

# Designing an autoverification system in Zagazig University Hospitals Laboratories: Preliminary evaluation on thyroid function profile

Amany Mohy-Eldin Sediq, Ahmad GabAllahm Hala Abdel-Azeez

From the Department of Clinical and Chemical Pathology, Faculty of Medicine, Zagazig University, Sharqia, Egypt

Correspondence: Dr. Amany Mohy-Eldin · Sediq Clinical and Chemical Pathology, Faculty of Medicine, Zagazig University, Sharqia 44519, Egypt · amany\_mohy2006@yahoo.com

Ann Saudi Med 2014; 34(5): 427-432

DOI: 10.5144/0256-4947.2014.427

**BACKGROUND AND OBJECTIVES:** The current practice in Zagazig University Hospitals Laboratories (ZUHL) is manual verification of all results for the later release of reports. These processes are time consuming and tedious, with large inter-individual variation that slows the turnaround time (TAT). Autoverification is the process of comparing patient results, generated from interfaced instruments, against laboratory-defined acceptance parameters. This study describes an autoverification engine designed and implemented in ZUHL, Egypt.

**DESIGN AND SETTINGS:** A descriptive study conducted at ZUHL, from January 2012-December 2013.

**MATERIALS AND METHODS:** A rule-based system was used in designing an autoverification engine. The engine was preliminarily evaluated on a thyroid function panel. A total of 563 rules were written and tested on 563 simulated cases and 1673 archived cases. The engine decisions were compared to that of 4 independent expert reviewers. The impact of engine implementation on TAT was evaluated.

**RESULTS:** Agreement was achieved among the 4 reviewers in 55.5% of cases, and with the engine in 51.5% of cases. The autoverification rate for archived cases was 63.8%. Reported lab TAT was reduced by 34.9%, and TAT segment from the completion of analysis to verification was reduced by 61.8%.

**CONCLUSION:** The developed rule-based autoverification system has a verification rate comparable to that of the commercially available software. However, the in-house development of this system had saved the hospital the cost of commercially available ones. The implementation of the system shortened the TAT and minimized the number of samples that needed staff revision, which enabled laboratory staff to devote more time and effort to handle problematic test results and to improve patient care quality.

Laboratory testing is a highly complex process. Traditionally, test result verification has depended on mental algorithms performed by a pathologist/medical technologist on a single analytical result or a group of results. This notion of human involvement in the total testing process was well addressed by Lundberg, who described it as a “brain-to-brain loop” with the laboratorian brain as a step among several steps in the loop that starts and ends with the clinician brain.<sup>1</sup> The purpose of the verification phase is to identify potential errors before results are released to the patient’s medical records.<sup>2</sup> When result verification is performed manually, it would be a time-consuming, tedious process with large interindividual variation that prolongs laboratory turn-

around time (TAT).<sup>3</sup> As an approach to overcome these demerits of manual result verification, and thus fulfilling total quality requirements, the concept of autoverification has been introduced into modern laboratory practice.<sup>4</sup>

The College of American Pathologists, in its laboratory general checklist for 2013, defined autoverification as the process by which test results are generated from interfaced instruments and are sent to the laboratory information system (LIS), where they are compared against laboratory-defined acceptance parameters. If the results fall within these defined parameters, the results are automatically released to the patient’s medical records without any additional validation by laboratory staff. Results that fail to fulfill these defined parameters are held for

review by laboratory staff prior to manual verification and reporting.<sup>4</sup> This means that autoverification software does not “make decisions”; it deals only with situations, data, and patterns of results that have been predefined by the laboratory.<sup>5</sup>

As in many laboratories, the current practice in Zagazig University Hospitals Laboratories (ZUHL) is manual review, verification, and the release of all results for later printing and delivery. These processes are performed by a qualified lab specialist. Revision of all results transmitted from different analyzers to the LIS is problematic. The challenge is to distinguish between straightforward cases (for immediate verification) and those that need further action (rerun, requesting more clinical data from the treating physician). This step is hindered by huge workloads, tight shift hours, and the lack of qualified staff, leading to the possible release of some undesired results. Such mistakes will lead to the loss of more time and effort when recollecting the results report and informing the treating physician about the correct results or increasing the TAT to allow staff to thoroughly inspect all results.

Autoverification could be the answer to these problems. It can be achieved through the use of information technology tools, but the laboratory is ultimately responsible for defining criteria that are implemented with these tools to make autoverification decisions. Although commercial autoverification engines are available, their processing algorithms and decision rules are considered proprietary and therefore cannot be modified by the users.<sup>6</sup> The commercial software deals with only the most basic levels of autoverification; eg, reference interval, instrument alarm, quality control (QC), etc. However, none of them deals with complex clinical data (clinical presentation, clinical history, drug history, etc). This can be achieved by developing an in-house autoverification engine that would work as a part of the LIS used in ZUHL. Another point to be considered was the financial cost of an autoverification engine. The commercial software is expensive and may not be affordable to ZUHL.

Thyroid function tests are 1 example of complex tests that the laboratory has to deal with. Laboratory testing of thyroid hormones is used to diagnose and document the presence of thyroid disease, a condition that often presents with vague and subtle symptoms. Familiarity with normal physiology and pathophysiology is important for proper use and selection of thyroid function tests. A number of medications have been shown to alter thyroid function. All these factors represent a challenge during their interpretation, making them a target for setting interpretation algorithms.<sup>7</sup> This manuscript gives a

detailed description of the autoverification system designed and evaluated by ZUHL, and also an evaluation of its impact on laboratory performance.

## MATERIALS AND METHODS

This study was conducted in the Clinical Chemistry Unit of ZUHL in cooperation with the LIS vendor (National Technology Egypt).<sup>8</sup> We designed an autoverification engine according to Clinical and Laboratory Standards Institute (CLSI) Auto10-A guidelines.<sup>2</sup> This engine was integrated into the ZUHL LIS (Laboratory Data Manager). The whole process was divided into 6 phases: engine design, algorithm and rule development, creation of programming rules, engine validation, multi-center evaluation of the rules and evaluation of impact of autoverification on TAT.

### Engine Design

A multilayer design for the engine was used with 2 prerequisite levels that must be fulfilled in sequential manner before results are processed in the engine core.

Level 1 comprised the following items:

- (1) QC check. The last QC results are transmitted from the analyzer to the LIS and are evaluated according to Westguard et al.<sup>9</sup> If any QC rule is violated, or there are no QC results within the last 24 hours, the engine will stop all cases from being autoverified.
- (2) Instrument flag. If any instrument flag was sent to the LIS along with a particular test results (for a certain case), the results for this case only will stop for later manual verification.
- (3) Reportable range check. The reportable range check is a filter to verify if the result is within the reportable range specified for this analyte in the LIS or not. If not, the engine will stop this case.
- (4) Reference interval check. If this check is activated, all abnormal results will stop, whereas the results that are within the reference interval will pass to the next level. Any of these checks can be selectively inactivated according to user settings.

Level 2 necessitates that all ordered tests (that are part of an algorithm) are completed. If any result is missing, the engine will stop the case. The engine core is based on the medical decision rules. The rule-based system was used in the engine design. A major advantage of this engine is the ability of the user to activate or inactivate any rule according to lab needs. The whole system can be inactivated if needed (eg, as part of annual revalidation plan, or due to a major breakdown of the system). **Figure 1** illustrates the engine design and logic in decision making.

*Algorithm and Rule Development*

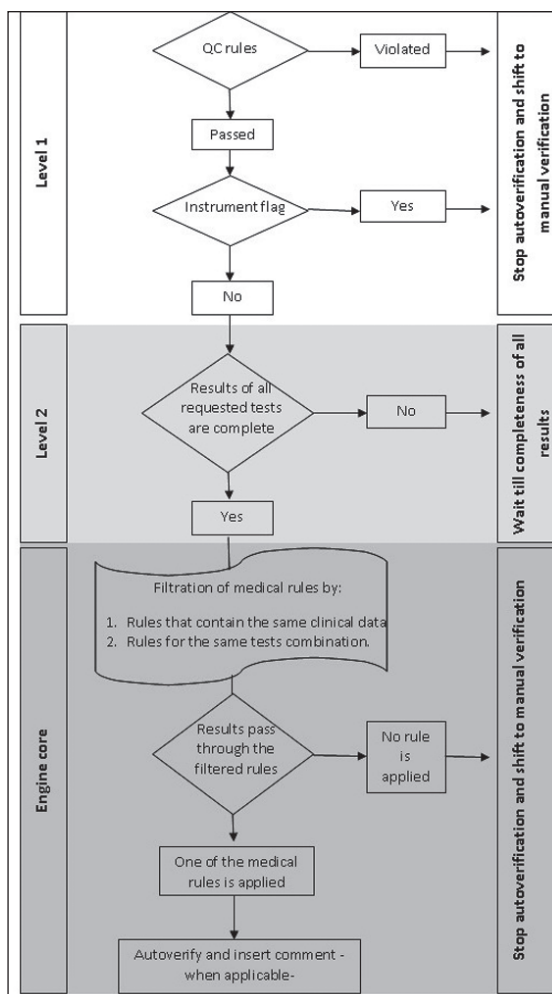
Thyroid function tests, including thyroid-stimulating hormone (TSH), total triiodothyronine (TT3), total thyroxine (TT4), free triiodothyronine (FT3), free thyroxine (FT4), antithyroid peroxidase antibody, thyroid-stimulating immunoglobulin, thyroid-stimulating hormone receptor antibody, and thyroglobulin antibody, were included in the algorithms.

Following CLSI Auto10-A guidelines,<sup>2</sup> all relevant data that could affect the interpretation of the aforementioned tests were gathered, mainly: (1) patient-related data and preanalytical variables such as age, sex, pregnancy status, clinical presentation, history of thyroid/pituitary disorders, drug history, and other medical conditions that may affect thyroid hormone levels, such as a history of vesicular mole or choriocarcinoma; (2) analytical variables—any possible method interference, measuring range of the method, or reference interval; and (c) postanalytical variables such as consistency check, limit check, and delta check. These data were thoroughly studied to decide which would be included during setting algorithms and how, and which would not.

The group of antithyroid antibodies indicates the immune nature of thyroid disease; hence, any positive result for a single test is considered positive for that group (named positive medical formula), while a negative medical formula is considered when all test results are negative. This led to a reduction in an enormous number of algorithms that can be generated using 9 tests. The engine will handle TSH, TT3, TT4, FT3, FT4, and medical formula results. Algorithms for thyroid function tests were created and then written down in 563 medical decision rules, 285 of which autoverify and 278 of which stop autoverification. **Figure 2** illustrates a simplified example of the used algorithms.

*Creation of programming rules*

Professional programmers and evaluators at National Technology performed the programming process of the engine. A comprehensive patient's sheet is programmed on the LIS and is linked to thyroid function tests. This sheet contains all clinical and medical data that are needed for proper interpretation of thyroid function tests. It must be completed during ordering of patient tests on the LIS. Inside the engine, each point on the sheet is number coded; this code is used for programming the medical decision rules. The programmed medical decision rules were revised to verify that they will produce the expected outcome. When necessary, changes were done to meet requirements. The logic used in programming is to stop any results unless stated otherwise in the programmed medical decision rules to ensure patient safety.

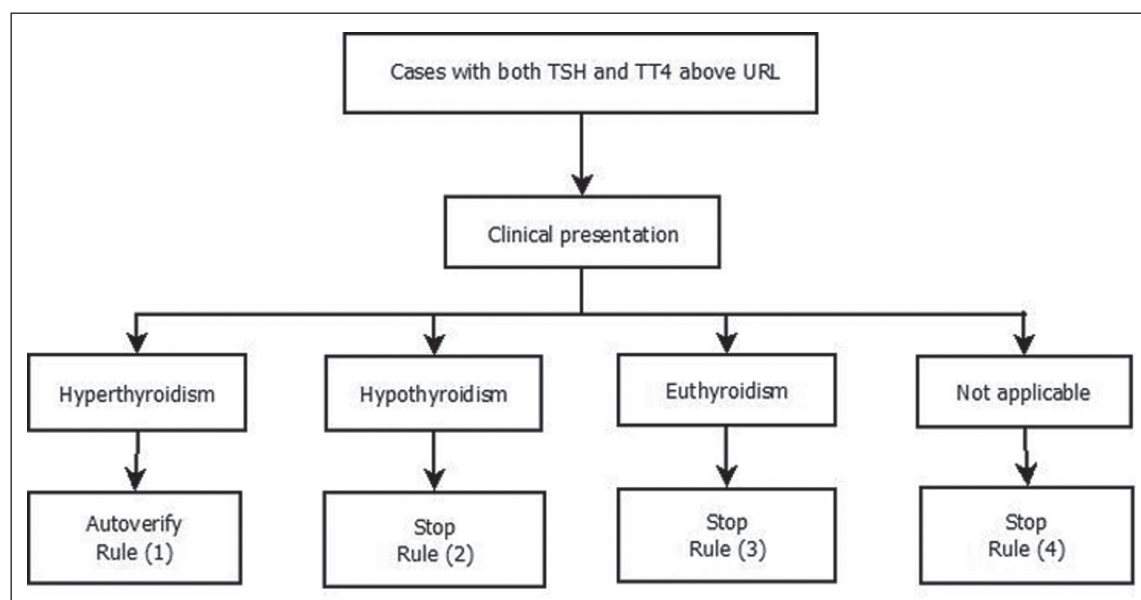


**Figure 1.** Illustration of the engine design and logic in decision making.

An autocoment system was developed to link the applied rule to a matched comment, when applicable. Still, the user has the ability to insert his/her comment for cases that are manually verified.

*Engine validation*

The whole validation process followed the recommendations of CLSI Auto-10 A.<sup>2</sup> The first step was to test simulated cases. These cases are used to verify that the programmed medical rules follow the expected logic and produce the expected outcome. One case for each rule was programmed. The total number of cases was 563; 285 (50.6%) end in autoverify and 278 (49.4%) end in stop autoverification. The simulated cases passed through the engine, and the results (whether Verify or Stop) were revised to match the expected outcome. The packages of simulated cases continuously pass through the engine every 15 minutes to test engine integrity.



**Figure 2.** Autoverification algorithm for TSH and TT4 profile. Using rule-based system 4 rules were generated. For example, rule (1) states that if both TSH and TT4 are above URL and the patient presents with symptoms suggestive of hyperthyroidism, then autoverify. A comment would be inserted according to the patient's thyroid status history. URL: Upper reference limit.

The second step of the validation process was to pass a part of the laboratory database (6 months of data) through the engine. The total number of database thyroid cases was 1673 cases. All these cases were manually revised and verified by expert lab specialists (MD in clinical laboratory medicine). The use of these cases challenged the engine with the same type and distribution of cases that are received by the lab and gave an idea about the expected autoverification rate of the engine. This testing process via a large number of results may detect problems with the autoverification engine that occur infrequently before the engine is connected online for a real-life evaluation. We ensured that the outcome matched what was written in the programmed medical decision rules.

After passing through the engine, all results are transmitted to the LIS. The report for autoverified results will contain "(Online Ver)" in place of the verifier's signature and the programmed comment, if any. Cases that fail to autoverify are color flagged on the program screen, waiting for lab staff to manually verify their results. In this case, the verifier name will be printed on the result report and saved within the LIS.

#### *Multicenter evaluation of the rules*

The previously mentioned simulated cases (n=563) (results report and clinical sheet) were evaluated by 4 clinical laboratory experts (professors in clinical laboratory medicine): A, B, C, and D, representing different independent Egyptian medical institutes. They were asked to

revise all the cases and record their opinion as to whether the data were sufficient for verification or not.

#### *Evaluation of impact of autoverification on TAT*

The engine was connected online with the LIS on October 2013 for 3 months to evaluate its real-time performance and its impact on TAT. The average of the reported lab TAT from the time of sample receipt by the lab to result verification on the LIS (T1), and of the segment of the TAT from the complete analysis until result verification on the LIS (T2), were obtained from the LIS after engine implementation, and they were compared with the same months in the preceding year.

All statistical calculations were done using SPSS version 17 (SPSS, Inc., Chicago, IL, USA) for Microsoft Windows.

## RESULTS

Cohen Kappa test for interobserver degree of agreement<sup>9</sup> was used to determine if there was agreement between each reviewer and the engine's judgment as to which situation results were to be autoverified and which situation results were to be stopped by the engine. The Kappa coefficient represents the observed agreement, which is above and beyond that due to chance. Kappa values were of moderate strength (range: 0.461–0.533) and Kappa coefficients were highly statistically significant from zero ( $P < .001$ ) (Table 1).

The 4 reviewers agreed with one another in 55.5% of

cases (312 cases). This percentage decreased to 51.5% (290 cases) when the engine decision was included. The chi-square statistic was not significant at  $P < .001$ . The difference occurred in 22 cases. Running the engine on database cases for the thyroid profile ( $n=1673$ ) resulted in 63.8% of cases being autoverified. Implementation of the engine in real time led to a reduction of T1 from 5 hours 21 minutes to 3 hours 29 minutes (34.9%), and of T2 from 3 hours 48 minutes to 1 hour 27 minutes (61.8%).

## DISCUSSION

Thyroid function tests were chosen for the initial evaluation of autoverification because they are a closely related group of tests that have a defined relation to other system functions. They have clearly defined reference intervals, and well-defined cutoff values for certain clinical abnormalities. They represent the main bulk (46%) of hormone cases in the ZUHL Clinical Chemistry Unit.

We had chosen the total agreement between the 4 reviewers as a benchmark in an attempt to maximize patient safety. The same approach was conducted in the VALAB system.<sup>11</sup> Overall agreement was observed between the 4 reviewers in 55% of cases. Although there was a statistically significant degree of agreement between the designed engine and each of the 4 independent expert reviewers (Kappa approximate significance,  $P < .001$ ), agreement was observed between the designed engine and the 4 reviewers in 51.5% of cases. The 4% difference in agreement between the reviewers and the reviewers with the engine is attributed to disagreement in 22 cases. Only in 1 case did the engine decide to autoverify while the 4 reviewers decided to stop verification. In reviewing this case, the concerned rule states that if TSH is  $< 0.01$  mIU/L, TT4 is within

reference interval, and TT3 is above the reference interval, then one should autoverify and comment (possible T3 toxicosis). High TSH alone was found in 9 of the cases; 4 cases had a T4 assay within normal limits, and another 4 cases had both T3 and T4 within normal limits in addition to high TSH. In the previously mentioned 17 cases, TSH values were below a fivefold rise of the upper reference limit (URL), and no clinical data were provided to support the diagnosis. Per Demers and Spencer,<sup>7</sup> we did not verify such cases, as TSH is not decisive.

Low TSH alone was found in 3 cases, and 1 case had a T4 assay within normal limits in addition to low TSH. In the previously mentioned 4 cases, TSH level was below the reference interval, but not  $< 0.01$  mIU/L (although we used a third-generation TSH assay that allowed for the determination of very low levels of TSH). Per Demers and Spencer,<sup>7</sup> we did not verify those cases, as the TSH value was not decisive of thyroid disorder; the reviewers verified them. So, in spite of these discrepancies, we did not change any rule in the engine to achieve the maximum possible patient safety. The highest risk associated with autoverification is releasing large numbers of results without proper review or editing. This will result in poor planning, implementation, or a failure to follow procedures. A lesser risk is to not release results that do meet the autoverification procedure. Other than affecting the TAT and workflow, the latter is a "safe" failure.<sup>5</sup>

The autoverification rate of archived database cases was 63.8%, close to that of the VALAB system, which has a mean rate of autoverification of about 50% to 90%;<sup>11</sup> however, VALAB was applied on a wider range of analytes. Our results were lower than that of the

**Table 1.** Degree of agreement between the engine and each expert reviewer.

Reviewer		Engine		Agreement <sup>a</sup>	Disagreement <sup>a</sup>	Kappa value <sup>b</sup>	Kappa approximate significance
		V	S				
A	V	260	25	412 (73.18%)	151 (26.82%)	0.461	$P < .001^c$
	S	126	152				
B	V	268	17	428 (76%)	135 (24%)	0.518	$P < .001^c$
	S	118	160				
C	V	210	75	413 (73.36%)	150 (26.64%)	0.467	$P < .001^c$
	S	75	203				
D	V	262	23	432 (76.73%)	131 (23.27%)	0.533	$P < .001^c$
	S	108	170				

<sup>a</sup>Data are presented as the number (% of total cases). <sup>b</sup>Strength of agreement:  $< 0.2$  poor; 0.2–0.4 fair; 0.41–0.6 moderate; 0.61–0.8 good; 0.81–1.0 very good. <sup>c</sup>Highly statistically significant degree of agreement. V: autoverify, S: stop autoverification.

DNSev system, which shows a verification rate of about 80%,<sup>12</sup> and of the system developed and described by Mu-Chin et al, which shows an overall autoverification rate for both immunoassay-related and biochemistry-related tests of 81.5%.<sup>13</sup> This difference could be explained by the difference in the studied group of tests and patient presentations.

The average of T2 for thyroid reports showed a reduction of 61.8% when comparing the same months in 2012 and 2013. This is the segment of TAT that is most affected with the application of autoverification. However, this improvement was masked while comparing average T1 values for the same months. The reduction in T1 was 34.9%. This reduced improvement was due to the increase in monthly workload for immunoassay analysis in the Clinical Chemistry Unit from 2068 samples to 6144 samples (a 297% increase). TAT is expected to further improve with full utilization of all rules programmed. In 2006, McFadden stated that with the application of autoverification, there could be up to a 44% savings in time—and consequently effort—of the lab staff.<sup>14</sup>

This study has some limitations. The limited connectivity between the LIS and the hospital information system disabled obtaining any clinical or drug history for the patients, and so it did not fully utilize the programmed rules. For the time being, we have overco-

methis defect by registering the previously mentioned comprehensive patient sheet. Moreover, the evaluation of the engine was performed on 1 group of tests only. We are currently working toward its evaluation on cardiac profile tests. We hope that we will soon be able to formulate medical rules for other groups of laboratory tests.

In conclusion, a rule-based autoverification system that utilizes clinical data, among other parameters, was developed and implemented in ZUHL. The verification rate of the system was comparable to that of commercially available software. However, the in-house development of the system saved the hospital the cost of commercially available ones. In spite of a huge increase in monthly workload, the system shortened TAT and minimized the number of samples that needed staff revision. This enabled laboratory staff to devote more time and effort to handle problematic test results and to improve patient care quality.

#### Acknowledgments

*The authors thank Prof Dr Abo-Hashem EM [CG11] (Mansoura University), Prof Dr El-Mogy F (Cairo University), Prof Dr Rizk MM (Alexandria University), and Prof Dr Zagloul MZ (Ain Shams University) for scientific revisions and decisions, and National Technology for technical support.*

## REFERENCES

1. Lundberg GD. How clinicians should use the diagnostic laboratory in a changing medical world. *Clin Chim Acta*. 1999; 280(1-2): 3–11.
2. Clinical and Laboratory Standards Institute. Autoverification of Clinical Laboratory Test Results: Approved Guideline (AUTO 10-A). 2006.
3. Food and Drug Administration. Keeping Blood Transfusions Safe: FDA's Multi-Layered Protections for Donated Blood. February 2002: FS 02-1.
4. College of American Pathologists. Laboratory General Checklist. Commission on Laboratory Accreditation. Laboratory Accreditation Program; 2013.
5. Duca DJ. Autoverification in a laboratory information system. *Lab Med*. 2002;33(1):21–25.
6. Fuentes-Arderiu X, Castiñeiras-Lacabra MJ, Panadero-García MT. Evaluation of the VALAB expert system. *Eur J Clin Chem Clin Biochem*. 1997;35(9):711–714.
7. Demers LM, Spencer C. The thyroid: pathophysiology and thyroid function testing. In: Burtis CA, Ashwood ER, Bruns DE, eds. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 4th ed. St Louis, MO: Elsevier Saunders; 2006:2053–2095.
8. National Technology [homepage on the Internet]. Available from: <http://www.nt-me.com>. Accessed December 19, 2014.
9. Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clin Chem*. 1981;27(3):493–501.
10. David Bowers. Measuring agreement. In: *Medical Statistics from Scratch- An Introduction for Health Professionals*. 2nd ed. West Sussex, UK: John Wiley & Sons Ltd; 2008:181–186.
11. Prost L. How autoverification through the expert system VALAB can make your laboratory more efficient. *Accred Qual Assur*. 2002;7:480–487.
12. Dorizzi RM, Caruso B, Meneghelli S and Rizzotti P. The DNSev™ expert system in the autoverification of tumor markers and hormones results. *Accred Qual Assur*. 2006;11:303–307.
13. Mu-Chin S, Huey-Mei C, Ni T, Chiung-Tzu H, Ching-Tien P. Building and validating an autoverification system in the clinical chemistry laboratory. *Lab Med*. 2011;42(11):668–673.
14. McFadden S. Decision time. *Adv Admin Lab*. 2006;15(11):64.