

Research Article

An Innovative Approach to Functionality Testing of Analysers in the Clinical Laboratory

Wolfgang Stockmann, Werner Engeldinger, Albert Kunst, and Margaret McGovern

Roche Diagnostics GmbH, Professional Diagnostics, Clinical Trials, 68305 Mannheim, Germany

Correspondence should be addressed to Wolfgang Stockmann, wolfgang.stockmann@roche.com

Received 12 November 2007; Accepted 19 December 2007

Recommended by Peter B. Stockwell

The established protocols for evaluating new analytical systems produce indispensable information with regard to quality characteristics, but in general they fail to analyse the system performance under routine-like conditions. We describe a model which allows the testing of a new analytical system under conditions close to the routine in a controlled and systematic manner by using an appropriate software tool. Performing routine simulation experiments, either reflecting imprecision or method comparison characteristics, gives the user essential information on the overall system performance under real intended-use conditions.

Copyright © 2008 Wolfgang Stockmann et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Conventionally, the evaluation of new analytical systems is conducted on the basis of established protocols related to analytical performance like those published by CLSI (NCCLS) [1], ECCLS [2], or other national organizations. Sometimes additional exploratory testing is performed in the hope of gaining some insight into the routine behaviour of the system. While the standard protocols produce indispensable information with regard to the quality characteristics, they fail to analyse the system performance under routine conditions. Similarly, random testing only gives a chance opportunity to detect system malfunctions.

Obviously, there is no easy way to experimentally test the course of events that lead to an erroneous assay result and to verify its incorrectness under nonstandardized, that is, routine conditions. This situation is caused by the increasingly complex interactions of hardware, software, and chemistry which are found on modern analysis systems. The manual generation of experiments that describe a sample sequence with variable specimens and request patterns is feasible, but it is cumbersome to produce and provides no information on the correctness of the measurements. A better approach can be obtained by developing a software tool which generates appropriate experimental request lists, allowing the testing of a new system under conditions close to the routine in a controlled and systematic manner, and which provides sufficient data reduction for the analysis of the results.

2. METHODS

We have integrated this functionality in our evaluation software tool Windows-based computer-aided evaluation (WinCAEv) [3, 4] in such a way, that the generation of simulation experiments, the transfer of requests to the instrument, the on-line data capture, and the result evaluation, can be easily achieved with the available programme functions [5]. The routine simulation (hereafter referred to as RS) module allows for the definition and generation of typical test request patterns.

A request list that reflects a routine laboratory workload can be simulated by WinCAEv using appropriately defined parameters. The required input data embraces typical test distributions, sample materials, and sample request profiles. As an alternative to this programme supported simulation, laboratory specific request lists captured electronically from the laboratory information system or directly from the routine analysers are automatically converted by WinCAEv to a corresponding worklist for the system under evaluation.

Three main types of RS experiments were designed to allow systematic testing of an analytical system. In this way, different types of routine situations can be modelled and the respective performance situation evaluated.

The *RS-Precision* experiment type is used to test for systematic and/or random errors using the imprecision characteristics of the system. The goal is to compare the analyte recovery and precision generated during randomized process-

TABLE 1: Basic structure of a routine simulation precision experiment.

	Sample no.	Material	Assays									
			A	B	C	D	E	F	G	H	I	J
Batch part	1	Pool A	x	x			x	x	x	x	x	
	2	"	x	x			x	x	x	x	x	
	...	"	x	x			x	x	x	x	x	
	...	"	
	...	"	x	x			x	x	x	x	x	
	11	"	x	x			x	x	x	x	x	
	12	Pool B		x	x	x	x		x			
	13	"		x	x	x	x		x			
	...	"		x	x	x	x		x			
	...	"				
	...	"		x	x	x	x		x			
	22	"		x	x	x	x		x			
	23	Pool C	x		x		x					x
	24	"	x		x		x					x
	...	"	x		x		x					x
	...	"
	...	"	x		x		x					x
	33	"	x		x		x					x

		n
Random part	$n+1$	Pool B		x		x			x			
	$n+2$	Pool A	x				x	x	x	x		
	$n+3$	Pool C	x		x		x					x
	$n+4$	Pool B		x			x					
	$n+5$	Pool A	x					x				
	$n+6$	Pool A	x							x	x	
	$n+7$	Pool B		x		x	x					
		x	x	x	x		x			
	x	x	x		x	x	x			
					x	x				
	x	x				x	x		x	
			x							x
				x						
		x			x		x		x	
	x	x				x	x		x	
		x			x	x		x		...

		$n+x$

ing with that produced during batch analysis. Pooled quality control and pooled human materials are used as samples. A typical request list is shown in Table 1.

In repetitions of the same experiment, routine provocations are introduced during the randomized processing to

further challenge the system's performance under various conditions. The type and number of provocations depend on the system under evaluation, but generally include items regularly encountered during operation in a routine laboratory, like calibration and quality control measurements, reagent

TABLE 2: Routine simulation series 1/2, cobas 6000.

Method	N	Categories						
		--- <-15%	-- <-10%	- <-5%	±5%	+ >5%	++ >10%	+++ >15%
Na	52				52			
K	60				60			
Cl	11				11			
ALP_2	26				26			
ALT	36			4	31		1	
AMYL_2	10				10			
AST	25				25			
CK	21				21			
LDH_2	34				34			
LIP	8				7	1		
BIL-T	36	5		6	20	3	1	1
CREA_2	35	1			32		2	
GLUC_3	44				44			
TP_2	64				64			
UREA	52				50	2		
LACT_2	10				10			
Ca	19				19			
Mg	16				15	1		
PHOS_2	15				15			
CRP	38				37	1		
DIG	9			1	3	1	1	3
Li	10				8	2		
AMYL_2 (urine)	10				10			
FT4	10				10			
TSH	8				6	2		
Pro-BNP	9				8	1		
TNT	9			1	8			
MYO_2	10	1			9			
βHCG	10				7	3		
TOTAL abs	697	7	0	12	652	17	5	4
TOTAL rel		1%	0%	2%	94%	2%	1%	1%

switchover or exchange, sample short, STAT analysis, provocation of various data flags, sample reruns, and so forth.

Errors related to instrument malfunctions or chemistry problems can be deduced from the experimental data by comparing the batch and random results. The mean, median, CV, relative 68%-median distance (md68% describes a robust measure of variation) [6], and minimum and maximum of the random part are compared with those from the batch part for every analyte measured. Random and/or systematic errors will result in significant deviations like elevated CVs and differences of the means. One can expect that the imprecision in a simulated routine run will result in somewhat higher CVs due to more interactions of the analytical system than during a standard batch run. Based on experience from various system evaluations, we use the following expected CV in the random part: $CV_{\text{exp,rand}} = CV_{\text{exp,ref}} + \Delta CV$, where we set $\Delta CV = 1/2(CV_{\text{exp,ref}})$; therefore

$CV_{\text{exp,rand}} = 1.5 CV_{\text{exp,ref}}$; $CV_{\text{exp,ref}}$ is the expected CV in the reference part.

Usually the routine simulation experiment is performed with many different methods, and the high number of results produced has to be assessed for relevant deviant results. This can easily be done by comparing the CV and the relative 68%-median distance.

The system handling of the routine provocations is assessed for correctness, and the analytical results produced during and after provocations are checked for marked deviations which may represent systematic and/or random errors.

Recently we extended this experiment in order to run the routine simulation precision experiment via a host download procedure (see below) so that the real routine request pattern is reflected and a simulation by the WinCAEv software is not necessary.

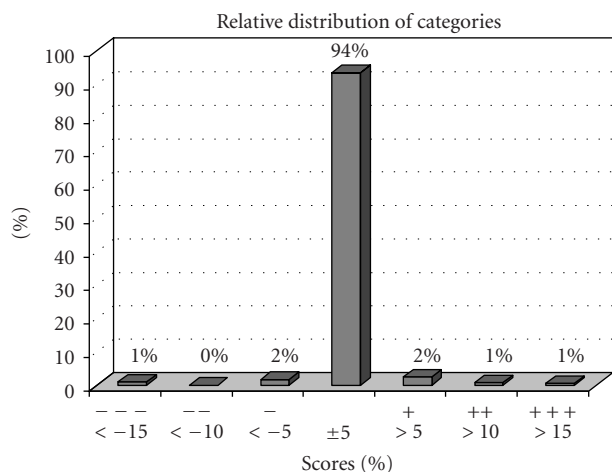


FIGURE 1: Routine simulation series 1/2, cobas 6000.

RS-Series 1/2 is used for the comparison of randomized test processing in two runs. Fresh human specimens are used as sample materials with request patterns reflecting the evaluation sites typical routine workloads. The sequence of sample processing is identical in both runs and the same samples are used, not placing fresh samples for the second run.

Random errors can be deducted from the experimental data by comparing the deviation of the second from the first run results. The results are grouped in 7 five percent categories between $\pm 15\%$ deviation. Each sample pair is categorized; a summary shows the number of samples per analyte in each category as well as a total statistic per category for the complete experiment (see Table 2 and Figure 1). Random errors will result in marked deviations between both run results for one or more samples.

RS-Method Comparison Download allows direct comparison of the routine analyser methods (reference data) with those of the instrument under evaluation processed in a randomized routine-like fashion. The test results and sampling patterns from the routine laboratory analyser(s) are electronically captured by WinCAEv via file import (host-download) or simply with a batch upload. Using the host-download option, sample identification numbers, requests, and results are exported from the laboratory host in a text format file (e.g., comma separated values (CSV)) and then imported in WinCAEv. No patient demographics are transmitted to WinCAEv. Method comparison statistics and graphs are generated per analyte and comparison instrument.

3. APPLICATIONS

Over the last decade, routine simulation experiments have become an integral part of inhouse and multicentre system evaluations at Roche Diagnostics. Here, we outline some typical areas of use based on practical experience, as well as examples of errors difficult to produce using conventional procedures yet observed using these experiments.

RS-Precision is an extremely effective means of testing the interaction of software with all other system components

under stressed conditions. During a multicentre study of Roche/Hitachi 917 in the early nineties for example, these experiments yielded CVs of up to 4% for test applications using low-sample volume ($2 \mu\text{L}$). An example is shown in Figure 2 for cholesterol. Of the 48 runs performed during this experiment, 83% (=40 series) were found with a CV higher than the expected 2%; 24 series had a relative 68-median distance of more than 2%. The difference between CV and md68% indicated that several series had clear deviant results. Further inhouse investigations revealed that a software malfunction in the sample pipetting process under certain conditions was the root cause for these conspicuous results. After correction of the software and repetition of this experiment, the CV of the cholesterol assay was in all cases below 2%.

On Roche/Hitachi 912, we found that introduction of STAT samples during operation led to intermittent incorrect data flags on STAT sample results when a sample material other than serum was selected. Investigation showed that if a STAT sample was requested by sample disk position only on the analyzer, and the sample type downloaded from the laboratory host is one other than serum or plasma, the sample was correctly measured for the specified biological material but the data flagging was done as if the sample was serum. In this case, an incorrect rerun of the sample was indicated by the generated data flag.

With the new generation of Roche systems MODULAR ANALYTICS from the late nineties, combining multiple analyser modules, this experiment became indispensable and gained many new areas of use. An Intelligent Process Manager distributes the sample carriers to the various analyser modules in a way that ensures most efficient operation, and background maintenance features on these systems allow the operator to perform maintenance on one or more modules while continuing routine operation on the other modules. The RS-Precision experiment allowed us to check these, among other complex functions, in a systematic manner under numerous simulated routine-like conditions. A typical provocation on such systems is the deactivation and reactivation of a module during routine operation. The goal is to check that the samples with requests for tests on the deactivated module(s) are handled correctly, and that the reactivated module performs as expected after return to operation. During the MODULAR ANALYTICS SWA (serum work area) evaluation, these provocations revealed sporadic errors after module reactivation like wrong reagent inventory and bottle change-over to standby reagents although current reagents were not empty.

Also, RS-Precision is an effective tool to test the interaction of reagents on a selective access analyser. During the recent multicentre evaluation of cobas 6000, a new reagent carryover—not observed on other Roche analysers—caused by the reagent probe was found for the creatinine enzymatic assay, when the creatine kinase reagent (CK) was pipetted just before the creatinine assay. Creatine phosphate of the CK reagent (vial 2) may partly be hydrolysed to creatine which can influence the creatine concentration of the enzymatic creatinine assay, when contaminated by the reagent probe. As shown in Table 3, the CV changes from 0.9% in the batch part to 2.1% in the random part. Looking to the single

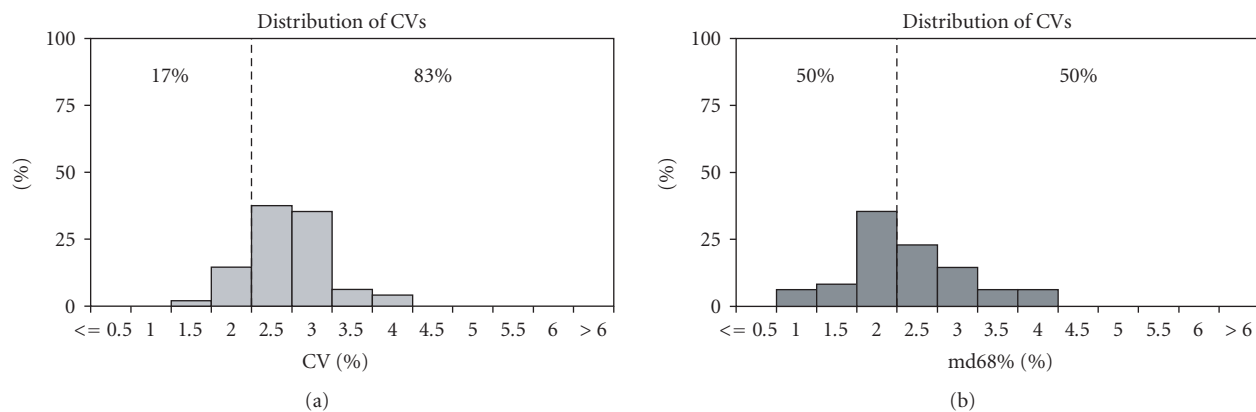
FIGURE 2: Routine simulation precision Roche/Hitachi 917, distribution of CV and md68% for cholesterol ($n = 48$ series).

TABLE 3: Routine simulation precision cobas 6000.

Test	[Study unit]	Material	N	Mean	CV [%]	CV diff. [%]	Remarks
Na	[mmol/L]	CSP	144	138.9	0.7	0.5	
K	[mmol/L]	CSP	144	5.0	0.8	0.5	
ALAT	[U/L]	CSP	53	88.6	1.0	0.3	
ASAT	[U/L]	CSP	18	105.8	0.9	0.1	
CK	[U/L]	CSP	11	308.4	1.1	0.3	
GGT_2	[U/L]	CSP	12	124.1	0.9	-0.3	
CHOL_2	[mmol/L]	CSP	70	3.6	0.9	0.3	
CREA_2 enz	[μ mol/L]	CSP	80	225.0	2.1	1.2	Outlier
GLUC_3	[mmol/L]	CSP	61	9.6	1.0	0.3	
TP_2	[g/L]	CSP	55	58.0	0.9	-0.1	
UREA	[mmol/L]	CSP	23	15.1	0.5	-0.3	
UA_2	[μ mol/L]	CSP	19	449.3	1.0	0.2	
Ca	[mmol/L]	CSP	39	2.9	1.1	0.6	
Fe_2	[μ mol/L]	CSP	75	26.6	1.6	0.9	
CRP	[mg/L]	HSP	70	14.5	1.2	0.3	
IGG_2	[g/L]	CSP	11	6.8	1.7	0.8	
IGA_2	[g/L]	CSP	11	1.6	1.4	0.2	
IGM_2	[g/L]	CSP	12	0.7	3.3	1.0	
RF.II	[kU/L]	HSP	10	19.2	2.4	1.0	
TRSF_2	[μ mol/L]	CSP	12	24.4	1.3	0.2	
DIG	[mmol/L]	CSP	10	4.6	2.2	1.0	
Li	[mmol/L]	CSP	10	1.6	1.5	0.4	
CREA_2 enz	[μ mol/L]	HUP	19	7474.3	2.5	1.7	
TP_U-CSF_3	[mg/L]	HUP	20	99.4	2.7	1.4	
FT4	[pmol/L]	CSP	26	16.9	1.6	1.0	
TSH	[mU/L]	CSP	31	1.2	1.6	0.6	
Pro-BNP	[pmol/L]	CSP	10	277.9	1.7	1.0	
TNT	[μ g/L]	HSP	10	0.3	2.0	0.9	
CKMBm	[μ g/L]	CSP	10	29.2	1.6	0.8	
PSA	[μ g/L]	CSP	10	20.9	1.3	0.9	
FPSA	[μ g/L]	CSP	10	5.1	1.3	0.4	
β HCG	[U/L]	CSP	0				Flag "FL"*

* Flag "FL" for expired reagent provocation, flag excludes values from statistical calculation.

results, the creatinine concentration of sample number 68 is increased and classified as an outlier (255 compared to the median of 225 mmol/L, which is increased by more than the fivefold md68 [20 mmol/L]). By monitoring the “cuvette history” data base of the analyser, one could confirm for that sample that the CK reagent was pipetted just before the creatinine assay. Consequentially, an extra probe wash cycle was installed on cobas 6000 for this reagent combination in order to avoid carryover. Table 3 shows also that the analyser works correctly after a provocation with an expired reagent pack (example β -HCG).

Conventionally used to test for reproducibility of results in two runs with the goal to check for random errors, the RS-Series 1/2 experiment can also be easily adopted to address numerous system specific functionalities. For modular systems, this experiment is used to compare consistency between modules for example [7]. On the other hand, during the cobas Integra 800 evaluation [8], it was used to check the analyser’s clot and bubble detection function. Samples with clots and bubbles were included in the first run and prior to the second run; bubbles were removed and those with clots were centrifuged. The results were checked for correct flagging of problematic samples in the first run and analyte recovery compared with the second run.

Having a practically automated procedure to immediately repeat the workload from one or more routine instruments on the system under evaluation and to directly compare the results, the *RS-Method Comparison Download* is an invaluable tool. It allows the investigator to assess the new system from an analytical performance as well as from an overall system performance perspective, under real life conditions.

A total of 187 method comparisons were processed for 50 analytes in nine laboratories under site-specific routine conditions during the MODULAR ANALYTICS evaluation [9] for example. Analysis of approximately 27 000 measurements for fresh human specimens gave the final proof that the system was ready for market launch.

4. CONCLUSION

Conducting such routine simulation experiments gives the manufacturer essential information on the overall system performance under real intended-use conditions. This innovative, realistic, and thorough approach to system testing has won the approval of principle investigators during international multicentre evaluations over more than ten years.

With the experience gained, we strongly recommend to focus more on routine simulation type experiments during the performance evaluation of a new analytical system, and above all to derive the various quality characteristics like imprecision and method comparison from these results.

ACKNOWLEDGMENT

A fundamental contribution to this topic was already made by Wolfgang Bablok, who died much too early (1998).

REFERENCES

- [1] CLSI evaluation protocols, <http://www.nccls.org/>.
- [2] R. Haeckel, E. W. Busch, R. D. Jennings, and A. Trucheaud, Eds., *Guidelines for the Evaluation of Analysers in Clinical Chemistry*, vol. 3 of *ECCLS Document*, Beuth, Berlin, Germany, 1986.
- [3] A. Kunst, A. Busse Grawitz, W. Engeldinger, et al., “WinCAEv— a new program supporting evaluations of reagents and analysers,” *Clinica Chimica Acta*, vol. 355S, (Abstr-No WP6.04), p. 361, 2005.
- [4] W. Bablok, R. Barembruch, W. Stockmann, et al., “CAEv— a program for computer aided evaluation,” *Journal of Automatic Chemistry*, vol. 13, no. 5, pp. 167–179, 1991.
- [5] W. Bablok and W. Stockmann, “An alternative approach to a system evaluation in the field,” *Quimica Clinica*, vol. 14, p. 239, 1995.
- [6] W. Bablok, R. Haeckel, W. Meyers, and W. Wosniok, “Biometrical methods,” in *Evaluation Methods in Laboratory Medicine*, R. Haeckel, Ed., pp. 203–241, VCH, Weinheim, Germany, 1993.
- [7] C. Bieglmayer, D. W. Chan, L. Sokoll, et al., “Multicentre evaluation of the E 170 Module for MODULAR ANALYTICS,” *Clinical Chemistry and Laboratory Medicine*, vol. 42, no. 10, pp. 1186–1202, 2004.
- [8] F. L. Redondo, P. Bermudez, C. Cocco, et al., “Evaluation of Cobas Integra® 800 under simulated routine conditions in six laboratories,” *Clinical Chemistry and Laboratory Medicine*, vol. 41, no. 3, pp. 365–381, 2003.
- [9] G. L. Horowitz, Z. Zaman, N. J. C. Blanckaert, et al., “MODULAR ANALYTICS: a new approach to automation in the clinical laboratory,” *Journal of Automated Methods and Management in Chemistry*, vol. 2005, no. 1, pp. 8–25, 2005.