measurements
- i = index of the individual measurement
- I = displayed measurement result
- $\bar{I}$ = mean of the displayed measurement results

Rounding Effect

The variance of the scale interval d is

$$v_d = \left(\frac{1}{12}\right) d^2 \tag{3}$$

This represents a rectangular distribution of $\pm 0.5d$. Because both the zero display—after zero-setting or tare balancing—and the measured value are rounded, the variance of the rounding error is doubled to

$$v_r = \left(\frac{1}{6}\right) d^2 \tag{4}$$

Performance/Linearity Errors

Normally, a weighing instrument is not ideally linear. This causes performance errors that can be evaluated with different test weights. More important for use in laboratories is the fact that due to this nonlinearity the same load will lead to different measurement results when the load is

applied with different tare loads. Therefore, the calibration procedure must include the influence of tare loads. A practical scheme is shown in Table 3.1 with a 400 g weighing instrument. The four calibration weights are to be about 25 percent, 50 percent, 75 percent and 100 percent of *Max* (the maximum capacity of the weighing instrument); the tare weight is to be in the range of 25 percent to 50 percent of *Max*. The performance is determined with 4 different calibration weights, $W_1 \ldots W_4$, after zeroing the display of the unloaded instrument, and with 2 calibration weights, W_5, W_6, with tare balancing between

Table 3.1. Performance Calibration, Single-Range Balances

Measurement No. (i)	*Tare Load*	*Calibration Load* (W_i)	*Displayed Net Value* (I_i)	*Error* (ΔI_i)
Zero setting	0 g	0 g	0.000 g	
1	0 g	100 g	100.002 g	+0.002 g
2	0 g	200 g	200.003 g	+0.003 g
3	0 g	300 g	300.003 g	+0.003 g
4	0 g	400 g	400.001 g	+0.001 g
Tare balancing	200 g	0 g	0.000 g	
5	200 g	100 g	100.000 g	0.000 g
6	200 g	200 g	199.998 g	−0.002 g

0.25 *Max* and 0.5 *Max*. W_5, W_6 are normally the same weights as W_1, W_2.

During calibration the ideal displayed values would be $I_i = W_i$, and the actual errors of measurement are

$$\Delta I_i = I_i - W_i \tag{5}$$

If the scale interval d of the balance is larger than the maximum permissible error of the calibration weight, W_i can be set to the nominal weight value; otherwise, the conventional mass value should be used. The relative error of the measurements is

$$\Delta I_{ir} = \frac{\Delta I_i}{W_i} \tag{6}$$

The mean relative error is

$$\overline{\Delta I_r} = \left(\frac{1}{6}\right)\left(\sum_{i=1}^{6} \frac{\Delta I_i}{W_i}\right) \tag{7}$$

Because the performance error from the ideal displayed value $I - W$ is not referred to for the correction, it must be included in the uncertainty calculation for unknown loads.

$$\Delta p = \left(\overline{\Delta I_r}\right)(I) \tag{8}$$

The variance of the relative performance errors is

$$v_p = \left(\frac{1}{5}\right)\left[\sum_{i=1}^{6}\left(\Delta I_{ir} - \overline{\Delta I_r}\right)^2 \times I^2\right] \tag{9}$$

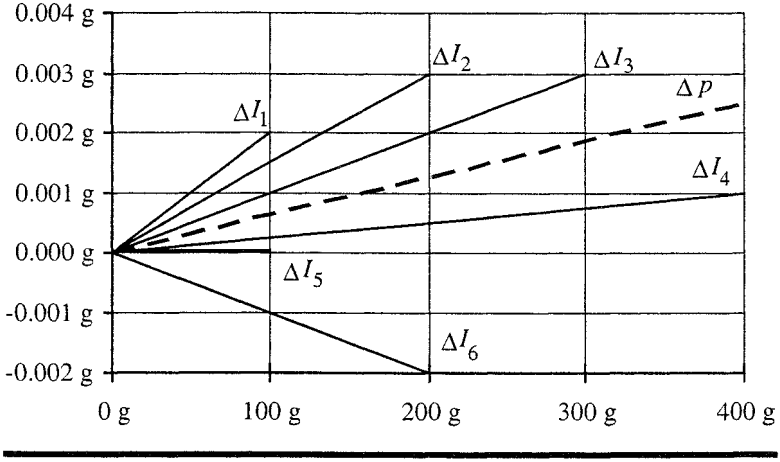
Figure 3.1 illustrates the results. Please notice that the horizontal axis is scaled by displayed net values, not gross load values.

Eccentric Load

A user will normally place the containers and goods on the pan in such a manner to obey the centre of gravity of the balance. But sometimes, especially if the weighing object is of larger dimensions, their centre of gravity will be off the centre of the pan. Therefore, an amount for this possible influence must be determined.

OIML R 76 gives good advice for the value of the test weight and the positions where to load the weight. A weight of $W_e \approx Max/3$ is used to determine the difference Δ in weight value between

Figure 3.1. Performance Measurement Errors and Mean Error for Displayed Net Values



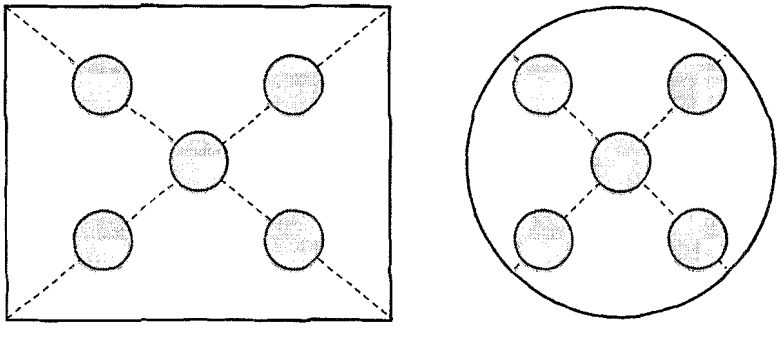
eccentric loads on four different positions and a centered load. The recommended positions on different pan shapes are shown in Figure 3.2.

The easiest way is first to load the test weight in the centre of the pan and tare. Then the weight is moved to the first off-centre position, and the first difference ΔI_1 is obtained from the display. The weight is then placed on the other three positions, and the other three differences, ΔI_2 to ΔI_4, can be read. Because of the small number of measurements, the variance is determined based on the highest absolute value:

$$v_e = \left(\frac{1}{3}\right)\left(\frac{|\Delta I|_{\max}}{Max}\right)^2 \times I^2 \qquad (10)$$

which represents a rectangular distribution of $\pm\Delta I_{\max}$.

Figure 3.2. Positions of Test Weight for Eccentricity Test



Calibration Weight

During the performance/linearity test, the weight W_4 will effect the calibration result more than the other three weights if all weights are of the same class, which, of course, is recommended. If the conventional mass values have been used for the calculation of the errors ΔI_i, the variance is determined from the uncertainty U of the calibration weight W_4:

$$v_w = \left(\frac{U}{k} \times \frac{1}{W_4}\right)^2 \times I^2 \qquad (11)$$

In accordance with OIML R 111, the coverage factor is $k = 2$.

If nominal mass values have been used, the variance is determined from the maximum permissible error of the calibration weight W_4:

$$v_w = \left(\frac{1}{3}\right)\left(\frac{\delta_m}{W_4}\right)^2 \times I^2 \qquad (12)$$

which represents a rectangular distribution of $\pm\delta_m$.

Temperature Influence

Manufacturers of weighing instruments specify the limits for changes in displayed values when the ambient temperature changes. This temperature coefficient, TC, is usually given in ppm/K. These changes can, of course, be compensated by

adjusting the weighing instrument. Many modern balances contain motor-operated internal weights that allow very easy span adjustment; some of them adjust automatically. Thus, only the maximum variation in temperature $\Delta T = T_{max} - T_{min}$ between consecutive adjustments is relevant. Note: This does not cover any time dependent effect. The variance is estimated according to

$$v_t = \left(\frac{1}{12}\right)(\Delta T \times TC)^2 \times I^2 \qquad (13)$$

which represents a rectangular distribution of $\Delta T = T_{max} - T_{min}$.

Single-Range Balances with High Resolution

Single-range balances with high resolution are characterised by the fact that the uncertainties of the test weights $W_1 \ldots W_4$ are larger than the scale interval d. Most analytical balances with scale intervals $d \leq 0.1$ mg belong to this category. It is obvious, therefore, that the accuracy of all weights used for a performance test will affect the uncertainty, and the calculation procedure must cover it.

Repeatability

Use the same procedure as given for single-range balances.

Rounding Effect

Use the same procedure as given for single-range balances.

Performance/Linearity Errors

To make things easier, another method has been developed and uses only two different test weights, W_1 and W_2. The weights normally are of class E_2. The scheme is shown in Table 3.2 with

Table 3.2. Performance Calibration, Single-Range Balances, High Resolution

Measurement No. (i)	*Tare Load*	*Calibration Load* (W_i)	*Displayed Net Value* (I_i)	*Error* (ΔI_i)
Zero setting	0 g	0 g	0.000 g	
1	0 g	400 g	399.9999 g	−0.0001 g
2	0 g	100 g	100.0002 g	+0.0002 g
Tare balancing	100 g	0 g	0.000 g	
3	100 g	100 g	100.0001 g	+0.0001 g
Tare balancing	200 g	0 g	0.000 g	
4	200 g	100 g	99.9999 g	−0.0001 g
Tare balancing	300 g	0 g	0.000 g	
5	300 g	100 g	99.9997 g	−0.0003 g
Tare balancing	100 g	0 g	0.000 g	
6	100 g	100 g	100.0002 g	+0.0002 g

another 400 g weighing instrument with a scale interval of 0.1 mg, which represents a resolution of four million digits.

After the unloaded balance has been set to zero, the first test weight $W_1 \approx Max$ is placed on the pan, and the first observation I_1 is obtained. Then this weight is removed from the pan, and the display is zeroed again. Now another test weight $W_2 \approx Max/4$ is placed on the pan, and the second observation I_2 is obtained. Then this test weight is removed, and a tare load $T_1 \approx Max/4$ is placed on the pan, and the display is set to zero by pressing the tare button. Again, the second test weight W_2 is placed on the pan, and the third observation I_3 is obtained. The same procedure is performed with two other tare weights, thus obtaining a fourth and fifth observation. At least the third measurement will be repeated and a sixth observation obtained. The sixth measurement represents the fact that such balances are more often used with small tare containers than with heavy ones.

Attention: It is very important that always the same weight piece is used as the second test weight W_2 and that it is not interchanged with the tare weight T_1 which normally is of the same nominal value. The tare weights can be of class F_1.

The actual errors of measurement are

$$\Delta I_1 = I_1 - W_1 \tag{14}$$

$$\Delta I_i = I_i - W_2 \quad \text{for } 2 \leq i \leq 6 \tag{15}$$

where W_1 and W_2 are the conventional mass values of the weights. The relative errors of the measurements are

$$\Delta I_{ir} = \frac{\Delta I_i}{W_i} \tag{16}$$

The variance and the mean performance error are obtained with equations 7, 8, and 9.

$$\overline{\Delta I_r} = \left(\frac{1}{6}\right)\left(\sum_{i=1}^{6} \frac{\Delta I_i}{W_i}\right) \tag{7}$$

$$\Delta p = \left(\overline{\Delta I_r}\right)(I) \tag{8}$$

$$v_p = \left(\frac{1}{5}\right)\left[\sum_{i=1}^{6}\left(\Delta I_{ir} - \overline{\Delta I_r}\right)^2 \times I^2\right] \tag{9}$$

Eccentric Load

Use the same procedure as given for single-range balances.

Calibration Weight

OIML R 111 shows the maximum permissible errors and uncertainties for test weights. The relative uncertainties $U^* = U/W$ for weights $W \geq 100$ g are constant in each class; for smaller weights, the relative uncertainties U^* are increasing. Therefore, the smaller test weight W_2 will normally influence the uncertainty of the

performance test more than the larger test weight W_1. The variance is determined from the uncertainty U of the test weight W_2:

$$v_w = \left(\frac{U}{k} \times \frac{1}{W_2}\right)^2 \times I^2 \qquad (17)$$

In accordance with OIML R 111, the coverage factor is $k = 2$.

Temperature Influence

Use the same procedure as given for single-range balances.

Multirange Balances

Multirange balances normally have two weighing ranges, with different scale intervals and different maximum loads. The displayed scale interval depends on the gross load on the pan. Normally, the two scale intervals have a ratio of $d_1/d_2 = 1/10$. There are some balances in the market that switch automatically from fine resolution to coarse resolution when the total load on the pan becomes larger than the maximum load limit of the first range, and they remain with the coarse scale interval even when the load is decreased to values lower than the maximum load limit of the first range.

Multirange balances should be calibrated in each range in like manner to single-range balances or single-range balances with high resolution.

Multi-Interval Balances

Multi-interval balances are characterised by the fact that the scale interval d is not constant but depends on the displayed weighing result. Up to a maximum load Max_1 in first weighing range, the display shows the smallest scale interval d_1. For larger indications, the scale interval changes to a larger value. There are balances with up to four different weighing ranges; the ratio between the different scale intervals is d_1/d_i = 1/2 or 1/2/5 or 1/2/5/10 or 1/10.

Another important fact is that after tare balancing, the zero point for all ranges is set to the value of the tare load. So the value of the scale interval depends only on the displayed net value.

When these balances are calibrated according to the procedure as described for single-range balances, the weighing errors are rounded to the adjacent scale interval. Especially in ranges with large scale intervals, the errors often will be rounded down to zero, which will lead to uncertainties with a confidence level less than the required 95 percent. The solution is to calibrate these balances in nearly the same manner as for single-range balances with high resolution.

Repeatability

Use the same procedure as given for single-range balances. The test weight W should be chosen less than Max_1.

Rounding Effect

Use the same procedure as given for single-range balances.

Performance/Linearity Errors

Use the same procedure as given for single-range balances with high resolution. The test weight W_2 should be chosen between $0.7\ Max_1 \leq W_2 \leq Max_1$. If the smallest scale interval d_1 of the balance is larger than the maximum permissible error of the weight W_2, it can be set to the nominal weight value; otherwise, the conventional mass value should be used.

Eccentric Load

Use the same procedure as given for single-range balances.

Calibration Weight

Use the same procedure as given for single-range balances, except instead of weight W_4, the weight W_2 must be inserted in the equations. If the conventional mass value has been used, the variance is determined from the uncertainty U of the calibration weight W_2:

$$v_w = \left(\frac{U}{k} \times \frac{1}{W_2}\right)^2 \times I^2 \qquad (17)$$

In accordance with OIML R 111, the coverage factor is $k = 2$.

If nominal mass values have been used, the variance is determined from the maximum permissible error δ_m of the calibration weight W_2:

$$v_w = \left(\frac{1}{3}\right)\left(\frac{\delta_m}{W_4}\right)^2 \times I^2 \qquad (12)$$

Temperature Influence

Use the same procedure as given for single-range balances.

Calculation of the Uncertainty

The ISO Guide to the Expression of Uncertainty in Measurements recommends giving an estimate for the expanded uncertainty. The uncertainty is yielded by the sum of the variances and the mean deviation of the characteristic curve as follows:

$$U = k \times \sqrt{s^2 + v_r + v_p + v_e + v_w + v_t} + |\Delta_p| \qquad (18)$$

In accordance with the ISO Guide to the Expression of Uncertainty in Measurements, the coverage factor again should be chosen to be $k = 2$.

To simplify application for the user, this function can be approximated by a linear equation that connects the initial value for the zeroed weighing instrument with the final value for the weighing instrument loaded to maximum capacity.

$$U \approx U_0 + \frac{U_{Max} - U_0}{Max} \times I \tag{19}$$

$$U_0 = U(I = 0) \tag{20}$$

$$U_{Max} = U(I = Max) \tag{21}$$

Single-Range Weighing Instruments (Example)

Uncertainty was determined for a calibrated 400 g weighing instrument with standard resolution. Calculations were made based on the following values:

Max = 400 g (from the data sheet)

d = 0.001 g (from the data sheet)

s = 0.0004 g (repeatability)

$|\Delta|_{max}$ = 0.002 g (eccentric load)

$\delta_m(W_4)$ = 0.0006 g (calibration weight 2×200 g, class E_2, less than d)

TC = 2 ppm/K (from the data sheet)

ΔT = 5 K (temperature variation at the place of installation)

These values lead to the following uncertainty components:

(1 ppm = 1 mg/1,000 g)

$s^2 = 0.16\ mg^2$

$v_r = 0.16\ mg^2$

ΔI_{ir} = 20, 15, 10, 2.5, 0, –10 ppm

$\overline{\Delta I_r}$ = 6.25 ppm

$$v_p = 190\ \text{ppm}^2 \times I^2$$

$$v_e = 8\ \text{ppm}^2 \times I^2$$

$$v_w = 8\ \text{ppm}^2 \times I^2$$

$$v_t = 100\ \text{ppm}^2 \times I^2$$

$$U = 2 \times \sqrt{0.32\ \text{mg}^2 + \left(306\ \text{ppm}^2 \times I^2\right)} + 6.25\ \text{ppm} \times I$$

$$U_0 = 2 \times \sqrt{0.32\ \text{mg}^2} \approx 1.2\ \text{mg}$$

$$U_{Max} = 2 \times \sqrt{0.32\ \text{mg}^2 + \left(306 \times 0.16\ \text{mg}^2\right)} + 6.25 \times 0.4 = 11.9\ \text{mg}$$

$$U \approx 1.2\ \text{mg} + \left(\frac{11.9 - 1.2\ \text{mg}}{400\ \text{g}}\right) \times I$$

$$U \approx 1.2\ \text{mg} + 0.027\ \text{mg/g} \times I$$

Table 3.3 shows the calculated values; Figure 3.3 shows a graphic illustration of the results. A comparison with the values using the approximating equation shows an acceptable degree of agreement.

Single-Range Balances with High Resolution

Equations (18) through (20) can be used in the same manner.

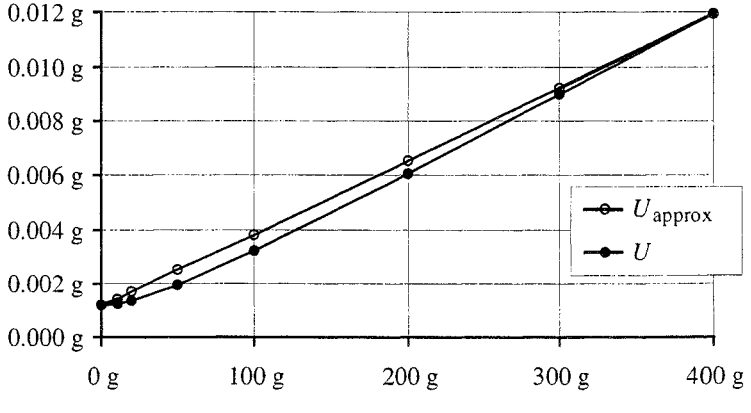
Multirange Balances

Equations (18) through (20) can be used in the same manner in each range.

Table 3.3. Uncertainty of Displayed Net Values

Displayed Net Values (I)	U	U_{approx}
0.000 g	0.0012 g	0.0012 g
10.000 g	0.0012 g	0.0015 g
20.000 g	0.0014 g	0.0017 g
50.000 g	0.0020 g	0.0026 g
100.000 g	0.0032 g	0.0039 g
200.000 g	0.0061 g	0.0066 g
300.000 g	0.0090 g	0.0093 g
400.000 g	0.0119 g	0.0119 g

Figure 3.3. Uncertainty of Displayed Net Values



Multi-Interval Balances

As the scale interval is not constant in the total weighing range, the equation (18) must be solved for each partial weighing range. Also, the approximations in equations (19) through (21) have to be performed in each partial weighing range. The experience of many calibrations has shown that it is sufficient to separate only the ranges from zero to Max_1 and Max_1 to Max. Thus, equations (19) through (21) must be altered as follows:

In the range $I \leq Max_1$,

$$U \approx U_0 + \frac{U_1 - U_0}{Max_1} \times I \tag{22}$$

$$U_0 = U(I = 0; d = d_1) \tag{23}$$

$$U_1 = U(I = Max_1; d = d_1) \tag{24}$$

In the range $I > Max_1$,

$$U \approx U_2 + \frac{U_{Max} - U_2}{Max} \times (I - Max_1) \tag{25}$$

$$U_2 = U(I = Max_1; d = d_2) \tag{26}$$

$$U_{Max} = U(I = Max; d = d_{Max}) \tag{27}$$

Uncertainty in Practical Use

As described before there are additional influences on the accuracy of weighing results which are not covered by the procedures described here. The three most important ones are as follows:

1. Rough handling by untrained persons causing

   –slight unvisible damages of the weighing system and/or

   –worse repeatability

2. Changes with time

3. Air buoyancy effects at desired higher accuracies for mass determination

The effects of points 1 and 2 can be evaluated if the balance is calibrated at different times and by different persons without previous adjusting. The calibrations will lead to different uncertainty results. A good estimate will be to take the worse result for the future.

Air buoyancy effects are systematic ones, and the related errors should be corrected for if necessary. The errors can be evaluated from the influencing parameters—the density of the weighed material and the density of the air. For materials with a density of 1 g/cm^3, the error at standard air density of 1.2 kg/m^3 is about 0.1 percent; for materials with higher densities, the errors are less; for materials with lower densities, the errors are higher.

Summary

The advantages of the procedures described in this chapter are as follows:

- They are based on the actual use of the weighing instrument being calibrated.
- They can be used for weighing operations that require tare balancing.
- Ambient conditions at the place of installation are taken into consideration.
- The uncertainty of measurement can easily be read from the diagram.

Reference

ISO Guide to the Expression of Uncertainty in Measurements. Geneva, Switzerland: International Organisation for Standardisation.

Chapter 4

SOFTWARE VALIDATION

B. A. Wichmann
National Physical Laboratory
Teddington, Middlesex, United Kingdom

Software is playing an increasingly important part in scientific instruments, and hence there is a need to ensure that the quality of the software is appropriate for the instrument and its application. This chapter provides a means for suppliers and users to assure themselves of the quality of the software.

Almost all of the current generation of scientific instruments contain a significant amount of software. Since it is hard to quantify the

reliability or quality of such software, two questions immediately arise:

1. As a user of such an instrument, how can I be assured that the software is of a sufficient standard to justify its use?

2. As a developer of such software, what development techniques should I use, and how can I assure my customers of the quality of the resulting software?

This good practice guide on software validation addresses these two questions. The intended readers are those responsible for software in scientific instruments and those using such software. Software written as a research project or to demonstrate the feasibility of a new form of measurement is excluded from the scope of this guide. The more specialised area of Programmable Logic Controllers is also excluded, since another guide is available covering this [1].

The chapter surveys current Good Practice in software engineering and relates this practice to applications involving scientific instruments. Known pitfalls are illustrated, along with suggested means of avoiding them. The general approach is a three-stage process as follows:

1. A risk assessment based on a model of an instrument with its software

2. An assessment of integrity required on the software, based upon the risk assessment (the Software Integrity Level)

3. Guidance on the software engineering methods to be employed (determined by the Software Integrity Level)

It must be emphasised that there is no simple universal method (silver bullet) for producing correct software, and, therefore, skill, sound technical judgement and care are required. Moreover, if it is essential for the quality of the software to be demonstrated to a third party, then convincing evidence is needed which should be planned as an integral part of the software development process.

Some check-lists are provided to aid in applying the guidelines in this chapter. To avoid complexity in the wording, it is assumed that the software is already in existence. It is clear that use could be made of the guide during the development, but it is left to the reader to formulate its use in that context. No consideration has been given here of the possibility of standardising this material or obtaining some formal status for it.

Software Use

The use of software either within or in conjunction with a scientific instrument can provide additional functions in a very cost-effective manner. Moreover, some instruments cannot function without software. Hence, it is not surprising that there is an exponential growth of software in this

area. Unfortunately, these changes can give rise to problems in ensuring that the software is of an appropriate standard. The problem with software is largely one of unexpected complexity. Software embedded with an instrument could be inside just one ROM (read only memory) chip and yet consist of 1 Mb of software. Software of such size is well beyond that for which one can attain virtually 100 percent confidence. This implies that one has to accept that there is a possibility of errors occurring in such software.

An area in which there has been a substantial effort to remove software errors is in safety applications; hence the methods used and specified in safety-critical systems are used here as an indication of what might be achievable, typically at a significant cost. For general advice in this area, see Safety-Related Systems [2].

An example area in which very high standards are required in software production is that for airborne flight-critical software. The costs for producing such software can easily be £500 per machine instruction—obviously too demanding for almost all software within scientific instruments. Hence the main objective behind this guide is to strike a balance between development cost and the proven quality of the software. The main approach taken here is one of *risk assessment* as a means of determining the most appropriate level of rigour (and cost) that should be applied in a specific context.

A major problem to be faced with software is that the failure modes are quite different than with a simple instrument without software. An example of this is that of the non-linear behaviour of software in contrast to simple measuring devices (see Appendix B1).

Requirements

There are a number of standards that specify requirements for software in scientific instruments which are collected here with an indication of the issues to be covered by this guide. The more relevant standards appear first.

ISO/IEC Guide 25

ISO/IEC Guide 25 [3] is equivalent to M10 (see below). Paragraph 10.6 states:

> Calculations and data transfers shall be subject to appropriate checks.

Paragraph 10.7 states:

> Where computers or automated equipment are used for the capture, processing, manipulation, recording, reporting, storage or retrieval of calibration or test data, the laboratory shall ensure that:
>
> 1. the requirements of this Guide [3] are complied with;

2. computer software is documented and adequate for use;
3. procedures are established and implemented for protecting the integrity of data; such procedures shall include, but not be limited to, integrity of data entry or capture, data storage, data transmission and data processing;
4. computer and automated equipment is maintained to ensure proper functioning and provided with the environmental and operating conditions necessary to maintain the integrity of calibration and test data;
5. it [the laboratory] establishes and implements appropriate procedures for the maintenance of security of data including the prevention of unauthorized access to, and the unauthorized amendment of, computer records.

ISO/IEC Guide 25 gives the most detailed indication of the requirements and is the most useful for both development and assessment.

EN45001

EN 45001 [4] is the European equivalent to ISO/IEC Guide 25. Section 5.4.1 states that:

> All calculation and data transfers shall be subject to appropriate checks. Where results are derived by electronic data processing techniques, the reliability and stability of the system shall be such that the accuracy of the

> results is not affected. The system shall be able to detect malfunctions during programme execution and take appropriate action.

EN 45001 gives a different gloss on the same area. Here, (numerical) stability and reliability are mentioned, but security and integrity are not. Again, following this guide should ensure compliance with this standard.

M10

M10 is the UK laboratory accreditation standard [5]. Section 8.6 states:

> The Laboratory shall establish procedures when using computer data processing to ensure that the collection, entry, processing, storage, or transmission of calibration or test data is in accordance with the requirements of this Standard.

Section 8.7 states:

> Calculations and data transfers shall be subject to appropriate checks.

In practice, these requirements are interpreted by the United Kingdom Accreditation Service (UKAS) assessor. It is hoped that this guide will aid in this interpretation and reduce the risk of the assessor taking a different view from the laboratory.

Weighing Machines

The WELMEC document [6] summarises the position for weighing machines. The requirements here derive from EU Directive 90/384/EEC, which has three relevant parts as follows:

> Annex I, No. 8.1: Design and construction of the instruments shall be such that the instruments will preserve their metrological qualities when properly used and installed, and when used in an environment for which they are intended. . . .
>
> Annex I, No. 8.5: The instruments shall have no characteristics likely to facilitate fraudulent use, whereas possibilities for unintentional misuse shall be minimal. Components that may not be dismantled or adjusted by the user shall be secured against such actions.
>
> Annex II, No. 1.7: The applicant shall keep the notified body that has issued the EC type-approval certificate informed of any modification to the approved type. . . .

Clearly, only part of these requirements is relevant to the software of instruments in general. The WELMEC Guide derives three specific requirements for software from the above directives as follows:

> Section 3.1: The legally relevant software shall be protected against intentional changes with common software tools.
>
> Section 3.2: Interfaces between the legally relevant software and the software parts not subject to legal control shall be protective.

> Section 3.3: There must be a software identification, comprising the legally relevant program parts and parameters, which is capable of being confirmed at verification.

In the context of instruments not within the ambit of legal requirements, there are two important principles to be noted from the above:

1. The handling of the basic measurement data should be of demonstrably high integrity.

2. The software should be properly identified. (This arises from configuration control with ISO 9001 [7], but there is no requirement that ISO 9001 be applied to such machines.)

IEC 601-1-4

IEC 601-1-4 covers the software in medical devices [8] and is used in Europe to support a directive and by the U.S. Food and Drug Administration (FDA) [9]. The standard is based on risk assessment with software engineering based on ISO 9000-3 [10]. The flavour of the standard can be judged from a few key extracts given below, as those relevant here:

> Section 52.204.3.1.2: Hazards shall be identified for all reasonably foreseeable circumstances including: normal use; incorrect use.
>
> Section 52.204.3.1.6: Matters considered shall include, as appropriate: compatibility of

system components, including hardware and software; user interface, including command language, warning and error messages; accuracy of translation of text used in the user interface and "instructions for use"; data protection from human intentional or unintentional causes; risk/benefit criteria; third party software.

Section 52.204.4.4: Risk control methods shall be directed at the cause of the hazard (e.g. by reducing its likelihood) or by introducing protective measures which operate when the cause of the hazard is present, or both, using the following priority: inherent safe design; protective measures including alarms; adequate user information on the residual risk.

Section 52.207.3: Where appropriate the specification shall include requirements for: allocation of risk control measures to subsystems and components of the system; redundancy; diversity; failure rates and modes of components; diagnostic coverage; common cause failures; systematic failures; test interval and duration; maintainability; protection from human intentional or unintentional causes.

Section 52.208.2: Where appropriate, requirements shall be specified for: software development methods; electronic hardware; computer aided software engineering (CASE) tools; sensors; human-system interface; energy sources; environmental conditions; programming language; third party software.

It can be seen that this standard is mainly system-oriented and does not have very much to say

about the software issues. However, the key message is that the level of criticality of the software must be assessed, and the best engineering solution may well be to ensure the software is not very critical. This standard covers only instruments which are on-line to the patient as opposed, for example, to those used to analyse specimens from a patient. Not all medical applications could be regarded as "scientific instruments", and, therefore, the relevance of this guide needs to be considered.

DO-178B

DO-178B is the civil avionics safety-critical software standard [11]. It is not directly relevant. However, if an instrument were flight-critical, then any software contained within it would need to comply with this standard. In practice, instruments are replicated using diverse technologies and hence are not often flight-critical. This standard is very comprehensive and specifies an entire software development process, including details on the exact amount of testing to be applied.

The conclusion for this guide is that this standard is only relevant for very high-risk contexts, in which it is thought appropriate to apply the most demanding software engineering techniques. This standard can be taken as an ideal goal, not achievable in practice, due to resource constraints.

IEC 61508 (draft)

IEC 61508 [12] is a generic standard for safety-critical applications. In contrast to DO-178B IEC 61508 allows for many methods of compliance. A catalogue is provided in Part 3 of the standard which handles the software issues. This catalogue is used here as a means of selecting specific methods that may be appropriate in some contexts. This generic standard is less demanding than DO-178B and hence could be applied without the same demands on resources. Guidelines for use within the motor industry have been developed from this standard [13].

The conclusion for this guide is that IEC 61508 is not directly relevant but could be applied in specific contexts. The catalogue of techniques provides a reference point to a wide variety of software engineering methods. For an analysis of this standard for accreditation and certification, UKAS has a feasibility study [14].

A very informative report [15], which has many parallels to this one, undertakes an assessment of devices which could be an instrument based on an early draft of IEC 1508 (now 61508).

For a detailed research study of assessing instruments for safety application by means of a worked example, see the SMART reliability study [16]. This study was based on (an earlier edition of) this guide, but enhanced to reflect the safety requirements. The issue of the

qualification of SMART instruments in safety applications is noted as a research topic in a Health and Safety Commission report [17].

Conclusion

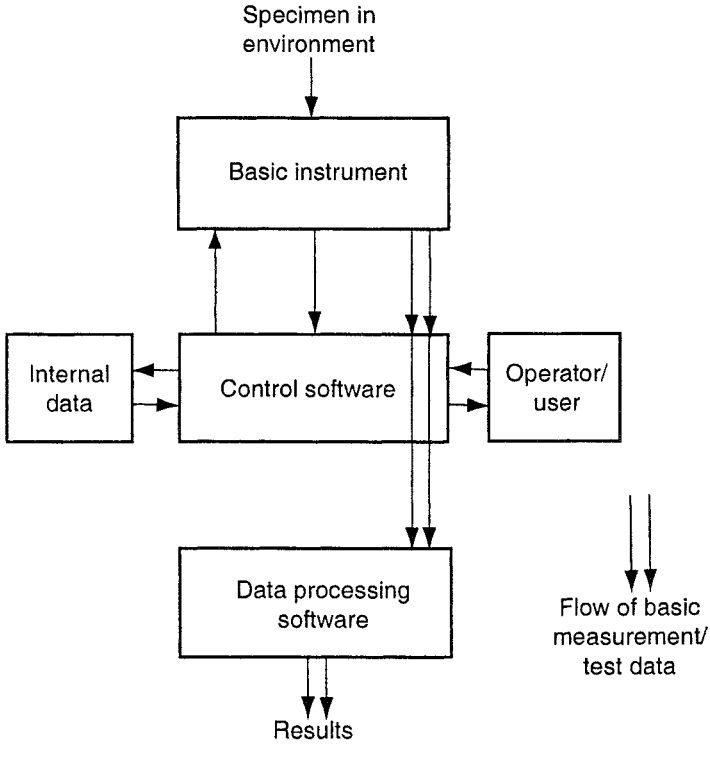
The conclusions from this survey of the standards are that they are broadly similar and aiming to meet all the requirements is a reasonable way of proceeding. One exception to this is that the very demanding requirements in DO-178B cannot be realistically merged with the others. Hence, if an instrument is required to meet the most demanding levels of DO-178B, then that cannot be expected of an instrument designed to satisfy the other standards mentioned here. Thus, this guide aims to provide advice on producing software which will satisfy any of these standards, with the exception of DO-178B (levels A and B).

An Instrument Model

In order to provide a framework for the discussion of the software for an instrument, a simple model is presented here. The components of the model in Figure 4.1 are as follows:

- *Basic instrument:* The basic instrument contains no software. It performs functions according to the control logic and provides output. This instrument itself is

Figure 4.1. Software Model for a Scientific Instrument



outside the scope of this guide, but it is essential that the properties of the instrument are understood in order to undertake an effective appraisal of the software. The basic instrument contains sensors and appropriate analogue/digital converters.

- *Control software:* This software processes output from the basic instrument for the purpose of undertaking control actions.

- *Data processing software:* This software performs a series of calculations on data from the basic instrument, perhaps in conjunction with the control software, to produce the main output from the instrument. (In the case of complex instruments, like coordinate measuring machines, the data processing software provided can include a programming language to facilitate complex and automatic control.)

- *Internal data:* These data are held internally to the instrument. A typical example would be calibration data. Another example might be a clock which could be used to "time-out" a calibration.

- *Operator/user:* In some cases, there is an extended dialogue with the user, which implies that the control function can be quite complex. This dialogue could be automated via some additional software (which therefore could be included or excluded from this model).

In this model, we are not concerned about the location of the software. For instance, the control software could be embedded within the instrument, but the data processing software could be in a personal computer (PC) or workstation. It is even possible for the subsequent data processing to involve several computers via a Laboratory Information Management System (LIMS). In applying this guide, you may have a choice in deciding

where to draw the line at the bottom of this diagram. For instance, one could decide to include or exclude a LIMS. If the LIMS is considered, then Fink et al. [18] on a medical application provides some useful insights. The integrity of the production of the test/measurement certificate should not be forgotten [19].

The basic measurement/test data is a key element in this structure. The major requirement is to show that the processing and entire flow of this data has suitable integrity. Note that in the area of legal metrology, the basic measurement data are converted into money (say, in a petrol pump), and this, therefore, has the same status as the basic measurement data.

Risk Factors

This section undertakes an analysis of a scientific instrument and its related software, the purpose of which is to make an objective assessment of the likely risks associated with a software error. For a general discussion on risk, see *Guidelines on Risk Issues* [20].

The first step in undertaking the risk assessment is to characterise the instrument according to aspects which influence the risk. These aspects are discussed below.

Criticality of Usage

It is clear that the usage of some instruments is more critical than others. For instance, a medical instrument could be critical to the well-being of a patient. On the other hand, a device to measure noise intensity is probably less critical.

To make an objective assessment of the level of criticality of usage, we need a scale for increasing criticality: critical, business-critical, potentially life-critical and life-critical. One of the major problems in this area is that a supplier may well be unaware of the criticality of the application. The user may well assume that the instrument is suitable for highly critical applications, while the supplier may well prefer to exclude such usage. For an example of this problem, see Appendix B4.

Legal Requirements

Several instruments are used in contexts for which there are specific legal requirements, such as with the WELMEC guide [6]. In this context, an instrument malfunction could have serious consequences. To make an assessment, one clearly needs to know if there are any specific legal requirements for the instrument and have a reference to these requirements. (It may be necessary to check what current legislation applies.)

Complexity of Control

The control function of the software can range from being almost non-existent to having substantial complexity. Aspects of the control will be visible to the operator in those cases in which operator options are available. Some control may be invisible, such as a built-in test function to help detect any hardware malfunction.

Many aspects of control are to make the device simpler to use and protect the operator against misuse which might be feasible otherwise. This type of control is clearly highly advantageous, but it may be unclear if any error in its operating software could produce a false reading. Hence aspects of the complexity of the user-instrument interface are considered here.

Very Simple

An example of a very simple control is when the instrument detects if there is a specimen in place, either by means of a separate detector or from the basic data measurement reading. The result of this detection is to produce a more helpful display readout.

Simple

An example of a simple control might be temperature control which is undertaken so that temperature variation cannot affect the basic measurement data.

Modest

An example of modest complexity arises if the instrument takes the operator through a number of stages, ensuring that each stage is satisfactorily complete before the next is started. This control can have an indirect effect on the basic test/measurement data, or a software error could have a significant effect on that data.

Complex

An example of complex control is when the software contributes directly to the functionality of the instrument. For instance, if the instrument moves the specimen, these movements are software controlled and have a direct bearing on the measurement/test results.

Complexity of Processing of Data

We are concerned with the processing of raw data from the basic instrument (i.e., the instrument without the software). In the case of software embedded within the instrument itself, the raw data may not be externally visible. This clearly presents a problem for any independent assessment; however, it should be the case that the nature of the raw data is clear and that the form of processing is well defined. Calibration during manufacture would typically allow for "raw data" to be displayed in appropriate units. (Subsequent to the calibration during manufacture, the raw data may not be available to the user.)

- *Very simple:* The processing is a linear transformation of the raw data only, with no adjustable calibration taking place.
- *Simple:* Simple non-linear correction terms can be applied here, together with the application of calibration data. A typical example is the application of a small quadratic correction term to a nearly linear instrument, which is undertaken to obtain a higher accuracy of measurement.
- *Modest:* Well-known published algorithms are applied to the raw data. The assumption here is that the algorithms used are numerically stable. For an example of a problem that can arise in this area, see Appendix B2.
- *Complex:* Anything else.

A risk assessment must be based on the visible information. Some complex devices may internally record information which is difficult or impossible to access. Examples of such information are the selection of operator options or low-level data within a complex system. For programmable instruments, it should be possible to reconstruct the program from a listing and repeat the execution from the data recorded from a prior execution.

INTEGRITY ASSESSMENT

We are now looking at the instrument as a black box but assume that some questions can be asked (and answered) which might not be directly apparent from the instrument. The underlying reasoning behind the questions is to assess the affects of the risk factors involved. If some key questions cannot be answered, then clearly any assessment is incomplete.

The first part is to characterise the instrument in terms of its risk factors. We now have a set of additional issues to resolve as follows:

- What degree of confidence can be obtained in the instrument merely by performing "end-to-end" tests, i.e., using the instrument with specimens of known characteristics? (Note that this type of testing is distinct from conventional black-box testing of the software because the software is only exercised in conjunction with the basic instrument.) Such tests regard the entire instrument as a black box and effectively ignore that software is involved. To answer this leading question, take into account the risk factors noted above. For instance, if complex software is being used which uses unpublished algorithms, then high confidence cannot be established.

- In the case in which the processing of the basic data is modest or complex, can the

raw data be extracted so that an independent check on the software can be applied?

- Has essentially the same software for the data processing been applied to a similar instrument for which reliability data is available? (Note that there is a degree of subjective judgement here, which implies that the question should be considered by someone who is suitably qualified.)
- For this instrument, is there a record available of all software errors located? Under what terms, if any, can this information be made available?
- To what extent can the complexity in the control software result in false measurement/test data being produced?
- If the operator interface is complex, can this be assessed against the documentation? How important is operator training in this?

At this point, sufficient information should be available to make an assessment of the integrity required of the software by taking into account all the factors above, including the target risk to be taken. This assessment should be as objective as possible but is bound to have a modest degree of subjectivity. If answering the questions above is straightforward and raises no problems, then the Software Integrity Level is the same as the

complexity of the data processing above. Hence, unless answering the above questions reveals additional problems, very simple complexity of data processing would have a Software Integrity Level of 1.

There are four levels of software integrity:

1. *Software Integrity Level 1:* The data processing software is very simple. No significant problems are revealed in the analysis.

2. *Software Integrity Level 2:* The data processing software is simple or some problems were encountered, and the processing is very simple.

3. *Software Integrity Level 3:* There is at least one major unquantifiable aspect to the software. This could be an inability to check the software since there is no facility to provide the raw data (combined with complex processing). Another possibility might be that the control software influences the basic measurement/test data in ways that cannot be quantified.

4. *Software Integrity Level 4:* Here we either have complex processing which is difficult to validate or processing of modest complexity with significant additional problems (or both!).

As the software integrity level increases, so should the risk of errors be reduced due to the

application of more rigorous software engineering techniques, which is the topic of the next section.

Computing the Software Integrity Level

It has been suggested that there should be an algorithm for computing the software integrity level from the information which has been requested previously. Each key factor is on a four-point scale, as is the resulting software integrity level. Hence one possibility is as follows:

$$\textit{Software Integrity Level} = \max(\textit{Usage, Control, Processing}) \textit{ levels}$$

This suggestion has not been developed further since it seems to be difficult to take into account all of the factors necessary. For instance, even the maximum function above is not quite correct, since the complexity in the control function could be offset by other factors.

On balance, it seems more appropriate for the factors to be determined, the check-lists used and a subjective judgement made (which could well be based on the formula above). The important aspect is to show how, and on what basis, the Software Integrity Level was determined. Some alternative methods are given in IEC 61508 [12].

Software Development Practice

The starting point is that the software development process being used should have a rigour to match the Software Integrity Level. This is the approach taken in several safety-critical software standards (IEC 61508 [12] and RTCA [11]).

For any Software Integrity Level, basic practices must be established, which could be a direct consequence of the application of ISO 9001 to software [7, 10] or from the application of other standards. It can be claimed that ISO 9001 itself requires the application of appropriate (unspecified) software engineering techniques, especially for the higher levels of integrity. However, even ISO 9000–3 does not even mention many such techniques, and hence we take the view that specific techniques should be recommended here, rather than depending on the general requirements of ISO 9001. In the pharmaceutical sector, specific guidance has been produced which is effectively a version of ISO 9000–3 oriented to that sector [21].

Theoretically, it is possible to undertake the testing of software to establish the actual reliability of the software. However, there are strict limits to what can be achieved in this area [22], and hence the approach taken here is the conventional one of examining the software development process. In practical terms, software testing is expensive, and hence the most cost-

effective solution uses other methods to gain confidence in the software.

In the United Kingdom, it is reasonable to expect suppliers to be registered to ISO 9001, which in the case of software implies the application of TickIT. If a supplier is *not* registered, then one lacks an independent audit on their quality system. ISO 9001 provides a basic standard for quality management, whereas in practice, companies will continually improve their system if the aim is high quality. In any case, improvements in the light of experience are essentially a requirement of ISO 9001. The standard implies a defined life cycle which is elaborated in ISO/IEC 12207 [23]. For those companies not formally registered to ISO 9001, we assume that a similar quality management approach is used.

The application of a quality management system to software should imply that a number of technical issues have been addressed and documented and that the following requirements are met:

- There should be documents demonstrating that a number of issues have been covered, such as design, test planning, acceptance, and so on. The acceptance testing should ensure that the operator interaction issues have been handled and validated.

- A detailed functional specification should exist. Such a specification should be sufficient to undertake the coding. This level of

information is typically confidential to the developer.

- There should be a fault reporting mechanism supported by appropriate means of repairing bugs in the software.

- The software should be under configuration control [24]. This implies that either the software should itself include the version number or the version can be derived from other information, such as the serial number of the device. In the case of free-standing software, it should be possible for users to determine the version number.

We assume that these requirements are met, whatever Software Integrity Level is to be addressed by the software.

For Software Integrity Level 1, the above requirements are recommended. For higher integrity levels, a recommendation for level n also applies to any higher level. In Table 4.1, we list the recommended techniques, which are all defined in Appendix C. Note that statement testing, equivalence partition testing, and structural testing are all (software) component test methods. The fact that a technique is recommended at a specific level does not (in general) imply that not applying the method would imply poor practice or that all the methods should be applied. For instance, the example given in Appendix C4, is a good choice precisely because other strong

Table 4.1. Software Engineering Techniques

Software Integrity Level	*Recommended Techniques*
2	Software inspection (C.1), Mathematical specification (C.7), Structural testing (C.2), System testing (C.5)
3	Regression testing (C.3), Equivalence partition testing (C.2), Independent audit (C.9), Numerical stability (C.6), Stress testing (C.10)
4	Statement testing (C.2), Formal specification (C.8), Static analysis (C.11), Accredited testing (C.4)

methods are not effective. Any design should involve a trade-off between the various relevant methods.

CONCLUSIONS

The attempt to answer the two questions posed at the beginning of this chapter is limited by the information available. One must accept that the user of an instrument may not be able to obtain information from the supplier to determine if appropriate software engineering practices have been used.

At this point, no consultation has taken place with instrument suppliers, so it is unclear if this guide reflects current practice, although it is

based on both established software engineering methods and information from appropriate standards. Many useful comments were obtained on a draft of this guide. Every attempt has been taken to reconcile the comments with this version.

ACKNOWLEDGEMENTS

Thanks to Bernard Chorley and Graeme Parkin for support and help with this project. Brian Bibb (Phillips Medical Instruments) provided some useful references, which greatly assisted this work. Peter Harris provided the material in sections on numerical issues (B.2, C.6 and C.12). Comments are acknowledged from Dr. Peter Vaswani (UKAS), Dr. John Wilson (BMTA), Brian Bibb (Elektra Oncology Systems Ltd), John Stockton, Bill Lyons (Claude Lyons Ltd), Dr. Richard Mellish (Medical Devices Agency) and Graeme Parkin.

Appendix 4.A
Check-Lists

The check-lists below ask key questions concerning the substance of this chapter. It is not usually possible to answer them with a simple yes or no. If any question is unclear, the main text of this chapter should be referred to. In some cases, the check-list questions do not relate to the larger issues discussed in the main text but to lesser issues which are known to have caused problems in the past.

4.A.1 Risk Factors

1. What is the criticality of usage? (critical/business-critical/potentially life-critical/life-critical)
2. Is the supplier aware of the criticality of your application?
3. Are there specific legal requirements for the instrument?
4. What are the consequences of an instrument malfunction?
5. Is independent evidence needed of the software development process?
6. Does the instrument require regulatory approval?
7. What is the complexity of control? (very simple/simple/modest/complex)

8. Does the instrument perform built-in testing?
9. Do the control functions protect the operator from making specific errors?
10. Can an error in the control software cause an error in the basic test/measurement data?
11. Is raw data available from the instrument?
12. Does the instrument contain local data, such as that derived from the last calibration?
13. Is the processing of raw data strictly linear?
14. Is the processing of raw data a simple non-linear correction?
15. Is the processing of data restricted to published algorithms?
16. Have the algorithms in use been checked for numerical stability?
17. Would a numerical error, such as division by zero, be detected by the software or would erroneous results be produced? This will typically depend on the programming system used to produce the software and can vary from no detection of such errors to elaborate indications of the exact point of failure. If no internal checks are applied, there is a greater risk of a programming error, resulting in erroneous results.

4.A.2 Integrity Assessment

1. What information is available from the instrument supplier or developer of the software?
2. What confidence can be gained in the software by end-to-end tests on the instrument?
3. Can raw data be extracted from the instrument?
4. Can raw data be processed independently from the instrument to give an independent check on the software?
5. Are software reliability figures available for an instrument using similar software (i.e., produced by the same supplier)?
6. Is a log available of all software errors? Has this log been inspected for serious errors?
7. Does the control function have a direct effect on the basic test/measurement data?
8. Has an assessment been made of the control software against the operating manual? If so, by whom?
9. Do operators of the instrument require formal training?
10. Have all the answers to the questions in this list been taken into account in determining the Software Integrity Level?
11. Has a list been made of all the unquantifiable aspects of the software?

4.A.3 Software Development Practice

These lists are increasing in complexity. Since the requirements at level n imply those at level $n - 1$, all the questions should be asked up to the level required.

4.A.3.1 Software Integrity Level 1

1. Is there design documentation?
2. Is there evidence of test planning?
3. Is there a test acceptance procedure?
4. Is there an error log?
5. What evidence is there of clearance of errors?
6. Is there a detailed functional specification?
7. Are security, usability and performance aspects covered in the specification?
8. How is configuration control managed?
9. Is there a defined life cycle?
10. How can the user determine the version number of the software?
11. Have all changes to hardware platform, operating system, compiler and added functionality been checked?
12. Have all corrections been checked according to the defined procedures?
13. Do all staff have the necessary skills and are these documented?

4.A.3.2 Software Integrity Level 2

1. Is software inspection used on the project? If so, to what documents has it been applied and what was the estimated remaining fault rate?
2. What alternatives have been used if software inspection was not applied?
3. Has a mathematical specification been produced of the main algorithms used for processing the test/measurement data?
4. Is the processing code derived directly from the mathematical specification?
5. What form of structural testing has been applied? What metrics of the level of testing have been produced?
6. What level of testing has been applied to the control and processing components of the software?
7. How has the completeness of system testing been assessed?
8. Has system testing covered all reasonable misuses of the instrument?
9. What records are available on system testing?
10. Are the security features consistent with any regulations or intended use?
11. Are the test strategies, cases and test completion criteria sufficient to determine that the software meets its requirements?

4.A.3.3 Software Integrity Level 3

1. Is regression testing applied? If so, at what point in development did it start?

2. For what components has equivalent partition testing been applied? Has the technique been applied to the components processing the basic test/measurement data?

3. Has an independent audit been undertaken? Have all problems identified been resolved? Did the audit apply to the basic quality management system and/or the software techniques?

4. Has the numerical stability of the main measurement/data processing routines been checked? Has the rounding error analysis been taken into account in formulating the mathematical specification of the routines?

5. Has stress testing been applied to the software? To what extent have the limits of the software been assessed by this testing? Has the stress testing revealed weaknesses in the system testing?

6. Have activities been undertaken in the development which are not auditable (say, no written records)?

7. Are known remaining software bugs documented? Are they adequately communicated to the user?

8. What methods have been applied to ensure that structural decay is avoided? (See Appendix B.3.)

4.A.3.4 Software Integrity Level 4

1. To what components has statement testing been applied? What coverage was obtained?
2. Has a formal specification been produced of any of the software components? Has this revealed weaknesses in the functional specification, testing and so on? Is it possible to derive an executable prototype from this specification to validate the equivalence partition testing?
3. What forms of static analysis have been undertaken?
4. Does accredited testing have a role in gaining confidence in the software? If a test suite is used for accredited testing, have all the results of other forms of testing been fed into this?
5. Is beta site testing undertaken?
6. Is memory utilization testing undertaken or can it be shown that such testing is not needed?

APPENDIX 4.B
SOME EXAMPLE PROBLEMS

A number of illustrative examples are collected here of problems that have been reported to NPL over a number of years. The exact sources are deliberately not given, even when they are known.

4.B.1 Software Is Non-linear

A simple measuring device was enhanced to have a digital display. This was controlled by an embedded microprocessor, with the code produced in assembler. The product was then subjected to an independent test. The testers discovered, almost by accident, that when the device should have displayed 10.000 exactly, the actual display was nonsense. The fault was traced to the use of the wrong relational operator in the machine code.

The example contrasts with pre-digital methods of recording measurements in which the record is necessarily linear (or very nearly linear). The example illustrates that the testing of software should include boundary conditions. However, only the most demanding standards actually *require* that such conditions be tested. For a four-digit display in this example, it should be possible to cycle through all the possible outputs to detect the error.

4.B.2 Numerical Instability

The repeatability standard deviation of a weighing balance was required as part of a reference material uncertainty estimate. Successive weighings of a nominal 50 g weight produced a set of 15 replicate values as follows: 49.9999 (1 occurrence), 50.0000 (5 occurrences), 50.0001 (8 occurrences) and 50.0002 (1 occurrence). The processing of the data used a built-in "standard deviation" function operating to single precision (8 significant figures). Because the data could be represented using 6 significant figures, the user anticipated no difficulties. The value returned by the function, however, was identically zero.

The reason is that the function implements a "one-pass" algorithm that, although fast to execute, is numerically unstable. The standard deviation computation in this algorithm is based on a formula involving the difference between quantities which are very close for the above data values, thus causing the loss of many figures. An alternative "two-pass" algorithm that first centres the data about the arithmetic mean and then calculates the standard deviation returns an answer for the above data that is correct to all figures expected. Unfortunately, the "one-pass" algorithm is widespread in its use in pocket calculators and spreadsheet software packages.

The example described above is concerned with the stability of the algorithm chosen for the required data processing. Numerical difficulties

may also arise from the improper application of good algorithms. In one example, the processing software was to be transferred from one (mainframe) platform to another (PC) platform.

Although the platforms operated to similar precisions, and the same numerically stable algorithm was used (albeit coded in different languages), the results of the processing agreed to only a small number of significant figures. The linear systems being solved were badly scaled and, therefore, inherently ill-conditioned, i.e., the solution unnecessarily depended in a very sensitive way on the problem data.

The lesson here is to ensure the required data processing is stated as a well-posed problem; then use a stable algorithm to solve the problem.

4.B.3 Structural Decay

A contractor is used to develop some software. The contractor has very high coding standards which include writing detailed flow diagrams for the software before the coding is undertaken. The contractor corrects these diagrams to reflect the actual code before delivery to the customer. It is satisfactory to use flow charts to generate a program. But once the program is written, these charts become history (or fiction), and only charts generated from the program source are trustworthy. The customer has tight deadlines on performing modifications to the software over the next five years. For the first

two amendments, the flow diagrams were carefully updated to reflect the changes to the code, but after that, no changes were made, making the flow diagrams effectively useless. As a result, the overall "design" provided by the contractor was effectively lost. The problem was that the "design" was not captured in a form that could be easily maintained.

The conclusion from this is that for programs which have a long life, one must be careful to capture the design in a format that can be maintained. Hence it is much better to use design methods which support easy maintenance—handwritten flow charts are exactly what is *not* needed!

A more serious example of the same aspect is the use of programming languages which do not support high-level abstraction, for instance C as opposed to C++.

4.B.4 Buyer Beware!

Professor W. Kahn is a well-known numerical analyst who has also tested many calculators over the years. Several years ago, he illustrated an error in one calculator in the following manner: Assume the calculator is used to compute the route to be taken by an aircraft flying between two American cities, then an error in the computation would result in the aircraft flying into a *specific* mountain.

Modern calculators are cheap and usually reliable. However, errors do occur. Hence, the use of such an instrument in life-critical applications needs serious consideration. If the same calculations were being performed by software within the aircraft, then the very demanding avionics standard would apply [11]. When used for a life-critical application, the same level of assurance should be provided by the calculator (however, it probably would not be cheap).

Appendix 4.C Recommended Software Engineering Techniques

Specific recommended techniques are given in this appendix. These are either defined here, or an appropriate reference is given. A good general reference to software engineering is the *Software Engineer's Reference Book* [25].

4.C.1 Software Inspection

Software inspection is a formal process of reviewing the development of an output document from an input document. It is sometimes referred to as Fagan inspection. An input document could be the functional specification of a software component, and the output document the coding. An excellent book giving details of the method and its practical application has been written by Gilb and Graham [26].

Software inspection is not universally applied, but many organisations apply it with great success. It tends to be applied if the organisation has accepted it and endorses its benefits.

4.C.2 Component Testing

Component testing is a basic software engineering technique which can (and should) be quantified. The software component to which the method is applied is the smallest item of software

with a separate specification (sometimes called a module). It is very rare for the technique *not* to be applicable for software development. The best standard, which is now available through the British Standards Institution (BSI), is the British Computer Society (BCS) standard [27]. The BCS standard allows for many levels of testing, four levels have been selected, as follows:

- *Structural testing:* Several forms of structural testing are defined in the standard but not to any specified level. In this context, we specify branch testing with 50 percent coverage.

- *Equivalence partition testing:* Undertaken to 100 percent coverage. This is complete functional testing at the component level. It is to be applied to those components handling the basic measurement/test data.

- *Statement testing:* 100 percent statement coverage for those components handling the basic measurement/test data. If a statement has not been executed, then a reason for this should be documented. Defensive programming techniques and the detection of hardware malfunctions give rise to statements that cannot be executed (but are quite acceptable).

- *Boundary value testing:* In this case, which can be seen as an addition to equivalence partition testing, values are chosen which lie on the boundary between partitions.

This form of testing is designed to show that the boundary cases themselves are correctly handled.

4.C.3 Regression Testing

Regression testing requires that tests be developed and used to retest the software whenever a change is made. Typically, sometime before the first release, a set of tests will be designed and run on the software. From that point on, all errors located should result in additions to the set of tests which would detect the bug.

To be effective, one needs a method of rerunning the set of tests automatically. The technique is very good for software that is widely used and for which initial bugs are not a major problem. The effect of the method is that subsequent releases of software should be very reliable on the unextended facilities.

4.C.4 Accredited Testing Using a Validation Suite

Accredited testing requires that a set of tests be developed (the validation suite) against which the software can be tested. This is appropriate for software having a detailed specification, such as compilers and communication software. Accredited testing ensures that the tests are run and the results are interpreted correctly, with the specific requirements of objectivity, repeatability

and reproducibility. The method is significantly stronger if the set of tests is updated regularly by means of regression analysis. This implies that errors in any implementation will result in tests being applied to all (validated) systems.

This form of testing provides an ideal basis for certification. An example of this form of testing for a scientific "instrument" is that being proposed in the area of nuclear medicine [28]. Gamma-camera pictures are taken of patients when they are treated with substances containing radioactive trace elements. The camera output is translated into a standard file format, but the difficult numerical part is the analysis of the picture to give the basic information for a medical diagnosis. Other strong methods, such as a mathematical specification, cannot be applied, and hence this method provides a means for the international nuclear medicine community to gain confidence in analysis software. Note that the application is potentially life-critical and the complexity of the processing of data is complex, which implies a high Software Integrity Level, say 3. At this level, the technique of accredited testing is not actually recommended (see Table 4.1), but it is one of the few methods which can provide reasonable assurance in this context. This method is made more effective by means of software phantoms, which are pictures where an agreed diagnosis is available (as least in the cardiac and renal areas), as explained in the reference above.

4.C.5 System-Level Functional Testing

System-level functional testing is based on testing the entire software system as a black box by a careful examination of the functionality specified and ensuring that every aspect of the functionality is tested. An ISO standard is based on application of this test method [29].

4.C.6 Numerical Stability

It is unreasonable to expect even software of the highest quality to deliver results to the full accuracy indicated by the computational precision. This would in general be possible only for (some) problems that are perfectly conditioned, i.e., problems for which a small change in the data makes a comparably small change in the results. Problems regularly arise in which the conditioning is significant and for which no algorithm, however good, can provide results to the accuracy obtainable for well-conditioned problems. A good algorithm, i.e., one that is numerically stable, can be expected to provide results at or within the limitations of the conditioning of the problem. A poor algorithm can exacerbate the effects of natural ill-conditioning, with the consequence that the results are poorer than those for a good algorithm. Software used in scientific disciplines can be unreliable because it implements numerical algorithms that are unstable or not robust. Some of the reasons for such failings are

- a failure to scale, translate, normalise or otherwise transform the input data appropriately before solution (and to perform the inverse operation if necessary following solution),
- the use of an unstable parametrisation of the problem,
- the use of a solution process that exacerbates the inherent (natural) ill-conditioning of the problem, and
- a poor choice of formulas from a set of mathematically (but not numerically) equivalent forms.

The development of algorithms that are numerically stable is a difficult task, and one that should be undertaken with guidance from a numerical analyst or someone with suitable training and experience. It requires that the intended data processing is posed sensibly and, if "off-the-shelf" software modules are used, that such software is appropriate.

There are established high-quality libraries of numerical software that have been developed over many man-years and cover a wide range of computational problems. Examples include the NAG library [30] (which is available in a number of computer languages and for a variety of platforms), LINPACK [31], and NPL libraries for data approximation [32] and numerical optimisation.

4.C.7 Mathematical Specification

Mathematical specification gives the output data values as a function of the input data values. This method is suitable for the simpler processing of basic measurement data and should clearly be expected. The mathematical function may not be the way the actual output is computed—for instance, the specification may use the inverse of a matrix, while the results are actually computed by Gaussian elimination. This method should avoid a common error of not specifying the exact effect of the end of a range of values. It is not easy to apply the method to digital images (for example), since the algorithms applied are quite complex so that any "complete" specification is likely to be very similar to the software itself.

The mathematical specification needs to be validated against the underlying physics. This includes establishing that the model describes the system sufficiently well and ensuring that the errors introduced by the system are fully understood.

4.C.8 Formal Specification

Several methods are available for providing a specification in a completely formal way which can handle most functional aspects of a specification. The best known methods are VDM [33] and Z [34]. For the author's personal views of this method, see Wichmann [35].

4.C.9 Independent Audit

In the United Kingdom, an independent audit to ISO 9001 is widely established. This provides evidence to third parties of a Software Integrity Level of 1. It does not require that stronger (and more expensive) techniques are applied or that the recommendations here are applied. Consequently, auditing to comply with the other standards mentioned in "Requirements" would be better.

4.C.10 Stress Testing

Stress testing involves producing test cases which are more complex and demanding than are likely to arise in practice. It has been applied to testing compilers and other complex software with good results. The best results are obtained when the results can be automatically analysed. For a paper on this method, see Wichmann [36].

4.C.11 Static Analysis/Predictable Execution

The static analysis technique determines properties of the software primarily without execution. One specific property is of key interest: to show that all possible executions are predictable, i.e., determined from the semantics of the high-level programming language in use. Often, software tools are used to assist in the analysis, typically using the programming language source text as input. In general, the analysis techniques

employed can be very minor (say, all variables are explicitly declared) or very strong (formal proof of correctness), but the goal of showing predictable execution should be cost-effective for high-integrity software. For a general discussion on static analysis, see Wichmann et al. [37].

4.C.12 Reference Test Sets

There is a growing need to ensure that software used by scientists is fit for purpose and especially that the results produced are correct to within a prescribed accuracy for the problems purportedly solved. Methodologies, such as those presented in Butler et al. [38], have been developed to this end. The basis of the approach is the design and use of reference data sets and corresponding reference results to undertake black box testing.

The approach allows for reference data sets and results to be generated in a manner that is consistent with the functional specification of the problem addressed by the software. In addition, data sets corresponding to problems with various "degrees of difficulty" or condition (Appendix C.6), and with application-specific properties, may be produced. The comparison of the test and reference results is made objective by the use of quality metrics. The results of the comparison are then used to assess the degree of correctness of the algorithm, i.e., the quality of the underlying mathematical procedure and its

implementation, as well as its fitness-for-purpose in the user's application.

The methodology has been applied successfully in particular areas of metrology. In dimensional metrology, for example, coordinate measuring machines (CMMs) are typically provided with software for least-squares (Gaussian) geometric element fitting. The methodology provides the basis of an ISO standard [39] for testing such software, and it is intended to base a testing service on this standard. Data sets have been developed in such a way that the corresponding reference results are known a priori. Consequently, there is no reliance on reference implementations of software to solve the computational problems, but the generation of the data sets is dependent on a set of simpler "core" numerical tasks that are well understood.

4.C.13 Back-to-Back Testing

In back-to-back testing, two comparable software systems are tested with the same input. The output from each test is then compared—identical results are not usually expected when numerical testing is undertaken. If the comparison can be automated, then it may be possible to run a large number of tests, thus giving a high assurance that the two items produce similar results. Of course, one of the items under test is likely to be a version of known characteristics, while another is the item being assessed.

In the SMART reliability study [16], this form of testing was used by testing a MatLab implementation against the C code within the instrument. The test cases used were those derived from boundary value/equivalence partition testing. This form of testing can also be applied at a higher level than just a single software component. Indeed, the standard method of calibrating instruments can be seen as a back-to-back test of a trusted instrument against one to be calibrated.

REFERENCES

1. *SEMSPLC Guidelines: Safety-Related Application Software for Programmable Logic Controllers.* 1996. London: IEE Technical Guidelines 8.
2. *Safety-related systems—Guidance for engineers.* 1995. London: Hazards Forum. ISBN 0 9525103 0 8.
3. ISO/IEC Guide 25. 1990. *General requirements for the competence of calibration and testing laboratories.*
4. EN 45001. 1989. General criteria for the operation of testing laboratories. Brussels: European Committee for Standardization.
5. NAMAS Accreditation Standard. 1992. M10. National Accreditation of Measurement and Sampling.
6. *Guide for examining software (non-automatic weighing instruments).* 1995. WELMEC 2.3. Brussels: European Committee for Standardization.
7. ISO 9001: 1994. *Quality systems—Model for quality assurance in design, production, installation and servicing.* Geneva, Switzerland: International Organisation for Standardization.
8. IEC 601-1-4: 1996. *Medical electrical equipment—Part 1: General requirements for safety 4: Collateral*

Standard: Programmable electrical medical systems. Geneva, Switzerland: International Electrotechnical Commission.

9. *ODE Guidance for the content of premarket submission for medical devices containing software.* Draft, 3 September 1996.

10. ISO/IEC 9000-3: 1991. *Quality management and quality assurance standards—Part 3: Guidelines for the application of ISO 9001 to the development, supply and maintenance of software.* Geneva, Switzerland: International Organization for Standardization.

11. *Software considerations in airborne systems and equipment certification.* 1992. Issued in the United States by RTCA, Inc. (Radio Technical Commission for Aeronautics) (document RTCA SC167/DO-178B) and in Europe by the European Organization for Civil Aviation Equipment (EUROCAE document ED-12B).

12. IEC 61508: 1998. *Functional safety: Safety-related systems.* Parts 1–7. Draft for public comment. (Part 3 is concerned with software which is the relevant part for this guide.) Geneva, Switzerland: International Electrotechnical Commission.

13. MIRA. 1994. *Development guidelines for vehicle based software.* Warwickshire, UK: The Motor Industry Software Reliability Association.

14. UKAS. 1997. *Study to investigate the feasibility of developing an* accreditable *product-based scheme of conformity assessment and certification based on satisfying the requirements of International Standard IEC 61508.* Available free on the Internet: http:/www.npl.co.uk/npl/cise/UKAS/index.html.

15. *Guidelines for assessment of software in microcomputer controlled equipment for safety-related systems.* 1995. NT TECHN Report 287.

16. Dobbing, A., N. Clark, D. Godfrey, P. M. Harris, G. Parkin, M. J. Stevens, and B. A. Wichmann. 1998. *Reliability of SMART Instrumentation.* National Physical Laboratory and Druck Ltd. Available free on the Internet: http:/www.npl.co.uk/ssfm/download/index.html.

17. HSC. 1998. *The use of computers in safety-critical applications.* Final report of the study group on the safety of operational computer systems. London: Health and Safety Commission.

18. Fink, R., S. Oppert, P. Collison, G. Cooke, S. Dhanjal, H. Lesan and R. Shaw. 1993. Data measurement in clinical laboratory information systems. In *Directions in safety-critical systems,* edited by F. Redhill and T. Anderson. Berlin: Springer-Verlag, pp. 84–95.

19. NAMAS. 1991. *Use of word processing systems in NAMAS accredited laboratories.* NIS41. National Accreditation of Measurement and Sampling.

20. *Guidelines on risk issues.* 1993. London: The Engineering Council. ISBN 0- 9516611-7-5.

21. ISPE. *Supplier guide for validation of automated systems in pharmaceutical manufacture.* Tampa, Fla., USA.: International Society for Pharmaceutical Engineering.

22. Littlewood, B., and L. Strigini. 1992. The risks of software. *Scientific American.* (November).

23. ISO/IEC 12207: 1995. *Information technology—Software life cycle processes.* Geneva, Switzerland: International Organization for Standardization.

24. NAMAS. 1993. *A guide to managing the configuration of computer systems (hardware, software and firmware) used in NAMAS accredited laboratories.* NIS37. National Accreditation of Measurement and Sampling.

25. McDermid, J. A., ed. 1991. *Software engineer's reference book.* Oxford: Butterworth-Heinemann.

26. Gilb, T., and D. Graham. 1993. *Software inspection.* London: Addison-Wesley. ISBN 0- 201-63181-4.

27. British Computer Society Specialist Group in Software Testing. 1998. *Standard for software component testing* (working draft 3.0). London: British Standards Institution.

28. Britton, K. E., and E. Vauramo. 1993. COST B2: The quality assurance of nuclear medicine software.

European Journal of Nuclear Medicine 20:815–816. Details available on the Internet at http:/www.phb-mpd. demon.co.uk/index.html.

29. ISO/IEC 12119: 1993. *Information technology—Software packages—Quality requirements and testing.* Geneva, Switzerland: International Organization for Standardization.

30. The NAG Fortran Library. Available from the Numerical Algorithms Group Ltd., Wilkinson House, Jordan Hill Road, Oxford OX2 8DR, UK.

31. Dongarra, J. J., C. B. Moler, J. R. Bunch, and G. W. Stewart. 1979. *LINPACK user's guide.* Philadelphia: Society for Industrial and Applied Mathematics.

32. Anthony, G. T., and M. G. Cox. 1987. The National Physical Laboratory's data approximation subroutine library. In *Algorithms for approximation,* edited by J. C. Mason and M. G. Cox. Oxford: Clarendon Press, pp. 669–687.

33. Dawes, J. 1991. *The VDM-SL reference guide.* London: Pitman Publishing. ISBN 0 273-03151-1

34. Spivey, J. M. 1988. *Understanding Z.* Cambridge: Cambridge University Press.

35. Wichmann, B. A. A personal view of formal methods. Available from NPL web site on www.npl.co.uk.

36. Wichmann, B. A. Some remarks about random testing. Available from NPL web site on www.npl.co.uk.

37. Wichmann, B. A., A. A. Canning, D. L. Clutterbuck, L. A. Winsborrow, N. J. Ward, and D. W. R. Marsh. 1995. An industrial perspective on static analysis. *Software Engineering Journal.* (March): 69–75.

38. Butler, B. P., M. G. Cox, S. L. R. Ellison, and W. A. Hardcastle, eds. 1996. *Statistics software qualification: Reference data sets.* London. Royal Society of Chemistry.

39. ISO/DIS 10360-6: 1996. *Method for testing the computation of Gaussian associated features.* Geneva, Switzerland: International Organization for Standardization.

Chapter 5

ERRORS ASSOCIATED WITH WEIGHING

Chris Jenkins
Kent Scientific Services
Kent, United Kingdom

The objective of this chapter is to explain the many and varied causes of errors inherent in weighing, with suggestions that will eliminate or minimise their effect. Hopefully, readers will be motivated enough to examine their own weighing schemes in order to minimise the effect of common weighing errors in weight determinations.

Even though the technology used in electronic weighing instruments is so different from

that of equal-armed weighing instruments, much of this chapter is relevant to all types of weighing instruments. Generally, this chapter will be of interest to readers who require better accuracies (such as users of balances of precision or mass comparators) as well as those requiring accuracies acceptable for trade purposes (as dictated by local metrological laws). Other works will assist readers who require coarser accuracies, and I guide them to the references mentioned below.

For no better reason than convenience, I have classified weighing errors by their causes, as follows:

- Errors caused by weighing instruments
- Errors caused by environmental effects
- Errors caused by other effects

Errors Caused by Weighing Instruments

Discrimination Errors

Discrimination errors are those caused by the inability of a weighing instrument to discriminate between weights of almost equal mass. The same indication will be given by the weighing instrument when it is loaded with either of two weights having a very slight difference in their masses. The cause is usually friction within the bearing

surfaces, usually at pivot points, and cannot be eliminated. Discrimination errors can be quantified at any point in the weighing range of a weighing instrument but are best found at those loadings that the weighing instrument will be used to weigh.

A suggested procedure for finding discrimination errors is as follows:

1. Note at which point on the weighing range discrimination errors need to be found.

2. Exercise the weighing instrument: load it with a load close to its capacity but not less than the maximum load that is to be weighed on the weighing instrument, record the indication given, remove the load, wait for an unladen indication, and repeat until four unladen indications have been recorded.

3. Place a large weight or weights on the pan of the weighing instrument to reach the relevant part of the weighing range and record the resulting indication.

4. Place a small known weight (taken from your most accurate weight set and equal in mass to about one-tenth of the weighing error acceptable to you at that loading) on the pan of the weighing instrument and record the resulting indication.

5. Add another small weight to the pan and repeat step 4 above until 10 small weights have been used.

6. Calculate the effect of each small weight by the equation

   effect n = (reading n) – (reading $n - 1$)

7. Calculate the errors caused by each small weight by the equation

   discrimination error n = (effect n) – (known mass of weight used)

8. Find the maximum of the discrimination errors found and use that value as the discrimination error at that loading.

Eccentricity Errors

Eccentricity errors are those evident when a load is placed off-centre on a pan of a weighing instrument. They are caused by the bearing surfaces within a pivot point taking a different relative position to each other than if the centre of gravity of the load was applied centrally on the pan. The additional moment caused by the off-centre loading results in a different relative position of the bearing surfaces, as the stay system and load levers of the weighing instrument balance the force on the pan against a resistant.

Eccentricity errors are easily reduced by restricting any loading to a central portion of the pan or by using a positioning plate over the pan so that individuals within groups of weights can be placed only at a designated position on the pan.

Eccentricity errors should be quantified between one-fourth and one-third of the capacity of the weighing instrument. Any loading at the edge of a pan at loads in excess of one-third capacity will almost certainly lead to permanent deformation of parts of the stay system of an electronic weighing instrument and will result in an unusable pan angle on an equal-armed weighing instrument.

The fact that eccentricity errors are determined should not be taken as a reason to ignore one of the basic tenets of good weighing practise—that all loads should be centred as far as possible on the pan of any weighing instrument being used. This is due to the fact that off-centre loading produces torque that stresses the stays and lever system of the weighing instrument being used. Also that torque ultimately needs to be counterbalanced by the resistant, and in some patterns of weighing instruments can effect the resulting indication.

A properly conducted eccentricity calibration is one of the best indicators of the general well-being of a weighing instrument. In the author's experience, those weighing instruments that give consistent and low error values during such a calibration will give good service afterward.

A suggested procedure for finding eccentricity errors is as follows:

1. Note at which load eccentricity errors need to be found.

2. Note at which positions on the pan that loads will be placed during normal use.

3. Exercise the weighing instrument: (load it with a suitable load, record the indication given, remove the load, wait for an unladen indication, and repeat until four unladen indications have been recorded.

4. Tare the instrument unladen and record the suitable indication produced as the "unladen 1" value.

5. Place the correct load on the pan of the weighing instrument so that its centre of gravity is above the centre of the pan and record the resulting indication as the "centre 1" value.

6. Lift (never scrape) the load and place it as close to the front edge of the pan as your positioning policy allows; record the indication produced as the "front" value.

7. Repeat step 5 above but with the load placed close to the left edge of the pan, recording the resulting indication as the "centre 2" value.

8. Repeat step 6 above, but with the load placed close to the back edge of the pan, recording the resulting indication as the "left" value.

9. Repeat step 5 above, recording the resulting indication as the "centre 3" value.

10. Repeat step 6 above, but with the load placed close to the right edge of the pan, recording the resulting indication as the "back" value.

11. Repeat step 5 above, recording the resulting indication as the "centre 4" value.

12. Repeat step 6 above, recording the resulting indication as the "right" value.

13. Repeat step 5 above, recording the resulting indication as the "centre 5" value.

14. Record an unladen indication as the "unladen 2" value.

15. Calculate the "unladen change" value by the equation

unladen change = (unladen 1) – (unladen 2)

16. Calculate the eccentricity error by the equation

eccentricity error = (front) – (½ centre 1) – (½ centre 2) – ⅕ unladen change)

17. Calculate the eccentricity error at the left hand side of the pan by the equation

eccentricity error = left – (½ centre 2) – (½ centre 3) – (⅖ unladen change)

18. Calculate the eccentricity error at the back of the pan by the equation

eccentricity error = back – (½ centre 3) – (½ centre 4) – (⅗ unladen change)

19. Calculate the eccentricity error at the right hand side of the pan by the equation

eccentricity error = right – (½ centre 4) – (½ centre 5) – (⅘ unladen change)

20. Find the maximum of the "front" value, the "left" value, the "back" value, and the "right" eccentricity errors and use that value as the eccentricity error at that loading.

Linearity Errors

Linearity errors are synomynous with accuracy errors. The linearity error at any load is the error between the indication given by the weighing instrument and the weight applied.

Linearity errors should be found at regular intervals over all parts of the weighing range that are used, using both increasing loads and decreasing loads. A large difference between an increasing load's linearity value and a decreasing load's linearity value, at the same load, indicates that there is play in the load lever and stay systems of the weighing instrument.

A suggested procedure for finding linearity errors is as follows:

1. Decide on the weighing range over which linearity errors need to be calibrated—the range between the minimum and maximum loads that will be applied.

2. Divide that weighing range by 10 and round that value to a value that minimises the number of individual weights that will be used during the calibration. If you round down, then you will increase the number of loads required for calibration; if you round up, then you will increase the weighing range for calibration.

Example (Step 2)

A 50 kg bench weighing instrument is to be used for weighing loads of not less than 4 kg and not more than 32 kg. Hence the weighing range over which linearity errors are to be calibrated is 28 kg.

Dividing 28 kg by 10 suggests an incremental difference value of 2.8 kg; if rounded up, it would be 3 kg. So appropriate loadings would be 3 kg, 6 kg, 9 kg, 12 kg, 15 kg, 18 kg, 21 kg, 24 kg, 27 kg, 30 kg, and 33 kg—11 different loads to apply.

If rounded down, it would be 2 kg, and the appropriate loadings would be 4 kg, 6 kg, 8 kg, 10 kg, 12 kg, 14 kg, 16 kg, 18 kg, 20 kg, 22 kg, 24 kg, 26 kg, 28 kg, 30 kg, and 32 kg—15 different loads to apply.

3. Exercise the weighing instrument: load it with a load equal to the maximum load that will be applied to it, record the indication given, remove the load, wait for an unladen indication, and repeat until four unladen indications have been recorded. During the "increasing" part of this calibration, the load on the instrument must always be increasing. Hence, a calibrator adding to the load applied to a comparator must remove those weights not needed from the instrument, before introducing other weights to it.

4. Use the most accurate weights available during the linearity calibration. Start by taring the instrument unladen and then

apply weights of nominal mass equal to the lightest linearity load previously determined; record the suitable indication as the "increasing load 1" value.

5. Remove any weights from the weighing instrument that will not be needed as part of the next linearity load. Then add any weights to the weighing instrument that will be needed as part of that linearity load; record the suitable indication as the "increasing load 2" value.

6. Repeat step 5 above until the suitable indication produced by the maximum load has been recorded as the "increasing load n" value, where n is not less than 11.

7. Now repeat the calibration but with a decreasing load. During this part of the calibration, the load on the instrument must always be decreasing. Hence, a calibrator reducing the load applied to a weighing instrument must introduce weights to it, before removing those weights not needed from it. Then add any weights to the weighing instrument that will be needed as part of the next linearity load. Remove any weights from the weighing instrument that will not be needed as part of the next linearity load record the suitable indication as the "Decreasing Load $n - 1$" value.

8. Continue this process until the load applied to the instrument has been reduced to the

minimum load at which the increasing loads calibration was started.

9. Calculate the difference between each pair of results and record them as

$$(\text{increasing diff. } n) = \{(\text{increasing load } n+1) - (\text{increasing load } n)\}$$

$$(\text{decreasing diff. } n) = \{(\text{decreasing load } n+1) - (\text{decreasing load } n)\}$$

10. Calculate the "change" values by taking the last determined mass of the incremental weight from the difference just calculated. The relationships are

$$(\text{increasing change } n) = \{(\text{increasing diff. } n) - (\text{mass of incremental weight})\}$$

$$(\text{decreasing change } n) = \{(\text{decreasing diff. } n) - (\text{mass of incremental weight})\}$$

11. By calculating the difference between the increasing change and the decreasing change, the backweighing difference value can be found using the relationship

$$(\text{backweighing difference } n) = (\text{increasing change } n) - (\text{decreasing change } n)$$

The most accurate and dependable weighing instruments will produce the smallest backweighing difference values.

Repeatability Errors

Repeatability errors are those errors caused by the inability of a weighing instrument to indicate

the same result when a single weight is repeatedly applied to it under constant conditions.

There are two schools of thought on how repeatability errors should be measured: (1) Simply take the difference between the maximum and the minimum values obtained during a calibration. (2) Examine the indications of the weighing instrument concerned, and analyse the variation within them. Once that is done, statistical methods can be used to predict their effects upon weighings.

High accuracy weights are not needed for repeatability calibrations, but it is vital that the mass of any weights used does not change during the calibration. A suggested procedure for finding the repeatability error is as follows:

1. Decide on the load at which the repeatability error is to be calibrated. This load should be close to the maximum load that will be applied to the weighing instrument and should consist of the fewest number of weights possible.

2. Exercise the weighing instrument: load it with a load equal to the maximum load that will be applied to it, record the indication given, remove the load, wait for an unladen indication, and repeat until four unladen indications have been recorded.

3. Load the appropriate weights onto the pan of the weighing instrument and record the suitable indication as the "repeatability 1" value.

4. Remove those weights from the instrument and await a suitable unladen indication. Then replace the weights on the pan of the instrument and record the suitable indication as the "repeatability 2" value on the record form.

5. Continue to repeat this process until the "repeatability 21" value has been recorded (i.e., repeat the process to produce 21 repeatability values).

6. Calculate the 10 correction values by the equation

$$\text{correction } n = \text{average of } \{\text{repeatability } (2n + 1) + \text{repeatability } (2n - 1)\}$$

7. Calculate the corrected weight value by subtracting the "Correction" value from the "Repeatability" value marked to the left of it. Continue until 10 such values have been calculated, using the relationship

$$\text{corrected } n = \{(\text{repeatability } 2n) - (\text{correction } n)\}$$

8. Calculate and record the average of the 10 "corrected 1" to "corrected 10" weight values and their sample standard deviation s where

$$s^2 = \frac{\sum \bar{x}^2 - \bar{x}}{9}$$

Length of Arm Errors

The length of arm error is an common cause of error in equal-armed weighing instruments. It is

the error caused by differences in the distance between each outer knife edge and the centre knife edge. If the working surface of any knife edge is not at an exact right angle to the long axis of the beam, then this error will be present. It will vary with the load applied to an equal-armed weighing instrument as the beam flexes under loading.

Whilst modern manufacturing techniques have minimised this type of error, it can never be ignored in high-accuracy weighing schemes. Many countries have legislated that a substitution weighing method must be used in legal metrology, allowing comparison weighing only at trade accuracies or worse. This is good practice and should adopted by users striving for perfection.

Rest Point Errors

Rest point evaluation is a quick tool for predicting the final rest point of the indicator of an equal-armed weighing instrument against its chart. However, it cannot predict correctly unless the first two extreme positions are ignored.

In theory, the predictor will work with any three consecutive readings, other than the first two. One way of eliminating error is to take the third, fourth and fifth extremes and predict a rest point. Then take the next three extremes and predict from them. If both predictions are the same (within very tight bounds), then the system is reliable on that weighing instrument. If the

predictions are not the same, then the cause must be found and remedied.

Relieving Mechanism Errors

Relieving mechanism errors are those caused by the relieving mechanism giving a throw to the beam during disengagement. If it is severe enough, then the operator will notice the beam bouncing off its stops during disengagement. If it is not noticed, the principal result is that it will effect the rest point predictor system so that it will not be reliable for several extremes.

It will also be evident as the beam is arrested that the beam will tend to be thrown to one side, often bouncing off its stop as it does so. Should any beam be thrown by the relieving mechanism, then the weighing instrument effected needs to be repaired immediately.

Hysteresis and Eddy Currents

Hysteresis is the inability of a loaded or stressed object to return to its unloaded or unstressed state immediately upon being relieved. Hysteresis will make itself evident if a weighing instrument is tested first with an increasing load and then without being unladen with a decreasing load.

Sensitivity Errors

An error of weights used in sensitivity determinations on equal-armed weighing instruments

should be known. The errors permitted in weights used in sensitivity determinations are high in proportion to larger weights. For example, a 1 mg Class F1 weight has a permitted error of 20,000 ppm (2 percent of its nominal mass), whilst a 1 kg Class F1 weight has a permitted error of 5 ppm (0.0005 percent of its nominal mass). Ignoring this fact could introduce such errors into the weight value per chart division used.

Binding Errors

Binding errors can occur in any weighing instrument and will be readily evident as repeated applications of the same weight produce different results. In bad cases, it will be possible to set the indicator to any desired indication. Sometimes binding will occur in one particular part of the weighing range and thus be less obvious. However, repeatability calibrations should identify any such effects.

Maintenance Errors

Incompetent service technicians is the kindest description that can be given to untrained and ignorant individuals who attempt to service weighing instruments without the required knowledge or skills. With an equal-armed weighing instrument, the incompetent service technician often seems to try to alter the sensitivity of the instrument. But the geometrical relationship between

the knife edges is critical to the correct operation of such instruments. Untrained service personnel can easily overadjust the sensitivity control devices so that the beam will be in neutral equilibrium at some loads, unstable equilibrium at others, but weigh light loads correctly.

Further reading on this relationship can be found in Notes on Applied Science No. 7 published by the NPL in 1954 [1]. Another source of information concerning trade equipment is the "Metcalfe Trilogy" [2].

Stay Shift Errors

Stay shift is a rare disorder caused by the sudden and violent overloading of the stay system. It results in damage to a stay, which allows that stay to catch on the stay system and to thus take one of two positions. The result is that the instrument will be capable of indicating two different indications with a single load.

The operator can use a fallible test to identify stay shift by placing a load close to but less than the capacity of the weighing instrument on its pan, waiting for stability, and then momentarily lifting it and replacing it on the pan. If the indication changes, then stay shift may be the cause.

Errors Caused by Environmental Effects

Vibrations

Vibrations will have their most serious effects if any part of a weighing instrument starts to resonate. As a rule of thumb, if a person standing next to a weighing instrument can hear anything, then the noise heard will effect the weighing instrument as a set of vibrations; the louder the noise, the more effect the vibrations will have.

A measure of how much effect the vibrations are having can be found by determining the difference between the repeatability error for the weighing instrument when the noise is heard and the repeatability error when it is silent (or at its quietest if the noise or vibrations cannot be stopped completely).

Certain types of electronic weighing instruments, those using microchips, may be susceptible to vibrations within buildings. Such weighing instruments use self-adjusting digital filters which can erroneously convert low-frequency interference into an indicated weight change.

Electromagnetic Interference

Electromagnetic compatibility (EMC) relates to the compatibility of a weighing instrument with electrical and electronic equipment of all descriptions to the extent that the weighing instrument is

internally protected from the effects of magnetic, electrical, magnetic and radio fields produced in any such equipment. Those weighing instruments that are not compatible with local electronic equipment will show obviously incorrect displays whilst being effected by the interference.

EMC is now controlled by legislation, so basic compatibility is ensured. Within the states in the European Economic Area, a manufacturer's CE mark is required to be shown on a new weighing instrument (along with other new electrical equipment) as proof that the marked item conforms with all the various European "New Approach" directives (sometimes called "100A" directives) on EMC.

Models of electronic weighing instruments built before about 1990 may well be effected by nearby electronic equipment, i.e., they are not compatible with the local electronic environment. Should a weighing instrument be so effected, then the first point of contact should be with the weighing instrument's manufacturer, or authorised repairer, who may be able to modify it so that it becomes compatible with local equipment.

Draughts

Draughts have a catastrophic effect weighing. With electronic weighing instruments, draughts increase the amount of time taken to weigh if it is possible to achieve a stable reading. With mechanical weighing instruments, draughts effect

the pans to make any rest point predictor unreliable; self-indicators will appear unable to find stability. Draughts cause changes in the air within the weighing chamber of modern high-accuracy weighing instruments.

Many laboratories accept that draughts cannot be eliminated completely and so condition the air in their weighing rooms to make any draught constant and unchanging. Therefore, the effect on a weighing instrument will also be constant and can be ignored whilst constant.

Unplanned draughts must be eliminated before accurate weighings can take place. Use pieces of soft tissue paper glued to rods to show where draughts flow around an effected weighing instrument. That airflow can then be diverted around the weighing instrument with barriers, or a shield can be placed around the whole weighing instrument.

Do not assume that the draught shielding incorporated into a weighing instrument will be sufficient, as the shielding protects only the weighing pan and the chamber immediately above it. Look for draughts that seem to aim themselves at the ventilation orifices of the effected weighing instrument or can be reflected off benches, plinths or walls into such orifices.

The effects of a draught playing directly on the weighing cell of a modern weighing instrument is that stability cannot be achieved unless the draught is constant. When the draught is not

constant, the weighing instrument will appear to display constantly changing values which appear to be close to the expected display value.

Leveling

Level is the most important attribute of a weighing instrument. Generally, weighing instruments are designed so that when loaded, forces act at perpendicular angles to the weighing system. In fact, factory adjustment will often be made by adjusting the housing of a weighing instrument until its displayed weight is maximised for the particular weight being weighed.

Be it electronic or mechanical, a weighing instrument cannot operate correctly when it is not level. I qualify that statement in the light of weighing instruments described as "level-proof", which weigh correctly when out of level to a degree that is within its manufacturer's specification.

Overloading

Overloading of weighing instruments usually occurs in two ways:

1. A load greater than the capacity of the weighing instrument is loaded onto it. This is bad practice because the load receptor systems will be stressed above their working limits. Eventually, of course, individual components will be deformed or will be

overstressed, leading to a shorter working life than would otherwise be achieved. Manufacturers of weighing instruments used in situations where they could well be overloaded will provide some protection in the form of stops, but it is unwise to test these too frequently.

2. The chassis of the weighing instrument becomes overloaded because a user is placing weights on the housing of the weighing instrument. Whilst it is bad practice to load a mass greater than the capacity of the weighing instrument on it, this can happen if the culture of the laboratory is allow operators to place weights that are not in use, but soon will be, on the housing of the weighing instrument.

As elsewhere in this chapter, I suggest that weights should be left to acclimatise in the weighing chamber in which they will be used so that there is no difference in temperature between them, the weighing instrument that they are used with, and the air in the weighing chamber at the time of calibration. The sensible solution appears to me that only a small mass of weights should be kept within the weighing chamber of the weighing instrument. How much that total mass is should be found be repeatability testing with various amounts of weights left on the housing.

Dust

Dust has no place in a high-accuracy weighing room, and all weighing rooms for Class F2 and finer calibrations should be protected by dust filters incorporated into their air-conditioning plant. A microscopic piece of dust could itself have a mass equal to a large part of the tolerance applicable to a single, high-accuracy weight. Thus, if it was present on a weight during calibration and subsequently fell off the weight, then that weight could easily be certified at an incorrect mass.

At Class M1 and coarser accuracies, dust is inherent in many of the materials used in the construction of weights. Cast iron invariably rusts if it becomes wet, no matter how well it is treated, and so will produce an amount of dust; paint will chip and flake from such weights when they are used. Further, cast iron weights have lead in their adjustment wells, and this will produce a dust if cold worked as part of any adjustment process.

Electrostatic Charges

Always present, but particularly noticeable when the humidity is low, electrostatic charges are caused by the interaction of charges on the weighing instrument being used and on the item being weighed. The materials most effected are

those of low conductivity: glass, plastics, filter materials, and some powders and liquids. Electrostatic charges are mostly caused by friction within powders and liquids during movement whilst being carried, friction between filters and their supports, and friction in air due to convection in a dry atmosphere. The result is either attraction or repulsion, which will dissipate over time; thus the effects of electrostatic charges on weighing instrument are characterised by drift of the indicated weight and by poor repeatability.

The simplest way of reducing the effects of electrostatic charges is to distance the item being weighed so that is further from the weighing instrument, thus reducing its electrostatic effects. Some weighing instruments are constructed so that pans hanging underneath them may be used for weighing.

One solution that will always eliminate electrostatic charges is to weigh items within a Faraday cage. Some weighing instruments are now constructed so that their casing is a Faraday cage. Another solution is to improve the surface conductivity of the item being weighed by wrapping it in metallic foil. If a non-hygroscopic substance is being weighed, then an open water container placed within the weighing chamber will increase humidity. Any condensate formed within the chamber will reduce the electrostatic charges there.

Instruments that neutralise electrostatic charges by blowing ions of the opposite charge

across the item to be weighed or into the weighing chamber (before weighing takes place) are commercially available. Some use a high-voltage ion source, whilst some use radioactive ion sources, often polonium or americium.

Humidity

Humidity can effect the operation of electronic weighing instruments, so operating humidity ranges are specified by manufacturers. If humidity is either less than the lower limit or greater than the upper limit, then arcing may occur within the weighing instrument, particularly in the coils. If no operating humidity range is specified by the manufacturer or that information is not available, then operating within the range of 30 percent to 70 percent is generally safe.

Some equal-armed weighing instruments are fitted with agate knife edges, which are particularly susceptible to absorbing moisture from the atmosphere. Such weighing instruments should not be used when relative humidity is outside the range of 30 percent to 70 percent, as moisture changes within their knife edges can lead to significant weighing errors.

Humidity must be measured and monitored when weighing at Class F2 or high accuracies, as humidity changes in a weighing room will alter the air density in that room enough to lead to a possible need to apply an air buoyancy correction to any weighing results achieved.

Temperature

The highest accuracy weighings will be undertaken in a laboratory with little or no change in temperature. These weighings benefit from the fact that all the standards, equipment and instruments used will be acclimatised at the same temperature.

A stable temperature in a laboratory, even though it is not at convention temperature, will lead to high accuracy results. Those results may have to corrected to take account of the variation from the convention temperature.

A laboratory with a constantly changing temperature will not be able to deliver the highest levels of accuracy for several reasons. If the temperature of the laboratory is constantly changing, then the temperatures of all the standards, equipment and instruments used will not be known accurately. Increased quantities will have to put into the relevant uncertainty budget to cover all eventualities. In these circumstances, it is possible that the item being weighed is at a different temperature to the weighing instrument. The air within the weighing chamber could alter so it is not a true sample of the atmosphere measured to find air density. Hence, an incorrect correction could be applied to the weighing results.

Temperature must be measured and monitored when weighing at Class F2 or high accuracies, as temperature changes in a weighing room will alter the air density in that room. Small

temperature changes will have a relatively serious impact on air density values.

Magnetism

When an item being weighed is magnetic or is magnetised, then it will effect the weighing instrument being used. By acting as a magnet, the item will effect the magnetisable parts of the weighing instrument and introduce errors. The effect will be a lack of repeatability without any drifting. It will be quite difficult to identify, as that loss in repeatability will probably not be noticed by users until after several weighings have taken place.

One method for identifying magnetic effects is to weigh the item, rotate it 120° relative to the load receptor of the weighing instrument, weigh it again, rotate it a further 120° relative to the load receptor and weigh it again. If any of the three results is substantially different to any of the others, then magnetism is suspected.

Some weighing instruments are more susceptible than others; an electromagnetic force–compensating weighing instrument will contain a strong permanent magnet whose effects cannot all be shielded completely. So weighing a magnetisable item on such a weighing instrument will lead to that item acting as a permanent magnet during the weighing process, with the unknown effect of that magnetism being included

as an unknown error in the instrument's indication.

The most obvious solution to the problem of weighing a magnetic item is to increase the distance between that object and the weighing instrument. Thus spacers could be used between the item and the pan of the weighing instrument, or a hanging pan beneath the weighing instrument could be used. Where increasing the distance is not viable, then shields, which are available in several shapes and made of highly permeable nickel-iron alloys, will reduce the effects of magnetism.

Errors Caused by Other Effects

Convention Mass

Convention mass is the mass of a weight under standardised conditions. All current mass calibration certificates which show a measured value for a weight will endorse that value by saying that it relates not to the weight calibrated but to a hypothetical weight of a density of 8,000 kg/m^3, which would balance the calibrated weight in air of a density of 1.2 kg/m^3 at 20°C (in some tropical areas, the temperature used is 27°C).

Thus an end user of such a calibration certificate who does not use a calibrated weight at the certified temperature will find a small error in

the measured value. The relationship between the masses of a standard weight and a test weight are

$$m_t\rho_s\,(\rho_t - \rho_a) = m_s\rho_t\,(\rho_s - \rho_a)$$

where m denotes mass and ρ denotes density, whilst the subscripts s denotes standard, t denotes test weight, and a denotes air.

Finding Mass

Finding mass by weighing can be complicated by a weight being calibrated and used in areas with differing g values or by a weighing instrument being calibrated and then used in an area with a differing g value. Such errors are caused by weight and weighing instrument users not appreciating the difference between mass and weight. Unfortunately, the units that both are measured in have the same names.

Often, weight is actually measured by those who need to measure it, whilst mass is often declared by an external source. Problems can occur when importing values from one laboratory into another, as mass certificates will certify not the mass of the item but its convention mass.

Any user of such a certified weight would need to be aware that he or she might need to make a correction from the declared measured value to an equivalent mass value in his or her own laboratory. That correction would depend on the temperature and air density (so atmospheric

pressure and humidity also relate to the it) at each instance of use.

Operator Errors

Operator errors are to be expected when trains of figures are written down, particularly if the writer is under pressure or otherwise stressed; a competent checking system is needed to minimise the number of errors that are not noticed.

However, certain numbers do seem to be recorded incorrectly more frequently. For instance, in my own laboratory, I take great care that numbers between 1100 and 1120 are recorded correctly by hand and that sets of values containing both positive and negative values are also rigorously checked.

The obvious answer to transcription errors is to use a computer to process results, and modern spreadsheet programs can be developed to check for correct sign and also to check for viability. A method of evaluating differences between operators is replicate testing; the two results are required to be within a specified value of each other; the uncertainty of calibration of the weight concerned is also a factor.

Known and Unknown Errors

The known errors of standards must be taken into account when making high-accuracy determinations. These values are best taken from the last calibration certificate.

Any unknown drift of standards must be taken into account, but this can only be done if a history of the weights involved has been built up over several calibration periods, using either time-based or user-based schemes.

Thus a weight that has previously averaged a mass loss of 2 μg a day if assigned a wear limit of 500 μg could be used for 250 days. In practice this would be reduced as a cautionary measure to, say, 200 days.

The assumption that individual standards within a group each have a statistically independent uncertainty is not valid. Invariably any set of standard weights used by a calibration laboratory will be traceable to the relevant national standards. Those national standards will themselves be traceable to a national kilogramme, which will itself be traceable to the International Prototype Kilogramme. So any error in the calibration of any standard in the chain of traceability will be imported to you and be accepted in good faith. But this will probably not exceed the declared uncertainty.

The safest way of combining the uncertainties of several weights, thus ensuring that the relevant uncertainty budgets are correct, is to use arithmetic addition of each individual uncertainty value, rather than taking the root mean sum of the squares of those individual uncertainty values.

Exercising a Weighing Instrument

Exercising a weighing instrument before use is essential as a means of minimising errors due to hysteresis and component start-up variations. The process is essentially a dummy calibration of a weight against itself so that any errors found are due to operator variation or to weighing instrument components.

As start-up errors can be expected from a weighing instrument, and these will probably exceed later process variations, the errors between the results of exercising can be assumed to be an overestimate of the actual error of the weighing instrument.

Weighing Schemes

Weighing schemes need to be chosen to deliver a weighing result that is at least as accurate as needed but at a minimum cost in terms of labour, equipment use and the material weighed. Remember that few weighing instruments will have as many as 50,000 weighings made using them without expensive repairs being required. Choosing a scheme that involves 7 weighings during each calibration will result in the weighing instrument being used to calibrate about 7,150 weights before a repair can be expected to be needed.

If any material is being weighed, then a trade-off between the cost of accuracy and a level of

accuracy must be decided. Often, local metrological laws will effectively dictate the accuracy of weighing equipment used. For instance, in the United Kingdom, a less accurate weighing instrument can be used to sell potatoes than those used to sell precious metals.

For trade weighings and those requiring lower levels of accuracy (say correct to within 100 ppm and coarser), a suitable weighing scheme will be based on comparison weighing. Comparison weighing consists of placing a standard on one pan of a weighing instrument, the item being weighed on the other, and being satisfied that they both weigh the same amount if the weighing instrument settles at horizontal equilibrium.

For calibrating weights to use in testing trade weighing instruments and those requiring higher levels of accuracy (say correct to within 50 ppm and finer), a suitable weighing scheme will be based on substitution weighing. Substitution weighing requires a third mass. The third mass is balanced against the standard, and a rest point is found. Then a small weight is added, and a second rest point is taken. The difference between the 2 rest points is assumed to be caused solely by the mass of the small weight used. Hence weight value per division is known. The small weight is then removed, and the standard is replaced by the test weight; the third weight is left untouched on its pan. The theory is that if the test weight has an identical mass to the standard, then it will cause the same deflection that the

standard caused. So a rest point is taken with the test weight substituted for the standard weight. The difference in the rest points they each cause is easily calculated in divisions, which can be converted to mass units using the weight value per division value found using the small weight.

In calibrations of weights at tolerance levels of 5 ppm or finer, a standard from a second set of standards will be used as a check on the wear of the standards.

Why You Must Not Touch Some Weights

Whenever a person touches a metal surface, he or she transfers dust, grease, sweat, moisture and oils to that metal surface—all having a strong hygroscopic effect. The quantities of each substance transferred will vary with the temperature and the individual.

To touch a weight is to transfer unknown amounts of those substances, which will increase the actual mass of the weight but not its certified mass; thus when used, an error will be introduced. If any evaporation takes place whilst the item is being calibrated, then there will be a large increase in the standard deviation of the weighing, as the weight of the item will be changing.

With Class M1 or coarser weighings, this amount of debris will usually be negligible compared with the tolerance, and so it is not objectionable to handle such weights, which includes cast iron weights and brass weights. For Class F2

weighings, that unknown transferred quantity can be a large proportion of the tolerance applicable to a test weight. For Class F1 and higher accuracies, the mass of a fingerprint can be at least equal to the tolerance permitted on any test weight. Thus such weights, which are made of non-magnetic austentite stainless steels, must not be handled. In fact, it is a good rule of thumb not to touch any stainless steel weight, even those calibrated at Class M1 or coarser accuracies, as to do so will probably mark them.

Enough Time to Warm Up and Stabilise

Whenever an electronic weighing instrument is reconnected to its power supply after a period of inactivity, there will be huge changes in temperature in the different-shaped components made of many different materials within the weighing instrument. Even within components there will thermally induced stresses, for example, within a load cell the strain gauge, the adhesive, and the billet they are attached to will all have differing coefficients of expansion. Thus a load cell cannot be expected to operate at its best immediately after being powered up.

Over a period of about 30 minutes after switching on, these variances will settle, and the components will find a thermal equilibrium. Until that is achieved, however, the weighing instrument will not be capable of consistent weighing.

With many high-accuracy weighing instruments, the display units can be switched off whilst leaving the weighing circuits powered. I have found that such weighing instruments can suffer from a lack of repeatability for 30 minutes after the display is switched on, and I assume that this is because the components within the display need to reach a thermal equilibrium before operating optimally and will produce eddy currents until that equilibrium is achieved. So I now leave my high-accuracy weighing instrument completely powered all the time.

Changes in g—the Gravitational Constant

It feels strange to be writing about changes to a constant! Around the globe, g varies between 9.77 and 9.83 m/sec^2, and so to name it the gravitational constant seems inappropriate. But locally it is a constant and will only change as described below.

The *Guide to the Measurement of Mass and Weight* [3] outlines the formulae used to calculate g. However, there will always be a component caused by local geological conditions, and national geological bodies will be able to advise on their magnitude.

For the user of weighing instruments, the changes in local g values can cause errors. As weight is a force, so weighing is the measurement of that force; the force is defined as mass multiplied by g, the gravitational constant. Thus

any change to g will also change the force that will be measured.

If a weighing instrument is calibrated and then moved to an area with a different g value, then it obviously will indicate weights incorrectly. An example is a weighing instrument set up to weigh 1,000 g precisely in Houston, Texas (g = 9.7928 m/sec^2), would indicate 1,002.686 g in Anchorage, Alaska (g = 9.8191 m/sec^2), if not recalibrated.

Also the local g value will be dependent on height above sea level; it will decrease by about 3.14×10^{-7} m/sec^2 for each metre gained in height. Thus, if a weighing instrument is calibrated to weigh a 1 kg weight and indicate 1,000 g precisely and then is taken to a higher floor, 10 m above, it would indicate about 999.9997 g if not recalibrated.

Evaporation

Whenever weighing a volatile liquid, there is a chance that some will evaporate during the weighing process. Use of a suitable beaker and suitable lid will reduce evaporation to minimal levels. Be sure that the beaker and lid have been cleaned and dried just before use.

Temperature-Stabilised Equipment

If all equipment used in a weighing has been allowed to temperature stabilise, then thermal effects will be minimised. If this has not been done, then items will change temperature relative to each other during weighing and introduce errors.

With weighings having uncertainties of 100 ppm or less, it is good practice to allow all items that will be used in the weighing to acclimatise for several hours. At uncertainties of 25 ppm to 10 ppm, the period should be at least overnight. At more accurate uncertainties, allow an acclimatisation period of at least 24 hours within the environmentally controlled weighing room that such weighings require.

Drift of Weighing Instruments

Be aware that all weighing instruments will drift slightly as different components within them heat and cool with temperature variations around them. Electric weighing instruments will also drift if there are fluctuations in the local electricity supply; in particular, sudden voltage changes will cause some taring circuits and some balancing circuits to be less efficient than otherwise.

Drift effects can be minimised by designing a weighing procedure that brackets each weighing of the item to be weighed by weighings of a standard. Thus the sequence of weighing will be "standard – test item – standard", with the error

in the average of the two standard values being a best estimate of the drift value when the test item was weighed. By subtracting that error from the test item value, a drift-corrected test item value is found. Always start and finish a sequence of weighings with the standard so that if n weighings of a test item are needed, then a total of $2n + 1$ weighings will be undertaken ($n + 1$ of which will be weighings of the standard used).

References

1. NPL. 1954 (reprinted 1962). Balances, weights and precise laboratory weighing. *Notes on Applied Science* No. 7. Middlesex, UK: National Physical Laboratory, Dept. of Scientific and Industrial Research.

2. Metcalf, T. J. 1969. *Weighing machines: Non-self-indicating mechanisms.* Essex, UK: Institute of Weights and Measures Administration (now known as the Institute of Trading Standards Administration).

   Metcalf, T. J. 1969. *Weighing machines: Semi-self-indicating and self-indicating mechanisms.* Essex, UK: Institute of Weights and Measures Administration (now known as the Institute of Trading Standards Administration).

   Griffith, E. H. 1970. *Weighing machines: Applications of electricity and electronics to weighing mechanisms.* Essex, UK: Institute of Weights and Measures Administration (now known as the Institute of Trading Standards Administration).

3. The Institute of Measurement and Control. 1997. *Guide to the Measurement of Mass and Weight.* London: The Institute of Measurement and Control.

Chapter 6

WEIGHTS AND WEIGHING MACHINES

M. J. Buckley
South Yorkshire Trading Standards Unit
Sheffield, United Kingdom

This chapter covers the calibration of weights and weighing machines, with special reference to the pharmaceutical industry. The main subjects covered are traceable measurements, explaining the concept of traceability from the International Prototype Kilogram to the shop floor, and the technical requirements which concern the calibration and use of weights and weighing machines. Quality specifications such

as ISO 9000, GLP (Good Laboratory Practice) and GMP (Good Manufacturing Practice) impose certain requirements on organisations which use weighing as part of their manufacturing or inspection process. The chapter will outline the requirements imposed on organisations which need to meet these standards and will also look at ways of improving quality in weighing.

Metrology and Standards

The science of metrology deals with the accurate measurement of such fundamental quantities as mass, length and time. It is also concerned with the direct derivatives of these quantities, such as area and volume, as well as certain other measurements, such as temperature, barometric pressure and humidity. Table 6.1 details the base units in the International System of Units (SI).

For the precision of any measuring system, there should be only one standard to which all others can be referred. In length measurement, all standards are defined in terms of the passage of light in vacuum during a certain time interval. As a result, the primary standard of length can be freely reproduced throughout the world. In the same way, the standard of time is defined in relation to the caesium 133 atom. However, with the unit of mass, reference is made to a material standard. Because of its nature, it has to be pre-

Table 6.1. SI Base Units of Measurement

Quantity	*Unit*	
Length	m	meter
Mass	kg	kilogram
Time	s	second
Electric current	A	ampere
Thermodynamic temperature	K	Kelvin
Amount of substance	mol	mole
Luminous intensity	cd	candela

served under the strictest conditions of custody, used only very rarely, and then solely for the purpose of comparing it with the secondary standards.

The International Prototype Kilogram

In order to measure a physical quantity, a reference quantity and a measuring instrument are required. The reference quantity is called the unit—the unit of mass is the kilogram. It is defined as the mass of the International Prototype Kilogram, which is held at the Bureau International des Poids et Mesures (BIPM) in Sèvres, near Paris. It is a cylinder of about 39 mm in height and 39 mm in diameter and is made of an alloy of platinum–iridium (90 percent Pt 10 percent Ir), with a density of about 21.5 g cm^{-3}.

In 1878 three platinum–iridium cylinders, which were later identified as KI, KII and KIII, were manufactured by Johnson, Matthey and Company, London. They were polished and adjusted by A. Collot of Paris and compared with the Kilogramme des Archives by four observers in 1880 at the Paris Observatory. The cylinder identified as KIII was found to be the closest to the Kilogramme des Archives and was sent to the BIPM, where in 1889 its use was sanctioned by the first Conférence Générale des Poids et Mesures (CGPM) as the International Prototype Kilogram. In 1882, a further 40 platinum–iridium cylinders were delivered from Johnson, Matthey and Company. These standards were polished, adjusted and then cleaned. They were then compared with the International Prototype Kilogram; as a result 34 standards were assigned to the signatories of the Convention du Mètre for use as national standards. Standards 9 and 31 were assigned to the BIPM as working prototypes, and KI and number 1 were retained as check (témoins) standards with the International Prototype. Prototype number 18 was assigned as the United Kingdom National Standard.

Calibrating the National Standards

Article 6 of the Convention du Mètre provides for periodic comparisons of the national prototypes with the International Prototype or the témoins. The first periodic calibration of the National Pro-

totypes was carried out between 1899 and 1911. This comparison was carried out without any preliminary cleaning. The results showed that the mass of most of the prototypes (used as national standards) had changed little compared to the value of 1889. In particular, the mass of those which had not been used was often found constant within a few microgrammes. Certain prototypes showed, however, visible signs of wear or accident.

The second periodic calibration in 1939 of National Prototypes was preceded by a comparison of the International Prototype with the check standards. All the standards were cleaned before weighing with a chamois skin impregnated with alcohol and then with redistilled benzene. The results of the measurements were not satisfactory. The four témoins and working prototype 9 showed an increase of mass of 30 μg to 80 μg relative to their values in 1889. Working prototype 31, whose base carried numerous lines and traces of rubbing, had conserved its original value. It was suggested that whilst the International Prototype is kept under three glass domes, of which the largest is furnished with a stopcock by which a partial vacuum is made and which stood on a ground glass plate, the other prototypes are only placed on their support under two domes which rest on a metal plate. Changes of air around these standards have been greater, and it was possible that deposits on the weights have occurred which, had they been cleaned in steam or alcohol

vapour, would have been removed. Studies were then made into the cleaning of the weights. It was decided to use steam to clean the weights, followed by dusting with a fine hair brush.

In 1946 a new comparison was made against the International Prototype, the six check standards and the two working prototypes. These comparisons now showed changes no greater than −30 to +40 μg from the 1889 values. Following further studies into the effects of air pollution and improved diamond machining techniques, several new prototypes have been recently constructed. In the last few years, the third verification of National Prototypes has been undertaken. On this occasion, the International Prototype Kilogram was itself cleaned. Subsequently, measurements were made over a long period, and the results were plotted to obtain a mass value for the International Prototype in a zero cleanliness state.

Other Materials and Other Definitions

Research is currently being pursued into the construction of mass standards in a material less expensive than platinum–iridium. In particular, stainless steel containing about 20 percent nickel and 20 percent chromium has been used for the construction of several National Prototypes. The density of this alloy is around 7.8 g cm^{-3}; the volume of these weights is about 82 cm^3 greater than that of the platinum–iridium prototypes. As

a result, when the mass of a stainless steel kilogram is determined by comparison with a platinum–iridium prototype, the different values of air buoyancy provides for a significant correction factor. This correction cannot be calculated to better than around 50 μg by reason of the limited precision with which the density of air is known as a function of ambient conditions of air pressure, temperature and humidity. With any material standard, there remains an element of doubt of the size of the physical quantity of the unit. The value of a material standard which represents a base unit may change with time. To attempt to overcome this, developments are taking place, under the co-ordination of the Comité Consultatif pour la Masse (CCM), to produce a new definition of the unit of mass in terms of a freely reproducible natural standard. These developments have not yet achieved the precision that is readily obtainable by existing methods of mass calibration:

- The kilogram is defined as a precise number of atoms or molecules, e.g., 2,741 . . . 1035 particles of orange-red radiation of krypton 86 corresponding to the transition of the levels $2p^{10}$ and $5d^5$.
- The kilogram is N_A kmol—1/12 times the mass of the atoms of the nuclide ^{12}C.

The established value of the Avogadro constant is given as

$$N_A = 6.0221367 \times 10^{-23}/\text{mol}.$$

Disseminating the Unit of Mass

In national laboratories, local trading standards authorities and in laboratories of industrial and research organisations, high-precision mass standards and test weights, known collectively as weights, serve as reference standards for the dissemination of the unit of mass or for high-precision weighing operations. To ensure the highest precision of mass standards, with corresponding uncertainties of measurement, a hierarchical structure of mass standards has been established. This structure ensures that the wear and possibility of damage to the primary standards is kept to a minimum.

The National Measurement System

The National Measurement System (NMS) is the technical and organisational infrastructure which ensures a consistent and internationally recognised basis for measurement in the United Kingdom. The central objectives of the NMS are

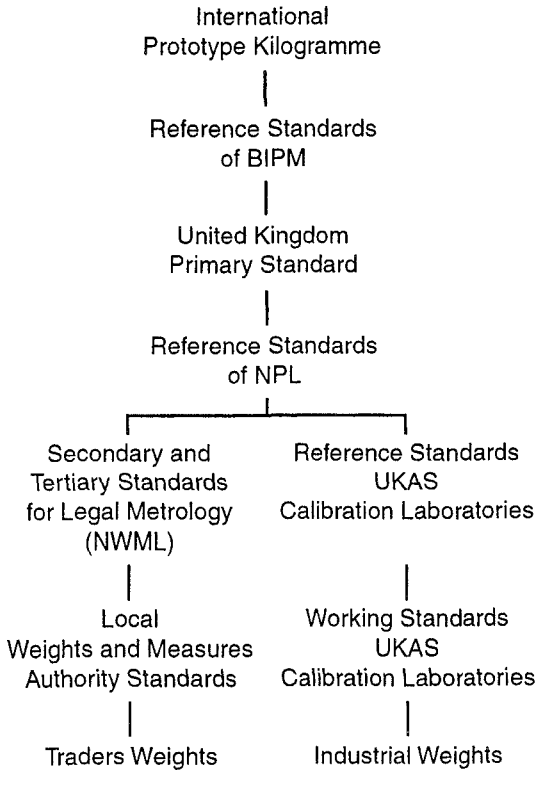
- to enable individuals and organisations in the United Kingdom to make measurements competently and accurately and to demonstrate the validity of such measurements, and
- to co-ordinate the UK's measurement system with the measurement systems of other countries.

At the heart of the NMS lies the physical measurement standards, which are realised in accordance with the internationally recognised SI definitions. For each unit, either a base unit or a derived unit as defined by the SI system, there exists a primary standard designated the UK National Standard. An unbroken chain of documented calibrations links the primary standard, through reference standards, to working standards for everyday use. The ability to relate an individual measurement back via successive calibrations to the national standard using recognised measurement procedures and practices is called "traceability" and is the principle on which the integrity of the NMS is founded (Figure 6.1).

UKAS, The United Kingdom Accreditation Service, formerly part of the National Physical Laboratory (NPL) and formed by the amalgamation of the British Calibration Service and NATLAS, is the means by which almost all industrial calibration and testing laboratories are able to become part of the NMS. UKAS is the main channel for the dissemination of measurement standards to British industry. Accreditation by UKAS is no trivial matter. The requirements include

- operation of an approved quality system and maintenance of an appropriate quality manual and operational procedures;
- establishment of traceability of measurements;

Figure 6.1. Traceability of the Unit of Mass



- participation in regular surveillance and reassessment exercises;
- maintenance of satisfactory laboratory accommodation, facilities and equipment;
- use of documented procedures and suitable verified equipment;

- use of appropriately qualified and trained staff to carry out the specified measurements; and
- maintenance of a satisfactory record system.

The existence of an internationally recognised NMS also creates opportunities for mutual recognition of accreditation organisations such as UKAS, which helps to overcome technical barriers to international trade. Such agreements ensure that a UKAS certificate will be accepted as equivalent to a certificate issued by an organisation which has been accredited in another country, thus removing the need for repeated testing or calibration. The NMS also helps the United Kingdom to assist in the establishment of similar measuring systems in developing countries.

Mass Standards and Weights

Both in legal and industrial metrology, weights are classified according to international standards—in this case the International Recommendation RI 111, issued by the International Organisation for Legal Metrology (OIML). Weights are classified not only on tolerance but also on material of manufacture, form and construction, surface finish, density and magnetic susceptibility. Furthermore, reference is made to "conventional mass" rather than true mass (or

mass in vacuum). Conventional mass is defined in terms of the mass required to balance at 20°C a hypothetical weight of density 8 000 kg m^{-3}, which is weighed in air at a density of 1.2 kg m^{-3}. The concept of conventional mass is one which is adopted in almost all measurements involving weight. Its use means that we can compare similar objects in average environmental conditions without making corrections for the density of the standard weight and the object being weighed, the air temperature, humidity or pressure. All weights and weighing machines are calibrated and used on this basis. The only exceptions to the use of conventional mass is in the calibration of primary standard weights and in the use of mass to determine liquid volume, force or pressure units, where reference is made to true mass (or mass in vacuum).

Table 6.2 shows the maximum permissible errors for weights of nominal mass 50 kg to 1 mg according to OIML RI 111.

OIML Recommendation RI 111 specifies not only the maximum permissible error from nominal but also the form, material and other requirements for each class of weights. For weights of nominal value 50 kg to 1 g the main requirements are as follows:

- *Class E1 and E2:* Integral stainless steel weights without markings or adjustment chambers
- *Class F1*: Stainless steel weights, which may have a screw knob

Table 6.2. Maximum Permissible Errors for Weights

Nominal Mass (g)	*Class E1 (± mg)*	*Class E2 (± mg)*	*Class F1 (± mg)*	*Class F2 (± mg)*	*Class M1 (± mg)*
50 000	25	75	250	750	2 500
20 000	10	30	100	300	1 000
10 000	5	15	50	150	500
5 000	2.5	7.5	25	75	250
2 000	1.0	3.0	10	30	100
1 000	0.5	1.5	5	15	50
500	0.25	0.75	2.5	7.5	25
200	0.10	0.30	1.0	3.0	10
100	0.05	0.15	0.5	1.5	5
50	0.03	0.10	0.30	1.0	3.0
20	0.025	0.080	0.25	0.8	2.5
10	0.020	0.060	0.20	0.6	2.0

(continued on the next page)

(continued from previous page)

Nominal Mass (g)	Class E1 (± mg)	Class E2 (± mg)	Class F1 (± mg)	Class F2 (± mg)	Class M1 (± mg)
5	0.015	0.050	0.15	0.5	1.5
2	0.012	0.040	0.12	0.4	1.2
1	0.010	0.030	0.10	0.3	1.0
0.5	0.008	0.025	0.08	0.25	0.8
0.2	0.006	0.020	0.06	0.20	0.6
0.1	0.005	0.015	0.05	0.15	0.5
0.05	0.004	0.012	0.04	0.12	0.4
0.02	0.003	0.010	0.03	0.10	0.3
0.01	0.002	0.008	0.025	0.08	0.25
0.005	0.002	0.006	0.020	0.06	0.20
0.002	0.002	0.006	0.020	0.06	0.20
0.001	0.002	0.006	0.020	0.06	0.20

- *Class F2*: Weights should be of stainless steel or chrome-plated brass (sometimes called miralloy)
- *Class M1:* Weights may be made of brass (which is free from corrosion or tarnishing) or of painted cast iron

Higher class weights are made from stainless steel, usually 25 percent Cr 20 percent Ni (weights of class E1 are made from a special stainless steel at a density of 8 000 kg m^{-3}), but weights of class F2 may also be made of chrome-plated brass. Brass, although permitted for weights of class M1, is not recommended on account of its tendency to chemical instability—it tarnishes easily in an ordinary atmosphere and being comparatively soft loses mass owing to wear. Lacquer is usually too hygroscopic to be satisfactory and in any case has a tendency to flake away from the metal. Larger weights of class M1 may be of painted cast iron. The OIML recommendation also specifies limits for surface finish, density and magnetic susceptibility. Weights of 500 mg or less are of flat sheet metal, polygonal wire or strip segment, all of a shape appropriate to their value. These weights may be of stainless steel, german or nickel silver, or tantalum. Aluminium should only be used for weights of 100 mg or less (10 mg or less preferred). Duplicate weights are often marked with a *o* or shaped to distinguish them from weights of a similar nominal value.

Construction and Shape

Since it is important to minimise all possible variations in mass due to changes in the surface condition of the weights (such as due to tarnishing, wear or possibly the porous or hygroscopic nature of the material), the shape of the weight is usually designed to give a minimum surface area consistent with convenience in lifting or handling the weight. Clearly, the simpler the shape, the better. The primary standards of mass are either cylindrical or show little modification of the simple cylindrical form. The height of the cylinder is approximately equal to its diameter.

Weights of the highest accuracy classes (E1 and E2), where stability of mass is most critical, are made in one piece with no detachable parts (commonly known as integral weights). Lower accuracy weights are constructed with a cavity, which allows the addition or subtraction of small fragments of homogeneous material in order to make the final mass equal, within certain limits, to its nominal value. Unnecessary edges and sharp angles should always be avoided, and all edges should be well rounded, especially on plated weights. Screw knobs, if fitted, should fit flush so as to prevent the harbouring of extraneous matter and dirt. Where the centre of the base is relieved, as in precision standards, the relief should be slight, and the rim should not be unduly narrow or convex.

Cast iron weights should be free from rust and surface corrosion and sealed to prevent the

ingress of moisture. This is usually achieved by the application of paint which is then baked on. All weights should be thoroughly clean and free from extraneous matter, both externally and within any adjusting cavity. A high degree of polish and finish has the advantage not only of presenting an attractive appearance but also of making foreign matter more obvious and might be expected to lead to greater stability because of the corresponding reduction in the intrinsic surface area.

The Adjustment of Weights

On account of its tendency to be converted into basic lead carbonate, lead should not be used for the adjustment to the nominal value of screw knob weights. For higher grade weights, the adjustment matter should be of the same material as the adjusted weight. In other cases, suitable adjusting materials are tin, brass, nickel–chromium and stainless steel. If a material of greater density is necessary, tantalum or gold cuttings are recommended. Fractional weights are adjusted on the same conventional density basis as used for the larger weights of the same set. Even when they have widely differing densities, fractional weights are not large enough to introduce serious buoyancy errors, provided that no aluminium weights larger than 100 mg are included.

Identification of Weights

An identifying serial number is usually found on the box containing the weights. Identification marks on weights should be restricted to the minimum necessary for purposes of recognition and should only be lightly inscribed by burnishing or depolishing with a stencil. Weights of classes E1 and E2 should bear no indication on the weight of its nominal value, but weights appearing two or three times in sequence should be distinguished from each other to facilitate the recognition of individual weights. Fractional weights in polygonal sheet form, or in strip or wire segment design, are so shaped that within each decade the denomination can be inferred from the shape of the weight:

- Triangle or 1 segment for 1, 10, 100 and 1000 mg
- Quadrilateral or 2 segments for 2, 20 and 200 mg
- Pentagonal or 5 segments for 5, 50 and 500 mg

Constitution of Sets

Decimally constituted weight sets are usually based on one of the following series:

- 5 2 2 1 (the most common series)

- 5 3 2 1
- 5 2 1 1

If the series 5, 3, 2 and 1 is used, the process of selecting the appropriate weights to achieve a balance is more difficult (i.e., it is greater than 10). The necessity of distinguishing between two weights of the same denomination when applying their individual values does not arise, but in some cases the 3 and 2 weights are not easy to distinguish. The last series (5, 2, 1 and 1) makes it more difficult to use as the total sum is less than 10.

Storage of Weights

Precision standards are normally maintained in special storage cupboards, where they can be protected from dust and atmospheric pollution by special glass covers. However, when transported for calibration or used outside the laboratory, weights should be contained in specially built boxes. Wood, particularly mahogany, is the traditional material for the manufacture of such boxes; acidic woods, such as oak, and the use of animal or vegetable glue should not be permitted. Some modern plastic materials are also used for weight boxes. Lining material, often desirable for the protection of the weights, should be free from loose fibres and thoroughly washed before use. Small weights should be secured in a small box to ensure that even the smallest of

weights are confined to their respective housings. Except when removing or replacing weights, boxes should always be kept closed.

Handling of Weights

Special pronged lifters are usually provided for some of the larger weights, but forceps are necessary for the smaller fractional weights. Any lifting device should be covered with a suitable material, such as chamois leather, so that the metal surfaces do not come into contact. Forceps tend to require frequent cleaning to prevent any dirt or dust being transferred to the weights. Large weights and weights made of cast iron may be handled with gloves made of a suitable material, such as chamois leather. Before weights are used, slight deposits should be removed with a soft camel-hair brush or with a lens brush with a bellows. Any other dirt, such as fingerprints due to accidental handling, will require removal with pure alcohol using a stick with a cotton-wool tip. This cleaning, however, may result in significant mass variations because the adsorption layer is changed—it will be necessary to redetermine the mass value.

Calibration of Weights

The value of any weight will change with time and with use. Weights may become lighter through wear and use or may become heavier

due to chemical depositions or atmospheric pollution. Accordingly, it is necessary to have weights of all classes calibrated prior to use and afterward at regular intervals. The periodicity of calibration varies with use and precision. In general terms, it is recommended that weights of classes E1 and E2 should initially be calibrated at intervals not exceeding two years, with all other weights being calibrated on an annual basis. After several calibrations, it should be possible to review this calibration interval based on the actual change in the mass value of the weight between calibrations. In general terms, the value of a weight should not change by more than one-third to one-half of the uncertainty of measurement with which the weight is normally calibrated. If the change in mass value is more than this, the weight is "out of control"—either the weight should be calibrated more frequently or, preferably, the cause for this large change should be identified and rectified. Weights submitted for calibration will normally be calibrated in a mass calibration laboratory. Because of instability in the environment and the lack of adequate mass comparators or high-accuracy balances which may be moved from site to site, it is not possible to establish a mobile calibration service for most types of weights.

Weights sent for calibration are first examined and if necessary cleaned. They are then placed in the calibration laboratory to stabilise for a minimum period of 24 hours. Larger

weights and weights of very high accuracy will require more time in order to acclimatise. Weights will also be checked for magnetic susceptibility, surface finish and density as appropriate. A modern calibration laboratory offering a full calibration service is likely to be equipped with perhaps 20 or more modern mass comparators and automated balances, virtually one comparator for each size of weight. Each mass comparator will be linked, where possible, to a computer for automatic recording of mass values—indeed, for the highest accuracy, the balance will be fully automatic in operation, with weight interchange and recording occurring under computer control. All E1 and E2 weights will be calibrated this way, usually overnight when environmental conditions are likely to be more stable. Where appropriate, weights will be adjusted so that the deviation from nominal is less than that specified for weights of the appropriate class. In all cases, a calibration certificate should be issued, showing the conventional mass value of the weights calibrated and the appropriate uncertainty of measurement. Before and after adjustment values should be stated if requested by the submitter.

The uncertainty of measurement quoted on a certificate is linked to the class of accuracy of the weight and its material, surface finish and construction. The uncertainty of measurement, which takes into account such factors as the uncertainty on the reference standards and

auxiliary equipment being used, as well as the random uncertainty of the weighing process, should not exceed (at the 1 kg level) one-third of the manufacturing tolerance for weights of class E1 and E2 and one-fifth of the tolerance for weights of other classes.

Weighing Machines

Weighing machines may take many forms and include laboratory balances, mass comparators, top pan electromagnetic force compensation balances (the most popular form of modern laboratory balance) and industrial machines (e.g., platform and counter machines, weighbridges and spring balances). All of these weighing machines may be collectively called non-automatic weighing machines (as opposed to automatic check-weighing machines, such as production line weighers). Some weighing machines will be "in use for trade", that is, subject to legal metrology control—in the United Kingdom this can be described as "weights and measures inspection and verification". This chapter will make no specific reference to this.

Weighing machines take many forms, but the most popular is the modern electronic balance, or electro-magnetic force compensation weighing machine. Such balances and weighing machines are relatively easy to use, providing a digital display of the mass value. They are best

used by being installed on-site in the place of use and then left in position. (Some balances are specifically designed for portable use and are designated as such.) The balance should be checked for level, if a spirit level is fitted. The weighing bench should ideally be made of brick (not reinforced concrete) with a granite top. Otherwise, a strong table free from vibration should be used, preferably solely as a weighing table. Balances should be placed if possible in a temperature-stable environment, free from drafts. If this is not possible, a cover or screen around the balance may improve weighing conditions. The balance should be left connected to the electricity supply at all times; if the balance is fitted with a standby mode, this should be activated when the balance is not in use. If the balance has an internal calibration cycle, this should be activated before weighing, each day or shift as appropriate. Modern balances are often fitted with automatic internal calibration cycles, and these should generally be left to operate, so that drift due to temperature or other effects is minimised. Most balances are also fitted with electronic filters, and these should be set (with the help of the manufacturer) to achieve the optimum performance of that balance in the environment in which the balance is used.

It should be noted that the use of an internal calibration cycle does not provide an alternative to a full balance calibration, nor does it invalidate

the results of such a calibration. It does, however, set the span or range of the balance to a standardised value depending on the ambient conditions occurring at that precise moment. When performing a full calibration cycle, therefore, it is imperative that this calibration cycle be activated prior to the calibration taking place. Most modern balances can be supplied with printers or computers with specialist software that can automatically be used to record weighings taking place on the shop floor or in the laboratory. This is an ideal way to keep records which can be used to meet the requirements of quality management systems. Where software is developed in-house, advice should be sought from the weighing machine manufacturer to ensure that adequate data safeguards are incorporated in the programmes being developed. Proprietary software supplied by major balance manufacturers already contains such safeguards.

Calibration of Weighing Machines

All weighing machines, with the exception of spring balances and some simple types of mechanical counter machines, will need to be calibrated in the place of use. In modern electronic weighing machines, this is particularly important because the calibration changes with location due to gravity. In any case, temperature, air pressure and humidity may affect the calibra-

tion, and vibration, thermal gradients and soundness of the floor or laboratory bench may all affect the displayed value. For laboratory balances with a resolution of less than 0.1 mg, it will also be necessary to measure the air pressure, temperature and possibly the humidity and to make air buoyancy corrections. The integrity of the power supply sometimes affects an electronic balance, but this can usually be remedied by a simple electronic smoothing device, such as sold for modern computers. Prior to calibration, electronic balances should be turned on for at least 30 min and ideally several hours prior to the calibration taking place. Some balances have a stand-by mode and should be left permanently connected to the mains supply when not in use.

The frequency of calibration will vary with use and the importance of any calibration uncertainty to the manufacturing process or to the test being undertaken. In general, a full calibration should be undertaken once or twice a year, with a daily, weekly or monthly intermediate check being undertaken by the balance or machine user. In companies where the balance is in full-time use, or where a high degree of certainty is needed, the machine should be calibrated more often. Daily or before use checks can be carried out using a single weight or a small number of weights. The results of these checks should be recorded, with action limits set relative to the importance of the measurement process. Where

this is not easily quantified, a limit of perhaps ±2 digital intervals may be used.

Weights and Other Equipment Used for Calibration

Weights used for the calibration of weighing machines should be appropriate to the accuracy of the machine being calibrated. The uncertainty of measurement on the weights used should be less than the resolution of the balance. However, this is not practicable for weighing machines with a resolution of less than 0.01 mg; in this case, weights of class E2 or E1 should be used. As a general rule, Table 6.3 gives the class of weights to be used for weighing machine calibrations.

The weights required for a calibration should cover the range of the weighing machine. Where a machine is used only over a limited range (for example, up to 100 g on a balance of capacity 200 g), it is possible to calibrate this machine only up to this point, but the machine should then be prominently labelled by the user to provide the operator with this information (e. g., this machine has been calibrated only up to 100 g). Modern electronic weighing machines often require a specific weight which is used to calibrate the electronic range. In some cases, the machine may be provided with an internal calibration weight, which may be of the nominal weight indicated or may be proportional to the indicated

Table 6.3. Weight Classes for Calibrating Weighing Machines

Nominal Capacity	*Resolution*							
	100 g	10 g	1 g	100 mg	10 mg	1 mg	0.1 mg	0.01 mg or less
Up to 200 g				M1	M1	F2	F1	E2 or E1
200 g to 1 kg			M1	M1	F2	E2	E1	E1
1 kg to 30 kg	M2	M2	M1	F2	E2	E1	E1	
30 kg to 100 kg	M2	M2	M1	F2	F1	E2		
Above 100 kg	M2	F2	F1	E2	E1			

value. This calibration cycle should always be activated prior to both calibration and use, but for calibration purposes, this should not be considered as sufficient to calibrate the weighing machine.

When in use weights should be handled with great care. They should never

- be touched with bare hands;
- handled with sharp or abrasive tools or implements which are not clean;
- placed on any surface (other than a balance pan or in their box) which is not suitably covered with acid-free tissue paper;
- slid across metal surfaces (e.g., scale pans);
- knocked together or come into contact with other objects; or
- cleaned (except by an approved method).

Weights should ideally be cleaned only by dusting with a clean fine brush. If you need to clean weights with a solvent cleaner (to remove grease or fingerprints), wash afterwards in distilled water, leave at least a week to stabilise and then have the weight recalibrated. Do not allow weights to come into contact with magnetic sources. Magnetic weights used on modern balances give unreliable results. If you calibrate

balances which have divisions of 0.1 mg or less, or microbalances, you will need to provide equipment for measuring air density. For routine work, you will need the following calibrated measuring instruments:

- Thermometer
- Barometer
- Hygrometer or humidity measuring device

Calibrating Balances and Weighing Machines

There are specific procedures for calibrating weighing machines. The range of tests that will be needed to adequately measure the performance of an individual weighing machine will depend on the design of that machine. However, the tests required are based on several general procedures. It is intended that all tests should be carried out in the environment in which the machine is normally used. The calibration of a weighing machine may vary with use, temperature and location (gravity). Most electronic weighing machines have devices for setting the stability level and the susceptibility of the weighing machine due to vibration. Calibration does not involve the adjustment or resetting of these levels or filters, although they may be reset, if requested by the balance user, prior to the commencement of the test. It is assumed that the

balance has been left on (to stabilise) for some time prior to the test and that the weights to be used in the calibration have been kept in this environment prior to the test commencing. The scale value of the weighing machine should be calibrated. Some machines have an built-in calibration cycle using an internal (uncalibrated) weight. Activate this prior to the test if this is the normal procedure (it would be rare if this was not the case). Otherwise, use an external weight to set the scale value.

It is essential that the balance is checked throughout its range for departures from linearity. This should be done at equally spaced intervals at a minimum of 10 points. This can be achieved by using a suitable set of calibrated weights, the values of which are known with an uncertainty less than the discrimination of the balance. At each calibration point, a weight (or a group of weights) of nominal value equal to the balance reading is weighed in the normal manner. A table of corrections is thus derived to cover the entire balance range. Weighing machines with internal weights should additionally be calibrated at each weight. Repeatability of reading indicates how well the balance will weigh and will depend on several factors. Repeatability tests are therefore essential in monitoring performance. Repeatability of reading is measured by taking a continuous series of readings (normally 10, but on large industrial machines this may be 5) and calculating the

standard deviation and the maximum difference between any two successive measurements.

Although modern weighing machines rarely suffer from hysteresis, a simple check can be carried out. This can be done by approaching a reading from below and above. The tare facility is essentially a mechanical or electronic zero offset. It may in some cases by necessary to re-establish the linearity of scale from the new datum. It is not essential to report the range of the tare facility on a calibration certificate, unless it has been graduated. The effects of off-centre loading (often called eccentricity) may depend on both the magnitude and position of the load and do not vary linearly with either. This test will therefore give only a general indication of how carefully a load must be positioned and to what extent the overall uncertainty must be increased to allow for this effect.

It is not necessary to use weights of accurately known mass for this test. Ideally, the test should be carried out at several loads within the range of the balance, but on machines of commercial design, the test is usually limited to one-fourth to one-third of the capacity of the weighing machine. To avoid possible damage to the weighing mechanism, it is advisable to seek guidance from the manufacturer's literature before deciding on the load to be used for this test. Measurements should be made with the weight placed at several positions on the pan (front,

back, left, right, centre—depending on the support of the load receptor).

The general tests applicable to most weighing machines have been described above. It should be noted that specific designs of weighing machines, as well as the use to which they are put, may dictate additional tests.

Chapter 7

WEIGHING IN THE PHARMACEUTICAL INDUSTRY

Ted Scorer
Glaxo Wellcome
Barnard Castle, United Kingdom

When a pharmaceutical product is prescribed by a doctor or a product, e.g., paracetamol (acetaminophen) tablets, and purchased over the counter at a pharmacy, the patient will be instructed as to the dosage to be taken, e.g., two 500 mg tablets to be taken every 4 hours. This is to ensure that the quantity of the compound will have the desired effect in the treatment of the patient. If the product is underweight, the dose

taken by the patient may be insufficient to achieve the desired effect. However, for some products, if the product is overweight, this may have unwanted side effects on the patient. In the pharmaceutical industry, this means that the dose weight of the final product is a critical factor. A great deal of effort is expended to ensure that the product has been correctly manufactured, filled and packaged.

The Pharmaceutical Industry

The pharmaceutical industry has traditionally been split into primary and secondary manufacturing.

Primary Manufacturing

Primary manufacturing is the bulk manufacture of pure, physiologically active compounds. For some compounds, e.g., penicillin antibiotics, this may involve a fermentation stage, or the starting compound may be naturally occurring. There may then be a number of chemical stages to convert the starting material into the final physiologically active compound. This compound will then undergo a number of purification stages and, finally, if the product is to be given intravenously, there will be a stage to ensure the product is sterile.

During this phase of manufacture, batch sizes can vary from a few grams to over a tonne. In production areas, weighing operations will be used to formulate the ingredients of a batch and, on completion of a stage, to assess the process yield. In a production area, most weighing instruments will be platform scales with a resolution in the range 1 g to 100 g. However, in the small-scale manufacturing of specialist products, balances with a resolution of 1 mg will be used. At all stages of the manufacturing process, regular testing will be performed to assess the purity of the product to ensure its suitability for use. The starting step of the majority of the testing will involve weighing the sample, and this will normally be performed on balances with a resolution of 0.001 mg to 1 mg.

Secondary Manufacturing

Secondary manufacturing is the conversion of pure, physiologically active compounds into a form that a patient can take as a dose (e.g., a tablet, capsule, injection, syrup, cream, ointment or inhaled product). For some products, this may be a relatively simple process. For example, for an injectable product, the compound may be dissolved in water, sterilised and then dispensed under sterile conditions into a vial, ampoule or syringe. For other products, the final product may involve a more complex formulation which

can involve one or more physiologically active compounds being mixed with a number of other compounds. For example, some compounds are so active that the required dose in a tablet form may be only a few micrograms. To allow a tablet to be of a size that is convenient for a patient to handle, the active compound will be formulated with an inert bulking agent, e.g., a starch. Once the product has been formulated, it will then be packaged into an appropriate container, e.g., a tube for a cream or ointment. The container will then normally be added to a carton, an instruction leaflet will be added and a number of cartons will be added to a box to aid in the distribution process.

In the secondary manufacturing process, weighing is a critical operation to ensure that the final dose weight is achieved. In filling/packaging areas, there will be several weighing instruments of various types. A typical pharmaceutical secondary manufacturing area has one of the highest ratios of weighing instruments to staff in all the industry.

Dispensary Operations

In the dispensary, there will be several weighing instruments with a variety of resolutions. A batch of 100 kg may require only a few grams of the physiologically active compound. A typical dispensary may contain up to 4 weighing systems: a platform scale with a resolution of 100 g,

a second platform scale with a resolution of 2 g, a balance with a resolution of 0.1 g, and a balance with a resolution of 1 mg. The operating procedure would instruct the operator which weighing instrument is to be used for each of the batch ingredients and what, if any, tolerance is permitted in the weight dispensed. Several suppliers now offer computerised dispensary software, and this is used in much of the pharmaceutical industry to allow the required level of control of this important step to be achieved.

Filling Operations

The formulated compound then moves to the manufacturing/filling stage, depending on the product presentation. This process will again involve the use of a variety of weighing instruments. Two typical processes are summarised here.

Bottle Filling. For the filling of a bottle of syrup, the fill weight will typically be in the range of 10–100 g and will operate at a line speed of 10–100 bottles per minute. In these situations, the filling line may incorporate an in-line check-weighing system with two load cells. The first will be incorporated prior to the filling station and will determine the tare weight of the bottle. The second will be located after the filling station and will record the gross weight. The net fill weight of the product is obtained by subtracting

the tare weight from the gross weight. This operation allows 100 percent testing to be achieved, and the use of an under- overweight reject gate allows the assurance of fill weight.

Vial Weighing. For the weighing of a vial for injection, the fill weight will typically range from 250 mg to 5 g, and line speeds will be on the order of 100–300 vials per minute. The weight tolerance will be such that the resolution of load cells is insufficient to achieve the required accuracy, and the speed of the line would prohibit 100 percent testing. In these applications, a byline balance is used by a line operator to perform periodic checks of the fill weight.

Packaging Operations

In the packaging stage, an in-line load cell may be present which will indicate that the carton contains all the required components. A balance will be used to ensure that the box contains the requisite number of cartons.

Laboratory Testing

At all stages of secondary manufacturing, extensive testing of the product is performed in the laboratory areas, and large numbers of balances are used with a weighing resolution of 0.0001 mg to 1 mg.

CHECK–WEIGHING

During secondary manufacture, check-weighing is the means used to ensure that the weight of the product is within the permitted limits. Check-weighing can be divided into two types: destructive and non-destructive.

Destructive Check-Weighing

In destructive check-weighing, the sample, having been weighed, cannot be returned to the process.

Fill Weighing

Fill weighing is typically employed on the check-weighing of vials, ampoules or bottles. The filled container is placed on the balance, and the balance is tared. The container is removed and emptied using a vacuum source to ensure that all the contents have been removed. The empty container is returned to the balance and, ignoring the minus sign, the fill weight is displayed.

Tablet Weighing

A container is added to the balance, and a number of tablets are individually added to the container. If the weight recording is being performed manually, the balance will be tared before each tablet

is added to allow the tablet weight to be seen. In the majority of the pharmaceutical industry, tablet weighing is performed using a software package. In this situation, the taring of the balance between additions is normally omitted, and the tablet weight is calculated by the weight increase after each addition. This allows a greatly reduced test time on the repeated taring operation, where the balance has to stabilise at each tare step. In applications where multiple tablet compression machines are being served by a central test point, the tablets must be destroyed on completion of the weighing. In areas where the check-weighing is being performed in the compression booth with a dedicated check-weigher, precautions, if appropriate, are taken to satisfy the regulatory authorities that no contamination can occur; thus it may be possible to return the tablets to the process.

Shot Weight Testing

For an aerosol or aqueous nasal spray product, the initial check-weighing will be to ensure that the total contents of the container, the physiologically active compound and, if applicable, the propellant, are within permitted limits. This will normally be performed using a pre-tare (see below, "Pre-Tare Weighing") weighing technique. In addition, it is necessary to test the weight of the shot to ensure the correct operation of the

actuator. A number of different tests exist for this type of weighing, but typically an aerosol is weighed, a number of shots is fired and the aerosol is reweighed. The average shot weight can then be determined. A number of shots is then fired, and the test repeated to confirm operation of the actuator in the middle and/or at the end of shelf-life.

Non-Destructive Check-Weighing

Destructive check-weighing can be a significant cost to the manufacturer. For example, a vial filling line with 12 filling ports will require that 12 vials be sampled every 15 minutes. If the filling line operates a double filling shift to permit 16 hours of filling per day, 5 days per week, for 45 weeks per year, 172,800 vials must be sampled per year. For many pharmaceutical products, a vial is worth several pounds, and so the cost of testing is very significant. As a result, a number of non-destructive techniques have been developed for many products.

Pre-Tare Weighing

Pre-tare weighing is typically used in bottle filling. A number of bottles are removed from the line, and a numbered collar is placed around the neck of the bottle. The bottle is then weighed to obtain the tare weight and is then returned to the

line for filling. After the filling operation, the bottle is again weighed, and the fill weight is calculated by the difference in weight. Pre-tare weighing cannot be used when filling sterile products because empty bottles that have been sterilised cannot be handled until they have been hermetically sealed.

Average Tare Weighing

Average tare weighing is used when the variation in weight of the container is small as compared to the fill weight. Typical applications are filling of dry products into sachets. At the start of each batch, a number of empty sachets will be weighed to determine the average weight. During production, the filled sachets are weighed, and the fill weight is calculated by subtracting the average sachet weight from the gross weight.

The Regulatory Environment

The pharmaceutical industry is one of the most regulated industries. For a company to launch a new pharmaceutical product, it must be granted a licence by the appropriate regulatory body. In the United Kingdom, it is the Medicines Control Agency (MCA); in the United States, it is the Food and Drug Administration (FDA). If a company is awarded a licence for the product, then

regulatory inspectors from these agencies (or other applicable agencies) must successfully perform a pre-approval inspection of the proposed manufacturing facilities before manufacturing can commence. This inspection will determine that the buildings and equipment are suitable for the task, that the staff are appropriately qualified and that written procedures are in place to cover all production and support operations. These procedures will include the operation, calibration and performance checking of all weighing instruments.

Specifying Weighing Instruments

When a user requires a new weighing instrument, he or she will be required to document a user specification. This will include the maximum load and resolution of the instrument and may specify the size of the weighing compartment and, for automated instruments, the speed of operation. Before the order is placed, a design qualification (DQ) will be performed to ensure that the weighing instrument ordered will meet the requirements for which it is being purchased.

Installing Weighing Instruments

When the instrument is delivered, it will be subjected to a series of documented actions. The installation qualification (IQ) will test the

functionality of the instrument. The operational qualification (OQ) will document the testing of the instrument's performance and will include a calibration of the instrument using weights traceable to a national standard. The instrument will then be subjected to a performance qualification (PQ) to demonstrate its functionality in the operating environment. IQ, OQ and PQ testing ascertain that the requirements of the user specification have been met.

Performance Checking

In the pharmaceutical industry, companies must maintain written procedures for checking the performance of all weighing instruments used in production and laboratory areas. This will specify the maximum interval at which checks should be performed and what testing is to be performed. Certified or check weights may be used for the testing.

Certified Weights

Certified weights are constructed to an approved design and tolerance as specified in OIML (International Organization for Legal Metrology) documentation. These weights must be certified in an accredited facility so that the certification is traceable to an international standard. Certification of the weights must be performed at appropriate intervals, e.g., every six months.

Check Weights

Check weights are constructed from an appropriate material, e.g., stainless steel. A value for the weight(s) will be known for a specified weighing instrument. For example, if a balance has been tested using a certified weight, then the check weight's value can be recorded for that instrument. If a check weight is to be used for more than one weighing instrument, then its weight must be obtained for each instrument. The value must be obtained at appropriate intervals when the instrument is tested with certified weights.

Testing of Precision

Precision testing will use one or more check weights to ensure that the weight displayed is within the permitted tolerance of the weight used for the test. Weighing equipment manufacturers will specify the precision for a balance; for a typical analytical balance, the precision will be similar to balance resolution. To specify a limit for routine testing of a balance, allowance will have to be made for any variation in the conditions of the test, which are typically set at five times the balance resolution. For some specific applications, it may be necessary to apply tighter limits; for some coarse applications, wider limits may be appropriate.

Hysteresis Testing

Hysteresis testing is used to ensure that when the load is removed from the weighing instrument, the display will return to zero. Typically, the limit for this test will be no more than twice the resolution of the weighing instrument.

Reproducibility Test

One weight should be added to the weighing instrument a minimum of five times. The maximum permitted deviation will be specified.

Linearity Test

Two or more weights are used to ensure that the equipment is linear over the range to be tested. For modern force compensation balances, the tests should be at approximately 25 percent and 75 percent of the maximum load of the balance—the area of the force compensation cell where the maximum deviation will occur.

Eccentricity Test

The eccentricity test, also known as corner load error, is used to determine the accuracy of the instrument if a load is applied off-centre. The test is performed by measuring a weight, typically at 50 percent of the balance load, on the centre of the weighing pan and then repeating the measurement at the periphery of the weighing pan. A maximum permitted deviation will be specified. For the manufacturers of weighing instruments,

very few will actually specify an error for eccentricity, although most instruments will have an in-house limit used by service engineers.

The tests to be performed and the frequency of the tests will vary depending on the type of weighing instrument. Some tests, such as the eccentricity test, will be performed infrequently, e.g., at six-month intervals, whereas the precision test may be performed several times per day.

Internal Calibration Weights

Many modern electronic balances contain one or more internal weights used to perform calibration and testing of the balance. Some models contain a temperature sensor. For these models, if the temperature changes by a specified amount, typically 1°C for a balance with a resolution of 0.1 mg, the internal calibration will be automatically instigated. For balances without a temperature sensor, or where the user has elected to disable this feature, a facility will be provided to allow the operator to initiate the balance test and/or calibration using the internal weights. Some balances contain multiple internal weights, allowing linearity checking to be performed, and some have the facility to use the internal weight to perform the reproducibility test. For balances containing these features, the frequency of manual checking may be reduced but cannot be completely removed, as there is normally no traceability of the internal check weights to international standards.

Non-Automatic Weighing Instruments Directive

As part of the harmonisation of regulations leading to the formation of a single market within the European Economic Community (EEC), later to become the European Union (EU), a number of directives were issued to harmonise regulations within the member states. The purpose of the regulations is to ensure that items approved for sale in one member state of the EEC would be acceptable in all other states.

The Non-Automatic Weighing Instruments (NAWI) Directive, 90/384/EEC, was approved by the council on 20 June 1990 and later amended by directive 93/68/EEC. Article 1(2)(a) of Directive 90/384/EEC defined six applications that would be covered by the directive:

1. Determination of mass for commercial transactions.
2. Determination of mass for the calculation of a toll, tariff, tax, bonus, penalty, remuneration, indemnity or similar type of payment.
3. Determination of mass for the application of laws or regulations, including expert opinions given in court proceedings.
4. Determination of mass in the practice of weighing patients for the purpose of monitoring, diagnosis and medical treatment.

5. Determination of mass for making up medicines on prescription in a pharmacy and determination of mass in analyses carried out in medical and pharmaceutical laboratories.
6. Determination of price on the basis of mass for the purposes of direct sales to the public and the making up of pre-packages.

In UK Statutory Instrument 1995 No. 1907, the Non-Automatic Weighing Instruments (EEC Requirements) Regulations 1995 came into force on 1 September 1995, although a derogation order meant that it would not be enforceable until 1 January 2003.

Legal Balances

In a number of EU states, the NAWI Directive extended the areas in which a "legal" or metrologically approved balance could be used. All balances currently sold within the United Kingdom must comply with EU low-voltage and radio-frequency (RF) interference regulations and thus are marked with the CE mark. Balances manufactured to be metrologically approved are in addition marked with the green M sticker. Such balances are known as "verified balances", where the accuracy of each model has been tested by an authorised body, and when deemed acceptable, the appropriate stickers are applied.

In order for a balance to be verified, the manufacturer must ensure that the relevant criteria

for approval are inherent in the design. Such requirements include, for example, seals to prevent an operator or non-qualified person from opening the balance, hatching over or inserting brackets around digits on the display for information only, e.g., 22.6(3) g may be reported only as 22.6 g. Designated mass units for legal metrology must be available only as grams and kilograms, not as ounces, pounds, pennyweights and so on; mg may be displayed only on balances with a resolution better than 0.1 mg. The manufacturer will then submit a typical model to a notified body, such as National Weight and Measures in the United Kingdom, or the PTB in Germany, for pattern approval. This consists of a series of metrological tests to prove the performance of the balance and its ability to cope with the environment to which a prospective owner may subject it. Such tests include out of level weighing, the ability to withstand changes in temperature and RF testing.

When pattern approval is given, it may well be that a 200 g analytical balance to 0.1 mg readability was submitted but may also include models in the same series with 100 g and 60 g capacity as long as the weighing systems and electronics are similar. When pattern approval is granted, the manufacturer can then make balances for approval. There are two stages of verification. Stage one is always done by the manufacturer at the place of manufacture. After satisfying the relevant criteria, a "blank"

metrology sticker is applied to the balance, indicating that it meets the requirements for legal metrology and can pass to the next stage. This sticker will have information such as the authorisation number given to the manufacturer by the notified body.

Stage two can be carried out only by a suitable qualified person. This may be the manufacturer or its approved agent or, in the United Kingdom, an approved trading standards officer. Stage two involves calibration and adjustment, if required, as well as other metrological data that may effect accuracy. When satisfied, the green M sticker may be applied. Where balances have internal weights and the manufacturer can demonstrate no deterioration during transport, then the second stage approval may also be performed at the place of manufacture. This means that the end user has only to unpack the balance, warm it up for the stipulated time and apply the internal weight to correct for any variation in regional gravity before the verified balance is ready for use.

Within the United Kingdom, most people noticed only the change to the labelling of prepackaged meat and vegetables sold in supermarkets (application 6 above). A much smaller number of people, myself included, were effected by application 5, where the determination of mass in a pharmaceutical laboratory required the introduction of legal balances into the laboratory and the addition of trading standards

officers to the list of those who would be involved in inspecting the laboratory.

By 1996, very few people in the pharmaceutical industry had encountered the NAWI Directive, and many of those who had would not spend a lot of effort on a requirement that was several years from becoming enforceable. As a result, in the earlier years of the NAWI Directive, very little progress was made. The major balance manufacturers, who had predicted significant orders from the pharmaceutical industry for legal balances, were left with unwanted stock.

The NAWI Directive and associated standards documents specify the requirements for an instrument to be metrologically approved. Companies that design and manufacture weighing instruments largely meets the requirements. For the instrument user, the regulations specify the minimum limits that can be applied to the testing of the instrument. To allow the applicable limit to be determined, you need to determine the class of the balance. The directive splits balances into four classes: special, high, medium and ordinary. Table 7.1 shows the accuracy classes, and the basic requirement is based on the verification scale interval.

Verification Scale Interval

The verification scale interval can initially be a confusing term, as many people assume that the verification scale interval is the same as the

Table 7.1. Accuracy Classes

Class		*Verification Scale Interval* (e)	*Number of Verification Scale Intervals*		
			Minimum Capacity	*Minimum Value*	*Maximum Value*
I	Special	$0.001\ g \leq e$	$100e$	50,000	No limit
II	High	$0.001\ g \leq e \leq 0.05\ g$	$20e$	100	100,000
		$0.1\ g \leq e$	$50e$	5,000	100,000
III	Medium	$0.1\ g \leq e \leq 2\ g$	$20e$	100	10,000
		$5\ g \leq e$	$20e$	500	10,000
IV	Ordinary	$5\ g \leq e$	$10e$	100	1,000

balance resolution. For most weighing instruments, the verification scale interval is one decimal place less than the resolution. The exception to the rule is that for a balance with a resolution of less than 0.1 mg, the verification scale interval will be 1 mg because it is considered that insufficient accuracy exists in check weights with a resolution of less than 1 mg. Table 7.2 illustrates examples of linking the resolution of a weighing instrument to the verification scale interval. From Table 7.2 it can be seen that the majority of balances in a pharmaceutical testing laboratory will be Class I with some in Class II.

Table 7.2. Verification Scale Intervals

Resolution (d)	*Verification Scale Interval* (e)
0.000001 g	0.001 g
0.00001 g	0.001 g
0.0001 g	0.001 g
0.001 g	0.01 g
0.01 g	0.1 g
0.1 g	1 g
1 g	10 g
2 g	10 g
10 g	100 g

Maximum Permitted Errors

Table 7.3 shows the maximum permitted errors at the installation of a balance. The limits for in-service testing are twice the limits at the time of installation. Two examples of this are as follows:

> Example 1: An analytical balance of Class I with a resolution of 0.00001 g is being tested with a load of 50 g. The verification scale interval will be 0.001 g, and the maximum permitted in-service error will be ±0.001 g.
> Example 2: A top pan balance of Class II with a resolution of 0.001 g is being tested with a load of 100 g. The verification scale interval will be 0.01 g, and the maximum permitted in-service error will be ±0.02 g.

In the pharmaceutical industry, a typical limit for a precision test will be set at five times the balance resolution. Using this limit in examples 1 and 2, the assigned limits would be ±0.00005 g and ±0.005 g respectively, which are significantly tighter than those required by the regulations.

A further anomaly exists within the pharmaceutical industry, where a verified balance must be used in those areas designated as quality control laboratories. Balances with resolutions of 0.00001 g as indicated in example 1 or even 0.000001 g (or 1 μg) are typically found in such

Table 7.3. Maximum Permissible Errors

Load (m)				*Maximum Permissible Error*
Class I	*Class II*	*Class III*	*Class IV*	
$0 \leq m \leq 50{,}000e$	$0 \leq m \leq 5{,}000e$	$0 \leq m \leq 500e$	$0 \leq m \leq 50e$	$\pm 0.5e$
$50{,}000e \leq m \leq 200{,}000e$	$5{,}000e \leq m \leq 20{,}000e$	$500e \leq m \leq 2{,}000e$	$50e \leq m \leq 200e$	$\pm 1.0e$
$200{,}000e \leq m$	$200{,}000e \leq m \leq 100{,}000e$	$200{,}000e \leq m \leq 10{,}000e$	$200e \leq m \leq 1{,}000e$	$\pm 1.5e$

areas, and under the NAWI Directive approval can be given only to, at best, the third decimal place of any such balance. In the earlier example of a verified balance in a jewellers shop, the jeweller is weighing gold indicating 22.6(3) g. The weight quoted to the customer must be 22.6 g, as digits below the verified digit are for indication only. What are operators supposed to report on a balance that indicates 2.263111 g where the last verified digit is the 3 [i.e., 2.263(111) g]? Until now, no absolute answer has been given.

Chapter 8

METHODS FOR ESTIMATING THE UNCERTAINTY OF ELECTRONIC BALANCE MEASUREMENTS*

John P. Clark
A. Harper Shull
Westinghouse Savannah River Company
Aiken, South Carolina

International and national regulations are requiring that testing and calibration laboratories provide estimates of uncertainty with their measurements. Many balance users are having questions about determining weight measurement

uncertainty, especially if their quality control (QC) programs have provided estimates of measurement system "bias and precision." Part of the problem is the terminology used to describe the quality of weight and mass measurements. Manufacturer's specifications list several performance criteria but do not provide estimates of the "uncertainty" of measurements made using an electronic balance. Several methods for estimating the uncertainty of weight and mass measurements have been described in various publications and regulations in recent years. This chapter will discuss the terminology used to describe measurement quality, i.e., "accuracy," "precision," "linearity," "hysteresis," "measurement uncertainty" (MU), and the various contributors to MU, and will discuss the advantages and limitations of various methods for estimating MU. The methods include using

- the manufacturer's specifications at specified conditions,
- balance calibration data, and
- measurement control program data.

Examples of each will be discussed.

Electronic balances have become so sophisticated that many calibrate themselves and appear to provide "error free" measurements. However, this is not actually the case. All measurements have error that obscures the true value. The error creates *uncertainty* about the quality of the

measured value. For a measurement to have value, it should have an estimate of the magnitude of error associated with it. An estimate of the error allows the user to evaluate the quality of the measurement. However, many measurement-generating organizations have questions about determining accurate estimates of the *uncertainty* associated their weight measurements.

The metrological community recognized the need for standardization in the area of measurement uncertainty over 20 years ago and developed standards and guides on the subject. Generally speaking, industrial calibration and testing organizations have lagged behind in reaching the same levels of understanding and implementation attained by legal and scientific metrology communities. (However, if they want to participate in international business, they are being required to catch up.) The specific area of weighing has its own MU concerns. Much has been written about the sources of error in weighing instruments, methods for controlling error, and methods for estimating the *uncertainty* of weight measurements. We are unaware of any consensus about the best method for determining estimates of weight MU. This chapter will review the sources of error and methods that are currently being used to "estimate" MU.

Westinghouse Savannah River Company (WSRC) calibrates more than 400 scales and balances at a government-owned, company-operated site near Aiken, South Carolina, called

the Savannah River Site (SRS). Regulations and the WSRC quality assurance (QA) program require all balances to be calibrated routinely with standards traceable to the national standards. Many of the site operations have quality control programs, which require operators to satisfactorily weigh two or more standards on a balance before making process measurements. Control charts are often used to record and evaluate the QC measurements. Biases are determined by subtracting the measurement means from the reference values of the mass standards. The variance of the measurements is used to estimate the precision of the balance. Locally, bias and precision statistics have been considered the MU information for balances.

The site also requires measuring and test equipment (M&TE) to be calibrated with standards having one fourth the uncertainty of the M&TE. Since weighing equipment specifications do not include the term *uncertainty*, many users look at the (manufacturer's) specifications (see Table 8.1) to find something similar and often use the standard deviation (s) of the repeatability as the uncertainty. This can cause a problem if the person calibrating the balance tries to ratio the balance standard deviation to a weight certificate's uncertainty statement, which is usually given as a standard uncertainty multiplied by a k factor of 2 or 3.

Weighing is considered one of the simplest measurements in a laboratory or production

Table 8.1. Mettler Toledo Balance Specifications

Model	*AT201*	*PR10003*
Readability	0.00001 g	0.001 g
Maximum capacity	205 g	10100 g
Linearity	0.12 mg	10 mg
within 10 g	0.03 mg	—
Repeatability (*s*)	—	2 mg
0–50 g	0.015 mg	—
50–200 g	0.04 mg	—
Temperature drift	1.5 ppm/°C	1 ppm/°C

facility. Procedures are required for all operations and provide basic techniques of operation and care but do not specifically address MU. Balance operator training covers good laboratory practices and basic principles for weighing. However, MU is not usually addressed.

The importance of knowing measurement quality is increasing. Measurement users are now requesting "uncertainty estimates" from measurement providers, rather than bias and/or precision estimates. This has caused some confusion, as MU is more than "bias and precision" estimates. Many balance operators are unaware of the vast amount of information metrologists have developed about every facet of weight and mass measurements.

Much of that information has been distilled and published in literature available from

balance manufacturers. Two such publications are quoted in defining some of the measurement quality terminology associated with weighing and manufacturers' specifications. Many of the definitions are similar to those in the international standards but have been adapted to weight or mass measurements. A good understanding of the basic terminology is essential in evaluating MU estimation methods.

Weight Measurement Terminology

The Sartorius Corporation's publication, *Fundamentals of Weighing Technology (Terms, Methods of Measurement, Errors in Weighing)*, is an excellent reference. Pages 23–28 provide the following explanations for some of the terminology used to describe various aspects of weight measurement quality [1]:

- **Accuracy:** A qualitative concept that defines the metrological extent to which the weight readouts of a weighing instrument approach the true values of the quantities being weighed. Accuracy is quantified by a weighing instrument's readability, standard deviation, resolution, accuracy class, or the uncertainty of measurement. Accuracy is validated with a calibration certificate.
- **Hysteresis:** At a constant load, the displayed value depends on the previous load.

Quantitatively, hysteresis is expressed as the difference between the readouts obtained when the same load is weighed once following a lighter load and once following a heavier load. In terms of weighing instruments, hysteresis occurs particularly with strain-gauge load cells and weighing instruments subject to mechanical friction.

- **Linearity error:** Also referred to in specifications as "linearity." It is the deviation from the theoretically straight-lined (linear) slope of two interdependent values. For weighing instruments, this means the positive or negative deviation of the readout from the actual load on the pan, when the zero point and the span have been correctly adjusted.

- **Repeatability:** The ability of a weighing instrument to display corresponding results under constant testing conditions, when the same load is repeatedly placed onto the weighing pan in the same manner. In general, the standard deviation or the difference between the largest and the smallest result for a defined number of measurements is used to specify repeatability.

- **Standard Deviation:** A mathematic quantity for evaluating a weighing instrument in terms of its reproducibility or repeatability. The standard deviation (s) is defined as:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - x_{\text{bar}}\right)}$$

where n is the number of individual results x_i and x_{bar} is the arithmetic mean of the individual results x_i. To determine the standard deviation with sufficient certainty, the number of times the measurement is repeated must be high enough (at least six times).

- **Uncertainty (of measurement):** The uncertainty of measurement (u) specifies the range for a measured value, within which the unknown, error-free result lies, usually with a statistical uncertainty of 95 percent. (This corresponds to $u = 2s$). An example of a weighing result expressed along with the uncertainty of measurement is $M = (394.27 \pm 0.02)$ g.

A second publication that provides a more comprehensive list of definitions of weighing terminology is the *Mettler-Toledo Glossary of Weighing Terms (A Practical Guide to the Terminology of Weighing)* [2]. It is the source of the following definitions:

- **Precision:** A qualitative term as a judgment regarding the metrological features of a balance. A better designation would be tolerance limits, standard deviation, and/or uncertainty.

- **Reproducibility:** Extent of the approximation between the results of measurements

of the same measured variable with the individual measurements being performed under different conditions with regard to, for example, the

- measurement method,
- observer,
- measurement equipment,
- measurement site,
- application conditions, and
- time.

A valid statement of reproducibility requires specification of the different conditions. Reproducibility can be specified quantitatively by the result scatter.

- **Uncertainty of measurement:** The uncertainty in the measurement of a result always includes random errors (mathematically expressed by the standard deviation or the confidence interval) of all individual variables, which are used to calculate the measurement result, as well as systematic errors, which have not been determined because they cannot be measured and can therefore only be estimated. It is always presupposed that those systematic errors that have been determined have also been corrected. Basically, the result of a weighing series y consisting of n individual weighings should be as follows:

$$Y = \bar{x}_E + u$$

where $\bar{x}_E$ is the mean error that is no longer afflicted by the detected systematic errors (e.g., air buoyancy) and u is measurement uncertainty. The uncertainty in the measurement of a specific measurement result can be characterized by the confidence interval of the mean value derived from n individual values:

$$u = \frac{t}{\sqrt{n}} \times S + |f|$$

where $|f|$ is the estimate of not detectable or not detected systematic error; $t/\sqrt{n}$ is the value that takes into account the distribution of the individual values and the number of weighings, and which can be taken from tables for the selected statistical certainty; and s is the standard deviation.

These weighing technology terms will be referred to later in this chapter. Additional terminology excerpts from the *ISO Guide to the Expression of Uncertainty in Measurement* [3] are listed to facilitate the evaluation of the different methods of estimating weight MU.

Classification of Components of Uncertainty

In general, terms that are specific to this Guide are defined in the text when first introduced. However, the definitions of six of the most important specific terms are given here for easy reference.

2.3.1 **standard uncertainty** uncertainty of the result of a measurement expressed as a standard deviation

2.3.2 **Type A evaluation (of standard uncertainty)** method of evaluation of a standard uncertainty by the statistical analysis of a series of observations

2.3.3 **Type B evaluation (of standard uncertainty)** method of evaluation of a standard uncertainty by means other than the statistical analysis of a series of observations

2.3.4 **combined standard uncertainty** standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities

2.3.5 **expanded uncertainty** quantity defining the interval about the result of a measurement, within which the values that could reasonably be attributed to the measurand may be expected to lie with a high level of confidence

NOTES

1. Expanded uncertainty is referred to as overall uncertainty in paragraph 5 of Recommendation INC-1 (1980).

2. To associate a specific level of confidence with the interval defined by the expanded uncertainty requires explicit or implicit assumptions regarding the probability distribution characterized by the measurement result and its combined

standard uncertainty. The level of confidence that may be attributed to this interval can be known only to the extent to which such assumptions may be justified.

2.3.6 **coverage factor** numerical factor used as a multiplier of the combined standard uncertainty in order to obtain an expanded uncertainty

NOTE—A coverage factor, **k**, is typically in the range 2 to 3.

Precision is a general term that encompasses the concepts of repeatability and reproducibility. There is an important distinction between repeatability and reproducibility that could have a monumental effect on the estimate of *u*. *Repeatability* defines the variability observed at one point in time by one operator recording a series of readings from placing one object on a balance several consecutive times. The standard deviation of this collection of measurements captures the rounding error and the ability of the balance to repeat the same "reading." *Reproducibility* defines the variability of a measurement system in measuring the same standard or object over a long period of time, under different environmental conditions and with different operators. All of these variables can contribute error to the measurement and expand *u*. Since measurements are collected over a long period of time, the *s* of the average measurement will include the environmental effects. The first two

methods that are used to estimate u use the *repeatability* statistic, rather than a *reproducibility* statistic. This will be seen later.

Quality Weighing

Prior to determining u of weight measurements, quality requirements must be established. Several activities or programs should be in place, including the following:

- Determination of accuracy/uncertainty requirements
- Proper selection of weighing instrument
- Maintenance and calibration of weighing instrument
- Proper location of weighing instrument
- Recognition of sources of weighing errors and applications of Good Manufacturing Practices (GMPs) and Good Laboratory Practices (GLPs)
- Application of proper weighing procedure
- Establishment and maintenance of a measurement assurance program

Weighing System Components

After determining the accuracy/uncertainty requirement and selecting the appropriate electronic balance, the other requirements for quality

weighing must be considered. The best way to do this is to look at everything that is involved in making the measurement. In addition to the balance, the standards, environment, operator, procedure, and material being weighed make up the measurement system. Each can contribute to the u of weight measurements. Within each component of the measuring system are potential sources of error that can make the measured weight uncertain. Each component has some error associated with it. However, there are usually less than six components that contribute >90 percent of the u to a measurement. Listing the components and estimating the amount of error each could contribute to the MU is considered an error budget.

The construction of a table of components that affect measurement quality is the first step in establishing an error budget. This requires an in-depth knowledge of the measurement system and/or investigative effort to quantify each contributor. This is an uncertainty analysis. The ultimate objective is to combine them into one statistic that estimates MU.

Table 8.2 is an example of a typical weight measurement system. This list has five main components in a system that involves making direct measurements. Other lists might contain the procedure or method used to make measurements, e.g., direct, indirect, or substitution weighing.

Table 8.2. Measurement System Components

Component	*Potential Error*
Balance	Calibration
	Repeatability
	Linearity
	Hysteresis
	Corner loading error
	Rounding
	Drifting
Standards	Wrong class/out of tolerance
	Magnetic
	Different temperature
Environment/installation	Temperature
	Humidity
	Vibrations
	Drafts
Operator	Inadequate training
	Biased data selection
	Poor techniques
Material weighed	Hygroscopic/evaporates
	Electrostatic charge
	Magnetic
	Density difference
	Temperature

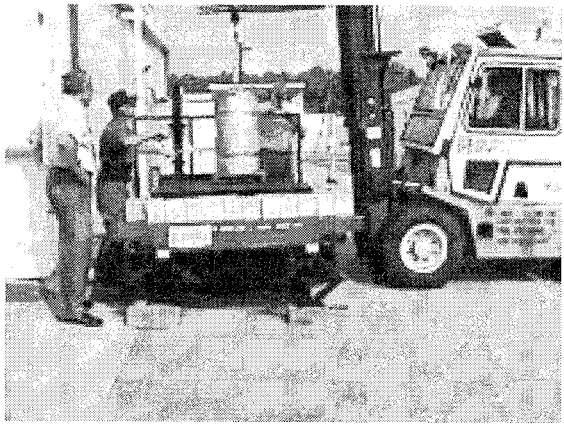
In making weight or mass measurements, an unknown object's weight is directly or indirectly compared to the known weight of a stainless steel standard or the force equivalent. The standards have u that is transferred to all the measurements made against them. This fact is the basis for guidelines or regulations that require using calibration or check standards having uncertainties that are much smaller than the measuring instrument. Variation in each of these components produces random and/or systematic variation in the measurand that causes u. To make the most accurate measurements, ways must be found to minimize the sources of random and systematic variation that affect the accuracy/uncertainty of the measurand.

After the sources of variation are minimized, it is still necessary to know the magnitude of the MU. Reliable estimates of MU allow users to determine if a measurement system is fit for its intended use.

Figure 8.1 shows a tank calibration being performed on trailer-mounted industrial balance outside on a warm day. The balance operator is using 50-pound Class F weights on a balance having 10,000 scale divisions. Many sources of error in the measurement system components will determine the final MU.

There are at least three methods currently being used to estimate MU: doing a type B uncertainty analysis, designing experiments or performing balance calibrations to empirically

Figure 8.1. Weighing System Located Outside



determine estimates, or determining estimates from measurement control program data and calibration certificates.

Method 1: Manufacturer's Specifications

Most manufacturer's specifications list several performance criteria but do not provide estimates of the balance "uncertainty" or accuracy. Balance users often use the *s* of the repeatability as the "uncertainty" of the balance. In specifying accuracy requirements for weighing, some authors have used *s* as a basis of calculating the minimum number of scale intervals allowed to attain a certain confidence level. Using the repeatability *s* as an estimate for MU fails to include the contributions from the other error sources given in a balance specification (see

Table 8.1). All of the performance specifications have error. They should be considered. Two examples of using this information are given.

Example 1. A Mettler Toledo publication, *Determining Weighing Uncertainty from Balance Specifications* provides a detailed method for using a balance's error sources to estimate the uncertainty of its measurements [4]. The author lists readability, repeatability, nonlinearity, sensitivity, and temperature coefficient of sensitivity as potential errors. They correlate to the balance and environment components given in Table 8.3 Systematic errors due to the balance's transfer characteristics are eliminated through adjustments after assembly or they are measured and stored in the balance and can be compensated by signal processing algorithms. Design or calibration compensates for some sources of deviations. This method is derived from the probability theory that allows variances from uncorrelated sources to be added. The resulting sum may be considered the variance of all the influences considered.

The paper [4] has three sections:

1. Sources of measurement deviations and uncertainties
2. Determination of the combined measurement deviation
3. Determination of the combined measurement uncertainty

Table 8.3. Uncertainty Components of an AT 201 Balance

Data Sheet Specifications	*Value (SPC)*	*Variance (SPC)*
Readability	0.01 mg	1×10^{-10} g^2
Repeatability < 50 g	0.015 mg	2.3×10^{-10} g^2
Repeatability 50–200 g	0.04 mg	1.6×10^{-9} g^2
Nonlinearity w/in 10 g	0.03 mg	9×10^{-10} g^2
Nonlinearity w/in 200 g	0.12 mg	1.4×10^{-8} g^2
Sensitivity accuracy	1.5 ppm	2.3×10^{-12} g^2
Temperature coefficient	1.5 ppm/K	2.3×10^{-12} g^2
Ambient temp. excursion	2 K	4 K^2

Balance manufacturers cannot control the error effects that the balance operator, measurement process, environment, and objects or material being weighed have on the total uncertainty of a measurement. Therefore, this method focuses on the balance's error contributions to the MU estimate and makes assumptions about the balance performance and some of the other sources of error. It is assumed that a balance performs within specifications, is operated according to good laboratory practices, and is used for determining small weights near capacity.

An example from the paper is paraphrased below. The assumptions minimize contributions from other components. These include making a small net weight near the capacity of balance so the repeatability specification for 200 g can be used. Also, no corrections are made for the

degrees of freedom. An AT201 model balance is used with a 200 g capacity and readability of 0.01 mg.

A sample of 1 g shall be weighed in a 190 g container. What is the resulting uncertainty of this weighing, conforming to a 95% confidence level? The formula valid for the combined normalized standard deviation for a single sample weighing is:

$$s_{rel} = [1/m^2(SPC_{RP}{}^2 + 2/3SPC_{NL}{}^2) + 1/3(SPC_{CAL,\,rel_2} + 1/3(SPC_{TCS} \times d_t)^2)]^{0.5}$$

$$s_{rel} = [1/(1\ g)^2\ (1.6 \times 10^{-9}\ g^2 + (2/3)\ 9 \times 10^{-10}\ g^2) + 1/3\ (2.3 \times 10^{-12}) + 1/3(2.3 \times 10^{-12} \times K^{-2} \times 4K^2)]^{0.5}$$

$$s_{rel} = [2.2 \times 10^{-9} + 1/3(2.3 \times 10^{-12} + 3.1 \times 10^{-12})]^{0.5} = (2.2 \times 10^{-9} + 1.8 \times 10^{-12})^{0.5} = 47 \times 10^{-6}\ g$$

Conclusion: The mass of a 1 g sample, weighed in a 200 g container, can be determined on this balance with a relative standard deviation of approximately 0.05 mg ($s_{rel} < 50 \times 10^{-6}$ g). Based on a confidence level of 95%, the corresponding expanded uncertainty would be ~0.10 mg ($u_{rel} = 2 \times S_{rel} = 1 \times 10^{-4}$).

These calculations were verified with the commercial software package "Uncertainty Analyzer" from the Integrated Sciences Group. Details about the software are available at http://www.isgmax.com. That program included degrees of freedom and produced a result that

agreed within 5 percent. This software is useful for calculating uncertainty estimates using type B error or heuristic estimates in a manner that complies with the requirements of MU standards.

Example 2: Page 48 in the Sartorius publication [1] gives the following example for calculating the uncertainty of measurement:

> Small amounts (approx. 5 g) are to be weighed on an analytical balance with a resolution of 0.1 mg. Ambient conditions are good (no incline, temperature difference of 5°C max; none of the containers or objects is electrostatically charged, nor is there any electromagnetic interference). The containers are small and must be correctly centered, as directed in the standard operating instructions.
>
> With the exception of the reproducibility/standard deviation, all values are maximum errors. If the equation of $u = 2\,s$ is used to express the maximum uncertainty of the reproducibility and if the air buoyancy has been corrected, the uncertainty of the measurement will be as follows:

Source of Measurement Uncertainty	*Standard Deviation*
The reproducibility/standard deviation is:	≤0.1 mg
The temperature coefficient for the sensitivity is $\leq 2 \times 10^{-6}$, as stated in the technical specifications. Hence, the error for 5 g and $^{\circ}\Delta T = 5^{\circ}C$ is $\leq 5\text{ g} \times (2 \times 10^{-6}/C) \times 5^{\circ}C =$	≤0.05 mg

The maximum linearity error is as stated in the technical specifications:	≤0.2 mg
The balance has been calibrated and adjusted with a standard E2 class weight of 200 g (maximum error of 0.03 mg). In relation to a 5 g load is:	≤0.0075 mg
The sample's density is 2.0 g/cm^3, with an uncertainty of 20%; the difference between the air buoyancy of the samples and the standard is thus 2.25 mg (systematic error). The uncertainty of this air buoyancy correction value due to a fluctuation in the density of 10% is:	≤0.225 mg
And the uncertainty due to the assumed fluctuation in the sample's density of 20% is:	≤0.45 mg

$$u = \sqrt{(2\times 0.1\text{ mg})^2 + (0.5\text{ mg})^2 + (0.0075\text{ mg})^2 + (0.225\text{ mg})^2(0.045)^2}$$
$$= 0.58\text{ mg}$$

However, if no correction is made for air buoyancy, a systematic error of 2.25 mg is added to the uncertainty of measurement "u" so that the total deviation can be as much as 2.83 mg.

The second example looks beyond the balance operation and the assumptions about the measurement process to point out a large systematic error in the direct weighing. In this case, the total uncertainty is more than five times the estimate derived from combining errors from the balance and material being weighed.

Method 2: Balance Calibration Data

Calibration of Non-Automatic Electronic Weighing Instruments [5] is a guideline document for DKD laboratories. (DKD Laboratories operate under the DKD standards writing group in Germany.) It has several procedures for calibrating balances with single, multiple, and adjustable ranges. An estimate of uncertainty can be determined from the collection of data from the linearity and corner loading tests, estimates of the temperature range and the temperature coefficient, the class of standards, the operating range, and the magnitude of the object being weighed. The guideline document is divided into a general part, a second part for single or multiple range balances with 1,000,000 or less scale divisions of 1 mg or larger, and a third part for single or multiple range balances with $> 1{,}000{,}000$ scale intervals and/or with scale intervals ≤ 0.01 mg.

Several examples are provided in the reference. The procedure recommends the minimum number of weighings that should be taken for each error source and provides examples of the mathematical relationship between the scale reading and the various sources of error. The examples use data from balance calibrations that are plugged into the math models to calculate the MU based on empirical data take from a balance in its environment and under conditions the operator would encounter. These estimates capture

more of the error sources than method 1 and verify the balance is performing within specified operating conditions.

At the SRS, a spreadsheet-based calibration procedure is used. A copy of the balance calibration report is shown in Figure 8.2. The report satisfies the WSRC QA requirements for calibration of M&TE. It includes the standards used and their uncertainties and determines a ratio of the standard's u at $k = 1$ to a maintenance limit, which is usually two scale divisions.

The calibration report includes graphs of the calibration measurement deviations and a summary of the observed bias and s observed at four points over the weighing range. An estimate of the balance u is given in the third group of numbers on the upper right side of the report. The MU estimate is given in scale divisions and as a relative percent of the mid-range reading. An estimate of the balance's combined uncertainty is determined in a spreadsheet using the following equation:

$$u = \sqrt{\left[\frac{\max(u_s)}{1000\times 2}\right]^2 + \left[\max\left(s_p{}^2, s_m{}^2\right)\right]^2 + \left|\frac{\max(\text{biases})}{\sqrt{6}}\right|^2 + \left|\frac{\max(\text{clbs})}{\sqrt{6}}\right|^2}$$

where u_s is the 2 std. dev. mg. uncertainty of a standard weight, s_p is a pooled std. dev. from replicate measurements, s_m is the maintenance

Figure 8.2. Standards Laboratory Balance Calibration Report

QA BALANCE CERTIFICATION

4/27/00

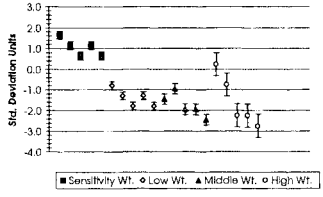
BALCAL Rev. 8.2, 4/25/00 — Reference Procedure: 13.8-3000, "Balance Pipet Calibration/Certification

Next Certification due date: 4/27/01

Analyst Information		Calibration Weight Set Data				Visual Inspection Parameters		
							YES	NO
Calibrator Initials:	GCC					Damaged		*
Calibrator ID #	W0770	SRS #	SL-346b	Exp 10/1/2001		Level	*	
Date:	8/22/00	Manufacturer	Mettler			Clean	*	
		Serial #	C26676		Bal./Wt.	Complete	*	
BALANCE INFORMATION				Uncert.(mg)	Uncertainty			
Lab / Room #	725-A/133		Nominal Wt.	App. Mass	at 2 SD	Ratio		
MODEL #	AE163HIGH	HighWt.	100.0	100.000593	0.033700	11.87	Function Tests	
Balance ID #	Surplused-163	HighWt2	50.0	50.000240	0.019200	20.83	PASS	FAIL
QA Limit (%)	0.10	MidWt.	100.0	100.000593	0.033700	11.87	Power *	
Sen. Limit (g)	0.0006	LowWt.	50.0	50.000240	0.019200	20.83	Load *	
1 Bal. Maint. SD	0.00020	Sensitivity	1.0	0.999771	0.004400	90.91	Taring *	
0.2 grams minimum quantity that can be weighed to maintain QA 95% Limits.							P/NDisplay *	

MEASUREMENTS				Cornerload Data (@ 50% Capacity)			Total Uncertainty		
High(e) Wt.	Middle Wt.	Low Wt.	Sensitivity	Position	Reading	T-Statistic	K- value (Std. Devs.)	Grams	Percent
150.00090	100.00030	50.00010	1.00010	Position	Reading	T-Statistic	1	0.00023	0.00023
150.00070	100.00040	50.00000	1.00000	Top	100.00070	2.29	2	0.00045	0.00045
150.00040	100.00020	49.99990	0.99990	Right	100.00030	0.30	3	0.00068	0.00068
150.00040	100.00020	50.00000	1.00000	Bottom	99.99960	3.19			
150.00030	100.00010	49.99990	0.99990	Left	99.99970	2.69			

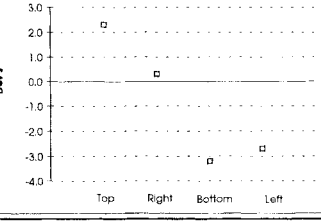
Plot of Bias Deviations in Balance SD's



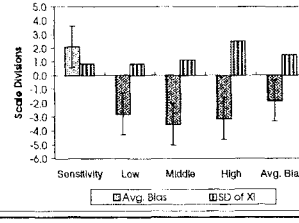
STATISTICS				
	High Wt(e).	Mid Wt.	Low Wt.	Sensitivity
AVERAGE	150.000540	100.000240	49.999980	0.999980
Bias (abs)	-0.000313	-0.000353	-0.000280	0.000209
% Bias	-0.000313	-0.000353	0.000560	0.020865
% RSD	0.000167	0.000114	0.000167	0.008367
SD(abs)	0.000251	0.000114	0.000084	0.000084
Total SD	0.000201	0.000201	0.000200	0.000200
Calculated t	Not Significant	Not Significant	Not Significant	Not Significant
Pooled Abs. Std. Dev., 16 df =		0.000149992	Tab. Chi Sq	32
Tabled t, 16 df (99% Conf.) =		2.583	Cal. Chi Sq	9.00
Tabled F, 2, 16 df (99% Conf.)=		6.23	F Statistic :	Not Significant

CONTROL LIMITS				
	Mean WT.	2 DEV'S	Low Limit	High Limit
High Wt.	150.00054	0.000540491	149.9999995	150.0010805
Middle Wt.	100.00024	0.000304136	99.99993586	100.0005441
Low Wt.	49.99998	0.000261081	49.99971892	50.00024108
Sensitiv. Wt.	0.99998	0.000260785	0.999719215	1.000240785

Plot of Cornerload Devs. from Average



Plot of Biases, 95% Maint. Error Bars & SDs.



ACCEPTANCE CRITERIA	
Summary:	Balance passes QA Limits
Linearity(Bias)	Balance passes
Precision (SD)	Balance passes
Corner Loading	Balance passes
Lack of Fit	Balance passes
Disposition	AS LEFT = AS FOUND

std. dev., biases are the differences between std. values and replicate averages, and clbs are the corner loading biases from the center of the balance. *u* is in grams.

This local balance calibration procedure has evolved over the past several years and attempts to provide as much useful information from the calibration effort as possible.

Method 3: Measurement Control Program Data

The first method estimates the MU for a target weight under given conditions using Type B estimates of uncertainty from the manufacturer's specifications and other sources. The method assumes the balance is operating within the manufacturer's specifications, and that other sources of *u* contribute minimally to the observed weight reading of a balance. If the assumptions are correct, this is a reasonable estimate of MU. If the other components of the measurement process are not adequately controlled and/or the balance is not functioning within the manufacturer's specifications, the MU estimate will be too small.

The second method is a combination of Type A and B estimates of MU from the balance and standards components. Experimental data are collected during the balance calibration over the operating range of the balance. If the balance performs outside the manufacturer's specifications, yet satisfies the customer's measurement quality requirements, the calibration process

provides objective data. This method captures the balance performance based on weighing standards. Calibrations are usually performed by an outside organization. Therefore, the estimate does not include potential sources of error from the balance operator, material being weighed, variations in the environment, and so on.

The third method uses measurement control program (MCP) data, which are based on the *reproducibility* of the measurement system to demonstrate measurement performance within predefined control limits. The control program should duplicate the measurement process. In the example given below, two measurement processes are considered. The first is the standards laboratory's mass calibration procedure. The second is a chemical laboratory's gravimetric pipette calibration program.

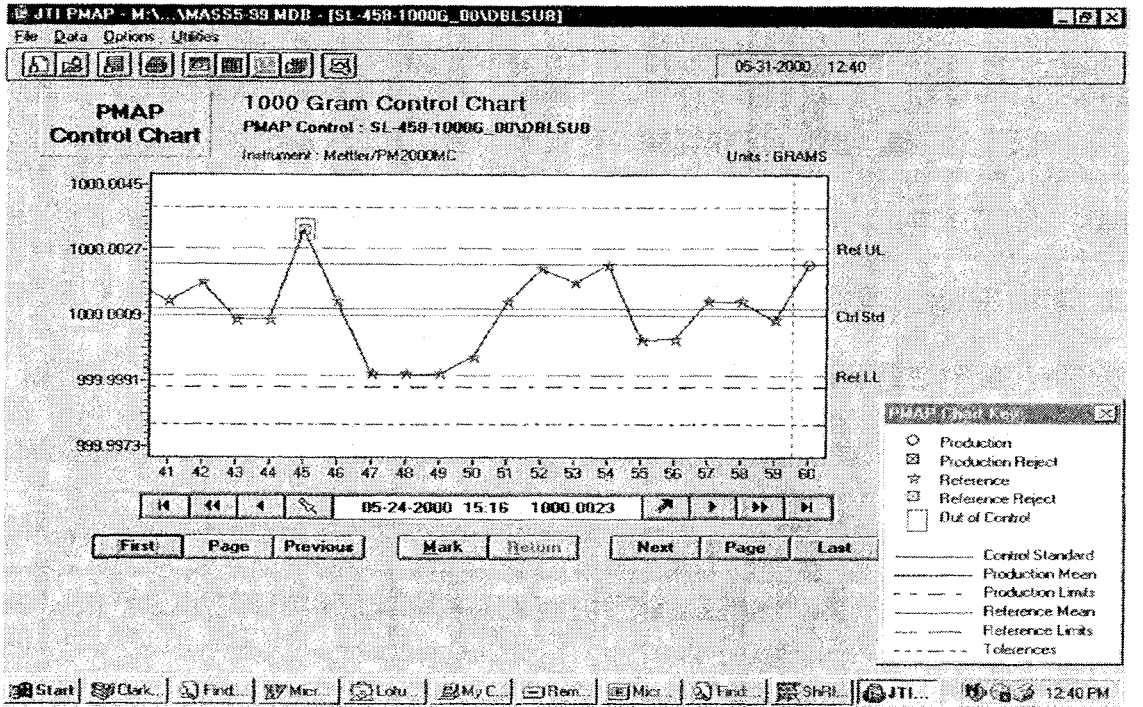
We have described in detail how to determine "real-time" estimates of MU in a paper presented at the "Weighing and Measurement in the Year 2000" international conference organized by the South Yorkshire Trading Standards Unit in 1997 [6].

Commercially available software, JTIPMAP™, was evaluated in a demonstration program for the Department of Energy and reported in document WSRC-MS-96–0032 [7]. (Excel Partnership Inc., JTI Systems has developed a process measurement assurance program (PMAP™) that provides statistical process control (SPC) for the measurement process and calculates estimates of MU

from the MCP data.) Details about the software can be found at http://www.jtipmap.com/index.htm. Some of the screen dumps from the software are shown below as examples of how control charts are set up and measurement control (MC) data are evaluated to maximize the information available in an MCP. This software, like any commercially available software, has its strengths and weaknesses. The SRS has several types of MCPs and uses commercial and custom-developed software in the different laboratories. None are endorsed or recommended as a standard for this or any other government-owned site. Many of the PMAP™ principles are applied in SRS MCPs.

Setting up a PMAP™ MCP involves selecting an artifact that represents the measurement process and calibrating it, if it has not been calibrated recently, and then using it as a check standard (CS) to qualify a measurement system prior to using it to generate measurements for a customer. Once a PMAP™ is established, MC data are entered each time an operator uses a measurement system. The results of each CS measurand are evaluated against up to three sets of control limits in the JTIPMAP™. They are shown on Figure 8.3. The computer-based control charts are dynamic. They are updated with each additional data point, and the current means of the control data are updated with each new computer entry. The charts are constructed to track data from two classes of operators. The first class is called the reference group. The metrologist or

Figure 8.3. QC Chart of a Weight Calibration



specialist who is well qualified usually generates this class of data. If production personnel, who are often less skilled than specialists, use the measurement system, their results are entered as production data.

The software is user-friendly and prompts for basic information in setting up the initial PMAP™ control charts for a measurement procedure. Figure 8.4 shows the first of four pages from the pull down menu to be completed. The first page prompts for information on method and measurement, specifies the number of significant figures, and indicates the units. Only the light boxes on the pages need to be completed at setup. Check or control standard measurements can be expressed as the following: actual values, a ratio of measured to reference values, or a deviation from the reference value.

Figure 8.5 shows page 2, which requires the following information for the control standard: the reference value (1 if normalizing control data), systematic uncertainty (drift for use and time) for the standard(s), variability in the number of *s* used in the control charts (random error or precision), and documentation of the identity of a CS that may be used off-line as an independent reference standard. The calibration values obtained on the CS be tracked on a separate PMAP™.

Page 3 in Figure 8.6 contains information on three sets of limits and scaling information for the PMAP™ control chart. Reference limits are

Figure 8.4. PMAP™ Chart Method and M&TE Basic Data

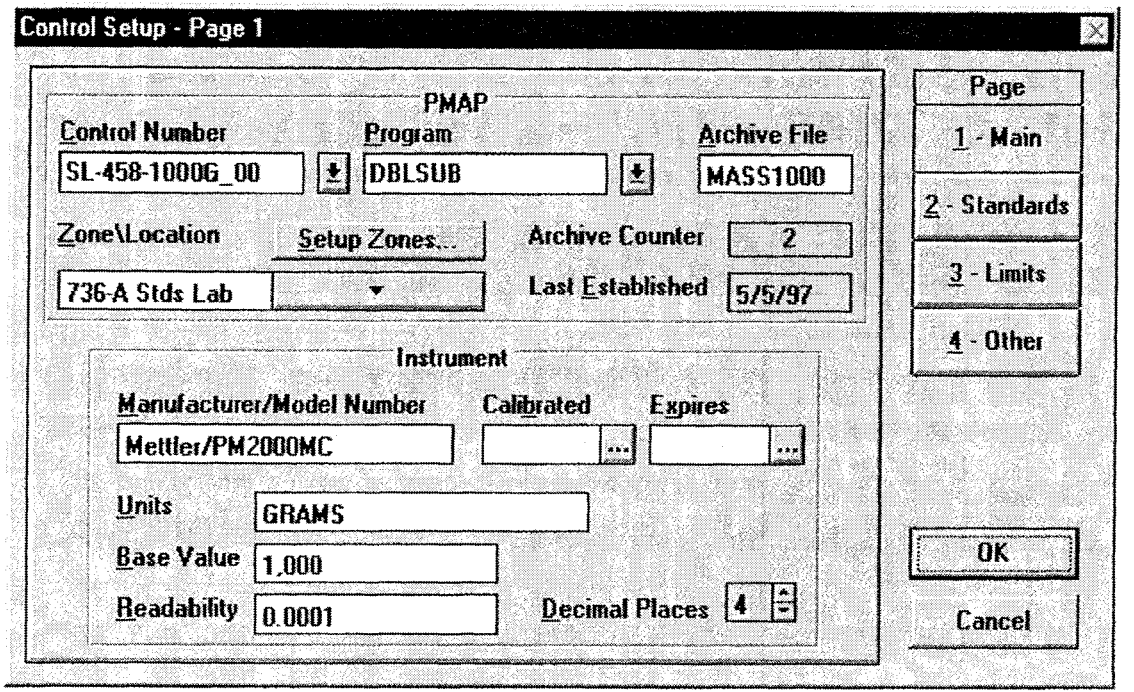


Figure 8.5. Control Standard Statistics

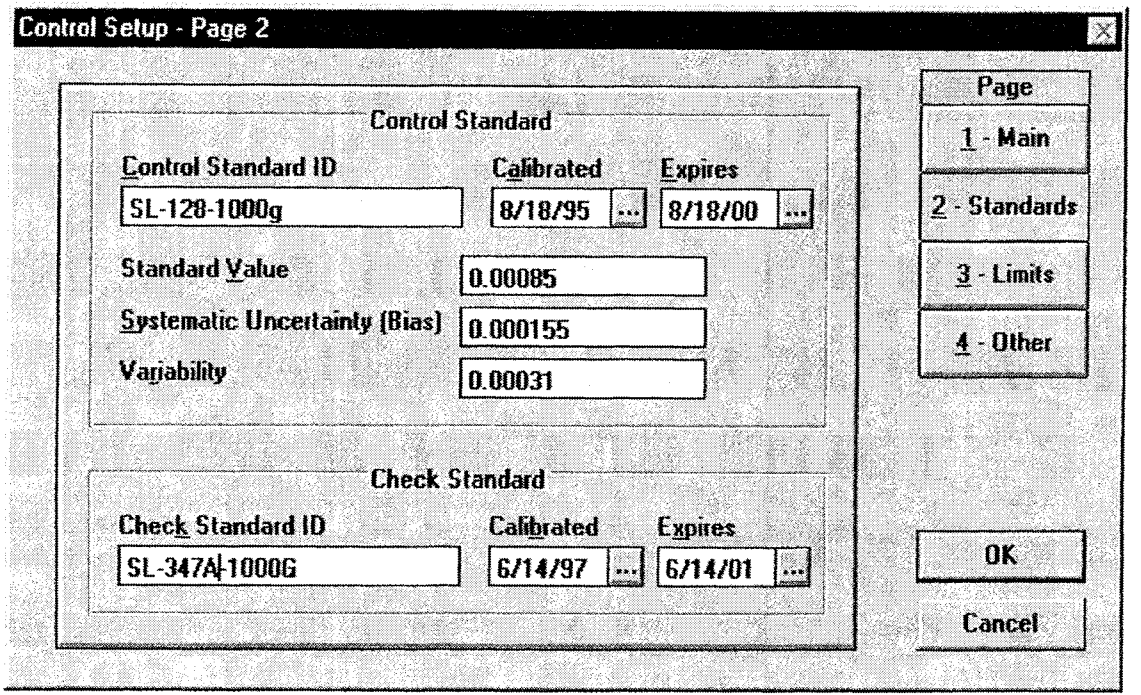
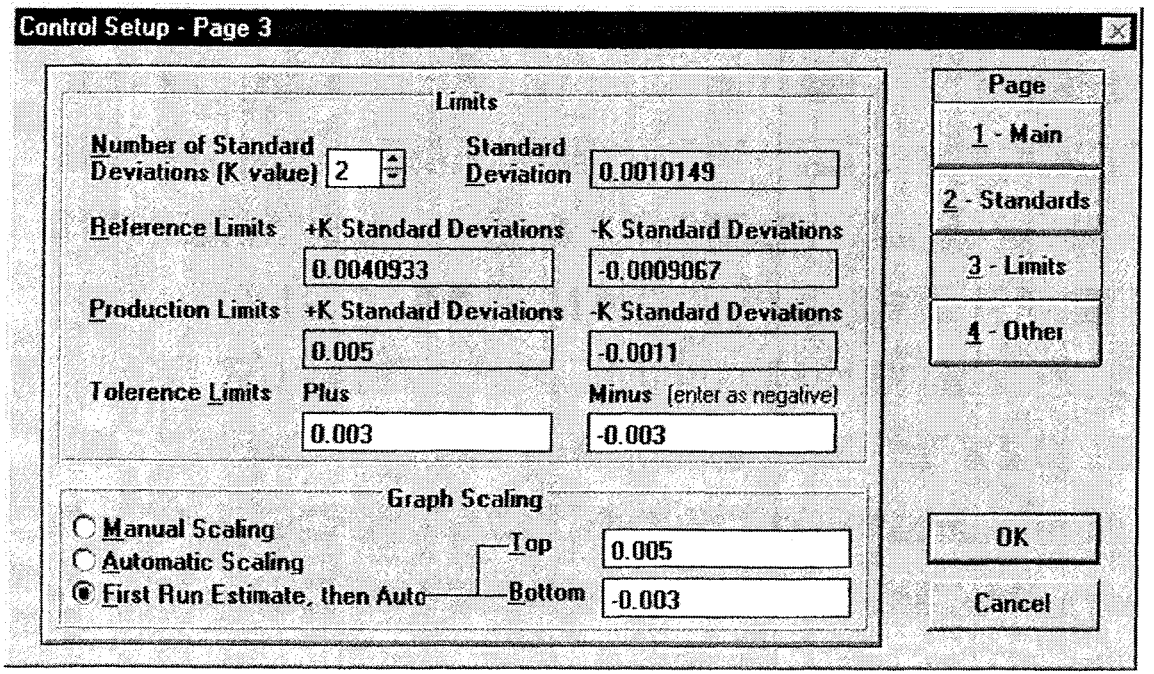


Figure 8.6. Control Chart Limits and Scaling



required (they are normally set by the best operators, who fully understand the effects of variables on the method is performance); production limits are optional (all other operators make production measurements under lab conditions); tolerance limits represent the maximum permissible error that can be tolerated by the customer.

Figure 8.7 shows page 4, which is used for inputting optional information for the graphs of the CS data. Each control chart will display the latest 20 measurements on the computer screen but will print up to 40 control measurements per page. Once this page is completed, all its information will be included on all charts. This information may include the metrologist in charge of the method and the person to be contacted if the method goes out of control. The method ID and related information should be included here, so it will provide all the information necessary on each and every control chart that is printed for hard copies that might be stored as permanent records.

Figure 8.8 shows the MU for the 1 kg mass standard calibration values. It can be calculated at any time with all data collected since the last calibration of the measurement system. The random s of the CS and the current set of data are squared, summed, and the square root taken. Equations are shown below in the excerpt from the PMAP™ operating manual [8]. This is an estimate of the standard uncertainty of the mass values obtained by a double substitution measurement system. It is multiplied by a coverage

Figure 8.7. General Information for Control Chart

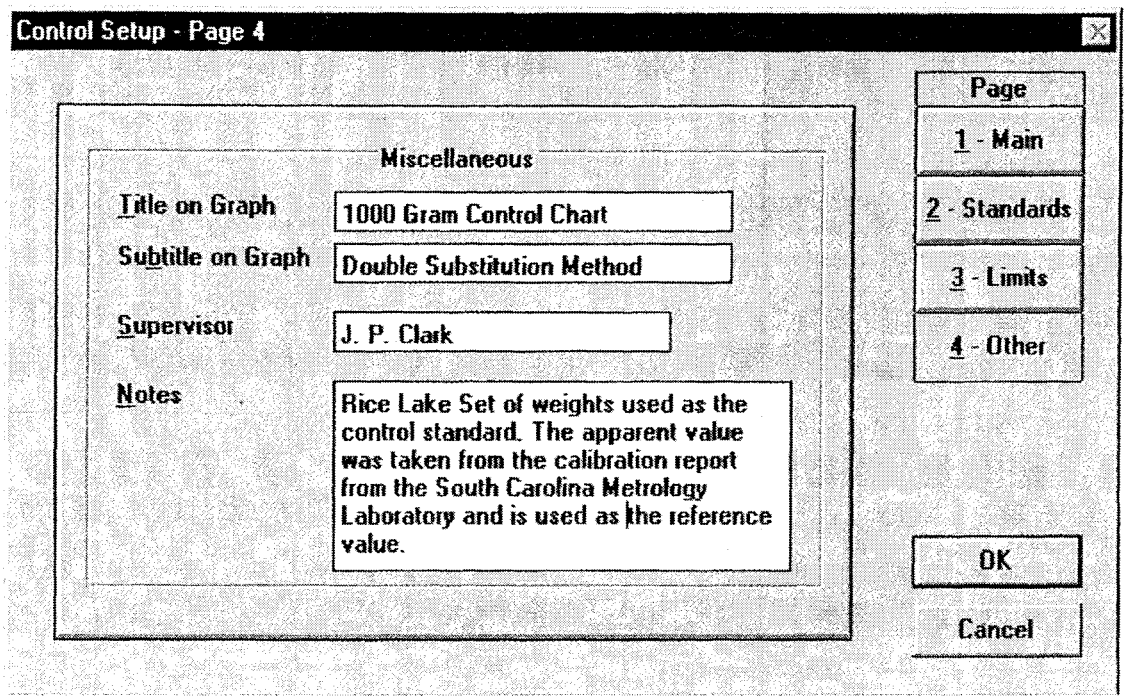
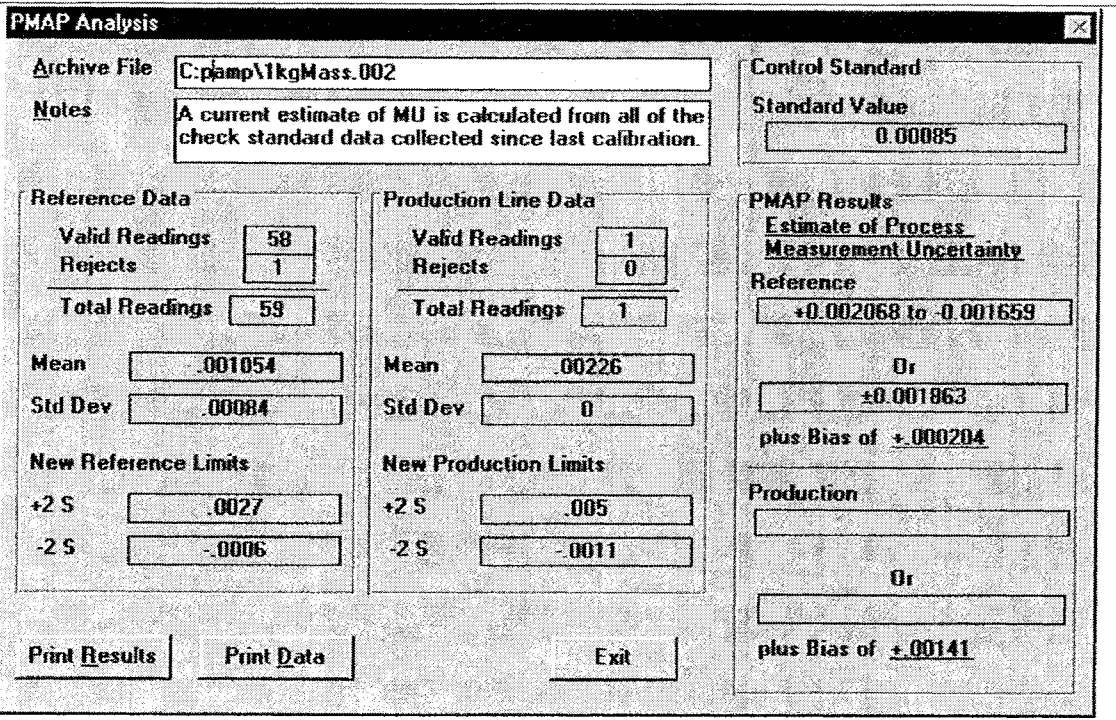


Figure 8.8. Real-Time MU Estimate for Calibrating 1 kg Mass Standards



factor of k = 2 that was specified on page three of the control chart setup.

Figure 8.8 shows that the "real-time" estimate of the potential uncertainty of the calibrated kilogram mass standard is +0.002060 and –0.001659 g under the conditions experienced in the Savannah River Standards Laboratory. These uncertainty values are asymmetric because they are the expanded u (± 0.001863 g at k = 2) added to or subtracted from the observed mean systematic error or bias of 0.000204 g. Using the larger value would give an uncertainty estimate that contained both the bias and k times the standard uncertainty. These statistics are based on 58 measurements since the last calibration before 10/1/97.

The next example comes from repeated gravimetric calibrations of a 100 μL pipette. The weighing system captures the error sources from an analytical balance and the air buoyancy corrections under different environmental conditions. Figures 8.9 and 8.10 are generated from the 2.1 version of the software. An analysis of the calibration data provides an estimate of the process MU and the bias. The program is capable of tracking two variables and calculating the total uncertainty of each. The program lists one variable as the "reference" data and the other as "production line" data. The confidence level can be set at k = 1, 2, or 3. Figure 8.10 sets k = 3, so the MU is 0.9067 μL for the reference data. No production data were generated.

Figure 8.9. Control Chart of the Calibrated Values of a 100 μL Pipette

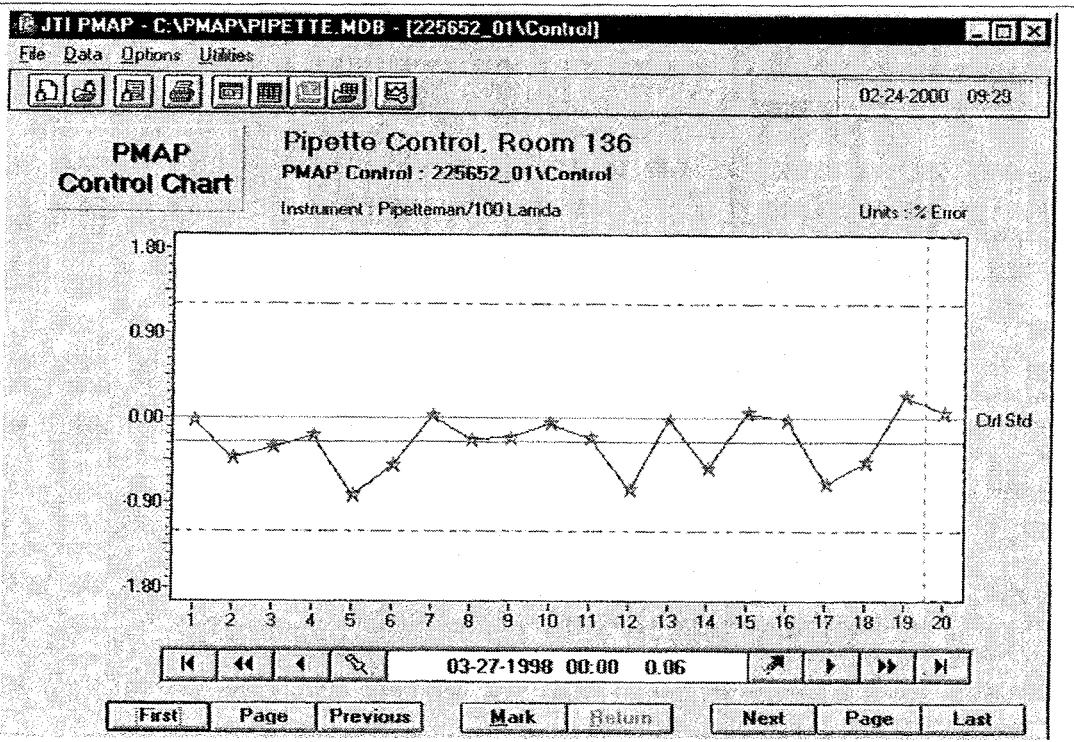
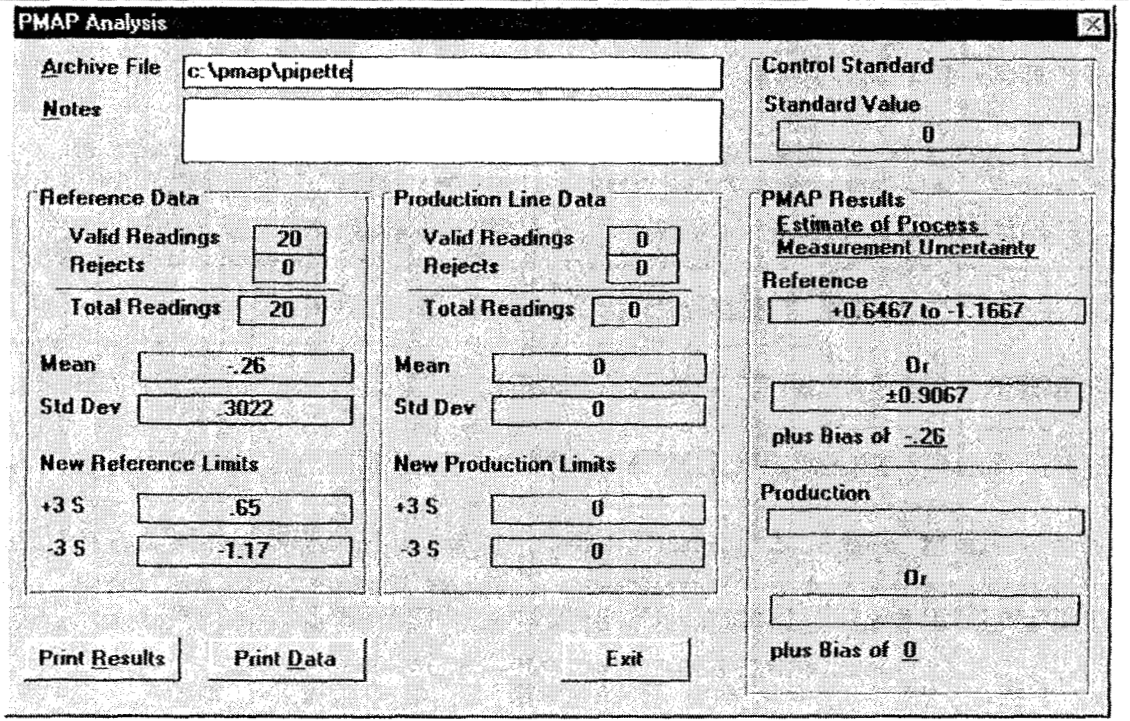


Figure 8.10. MU Estimate of the 100 μL Pipette



There are several advantages to using a MCP to capture the total variation of the measurement process. Many people do not know what variables add uncertainty to their measurement process, so they underestimate the MU of their processes. This method provides a moving picture of MU for the measurement process rather than the snapshot picture at time of calibration or from a calculation based on several assumptions.

Calibrate.

x_i = each individual value under test (for i = 1 to n)

n = number of values under test

$\bar{x}$ = mean

s = JTIPMAP™ estimate of standard deviation

k = coverage factor

$$\bar{x} = \frac{\Sigma x_i}{n}$$

$$s = \frac{\Sigma(x_i - \bar{x})^2}{\sqrt{(n-1)}}$$

$+k$ standard deviation limit = $\bar{x} + ks$ Upper k standard deviation limit

$-k$ standard deviation limit = $\bar{x} + ks$ Upper k standard deviation limit

$\pm 2k$ standard deviation is expected to contain 95.4 percent of the values if all test conditions are constant.

$\pm 3k$ standard deviation is expected to contain 99.7 percent of the values if all test conditions are constant.

Process Measurement Uncertainty.

$$(\bar{x} - \text{standard value}) \pm \left\{ k \times \left[\sqrt{s^2 + \left(\frac{\text{variability}}{k} \right)^2} \right] + \left| \text{systematic uncertainty} \right| \right\}$$

where $\bar{x}$ – standard value is the JTIPMAP™ systematic error, variability sum is the variability portion of the control standard uncertainty as input of setup controls, and systematic uncertainty is the systematic portion (bias or drift for use/time) of the control standard uncertainty as input on page 2 of setup controls.

Summary of the Third Method

The software performs many other useful functions, such as keeping a history file of the calibration data that can be used determine optimum calibration intervals. The software also includes statistical tests with built-in critical values to test bias and s estimates of various calibration

periods. This feature identifies significant changes that may be taking place in the control data or between previous control periods.

After the uncertainty of the measurement system has been determined, the user can use that estimate to make sure the right measurement instrument has been selected and that the right standards and methods are also selected to provide measurements that are fit for purpose.

By using PMAPs™ to estimate MU, the variation produced by the influence factors that make measurements uncertain will be captured. The first two methods provide estimates of uncertainty but are more like snapshots versus a video of a measurement process. The third method provides a liberal estimate of MU that allows the measurement organization to base their estimates on real data collected under actual operating conditions.

Summary

Knowledge of the quality of measurements is essential for managing processes. All measurements are estimates and have error that causes uncertainty about the true value. Quantifying the sources of uncertainty associated with a measurement involves determining an estimate of the MU. Three methods are commonly used for estimating the uncertainty of electronic balance measurements.

The first method involves using the manufacturer's specifications, the known operating range for temperature, and the nominal target weight. This method assumes all the other variables are controlled and the balance functions within the specifications. If the assumptions are incorrect, the estimate will be wrong.

The second method involves collecting calibration data that captures variables that produce random and systematic errors in the measurement process. This method is a snapshot in time and does not capture other sources of variation in the workplace that influence the quality of measurements. This method validates the performance of the weighing instrument and its contribution to MU. It is an economical method that uses data collected during calibration.

The third method involves using a well-designed measurement control program using replica artifacts, for check standards, and duplicating the measurement process to generate data that can provide "real-time estimates" of uncertainty associated with the measurement process. The latter method not only collects data for estimating uncertainty but can also provide assurance the measurement system is performing within statistical control limits at the time process measurements are made. This method does not require the depth of knowledge the others require in coming up with realistic estimates of MU for weight values if computer software like JTIPMAP™ were used.

A major difference in the methods is that the third method estimates the MU based on the *reproducibility* of the measurement system, not just balance sources of uncertainty considered and assumptions made that the influence factors are controlled or are negligible. The first two methods for estimating MU are based more on *repeatability*. The measurement organization needs to have MU estimates that will include all sources of variation that can make their measurements uncertain. The third method does this.

REFERENCES

1. Berg, C., W. Langner, B. Schubart, and S. Weyhe. 1996. *The fundamentals of weighing technology (terms, methods of measurement, errors in weighing)*. 1st English Translation. Goettingen, Germany: Sartorius AG.
2. Bitéry, L., M. Kochsiek, and J. Perton. 1997. *Mettler-Toledo glossary of weighing terms (a practical guide to the terminology of weighing)*. Greifensee, Switzerland: Mettler-Toledo GmbH.
3. ISO. 1992. *Guide to the expression of uncertainty in measurement*. Technical Advisory Group 4 (TAG 4) and Working Group 3 (WG 3). Geneva, Switzerland: International Organization for Standardization.

---

*The information contained in this chapter was developed during work under Contract No. DE-AC09-96SR18500 with the U.S. Department of Energy. By acceptance of this chapter, the publisher and/or recipient acknowledges the U.S. government's right to retain a nonexclusive, royalty-free license in and to any copyright covering this chapter along with the right to reproduce and to authorize others to reproduce all or part of the copyrighted paper WSRC-MS-99-00958.

4. Reichmuth, A. 1999. *Determining weighing uncertainty from balance specification.* Greifensee, Switzerland: Mettler-Toledo pp. 13–14.

5. *Calibration of Non-Automatic Electronic Weighing Instruments*, DKD-R-7-1, provisional until EA-R2 is augmented. Issued 1998.

6. Clark, J. P. 1997. Real-time estimates of measurement uncertainty. *Proceedings of Weighing & Measurement in the Year 2000 Conference* in Sheffield, England, pp. 151–164.

7. Clark, J. P. and A. H. Shull. 1996. Uncertainty demonstration program on MC&A measurement systems. *Proceedings of the INMM 36th Annual Meeting*, in Naples, Fla.

8. *User Manual*, PMAP™ SOFTWARE Version 1.2 for Windows. 1996. Alburquerque, N.M., USA: JTI Systems, Inc., p. C-1.

INDEX