Accred Qual Assur (2003) 8:291–295
DOI 10.1007/s00769-003-0635-7

PRACTITIONER'S REPORT

Orna Dreazen
Etty Feller
Moshe Leshno

Proficiency testing as a tool for assessing the medical and economic impact of laboratory results: The blood coagulation case

Received: 26 December 2002
Accepted: 23 March 2003
Published online: 15 May 2003

Presented at the 14th International Conference of the Israel Society for Quality, 18–21 November, 2002, Jerusalem, Israel.

O. Dreazen (✉) · E. Feller
Israel Laboratory Accreditation Authority,
2 Habonim St., Ramat-Gan 52522, Israel
e-mail: ornad@israc.gov.il
Tel.: +972-3-5751690
Fax: +972-3-5751695

M. Leshno
Tel Aviv University,

Abstract An international commercial proficiency testing scheme was used to evaluate the impact of laboratory results on clinical decisions. The affect on atrial fibrillation was chosen as a model with 16 Israeli laboratories participating in an international study. A Markov model was constructed to evaluate the impact of any inaccurate results on the clinical outcomes. From the proficiency test study and the Markov model, 13–21% of the results were inaccurate and would have yielded erroneous medical decisions.

Keywords Proficiency testing · Atrial fibrillation · Accuracy · Variation · Markov model

Introduction

Tests results from laboratories of all types comprise an essential part of the medical decision making process. Dr. David W. Secombe wrote "Although laboratories may account for less than 5% of the healthcare budget, the results they generate have a direct impact on approximately 75% of clinical decisions that are made" [1]. The modern world challenges laboratory medicine through two major trends:

I People move rapidly from one place to another and use different medical systems as well as laboratories.
II Technology is moving more and more tests to the point of care, the emergency room and doctor's office, and into the patient's home.

Some estimates have suggested that in 5 years, 80% of all routine testing will be done at the point of care. With these trends, there is a growing need for high quality, accurate and standardized laboratory results. These requirements also increase the awareness and need for a comprehensive infrastructure that can manage and support the quality objectives of such testing.

In the effort to control costs and utilization, we are in fact seeing an increasing awareness of the benefits of the implementation of quality assurance programs in medical laboratories. We have witnessed the introduction of national and international programs. The most widely used in many types of laboratories is accreditation according to ISO/IEC 17025 [2]. Recently, a large group of laboratory medicine experts have written a dedicated standard, ISO 15189: 2003, for the medical field, based on ISO/IEC 17025: 1999 and ISO 9001: 2000 [3]. ISO 15189: 2003 will be widely used by accreditors around the world for accreditation and regulatory purposes.

This standard specifies all elements required by accreditation authorities in order to build trust between the medical laboratory and its clients. One of the most important elements is the use of methods validated by collaborative trails. Thus, conforming to the protocol for

the design, conduct, and interpretation of collaborative studies. Such studies assist laboratories in demonstrating performance, reliability [4], and comparability.

ISO/IEC Guide 43-1 [5] defines proficiency testing as "Determination of laboratory testing performance by means of interlaboratory comparisons". Interlaboratory comparisons may be used in many instances such as: identifying problems and initiating remedial actions, providing additional confidence to laboratory clients, identifying interlaboratory differences, etc.

There are many debates about the differences between, and benefits of, educational and regulatory proficiency testing schemes. Each has its own limitations and advantages, as pointed out at the international conference on proficiency testing for medical laboratories [6]. An educational proficiency testing external quality assessment program (PT/EQA) reaches a wider audience than regulatory PT/EQA and encourages discussion. On the other hand, the response from participants may be lower compared to participation in regulatory schemes, as educational objectives are not always well understood. Also since there is no grading of samples, participants may decide not to reply. Nagy and Collins [7] observed that proficiency testing has both low sensitivity and specificity for the identification of incompetent practitioners. O'Leary [8] claimed that "Although proficiency testing programs should, in principle, contribute to the reduction of cervical cancer mortality, the data does not provide convincing evidence for such a result". It seems that EQA providers gather a lot of helpful information which is not always used for the education and learning process of participating laboratories and regulators.

Israeli medical laboratories are required to participate in proficiency testing schemes regularly. In most cases each laboratory analyzes its own results and learns its own lessons. We wanted to ask ourselves whether the patient's well being would benefit from a nationaly coordinated program to enhance the learning process, and possibly promote practitioners' collaboration so as to increase accuracy and precision, and be a platform to support the quality objectives of testing. As a model we chose the performance of laboratories in prothrombin time international normalized ratio (PT/INR) tests.

PT/INR is a test routinely performed by most medical laboratories. Normal prothrombin time (PT) varies between 10 and 12 s, depending on the tissue factor reagent and other technical details. The therapeutic range of PT depends on the thromboplastin used in each laboratory. The international normalized ratio (INR) which normally ranges between 0.9 and 1.1 has been introduced by the World Health Organization (WHO) to standardize control of anticoagulant therapy internationally. The INR is the ratio of patient PT to the control PT raised to the power of the international sensitivity index (ISI), which is determined by comparing each reagent with WHO thromboplastin:

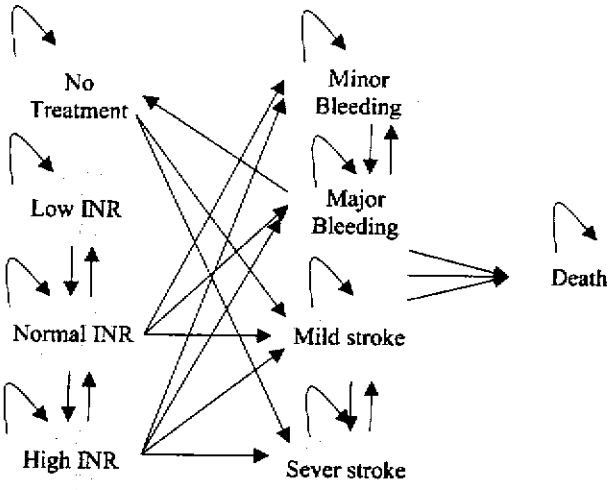

Fig. 1 Markov model

$$\mathrm{INR} = \left(\frac{\mathrm{PatientPT\,(sec)}}{\mathrm{ControlPT\,(sec)}}\right)^{\mathrm{ISI}}$$

These PT/INR results depend heavily on the nature of the control PT used in the test.

Atrial fibrillation (AF) is a common arrhythmia with a prevalence of 4–5% in those aged over 65 years [9]. It is associated with an increased risk of congestive heart failure, an annual incidence of thromboembolic stroke of 4–6%, and decreased longevity [10, 11]. The overall costs for cerebrovascular events are indeed high [12]. Treatment with warfarin provides a high, 70–80%, relative risk reduction in stroke compared with a placebo. However, warfarin is associated with major hemorrhage at a rate of 1.3% per year. Warfarin has a substantial reduction in stroke risk only when an INR of 2.0–3.0 is maintained. The risk of bleeding increases dramatically when the INR increases above 3.0.

Methods

Sixteen laboratories in Israel participated in proficiency tests for blood coagulates. The samples were purchased from the United Kingdom National External Quality Assessment Scheme (NEQAS). Each laboratory received five different samples obtained by plasmapheresis from patients on oral anticoagulant therapy (warfarinized plasma samples), analyzed them and returned the results to NEQAS for comparison and statistical analyses.

Only the results of 3 samples for which the median results generated by over 400 laboratories were in the range, subject for medical intervention, were used for final analysis. Each participant was provided with its own results relative to other participating laboratories.

In addition, we constructed (using DATA software, version 4.0, [13]) a Markov model to simulate the course of a 65–70-year-old patient with AF (Fig. 1).

Table 1 Basic assumptions used in this model: INR, international normalized ratio

Variable	Estimation
Effectiveness of warfarin in reducing stroke	68%
Effectiveness of warfarin when INR is high	75%
Effectiveness of warfarin when INR is low	34%
Annual risk of stroke w/o warfarin	4%
Annual risk of bleeding given normal INR	1%
Probability of fatal bleeding given bleeding	20%
Probability of major bleeding given bleeding	11%
Probability of minor bleeding given bleeding	69%
Annual cost of prophylaxis warfarin including monitoring	$800
Cost of sever stroke	$34,200

The accuracy rate in the INR results was based on the comparison between the actual result and the overall median results of over 400 laboratories participating in the proficiency testing. Other parameters of the model were based on the literature (Table 1). Using this model we evaluated the impact of the inaccurate INR results vis a vis the cost (US$), and ultimately the clinical outcomes. The following clinical outcomes were considered:

- Quality adjusted life years (QALY) has emerged as a preferred outcome measure, and evaluates both longevity of life as well as quality of life.
- Life time risk of sever stroke.
- Life time risk of death caused by stroke or bleeding.

The cycle length in the Markov model was 4 weeks and the following health states were used in the Markov model:

- INR level is low (<2)
- INR is level normal (2–3)
- INR level is high (>3)
- Minor bleeding
- Major bleeding
- Fatal bleeding
- Minor to mild stroke
- Major stroke
- Fatal stroke
- Death.

The basic assumptions in the model were based on the literature and are presented in Table 1.

Results

The range between the highest and lowest results for each sample varied between 39% in the best case and 62% in the worst case. These ranges are shown in Table 2.

It is interesting to note that the average and median calculated for 16 Israeli laboratories matched well with the overall median of laboratories participating in these schemes.

Studies 131, 132 (two samples) were used in our Markovian model. Using the overall median as "the true value" of the analyzed samples, in 13–21% of laboratories, the INR results were inaccurate in a sense that although the INR was within its normal range in 13–21% of laboratories, the INR resulted above the normal range (INR>3), and would have yielded an erroneous medical decision. We defined an accuracy rate as the probability of obtaining an INR result above the normal range given that the true level of the INR is within the normal range of INR (INR between 2 and 3). For the base assumption of the model we used an accuracy rate of 80%.

Table 2 Results obtained by Israeli laboratories

Survey	131	131	132	133	133
Sample No.	1/37	1/38	2/01	2/9	2/10
	2.43	3.8	2.41	1.67	3.02
	3.55	5.1	3.05		
	2.95	4.58	2.67	1.6	3.7
	2.85	4.24	2.51	1.63	3.7
	3.02	4.8	2.85	1.82	3.17
	2.9	4.36	3.49	1.6	3.3
	2.64	4.09	2.74		
	2.54	3.91	2.5	1.6	2.88
	2.54	4.3	2.15	1.34	3.05
	3.24	4.96	2.54	1.7	3.52
	3.1	5.24	2.71	1.7	4.01
	2.26	3.43	2.51	1.64	3.18
	3.1	5.45	2.71	1.54	3.69
	2.88	5	2.84	1.65	3.83
	2.78	4.83	2.51	1.47	3.46
		2.71	1.49	3.19	
AVE[a]	2.85	4.54	2.68	1.60	3.41
SD[b]	0.34	0.58	0.30	0.12	0.34
Median[c]	2.88	4.58	2.69	1.62	3.38
No.[d]	419	417	437	430	432
Median[e]	2.88	4.70	2.50	1.54	3.36

[a] Average of tests results performed by Israeli laboratories.
[b] Standard deviation of tests results performed by Israeli laboratories.
[c] Median of tests results performed by Israeli laboratories.
[d] Number of overall participants in this study.
[e] Median international normalized ratio (INR) obtained by overall participants.

An accuracy rate of 80% with the INR test resulted in 12.92 QALY compared to 13.12 QALY when the accuracy rate is 100%. The QALY gain for improving the accuracy rate from 80% to 100% is 0.2 QALY. Table 3 summarizes the QALY and QALY gained, the lifetime risk for sever stroke, the rate of death caused by stroke or bleeding, and estimated costs for various accuracy rates of the INR tests. Note that we did not assume that increasing the accurate rate of INR would increase the cost of the laboratories.

The model revealed that an accurate rate of INR close to 100% would be beneficial and result in an QALY gain. This achievement may be compared to a conventional glycemic control vs. an intensive glycemic control, or an intensive hypertension control vs. a moderate hypertension control [13].

Table 3 Effect of accuracy of the international normalized ratio (INR) on quality adjusted life years (QALY), patient risk costs

	Without treatment	Accuracy of INR			
		70%	80%	90%	100%
QALY	10.72	12.84	12.92	13.06	13.12
QALY gained/loss	2.4	0.28	0.2	0.06	–
Life time risk of sever stroke	20.1%	9.2%	8.92%	8.56%	8.13%
Rate of death caused by stroke/bleeding	40.5%	23.0%	22.3%	21.4%	20.4%
Average costs ($)	9,900	8,544	8,517	8,484	8,444

Discussion

This work was performed in order to estimate the impact of a laboratory's results on medical decisions and economic aspects.

The model of AF affected by PT/INR tests was chosen for the following reasons:

- The existence of clear and universal criteria according to which the treatment of the patient is determined.
- There are no (accept in extreme cases) clinical signs to support laboratory results.
- Patient follow-up is done by testing every 4 weeks.
- The prevalence of AF is relatively high (4–6%) in this age group (>65-years old).
- The test is frequently performed in both major as well as peripheral laboratories.
- In most cases increasing or decreasing of drug dosage is decided upon by a nurse or sometimes even by the patient, without physician involvement.

For these reasons it is a good model for estimation of the impact of laboratory results on medical decision.

Two parameters were considered among the participating Israeli laboratories: accuracy and variation. Inaccurate results would cause erroneous treatment (decrease or increase of the warfarin dosage, respectively), which can be severe and harmful to the patients. Participating in a proficiency scheme as above on a regular basis can be used as a tool in evaluating the complete process as well as an educational training tool for corrective actions where necessary. Can these parameters be a controlled using proficiency testing schemes?

In their review paper discussions Libeer et al. [6] describe working groups addressing issues associated with the development and management of proficiency testing programs, and with their optimal use.

The authors mention that educational proficiency testing programs should encourage discussion. Among the essential elements for PT/EQA to be a useful educational experience are:

(1) Focused objectives and clear standardized instructions.
(2) Interlaboratory comparison of results by methods and timely feedback.
(3) Knowledge of differences between disciplines, tests methods, and clinical practice patterns.

A Markov model was constructed to simulate the course of a 65–70-year-old patient with AF. The route of treatment depending on the INR level. The model enables calculation of life time expectancy , life time risk for sever stroke, rate of death caused by stroke/bleeding, and costs as a function of the accuracy rate of laboratories INR results. An accuracy rate of 80% with the INR test resulted in 12.92 QALY compared to 13.12 QALY when the accuracy rate is 100%. The QALY gain for improving the accuracy rate from 80% to 100% is 0.2 QALY. The rate of death caused by stroke/bleeding increases from 20.4% to 22.3% when the accuracy rate of INR results varies from 100% to 80%. The average cost of AF decreases with the increase in the accuracy rate of INR.

The model revealed that by achieving an accurate rate of INR close to 100% benefits could be gained, and the QALY gain would be compared to a conventional glycemic control vs. an intensive glycemic control, or an intensive hypertension control vs. a moderate hypertension control.

Conclusions

We conclude that although the economic as well as the clinical impact of the accuracy rate of INR is not enormous, improving the accuracy rate of the INR result with an increase of life expectancy will decrease the risk of death caused by stroke or bleeding and decreases costs, and an improvement of the accuracy rate is called for.

Proficiency testing schemes can be (and possibly should be) used for standardization in a region or a country.

Use of different equipment, calibrators, and standard operating procedures ultimately leads to a variation between results of medical laboratories. There are many reasons for this variability in technologies and reagents. For example:

- Laboratories like to use advanced equipment and not keep to old techniques and sometimes slow machinery.
- Laboratories do not like to depend on one reagent provider.
- Economic advantages are gained by competition between suppliers.

And more.

The freedom to choose new types of laboratory equipment together with point-of-care testing instruments works in some ways against the patient. Because producing variable results with different testing expands uncertainties. In most cases clinicians are not aware of these variations in accuracy and precision; however, being aware would not solve the problem. It is the responsibility of laboratory medicine community to make every effort to standardize their results, and give aid and support to clinicians in their everyday work.

Success in a proficiency testing event does not demonstrate that a laboratory is performing well in everyday practice, but rather that the laboratory has the capability to perform well under circumstances similar to those of the proficiency testing activity [8]. Proficiency testing programs cannot make up for lack of attention to all other quality assurance activities, but instead complement them.

In order to support national standardization, the information gathered by all laboratories in a country should be analyzed and used to generate relevant performance criteria.

Working with a unified calibrator and in a PT/EQA scheme could enhance harmonization and benefit patients.

Nevertheless, we believe that using proficiency testing is advantageous and can at least be used as a good indicator for these parameters.

References

1. Applications of the Internet to PT Accreditation. (2002) Accred. Qual. Assur 7/8–9:363–364
2. ISO/IEC 17025 (1999) General requirements for the competence of testing and calibration laboratories. ISO, Geneva, Switzerland
3. ISO 15189 (2003) Medical laboratories—Particular requirements for quality and competence. ISO, Geneva, Switzerland
4. Horowitz W (1988) Pure Appl Chem. 60:855–867
5. ISO/IEC Guide 43–1 (1997) Proficiency testing by interlaboratory comparisons—Development and operation of proficiency testing schemes. ISO, Geneva, Switzerland
6. Libeer JC, Richardson H, Bullock D, Heuck C, Tholen D, Martin R, Carter J, Boone EJ, Noble M (2002) Accred Qual Assur 8–9:320–334
7. Nagy GK, Collins DN (1991) Acta Cytol 35:3–7
8. O'Leary TJ (2002) Accred Qual Assur 7:357–361
9. Sudlow M, Thomson R, Thwaites B, Rodgers H, Kenny RA (1998) Lancet 352:1167–1171
10. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D (1998) Circulation 98:946–952
11. Kannel WB, Abbott RD, Savage DD, McNamara PM (1982) N Engl J Med 306:1018–1022
12. Gustafsson C, Asplund K, Britton M, Norrving B, Olsson B, Marke LA. (1992) Br Med J 305:1457–1460
13. DATA software, Version 4.0 (2002) TreeAge Software Inc., Williamstown, MA, USA
14. The CDC Diabetes Cost-effectiveness Group (2002) JAMA 287:2542–2551