

January 2015

ISSN 1650-3414

Volume 26 Number 1



Communications and Publications Division (CPD) of the IFCC

Editor-in-chief : Prof. Gábor L. Kovács, MD, PhD, DSc

Department of Laboratory Medicine, Faculty of Medicine, University of Pecs, Hungary

e-mail: ejournal@ifcc.org

The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine



In this issue

Foreword of the editor

Gábor L. Kovács

3

Measuring the impact of laboratory medicine on clinical management and patient outcomes

Mike Hallworth

4

Diagnostic errors and laboratory medicine – causes and strategies

Mario Plebani

7

Towards better test utilization – strategies to improve physician ordering and their impact on patient outcomes

Danielle B. Freedman

15

Collaborating with international clinical organizations

Howard A. Morris

31

The impact for patient outcomes of failure to follow up on test results. How can we do better?

Joanne Callen, Andrew Georgiou, Julie Li, Johanna I Westbrook

38

The clinical and health economic value of clinical laboratory diagnostics

Bruce Jordan, Cheryl Mitchell, Andy Anderson, Norbert Farkas, Richard Batrla

47

Evaluating biomarkers for guiding treatment decisions

Patrick M. Bossuyt, Parvin Tajik

63

Book review — A Practical Guide to ISO 15189 in Laboratory Medicine

Joseph B. Lopez

71

Book review — Practical Clinical Chemistry: Core Concepts

Barna Vasarhelyi

73

Foreword of the editor

Editor in Chief: Gábor L. Kovács, MD, PhD, DSc

This themed issue of the journal is focused around the impact of laboratory medicine on clinical management and patient outcomes. Mr. Mike Hallworth (UK) was asked to guest edit the issue. Mike Hallworth MA MSc MCB FRCPath has recently retired from the post of Consultant Clinical Scientist to the Shrewsbury and Telford Hospital NHS Trust, based at the Royal Shrewsbury Hospital, Shrewsbury, Shropshire, UK. He is a past President of the European Communities Confederation of Clinical Chemistry and Laboratory Medicine (EFLM) and past Chairman of the UK Association for Clinical Biochemistry. Mike was awarded the

UK Healthcare Scientist of the Year Award in November 2008 by the UK Chief Scientific Officer, Professor Sue Hill, and was the 2011 winner of the EFCC-Roche European Scientific Award for Laboratory Medicine. He has been vice-Chair and Chair of the AACC's Annual Meeting Organizing Committee (in 2007 and 2010, respectively), and is the Chair of the IFCC Task Force on the impact of laboratory medicine on clinical management and outcomes. He is Co-Editor-in-Chief of a new journal, Practical Laboratory Medicine (Elsevier). Mr. Hallworth invited a number of internationally renowned laboratory scientists to discuss the topic.

Measuring the impact of laboratory medicine on clinical management and patient outcomes

Guest Editor: Mike Hallworth

Chair, IFCC Task Force on the impact of laboratory medicine on clinical management and outcomes (TF-ICO)

All who work in laboratory medicine have anecdotal evidence of the value of laboratory medicine in delivering safe and effective patient care and improving individual patient outcomes by enabling faster, more accurate diagnosis and effective treatment. However, systematic evidence of the contribution of laboratory medicine to the clinical process has been much harder to obtain – understandably so, in view of the multitude of factors that are involved in reaching a diagnosis or planning treatment for an individual. Laboratory medicine has also had a broader impact upstream of diagnosis and management, playing a key role in areas such as risk assessment and screening of healthy subjects for latent disease. These areas are becoming increasingly important with the recognition that early diagnosis and intervention reduces overall healthcare costs for a wide range of common diseases.

The so-called “70% claim” is commonly cited to indicate the value of laboratory medicine. It occurs in various forms, most commonly that “Laboratory medicine data influences 70% of clinical decisions” (1), or minor variations around this figure. Unfortunately, the data on which this claim was based represents unpub-

lished studies and anecdotal observations (2), and cannot now be objectively verified.

We need more specific and evidence-based measures of the added value of laboratory medicine, which in turn require better designed studies and better use of existing biomarkers. The IFCC Task Force on the Impact of Laboratory Medicine on Clinical Management and Outcomes was established by the Executive Board in 2012 to evaluate the available evidence supporting the impact of laboratory medicine in health care, and to develop the study design for new and prospective studies to demonstrate the contribution made by laboratory medicine to improving outcomes.

The Task Force has recently published its report (3), which summarizes the existing evidence and indicates the gaps in our understanding. It also identifies deficiencies in current utilization, suggests potential solutions and offers a vision of a future in which laboratory medicine is used optimally to support patient care. This special issue of eJIFCC explores the central issues in more detail, with contributions from acknowledged experts in the field.

Rapid, accurate diagnosis of the patient’s presenting condition is essential to obtaining the best outcome, and there has been much

emphasis on recent years in reducing diagnostic error. The work of Plebani's group has clearly shown that, where diagnostic error arises from laboratory testing, the pre- and post-analytical phases are much more vulnerable to error than the actual analytical phase (4), which implies that laboratories need to refocus their efforts on error reduction toward the total testing process rather than simply on the analytical aspects of their work. Mario Plebani develops these ideas in the first article in this issue *"Diagnostic Errors and Laboratory Medicine – causes and strategies"*. He emphasises the importance of focusing on appropriate test utilization and accurate result interpretation to reduce the overall risk of laboratory-related diagnostic errors and improve patient care.

Danielle Freedman takes up this theme in our second article *"Towards better test utilization – strategies to improve physician ordering and their impact on patient outcomes"*. She discusses the factors that influence test ordering by physicians, and describes proven strategies for achieving change which improve laboratory utilization and have a direct effect on patient outcomes. Influencing the behaviour of individual physicians is important, but physicians are increasingly reliant on evidence-based international guidelines for effective diagnosis and management of disease, and the laboratory community must ensure that it is represented when these guidelines are prepared if the uses and limitations of laboratory tests are to be properly understood. Howard Morris' article *"Collaborating with International Clinical Organizations"* describes IFCC's role in working with international clinical organizations to enhance the effective translation of developments in laboratory medicine to improve patient care and clinical outcomes, and ensure their adoption into routine clinical practice via inclusion in relevant clinical guidelines.

However good a laboratory test, it cannot affect the individual patient outcome if the result never reaches the clinician who is responsible for delivering care. Joanne Callen and colleagues address the topic of *"The impact for patient outcomes of failure to follow up on test results. How can we do better?"*, and outline potential solutions to the widespread problem of missed results. Solving that problem requires the laboratory to get involved in establishing and maintaining resilient governance approaches, and creating a culture dedicated to ensuring reliable and safe patient care.

Having explored in detail what needs to be done to ensure that laboratory tests are ordered and used appropriately, the other two presentations in this issue focus on how the value of laboratory medicine can be measured and demonstrated. Bruce Jordan and colleagues discuss *"The clinical and health economic value of clinical laboratory diagnostics"*, using as exemplars three disease areas that represent substantial health care burdens for society – heart failure, Alzheimer's disease and asthma. Finally, Patrick Bossuyt and Parvin Tajik's article *"Evaluating biomarkers for guiding treatment decisions"* presents a theoretical framework for evaluating treatment decisions and summarizes study designs for evaluating treatment selection markers. It is vitally important that new markers receive robust outcome-based evaluations before they are introduced into clinical practice, in exactly the same way that new drugs are evaluated before they are licensed. The European Group on Tumor Markers has recently published a proposal on evaluation of new tumor markers (5), which describes a four-phase approach, similar to the process used by the FDA and others for the evaluation of new drugs.

The report of the IFCC Task Force (3) concludes that work is required in five areas to ensure that laboratory medicine is firmly focussed on improving outcomes:

1. Improved utilization of existing and new tests. This requires determination of optimum testing strategies based on patients' presenting complaints, development of interventions to support appropriate test ordering/requesting, proper sample collection, transport and storage, effective strategies for transmission of test results, agreement on clinically-appropriate triggers for critical result notification and consultative services and comments to ensure that results are properly applied.
2. Defining new roles for laboratory professionals that are focussed on optimizing patient outcomes by adding value at all points of the diagnostic brain-to-brain cycle and auditing the effectiveness of these roles and the overall diagnostic process.
3. Development of standardized protocols for prospective patient-centred studies of biomarker clinical effectiveness or extra-analytical process effectiveness.
4. Benchmarking of existing and new tests in specified situations with commonly accepted measures of effectiveness including post-implementation audit. This must include the effects of pre- and post-analytical components of the testing process, and must consider the overall impact of the testing process on all relevant clinical outcomes.
5. Agreed definition and validation of effectiveness measures and use of checklists for articles submitted for publication.

Laboratory doctors and scientists of the future must be involved in producing guidelines for investigation, advising clinical staff on the best strategy for individual clinical presentations and the further tests needed to confirm a diagnosis, and ensuring that results are not misinterpreted or missed and that resources (human, technical and financial) are used to do the right test on the right person at the right time. It's a daunting challenge, but getting this right means better use of tests, better patient care, lower health care costs, improved job satisfaction for laboratory workers and enhanced ability to recruit and retain good scientists in laboratory medicine. That's a goal worth working for, and the Editors hope that the Task Force report and the contents of this special issue will inspire and equip laboratorians across the world to rise to the challenge!

REFERENCES

1. Beastall GH. Adding value to laboratory medicine: a professional responsibility. *Clin Chem Lab Med* 2013; 51: 221-7.
2. Hallworth MJ. The 70% claim – what is the evidence base? *Ann Clin Biochem* 2011; 48: 487-8.
3. Hallworth MJ, Epner PL, Ebert C, Fantz CR, Faye SA, Higgins TN, et al. Current evidence and future perspectives on the effective practice of patient-centered laboratory medicine. *Clin Chem* 2015; 61: in press.
4. Plebani M. Detection and prevention of errors in laboratory medicine. *Ann Clin Biochem* 2010; 47: 101-10.
5. Soletormos G, Duffy MJ, Hayes DF, Sturgeon CM, Barak V, Bossuyt PM et al. Design of tumor biomarker-monitoring trials: A proposal by the European Group on Tumor Markers. *Clin Chem* 2013; 59: 52-9.

Diagnostic errors and laboratory medicine – causes and strategies

Mario Plebani

Department of Laboratory Medicine, University-Hospital, Padova, Italy

ARTICLE INFO

Corresponding author:

Prof. Mario Plebani
Department of Laboratory Medicine
University-Hospital of Padova
35128 Padova, Italy
Phone: +390498212792
Fax: +39049663240
E-mail: mario.plebani@unipd.it

Key words:

quality, patient safety,
laboratory medicine, diagnostic
errors, brain-to-brain loop,
laboratory-associated errors

ABSTRACT

While the frequency of laboratory errors varies greatly, depending on the study design and steps of the total testing process (TTP) investigated, a series of papers published in the last two decades drew the attention of laboratory professionals to the pre- and post-analytical phases, which currently appear to be more vulnerable to errors than the analytical phase. In particular, a high frequency of errors and risk of errors that could harm patients has been described in both the pre-pre- and post-post-analytical steps of the cycle that usually are not under the laboratory control. In 2008, the release of a Technical Specification (ISO/TS 22367) by the International Organization for Standardization played a key role in collecting the evidence and changing the perspective on laboratory errors, emphasizing the need for a patient-centred approach to errors in laboratory testing.

A further step in the journey towards improved understanding of the issue is the recent demonstration that errors in laboratory medicine are part of a much wider issue, commonly known as “diagnostic error”, thus definitively linking laboratory-associated errors to patient safety problems. The current awareness of the nature of laboratory testing-associated errors, in particular the link between appropriate test ordering and result interpretation/utilization, and their potential in reducing diagnostic errors, should herald a change in the old paradigm which was focused only on

errors detected within the laboratory walls. Evidence-based quality indicators represent a formidable tool for improving quality and decreasing the risk of errors in the total testing process.

INTRODUCTION

During the past decade, after the publication of the Institute of Medicine (IOM) report, *To Err Is Human* (1), patient safety has finally become the object of medical and public attention. Compared with other types of medical error, however, errors in laboratory medicine have received little attention. The reasons for this neglect are complex, but the difficulties largely arise from the number of steps and the time lapse which separate laboratory testing, physicians' actions and patient outcomes (2). Moreover, usually only the analytical phase falls under laboratory control, while the pre- and post-analytic phases are the responsibility of stakeholders other than the laboratory such as the clinician, the nurse, the patient and others involved in patient identification, data entry, specimen collection and transport. In addition, most of the many different terms used in the literature to define errors in laboratory medicine (e.g. mistakes, blunders, defects, outliers, unacceptable results, quality failure) have negative connotations involving blame, individual failure and culpability and, even worse, pertain to studies focusing on a limited number of total testing process (TTP) steps. Taken together these are the "reasons for neglect" for errors in laboratory medicine, and should explain why the patient-centred viewpoint has been taken into account only in recent years (3).

A brief history of errors in laboratory medicine

Initial studies, starting from the seminal paper by Belk and Sunderman in 1947 (4), as well as other articles published before the 1990s,

focused only on the analytical phase and demonstrated high rates and severity of analytical errors. However, despite the limited study design, they provided a wide range of opportunities to improve analytical performance, including the development of external quality assurance programs (EQA) and improved rules for internal quality control (IQC).

In the late nineties, a body of evidence was accumulated which documented: a) a dramatic decrease in the analytical error rates from 162,116 errors per million laboratory tests (parts per million, ppm) to 447 ppm (5, 6); b) high rates of errors in the pre- and post-analytical steps (7-9); and c) the risk of adverse events and inappropriate care due to laboratory errors, mainly for errors in pre-pre-analytical steps (10, 11).

In fact, over the past decades, a ten-fold reduction in the analytical error rate has been achieved thanks to improvements in the reliability and standardization of analytical techniques, reagents, and instrumentation. In addition, advances in information technology, quality control and quality assurance methods have made a valuable contribution to error reduction. However, although the state-of-the-art highlights that pre- and post-analytical phases are more vulnerable to errors, there is still evidence indicating that analytical quality remains a major issue. In particular, a relatively high frequency of analytical errors has been documented for immunoassays with associated adverse clinical outcomes, sometimes resulting in grossly erroneous results (2). The issue of analytical interference does not only affect immunoassays. As an example, monoclonal proteins may affect many laboratory measurements, including glucose, bilirubin, C-reactive protein, creatinine and albumin. The frequency of this type of error is variable and probably underreported (12). The lack of inter-changeability between different methods

and clinical laboratories, although not considered an “analytical error” in the strict sense, may also confound both clinical reasoning and patient management. This, in turn, is the main driver for the increasing awareness and concern regarding the need of standardization and harmonization projects in laboratory medicine (13).

Pre- and post-analytical phases

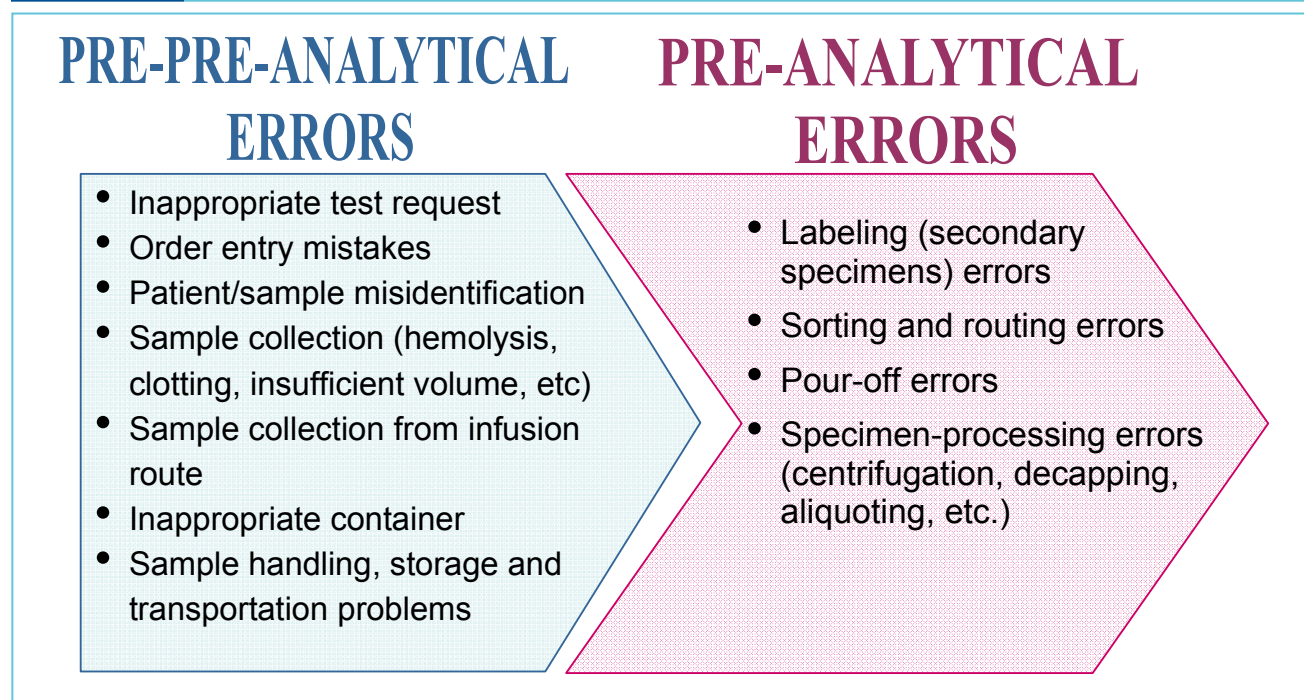
While the frequency of laboratory errors varies greatly, depending on the study design and the specific steps of the total testing process (TTP) investigated, a series of papers published between 1989 and 2007 drew the attention of laboratory professionals to the pre-, and post-analytical phases, which currently appear to be more vulnerable to errors than the analytical phase. In particular, two papers published in 1997 and 2007 (7, 8) used a study design that allowed us to investigate most TTP steps in the same clinical context (stat

laboratory). In both studies, the pre-analytic phase had the highest error rate, the most frequent problems arising from mistakes in tube filling, inappropriate specimen containers, and requesting procedures. Identification errors were noted too, although the appropriateness of test request was not considered in the study design. Further studies confirmed these data and, currently, pre-analytical errors or more accurately pre-pre-analytical errors are estimated to account for up to 70% of all mistakes made in laboratory diagnostics, most of which arise from problems in patient preparation, and sample collection, transportation, preparation for analysis and storage (9-11), as shown in Figure 1.

Laboratory errors and risk management

From a risk management viewpoint, the great majority of laboratory errors have little direct impact on patient care but provide important learning opportunities. In fact, any error,

Figure 1 Most frequent sources of errors in the pre-pre- and pre-analytical steps (accounting for 48%-62% of total errors in laboratory medicine)



regardless of its apparently trivial nature, might indicate weaknesses in policies and procedures that may not lead to adverse events in their particular context, but might cause the patient harm in slightly different circumstances (14). The lesson we learnt is that the entire system (TTP) should be designed to consider not only the real patient harm sustained, but also the potential worst clinical outcome if such an error were to recur.

In 2008, the release of a Technical Specification (ISO/TS 22367) by the International Organization for Standardization played a key role in collecting the evidence and changing the perspective on laboratory errors, defining laboratory error as “failure of planned action to be completed as intended, or use of a wrong plan to achieve an aim, occurring at any part of the laboratory cycle, from ordering examinations to reporting results and appropriately interpreting and reacting to them” (15). In addition, according to this Technical Specification (15), any clinical laboratory should employ processes for: a) identifying high risk processes where the potential error could lead to a safety risk for patients; b) detecting actual incidents associated with deviations from standard requirements; c) estimating and evaluating the associated risks to patient safety; d) controlling the risks; and e) monitoring the effectiveness of the measure taken.

This inspired a patient-centred evaluation of errors in laboratory testing and an increased concern to identify weaknesses and vulnerability in procedures and processes, so that corrective and preventive actions can be activated before any adverse event or patient harm may occur.

A further step in the journey towards a better understanding of the issue is the recent proof that errors in laboratory medicine are part of a much wider issue, commonly known as

“*diagnostic error*”, thus definitively linking laboratory-associated errors to patient safety problems, as shown in Table 1.

Diagnostic errors and laboratory testing

Diagnostic errors have been defined as “errors in which *diagnosis* was unintentionally *delayed* (while sufficient information was available earlier), *wrong* (another diagnosis made before the correct one), or *missed* (no diagnosis made) as judged from the eventual appreciation of more definitive information (e.g., autopsy studies)” (16). The evidence on the importance of and direct link between diagnostic errors and errors in laboratory medicine derives from a series of studies with a clinical starting point. In particular, studies performed on the pre-pre-analytical phase (initial procedures performed outside clinical laboratory or, at least in part, beyond the control of laboratory personnel) confirm that failure to order appropriate diagnostic tests (laboratory tests included) makes up 55% of observed breakdowns in missed and delayed diagnosis in the ambulatory setting (17-19) and 58% of errors in emergency departments (20).

Incorrect interpretation of diagnostic or laboratory tests in the end stages of the TTP loop was found to underlie a large percentage of errors in the ambulatory setting and in emergency departments. Failure to inform patients of clinically significant abnormal test results or to record the delivery of relevant information is relatively common, occurring in 1 out of every 14 tests; for example, patients not being informed of a total cholesterol value of 8.2 mmol/L (318 mg/dL), hematocrit of 28.6% or a potassium level of 2.6 mmol/L. The overall rate of failure to inform the patient or to record communication of information was 7.1%, in different practices, ranging from 0 to 26% (21). As revealed in a systematic review of the literature, failure to follow-up test results markedly

Table 1 The journey towards a patient-centered view of errors in laboratory medicine

Years	
1950-1990	ANALYTICAL ERRORS ↓
1990s	ERRORS IN CLINICAL LABORATORIES (including pre- and post-analytical phases) ↓
2000s	ERRORS IN THE TOTAL TESTING PROCESS (including pre-pre- and post-post-analytical phases)
	AND
Today	TESTING-RELATED DIAGNOSTIC ERRORS

compromises patient safety, yet the rate of abnormal laboratory results (for INR and PSA) without follow-up ranges from 6.8% to 62%. (22). Further evidence of inappropriate response to laboratory information is provided in a study evaluating the prescription of potassium in cases of hyperkalemia (23). Moreover, findings in another study (24) showed that over 2% (2.6% in 2000, 2.1% in 2007) of patients with thyrotropin (TSH) levels exceeding 20 mU/L were not followed up. Yet another study revealed that of 1,095 discharged patients, almost half had pending laboratory and radiology test results, 9% of which potentially required action (25). In another study, approximately one-third of sub-acute care patients had laboratory tests (microbiology tests in

particular), which were pending at discharge, but few of these cases were recorded in hospital discharge forms (26). Overall, data reported demonstrate that the initial and final steps of the TTP process, above all test requesting and reaction to laboratory results, are not only more error-prone than all the other steps, but are also the most important causes of potential adverse outcomes for patients. Moreover, the data confirm that a significant number of failures occur in the interface between clinical practice and laboratories, thus emphasizing the need for laboratory professionals and physicians to “understand their mutual ownership and work together to ensure that patients are more safe” (27).

Towards a patient-centred approach to laboratory-associated errors

The awareness of the current nature of laboratory testing-associated errors, in particular the link between appropriateness in test ordering and result interpretation/utilization, and their potential in addressing diagnostic errors, should herald a change in the old paradigm which was focused only on errors detected within the laboratory walls. In order to translate the concept of “patient-centred care” from theory to practice it is of the utmost importance to investigate, and improve upon, not only those procedures and processes performed under the direct control of the clinical laboratory, but also the initial and final steps of the testing cycle that are usually managed by other healthcare personnel. Projects aiming to improve quality and patient safety must therefore be based upon a total quality perspective, in particular the accreditation of clinical laboratory services according to the International Standard ISO 15189:2012 (28) and the search for valuable quality indicators (QIs) for all phases of the testing process. In particular, the identification and implementation of valuable QIs are requested as mandatory for clinical laboratory accreditation according to the International Standard (ISO 15189:2012). In this document quality indicators are defined as “a measure of the degree to which a set of inherent characteristics fulfils requirements” and “can measure how well an organization meets the needs and requirements of users and the quality of all operational processes” (28). The second definition is emphasised in the context of the present paper, and specifically the fact that “all operational processes” requires the inclusion of pre- and post-analytical steps. However, a major problem is the lack of consensually defined QIs, particularly for extra-analytical phases. The International Federation of Clinical Chemistry

and Laboratory Medicine (IFCC) launched in 2004 a new project, implementing a Working Group on Laboratory Errors and Patient Safety (WG LEPS) that promoted and developed a model of quality indicators (MQI) (29, 30). This model is divided into process and outcome measures, mainly based on measures of the pre-, intra- and post-analytical procedures and processes, and has been revised in a Consensus Conference organized to establish a list of QIs that should be evidence-based, feasible for most laboratories around the world and actionable (31). The list of QIs is available on line at www.ifcc-mqi.com

CONCLUSIONS

According to recent data from malpractice claims, diagnostic errors appear to be the most common, most costly and most dangerous of medical mistakes both in inpatients and outpatients (32, 33). Failure in the ordering of appropriate laboratory test and the application of laboratory test results are major contributors to diagnostic errors, along with residual problems in test performances (analytical errors) (34). Therefore, the main message is the need to improve the quality of laboratory services, avoiding errors and improving patient safety, employing a global approach across the TTP, according to the seminal concept of the brain-to-brain loop (35). The use of a consensually-defined list of evidence-based QIs to be applied in the accreditation programs of clinical laboratories according to the current International Standard (ISO 15189:2012) is an effective tool for improving quality, decreasing the risk of errors and increasing patient safety.

REFERENCES

1. Institute of Medicine (US) Committee on Quality of Health Care in America; Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health

- System. Washington (DC): National Academies Press (US); 2000.
2. Plebani M. The CCLM contribution to improvements in quality and patient safety. *Clin Chem Lab Med* 2013;51:39-46.
3. Plebani M. The detection and prevention of errors in laboratory medicine. *Ann Clin Biochem* 2010;47:101-10.
4. Belk WP, Sunderman FW. A survey of the accuracy of chemical analyses in clinical laboratories. *Am J Clin Pathol* 1947;17:853-61.
5. Steindel SJ, Howanitz PJ, Renner SW. Reasons for proficiency testing failures in clinical chemistry and blood gas analysis: a College of American Pathologists Q-Probes study in 665 laboratories. *Arch Pathol Lab Med* 1996;120:1094-101.
6. Witte VanNess SA, Angstadt DS, Pennell BJ. Errors, mistakes, blunders, outliers, or unacceptable results: how many? *Clin Chem* 1997;43:1352-6.
7. Plebani M, Carraro P. Mistakes in a stat laboratory: types and frequency. *Clin Chem* 1997;43: 1348-51.
8. Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. *Clin Chem* 2007;53:1338-42.
9. Plebani M Exploring the iceberg of errors in laboratory medicine. *Clin Chim Acta* 2009;404: 16-23.
10. Howanitz PJ. Errors in laboratory medicine: practical lessons to improve patient safety. *Arch Pathol Lab Med* 2005;129:1252-61.
11. Carraro P, Zago T, Plebani M. Exploring the initial steps of the testing process: frequency and nature of pre-analytic errors. *Clin Chem* 2012;58:638-42.
12. Dalal BI, Bridgen ML. Factitious biochemical measurements resulting from hematologic conditions. *Am J Clin Pathol* 2009;131: 195–204.
13. Plebani M. Harmonization in laboratory medicine: the complete picture. *Clin Chem Lab Med* 2013;5:7411-51.
14. O’Kane M. The reporting, classification and grading of quality failures in the medical laboratory. *Clin Chim Acta* 2009; 404:28-31.
15. ISO/TS 22367:2008. Medical laboratories- reduction of error through risk management and continual improvement. Geneva, Switzerland: International Organization for Standardization, 2008.
16. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493-9.
17. Hickner J, Graham DG, Elder NC, Brandt E, Emsermann CB, Dovey S, et al. Testing process errors and their harms and consequences reported from family medicine practices: a study of the American Academy of Family Physicians National Research Network. *Qual Saf Health Care* 2008;17:194-200.
18. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med* 2006;145:488-96.
19. Wahls TL, Cram PM. The frequency of missed test results and associated treatment delays in a highly computerized health system. *BMC Fam Pract* 2007;8:32-42.
20. Kachalia A, Gandhi TK, Pupolo AL, Yoon C, Thomas EJ, Griffey R, et al. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med* 2007;49:196-205.
21. Casalino LP, Dunham D, Chin MH, Bielang R, Kistner EO, Karrison TG et al. Frequency of failure to inform patient of clinically significant outpatient test results. *Arch Int Med* 2009;169:1123-9.
22. Callen JL, Westbrook JI, Georgiou A, Li J. Failure to follow-up test results for ambulatory patients: a systematic review. *J Gen Intern Med* 2012;27:1334-48.
23. Schiff GD, Aggarwal HC, Kumar S, McNutt RA. Prescribing potassium despite hyperkalemia: medication errors uncovered by linking laboratory and pharmacy information systems. *Am J Med* 2001;109:494-7.
24. Schiff GD, Kim S, Krosnjar N, Wisniewski MF, Bult J, Fogelfeld L, et al. Missed hypothyroidism diagnosis uncovered by linking laboratory and pharmacy data. *Arch Intern Med* 2005;165 (5):574-7.
25. Roy CL, Poon EG, Karson AS, Ladak-Merchant Z, Johnson RE, Maviglia SM, et al. Patient safety concerns arising from test results that return after hospital discharge. *Ann Intern Med* 2005;143:121-8.
26. Walz SE, Smith M, Cox E, Sattin J, Kind AJ. Pending laboratory tests and the hospital discharge summary in patients discharged to sub-acute care. *J Gen Intern Med* 2011;26:393-8.
27. Graber ML. The physician and the laboratory. Partners in reducing diagnostic error related to laboratory testing. *Am J Clin Pathol* 2005;124 (Suppl 1):S1-S4.
28. ISO 15189:2012. Medical laboratories - Requirements for quality and competence. Geneva, Switzerland: International Organization for Standardization, 2012.
29. Sciacovelli L, O’Kane M, Skaik YA, Caciagli P, Pellegrini C, Da Rin G, et al. IFCC WG-LEPS. Quality Indicators in Laboratory Medicine: from theory to practice. Preliminary data from the IFCC Working Group Project “Laboratory Errors and Patient Safety”. *Clin Chem Lab Med* 2011;49:835-44.

30. Plebani M, Chiozza ML, Sciacovelli L. Towards harmonization of quality indicators in laboratory medicine. *Clin Chem Lab Med* 2013;51:187-95.
31. Plebani M, Astion ML, Barth JH, Chen W, de Oliveira Galoro CA, Escuer MI, et al. Harmonization of quality indicators in laboratory medicine. A preliminary consensus. *Clin Chem Lab Med* 2014;52: 951-8.
32. Saber Tehrani AS, Lee H, Mathews SC, Shore A, Makary MA, Pronovost PJ, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National Practitioner Data Bank. *BMJ Qual Saf* 2013;22:672-80.
33. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013;173:418-25.
34. Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. *BMJ Qual Saf* 2013;22:ii6-ii10 doi: 10.1136/bmjqs-2012-001621.
35. Plebani M, Laposata M, Lundberg GD. The brain-to-brain loop concept for laboratory testing 40 years after its introduction. *Am J Clin Pathol* 2011;136:829-33.

Towards better test utilization – strategies to improve physician ordering and their impact on patient outcomes

Danielle B. Freedman

Luton & Dunstable University Hospital NHS Foundation Trust

ARTICLE INFO

Corresponding author:

Danielle B. Freedman
Director of Pathology
Luton & Dunstable University Hospital
NHS Foundation Trust
Lewsey Road, Luton, Bedfordshire
LU4 0DZ, UK
E-mail: danielle.freedman@ldh.nhs.uk

Key words:

test utilization, test ordering, physician education, computerized physician order entry, guidelines, clinical outcomes

ABSTRACT

Laboratory medicine is the single highest volume medical activity in healthcare and demand for laboratory testing is increasing disproportionately to medical activity. It has been estimated that \$6.8 billion of medical care in the US involves unnecessary testing and procedures that do not improve patient care and may even harm the patient. Physicians face many challenges in accurately, efficiently and safely ordering and interpreting diagnostic tests. In order to improve patient outcomes, laboratory tests must be appropriately ordered, properly conducted, reported in a timely manner, correctly interpreted and affect a decision for future diagnosis and treatment of the patient.

This paper discusses factors influencing test ordering by physicians, strategies for modifying physicians' ordering patterns, and ways to implement policies to improve laboratory utilization and thereby improve patient outcome.

Successful management of laboratory test utilization requires the entire laboratory team to use their skills and knowledge to identify utilization issues, implement a programme that will achieve more effective testing and establish appropriate processes from the beginning to the end of the test cycle.

Laboratory medicine is the single highest volume medical activity in healthcare and demand for laboratory testing is increasing disproportionately to medical activity. Over the past 20 years, the number of laboratory tests available to clinicians has more than doubled, to at least 3,500 tests (1). The global IVD market, valued at \$49 billion in 2012, is expected to grow by 7% over the period 2012-2017, and represents 3-5% of all healthcare costs (2).

A major component of US healthcare expenditure is an estimated \$65 billion spent each year to perform more than 4.3 billion laboratory tests (3) but it has been estimated that \$6.8 billion of medical care in the US involves unnecessary testing and procedures that do not improve patient care and may even harm the patient (4). Physicians face many challenges in accurately, efficiently and safely ordering and interpreting diagnostic tests. (The term 'ordering' will be used throughout this paper for consistency. However, tests are 'requested', not 'ordered' in many countries, and 'requesting' better reflects the collaboration between clinician and laboratory). To improve patient outcomes, laboratory tests must be appropriately ordered, properly conducted, reported in a timely manner, correctly interpreted and affect a decision for future diagnosis and treatment of the patient (5).

However, the use of laboratory diagnostics varies between countries and in the US it was 5 times greater (as a proportion of medical expenditure) than in the UK in 2006 (2). Large differences between individual practitioners in laboratory utilization have been reported in several countries (6-9) The recent publication in England of the 'National Health Service Atlas of Variation' (10) demonstrated the variation in ordering rates for diagnostic tests across 151 primary care organizations. There may be valid reasons to explain some of the observed variation, such as different populations or case mix, incidence of deprivation, disease prevalence, local policy decisions

on specific services and the availability of relatively new or high-technology tests. However, despite these factors, the variation in ordering rates is so large that it must reflect considerable differences in the individual ordering patterns of doctors within each primary care organization. An example is shown in Figure 1 for B-type natriuretic peptide (BNP). This test has been advocated for many years as a first line screening test for patients with symptoms of heart failure. UK national guidance commends its use (11) and recommends that the test is used to support the decision-making process as to whether a patient should be referred for echocardiography and/or to a specialist cardiologist.

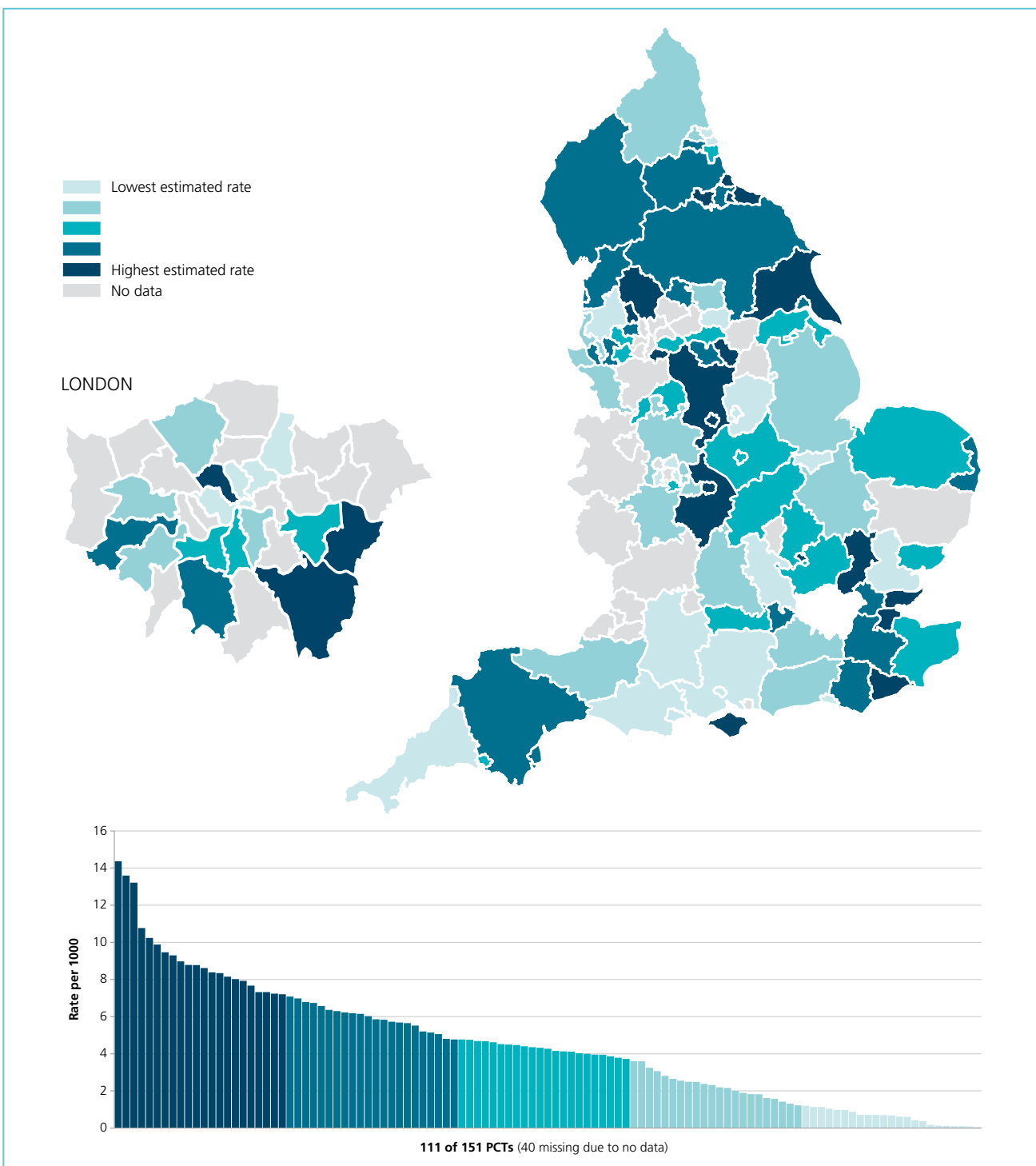
Figure 1 shows an 89-fold difference in ordering rates for BNP between different primary care organizations. This may represent failure of guideline uptake or the unavailability of the test in some areas due to cost pressures. Variation in utilization of this test can have a real impact on patient care and subsequent morbidity and mortality.

There are many factors which determine a physician's test ordering practices. In literature surveys (12, 13), physicians mostly cite fear of legal (malpractice) complaints as the primary driver of over-testing. A recent article by Hoffman et al. (14) states that the main driver of over-diagnosis and over treatment is zero tolerance for error and uncertainty. Addressing the widespread intolerance of uncertainty requires a cultural change both within the medical profession and by the public.

This paper will examine the following:

- Factors influencing test ordering by physicians;
- Strategies for modifying physicians' ordering pattern;
- Ways to implementing policies to improve laboratory utilization and thereby improve patient outcome.

Figure 1 Brain natriuretic peptide (BNP or NTproBNP) ordering rates across primary care organizations (primary care trusts [PCTs]) in England in 2012



Map colours show relative density of ordering, bar chart shows ordering rates by organization (orders per 1000 patients served by organization). Crown copyright and database rights 2013 Ordnance Survey. 100039906.

Reproduced with the kind permission of Right Care, and first published in the NHS Atlas of Variation in Diagnostic Services, p155. <http://www.rightcare.nhs.uk/atlas/>

FACTORS INFLUENCING TEST ORDERING BY PHYSICIANS

Users of the clinical laboratory want information to allow them to make better decisions about patients. They want to be assured that the investigations they order will be quick, accurate and inexpensive and they want 'new' tests to be readily available. They want to be able to do the right investigation on the right patient at the right time, with results reaching the right clinician at the right time and in the right format and medium. In addition, availability of the right interpretation is essential to ensure the optimum patient outcome. Hopefully, the clinician is also concerned with patient safety, clinical accountability and clinical governance.

However, the clinician faces huge problems in getting test ordering right. There are too many tests, they have different names, they are reported in different units, there are different reference intervals between laboratories, there are different decision limits and guidelines are often inconsistent. Clinicians want tests with high diagnostic accuracy, good predictive value and proven clinical utility in decision making.

Two literature reviews (15, 16) are in broad agreement on the reasons for ordering diagnostic tests. These include diagnostic factors, such as rule-in or rule-out disease, therapeutic and prognostic factors, such as help in deciding on appropriate treatment, as well as patient-related factors such as patient reassurance, doctor-related factors such as clinical experience, confidence in clinical judgement and fear of litigation and policy and organization-related factors, such as test availability, institutional policies and clinical guidelines, and the use of structured test ordering forms (17).

FAILURE OF GUIDELINE IMPLEMENTATION

Despite the clear recommendations of the UK National Institute for Health and Care Excellence

(NICE) guidelines for use of the CA125 antigen in detection of ovarian cancer (18), the data from the NHS Atlas of Variation (10) still demonstrates vast variation in the use of CA125 in UK primary care organizations. There is a need to target low-ordering areas, emphasising the importance of guideline implementation. It would be of great interest to determine if the areas with low CA125 ordering rates have higher morbidity and mortality for ovarian cancer. Schulenberg-Brand et al. (19) investigated the impact of local guidance on tumour marker ordering within a single surgical department in the UK through an audit process and found a significant rate of inappropriate ordering underpinned by an apparent lack of knowledge about the correct use of the test. For example, 33% of CA125 orders were made on male patients!

The use of faecal calprotectin to distinguish between irritable bowel syndrome (IBS) and inflammatory bowel disease (IBD) is well established (20). The test distinguishes patients with symptoms of functional IBS from those with organic symptoms of IBD with greater than 95% sensitivity and specificity. A normal faecal calprotectin result excludes IBD and removes the requirement for endoscopy. In our hospital, over the past 12 months, this has resulted in a 70% reduction in the number of endoscopy procedures. This does not only benefit patients but also provides a significant financial saving, since the cost of an endoscopy is \$908 whereas a faecal calprotectin test costs \$80.

In addition, calprotectin predicts clinical relapse in IBD with 90% sensitivity and 83% specificity. This again influences patient outcome by enabling treatment to be started earlier, thus resulting in improved outcomes. Despite this, calprotectin testing in the UK has not yet been implemented nationwide (10). For primary care organizations in England, the estimated annual rate of use for calprotectin tests ranges from 0.01 to 5.1 per thousand practice population,

a 446-fold variation. Patchy uptake in primary care, despite the evidence of clinical utility, probably indicates a lack of understanding of the value of the test or its lack of availability from local laboratory services. This may be because secondary care providers are reluctant to lower the rates of endoscopy for financial reasons.

Failure of uptake of guidelines is a problem that spans all specialities and sectors of healthcare. A recent article by Misra et al. (21) confirmed the findings from Cabanagh et al. (22) in 1999, showing little change over 15 years. The barriers to guideline adherence include:

- a. Lack of awareness of the existence of guidelines or unfamiliarity with the guideline content;
- b. Lack of agreement with the specific guideline and/or lack of agreement with guidelines in principle;
- c. Inertia of previous practice;
- d. The guideline is contradictory to established practice or difficult to follow/use. There may be patient reluctance to comply with guideline;
- e. External barriers such as resource availability, practice constraints and lack of time.

There have been several national initiatives to try to reduce over-diagnosis and change physician behaviour and adherence to guidance. In the UK, these include the NICE initiative of a 'do not do' recommendation database, comprising tests or procedures with limited or no value that should not be used (23). In the US, the 'Choosing Wisely' campaign (24) aims to help healthcare practitioners, patients and other stakeholders develop sustainable solutions to stop the over-use and misuse of medical tests and procedures that provide little or no benefit. In addition, a group from the Australian Government Department of Health identified potentially unsafe,

ineffective or inappropriate services listed on the country's Medicare Benefit Schedule (25).

The US National Physician Alliance (NPA) have created a project entitled 'Promoting good stewardship in clinical practice' that aimed to develop a list of the top five activities in family medicine, internal medicine and paediatrics where the quality of care can be improved. As part of the list for internal medicine, they recommended not obtaining blood chemistry panels or urinalysis for screening asymptomatic healthy adults, and only screening for type II diabetes mellitus in asymptomatic adults with hypertension (26).

Improving adherence to clinical guidelines requires targeting, proper dissemination and education. As we will see later, there is considerable overlap between successful implementation of guidelines or strategies and improving ordering behaviour. Guidelines should be written, published and disseminated, but it is essential that proper implementation strategies are devised and delivered, as implementation is crucial to ensuring a positive impact on patient outcome.

INAPPROPRIATE LABORATORY UTILIZATION

An analysis of 307 malpractice claims in the US (27) studied the principal areas of faulty processes which led to misdiagnosis in patients. The top cause, found in 55% of patients, was the failure to order the appropriate diagnostic/laboratory test. There is growing recognition that errors in test selection (inappropriate ordering) and result interpretation can have significant or adverse clinical consequences to patients and financial consequences to healthcare institutions (28).

As Moynihan et al. (29) have written, "Medicine's much heralded ability to heal the sick is fast being challenged by its propensity to harm the healthy. Too many people are being

over-dosed, over-treated and over-diagnosed.” They state that \$200 billion may be wasted on unnecessary treatment every year in the US, for example screening programmes detecting early cancers that will never cause symptoms or death.

Five per cent of all healthy patients will get abnormal test results and false findings or trivial abnormalities can lead to unnecessary further testing and expensive and potentially risky interventions, leading to poor patient outcomes. Causes of over-utilization include patient pressure, duplicate ordering, lack of understanding of the diagnostic value of a test, ordering the wrong test, failure to understand the consequences of over-utilization, defensive testing, perverse financial incentives and ‘availability creates demand’ (where the key driver is technological advance). Some of the consequences of over-utilization include incorrect diagnosis and treatment, incorrect test ordering which delays the actual diagnosis, increased length of hospital stay, unnecessary blood loss, increased resource utilization and, most important, unnecessary patient alarm.

Moynihan et al. (29) point out that the concern about over-diagnosis does not preclude awareness that many people miss out on much needed healthcare. In fact resources wasted on unnecessary diagnoses and care can be much better spent treating and preventing genuine illness.

Van Walraven and Naylor, in their systematic review in 1998 (30), concluded that the frequency range of inappropriate testing was between 5% - 95%. This was a review of North American studies, but similar non-American studies (UK, Netherlands, Australia, Canada, Egypt and Thailand) reported inappropriate testing rates between 10% - 50%. A study of hospitalized patients in our institution demonstrated that 34% of orders were inappropriate (unpublished

data). Zhi et al. in their systematic review of the literature from 1997-2012 (31) found the overall mean rate of over-utilization of testing to be 20.6% (95% CI = 16.2-24.9%), with over-utilization of low volume tests higher at 32.2% (95% CI = 25.0-39.4%).

Laposata (28) has shown that the highest incidence of error in laboratory testing is in test selection by clinicians and interpretation of test results by clinicians. This confirms the work of Plebani (32) in his review of the literature: up to 68% of laboratory testing errors occur in the pre-pre-analytical phase which includes inappropriate test orders. Reviewing the diagnostic error and testing literature, Epner et al. (33) identified 5 causes of diagnostic error and harm relating to the testing process. They called this the ‘five cause taxonomy of testing-related diagnostic error’ and it includes both ordering an inappropriate test and not ordering an appropriate test.

There is often little thought given to the patient’s views and the non-clinical outcomes. A reduction in inappropriate ordering will reduce the need for some phlebotomy episodes and reduce the associated discomfort and inconvenience such as time off work, as well as minimizing potential patient anxiety. It must be recognised that inappropriate testing will impact on follow-up, by leading to false positive results, and unnecessary further interventions such as referral and further invasive investigations.

There is no point in ordering a test if no-one looks at the results and/or acts on them. The issue of failure to follow-up tests which have been ordered is addressed by Callen et al. elsewhere in this issue of eJIFCC.

In a very recent publication ‘Protecting Resources, Promoting Value: A doctors’ guide to cutting waste in clinical care’ (November 2014) from the UK Academy of Medical Royal

Colleges (34) there is a guide/toolkit to help doctors and other clinicians to use resources in the most effective way to provide the best possible quality and quantity of care for patients. It promotes the identification of tests or procedures whose necessity should be questioned.

The emphasis of laboratory utilization programs should never be exclusively on reducing the number of tests. It is imperative to consider clinical outcomes and the changes to patient management. Zhi et al. (31) found the mean rate of under-utilization of testing in their systematic review to be 44.8% - more than twice the rate of over-utilization. Missed tests may have a significant impact on patient outcome. In a study looking at the effect of HbA1c ordering frequency, Fu and his colleagues showed lower frequency of HbA1c monitoring is significantly associated with poorer glycaemic control. To achieve HbA1c concentrations below a target of 53 mmol/mol the optimal testing frequency was 4 times per year (35).

STRATEGIES TO IMPROVE PHYSICIAN ORDERING

There have been several studies and audits published which describe initiatives to change ordering behaviour. An article in Bandolier (36) identified 49 studies reporting interventions which were designed to change physicians' ordering practice. The studies used a range of single or combined interventions which included: educational initiatives, guideline dissemination, Computerized Physician Order Entry (CPOE) design with algorithms, clinical pathway analysis, activity utilization and cost information, vetting of orders and restricting tests to ensure the appropriate test repertoire.

Successful intervention strategies included: 1) educational initiatives aimed at predisposing factors; 2) targeting re-enforcing factors by provision of activity and costing data; and 3) targeting enabling factors such as limiting the number of tests allowed by deleting tests from the laboratory repertoire or specialist vetting of orders. A summary of strategies can be found in Table 1.

Astion has described the factors he feels improve laboratory utilization (37). He describes both physician education and patient education as weak interventions in isolation. CPOE can improve laboratory utilization if thoughtfully implemented and education can be made more effective by combining it with other methods that make the desired behaviour more likely, including CPOE, use of formularies, implementing higher levels of approval for some tests and the use of physician utilization reports with performance feedback. As in many other studies, the best approach to improving laboratory utilization combines multiple interventions (17).

It is imperative that these interventions remain in place, or ordering behaviour will drift back to the initial condition. In a cluster randomized trial by Thomas et al. (38), the effect of enhanced feedback and brief educational reminder messages on 9 tests ordered in primary care over a 12 month period achieved a reduction of around 10% in the number of orders when used alone but when the initiatives were used together (in combination) they demonstrated a larger reduction - greater than 20% of total tests ordered.

Figure 2 shows the 'test cycle' and highlights the points at which the laboratory clinicians can become engaged in managing appropriate test utilization. The patient must always be the focus of all processes and outcomes.

Table 1 Summary of intervention strategies
to improve physician ordering behaviour

Pre-laboratory	Laboratory	Post-laboratory
Educate and engage with users regarding testing, including presentations to clinical teams	Withdraw outdated tests	Perform clinical audit with adherence to local and national clinical guidelines
Make formal contributions to training, induction, undergraduate curriculae, guideline pathway development	Harmonize nomenclature	Perform audit of the impact of test results on patient pathways and outcomes, including whether results were reviewed, whether action was taken and the outcome
Develop laboratory formulary in conjunction with users	Standardize units	Include information on test appropriateness as part of report
Develop CPOE systems including test repertoire available to all who order tests. Use disease-specific or question-specific profiles	Standardize reference intervals	
Ensure targeting of clinical guidelines	Harmonize laboratory profiles	
	Apply minimum re-testing intervals between repeat tests	
	Use reflex/reflective testing when appropriate	
	Introduce vetting of selected (esoteric) tests by senior laboratory staff	

MEDICAL EDUCATION

One of the biggest areas of concern is the level of education of junior doctors (interns) about laboratory medicine, which has decreased in many countries. Khromova and Gray (39) surveyed junior

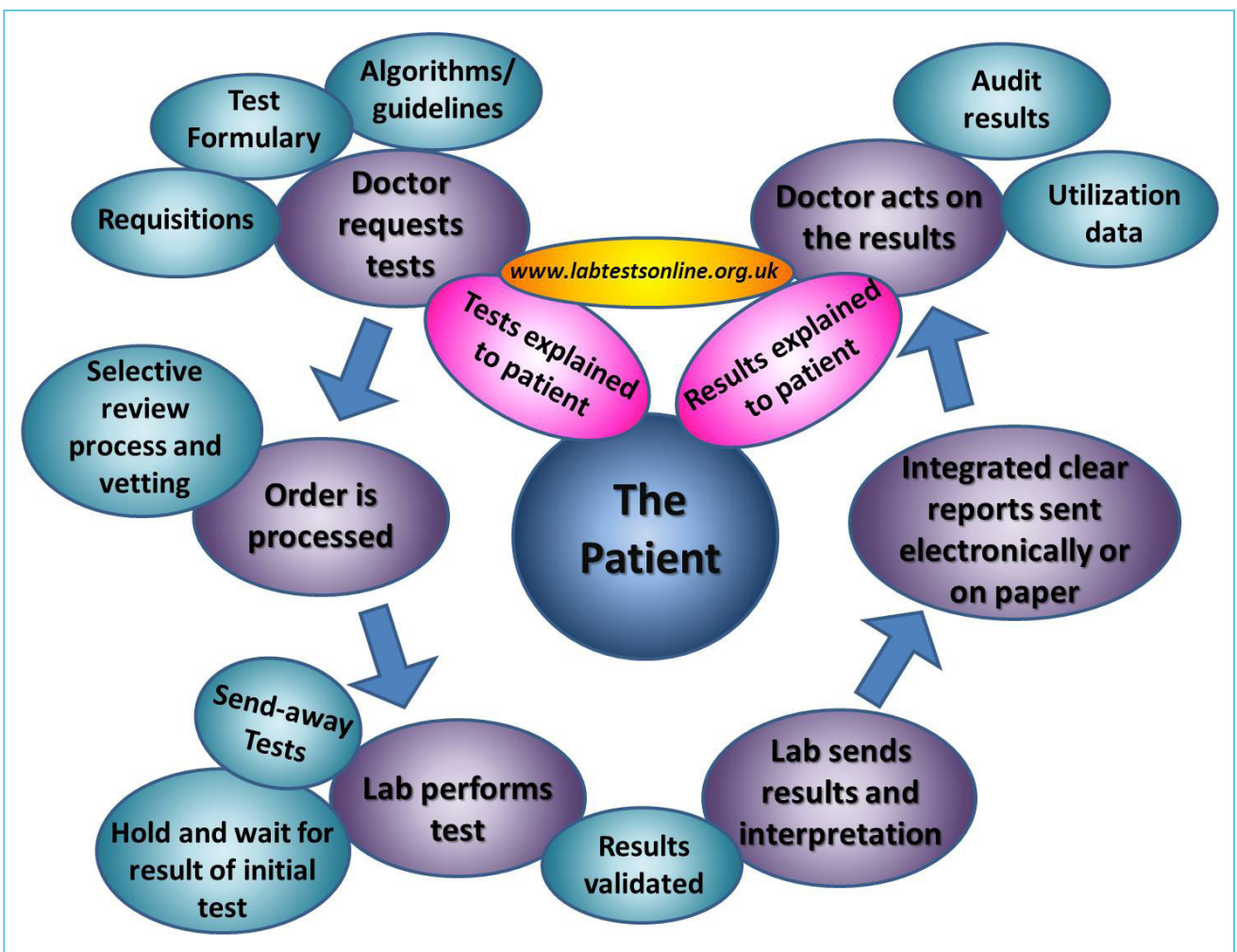
medical staff in Sheffield and found they lacked confidence in both ordering and interpreting basic clinical chemistry tests, such as serum protein, magnesium and phosphate. Up to 75% of the junior doctors felt they needed further teaching in

relation to investigations that they were not confident in ordering or interpreting. A study of interns in two teaching hospitals in Cape Town, South Africa (40) demonstrated similar results. The study concluded that junior doctors felt unprepared for their roles and needed more exposure to laboratory medicine in training, and more instruction on the basics of rational ordering of laboratory tests.

In a study of final year medical students at Oxford in 2010 (41), Clarke and Littlewood explored their attitudes and their competence in haematology. Haematology was viewed as

a particularly difficult speciality, but was nevertheless a popular and interesting career choice. A worrying lack of important clinical knowledge as the students began their internship was found. The study demonstrated the students' relative lack of both confidence and competence in managing blood disorders, for example, only one third of the final year medical students knew the value of the International Normalized Ratio (INR) in deciding whether to administer vitamin K to a patient. As a result of this survey, the British Society of Haematology reviewed the curriculum for undergraduate haematology teaching.

Figure 2 The 'Test Cycle'



The 'test cycle' above shows the points at which laboratory clinicians can become engaged in managing appropriate test utilization. The patient must always be the focus of all processes and outcomes.

There is very little hard evidence in the literature to demonstrate the impact of the knowledge of basic science by junior doctors on patient outcome (42), although there is no shortage of anecdotal and circumstantial evidence. A survey of 300 participants attending an AACC Annual Meeting in 2012 showed that the single most important issue identified as leading to ineffective test ordering was inadequate teaching about laboratory medicine in Medical School (unpublished data). 94% of respondents rated this as 'highly important' or 'important'.

A survey by Laposata (28) showed that every US medical school teaches more than 100 hours of anatomic pathology, whilst only 9% have a separate and distinct course in laboratory medicine. The mean time spent teaching medical students on the appropriate selection of laboratory tests and the correct interpretation of results over the entire 4 year curriculum was 10 hours and it was less than 5 hours in many of the institutions. The survey showed that completion of anatomic pathology training required passing an examination, but there were no examinations for laboratory medicine, despite it forming a much greater part of the experience of most physicians.

A recent survey from the Clinical Laboratory Integration into Healthcare Collaborative (CLIHCC) found that primary care physicians are uncertain about the right test to order 14.7% of diagnostic encounters and are uncertain about the correct interpretation of test results in 8.3% (1). With more than 500 million primary care visits per year in the US, the data indicates that approximately 23 million times per year, primary care physicians are not certain about the best use of the diagnostic test. Inadequate education in laboratory medicine must be seen as a patient safety issue.

LABORATORY FORMULARIES

A laboratory (test) formulary is analogous to the pharmaceutical formulary present in most institutions and can be used in many ways. It may simply outline what tests a clinician may order or what tests are permitted to be sent to outside (reference) laboratories. A test formulary requires an understanding of the clinical value of the test, the financial impact and whether or not there is a history of the test being poorly utilized. Many laboratories now have their own laboratory formulary to help the clinician to select the right test in specific situations. An example is Brigham and Women's Hospital, Boston, MA, USA (43). The trend is to use the laboratory formulary to reduce inappropriate ordering of expensive molecular and genomic tests. Effective laboratory formularies need to be developed with full involvement of laboratory staff, physicians and other stakeholders.

HARMONIZATION OF NOMENCLATURE

CLIHCC have also examined the issue of the wide inconsistencies in test nomenclature as a significant barrier to physicians ordering the correct test. For example, there are at least 18 different titles for vitamin D related tests in the US. In the UK, the National Laboratory Medicine Catalogue (NLMC) has the long-term objective that each test 'name' represents a single pathology test concept and each concept is represented by just one name. The NLMC aims to standardize ordering, reporting and analysing of pathology tests to ensure that the right patient gets the right test at the right time. There is a standardized list of pathology tests that have been validated for use within the UK NHS. This list is provided in an XML format and may be used within Laboratory Information Management Systems (LIMS),

electronic patient records and pathology order communications (44).

HARMONIZING COMMON LABORATORY TEST PROFILES

A common source of physician confusion is that different laboratories provide different 'profiles' of tests to answer the same clinical question. This variation is often for historical reasons. In the UK, it has been revealed by a national pathology benchmarking initiative (45), which showed 12 different profiles for liver function tests among 50 laboratories subscribing to the initiative. As a consequence, the UK Association for Clinical Biochemistry and Laboratory Medicine, have produced proposals for a consensus view on profile composition, e.g. liver panel: bilirubin, alanine transaminase, alkaline phosphatase, albumin (46). As well as removing confusion, harmonizing profiles can save money and reduce further investigations instigated as a result of clinically irrelevant minor abnormalities in irrelevant tests.

Laboratories must increase their efforts to engage with the test user to provide the appropriate tests in any clinical situation, whatever the core profiles contain.

COMPUTERIZED PHYSICIAN ORDER ENTRY (CPOE)

CPOE can be a blessing or a curse, depending on how it is implemented. The worst case scenario is an electronic test order form in which the full menu of laboratory tests, from the most common to the most esoteric is made readily available to all practitioners, and repetitive interval-based testing (e.g. daily thyroid function testing) is easy to instigate. This is a recipe for laboratory mis-utilization, and the laboratory involved would have to bear responsibility for the resulting situation.

If CPOE is implemented with a strategy that prompts physicians with relevant information at the time of test ordering, it has been shown to decrease utilization of some commonly ordered tests in the in-patient setting. In one study, physicians were prompted electronically as to whether they wanted to continue their daily metabolic panel order after the patient had been in hospital for 72h (47). The effect of this was to reduce testing by 24% with no change in patient outcome. Design of the electronic order form is crucial: following literature searches showing that gamma-glutamyl transferase (GGT) need not form part of the routine liver panel, local experience of removing the GGT tick box for a 12 month period during 2010/2011 reduced GGT ordering by almost 50% (unpublished data).

As yet, there is limited published evidence on the impact of CPOE on clinical outcomes. However, the potential of the approach has been investigated in the context of imaging (48). It was proposed that a system linking electronic ordering of imaging orders to best practice diagnostic pathways represented the way to maximize appropriate referrals.

A systematic review in 2006 identified 19 studies of the impact of CPOE on laboratory testing (49). Eleven of these compared CPOE (with and without decision support) to no CPOE for laboratory testing in a range of countries (South Korea, USA, UK, Canada, Norway), and eight studies compared CPOE with and without specific decision support (all in USA). Eight of the first group of studies and all of the second group considered outcomes that could be specifically related to appropriateness issues such as clinical indicators, length of stay or appropriateness of stay. The CPOE systems (both with and without decision support) showed an overall trend towards reduced test volume and cost, when compared to no CPOE. Overall, fewer tests and (when measured) fewer inappropriate tests were performed in the decision support group.

In addition, the decision support group showed a significant reduction in the median time to appropriate treatment for critical results reported in one of the randomized controlled trials.

Four of the studies found that CPOE systems combined with decision support improved adherence to guidelines provided on the system. One of the advantages of CPOE is the ability to link electronically to relevant knowledge resources, but care is needed not to make the ordering process unwieldy.

In a UK study of implementation of CPOE (50), it was shown to be associated with a reduction in the proportion of outpatient appointments at which full blood count, urea and electrolytes and urine culture tests were ordered and at which full blood count tests were repeated. However, the system was associated with an almost 4-fold increase in the use of urea and electrolytes testing amongst day case patients.

A recent publication from Turner et al. (51) looked at pre-analytical errors from primary care during two six-month periods pre- and post-implementation of electronic ordering. Outcomes measured included whether there was correct information on the sample, whether the correct sample was received and whether clinical history was provided. There was a marked decrease in the number of pre-analytical errors following the introduction of electronic ordering (2764 pre-implementation versus 498 post-implementation). The error rate dropped across all general practices: pre-implementation error rates ranged up to 5.7% of orders, post-implementation error rates were less than 0.6%.

In 2014, the Association for Clinical Biochemistry and Laboratory Medicine and the Royal College of Pathologists in the UK proposed 'National Minimum Retesting Intervals in Pathology' (52). The recommendations cover minimum intervals before retesting for common

tests in clinical biochemistry, therapeutic drug monitoring, haematology and immunology in specified clinical situations, supported by an evidence base. In this context, electronic ordering has an advantage over laboratory-based interventions as it can prevent inappropriate repeat orders at the source prior to phlebotomy, minimising the inconvenience for patients and the burden on phlebotomy staff and laboratory reception staff. Electronic orders can also provide links to external sources such as diagnostic algorithms and other resource sites such as Lab Tests Online (www.labtestsonline.org.uk), which help the primary care physician choose the correct test and explain the result to the patient.

Epner and Astion have reported on the use of CPOE to reduce diagnostic errors, particularly: the use of CPOE templates in a specific care testing environment, e.g. diabetes care; the incorporation of reflex testing strategies, e.g. autoantibody panel after positive ANA test; decreasing the number of synonyms for the same test; and restriction of ordering of specific tests to a defined set of physicians or specialists, e.g. medical geneticists (53).

VETTING (RESTRICTION) OF TESTS

Laboratories have the option to vet high cost, low volume tests, often those referred to specialist laboratories, on an individual basis. Fryer et al. reduced the number of urine toxicology screens from 30 to less than 5 orders per month (54), resulting in an annual saving of around \$48,000. In another study, consultant level restriction of a specific test, C-reactive protein (CRP), led to an 85% reduction in test volume (55).

In our experience, restriction of CRP and ESR ordering with the use of computerized decision support led to a 17% reduction of orders (unpublished data). Application of

knowledge-based rules significantly improves the appropriateness of test ordering. Sometimes it may be appropriate to use a 'send and hold' process in which a specimen is sent to the laboratory but the test is not performed until another initial test result comes back. Flow cytometry is a good example, along with molecular assays and genetics studies in haematological diseases, which can be held until the bone marrow aspirate and biopsy is viewed by the pathologist and then sent for testing in appropriate cases. Using fixed rules as a form of vetting, Srivastava et al. prospectively measured the efficiency and effectiveness of reflex and reflective testing in specific clinical scenarios (56). These approaches improved the diagnosis of hypovitaminosis D, hypomagnesaemia, hypothyroidism, hyperthyroidism and haemochromatosis, improving both the clinical utility of the laboratory service and the patient outcome.

PROVIDING COST INFORMATION ON LABORATORY TEST ORDERING

Healthcare budgets worldwide are facing increasing pressure to reduce costs and improve efficiency whilst maintaining quality. Pathology investigations cost the UK National Health Service £2.5 billion per year. A review commissioned by the UK Department of Health estimated that 20% of this could be saved by improving utilization of pathology services, despite the annual increase of 8-10% in workload.

The review estimated that 25% of pathology tests were unnecessary, representing a huge potential waste of resource (57).

A controlled clinical trial at John Hopkins' Hospital displayed 'fees' for 61 random laboratory tests in their CPOE. In the 'active arm', there was an 8.59% decrease in the number of tests

per patient. In the 'control arm', there was a 5.64% increase (58).

In a similar study involving 215 primary care physicians in Massachusetts, Medicare reimbursement rate for 27 laboratory tests was displayed. In the intervention group there was a significant decrease of 19% in ordering rates compared to control physicians for 5 tests. In addition, the majority (81%) of physicians reported that the intervention improved their knowledge of the relative cost of laboratory tests (59).

CONCLUSION

Successful management of laboratory test utilization requires the entire laboratory team to use their skills and knowledge to identify utilization issues, implement a programme that will achieve more effective laboratory testing and establish appropriate processes from the beginning to the end of the test cycle. This is not easy and requires interactions with our clinical colleagues that some laboratory workers may find uncomfortable - questioning clinicians, and advising that they should not order a particular test but another test is more appropriate. Generally, clinicians have few direct incentives to restrict laboratory utilization, and are not being trained to do so. It is disappointing that there is so little literature on the effectiveness of appropriate test utilization on patient outcomes, as well as on cost effectiveness across the whole patient pathway.

"We need to recognise that the target of requesting of the test and of the results should be the patient. It is the person who actually, in the end, is going to have to change their lives and start adopting new behaviours...." (Goetz [60], adapted).

REFERENCES

1. Hickner J, Thompson PJ, Wilkinson T, Epner P, Shaheen M, Pollock AM et al. Primary care physicians' challenges in ordering in clinical laboratory tests and interpreting results. *J Am Board Fam Med* 2014; 27: 268-74.
2. Research and Markets. In vitro diagnostics (IVD) market, techniques and applications. Dublin: Research and Markets, 2013. http://www.researchandmarkets.com/publication/pfdniq/in_vitro_diagnostic_ivd_market_technique. Accessed November 2014.
3. Alexander B. Reducing healthcare costs through appropriate test utilization. *Critical Values* 2012; 5: 6-8.
4. Holladay EB. Test right. *Critical Values* 2012; 5: 3.
5. The Lewin Group. The value of laboratory screening and diagnostic tests for prevention and health care improvement, 2009. <http://www.lewin.com/publications/publication/403/>. Accessed November 2014.
6. Mogyrosy Z, Mogyrosy G. Practice pattern and geographic variation in test ordering: a literature review. *Orvosi Hetilap* 2006; 147: 25-31.
7. Larson A, Palmer M, Hulten G, Tryding N. Large differences in laboratory utilization between hospitals in Sweden. *Clin Chem Lab Med* 2000; 38: 383-9.
8. Smellie WS, Galloway MJ, Chinn D. Benchmarking practice use of pathology services: a model for monitoring change. *J Clin Pathol* 2000; 53: 476-80.
9. Smellie WS, Clark G, Nulty CA. Inequalities of primary care microbiology testing between hospital catchment areas. *J Clin Pathol* 2000; 56: 199-203.
10. Public Health England. NHS Atlas of Variation in diagnostics services. November 2013. <http://www.rightcare.nhs.uk/index.php/atlas/diagnostics-the-nhs-atlas-of-variation-in-diagnostics-services/>. Accessed November 2014.
11. Clinical Guideline 108: Chronic heart failure: management of chronic heart failure in adults in primary and secondary care. National Institute for Health and Care Excellence, 2010. <https://www.nice.org.uk/guidance/cg108>. Accessed November 2014.
12. Kanzaria HK, Hoffman JR, Probst MA, Berry S, Brook RH. Emergency physician perceptions of medically unnecessary advanced diagnostic imaging. (abstract) *Acad Emerg Med* 2014; 21 (Suppl. 1).
13. Studdert DM, Mello MM, Sage WM, DesRoches CM, Peugh J, Zapert K, Brennan TA. Defensive medicine among high risk specialist physicians in a volatile malpractice environment. *JAMA* 2005; 293: 2609-17.
14. Hoffman JR, Kanzaria HK. Intolerance of error and culture of blame drive medical excess. *BMJ* 2014; 349: g5702.
15. Whiting P, Toerien M, de Sallis I, Sterne AC, Dieppe P, Egger M, Fahey T. A review identifies and classifies reasons for ordering diagnostic tests. *J Clin Epidemiol* 2007; 60: 918-9.
16. Sood R, Sood A, Ghosh AK. Non-evidence-based variables affecting physicians test-ordering tendencies: a systematic review. *Neth J Med* 2007; 65: 167-77.
17. Smellie W. Demand management and test request rationalisation. *Ann Clin Biochem* 2012; 49: 323-36.
18. Clinical Guideline 122: The recognition and initial management of ovarian cancer. National Institute for Health and Care Excellence, 2011. <https://www.nice.org.uk/guidance/cg122>. Accessed November 2014.
19. Schulenburg-Brand D, Kumar N, Zouwali S. The impact of local guidelines on the tumour marker requesting patterns of a general surgery department. *Ann Clin Biochem* 2013; 50: 438-42.
20. Diagnostics Guideline 11: Faecal calprotectin diagnostic tests for inflammatory diseases of the bowel. National Institute for Health and Care Excellence, 2013. <https://www.nice.org.uk/guidance/DG11>. Accessed November 2014.
21. Misra S, Barth JH. Guidelines are written, but are they followed? *Ann Clin Biochem* 2013; 50: 400-2.
22. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PC, Rubin HR. Why don't physicians follow clinical practice guidelines? *JAMA* 1999; 282: 1458-65.
23. 'Do not do' database. National Institute for Health and Care Excellence. <http://www.nice.org.uk/savings-andproductivity/collection>. Accessed November 2014.
24. Morden NE, Colla CH, Sequist TD, Rosenthal MB. Choosing Wisely – the politics and economics of labeling low-value services. *N Engl J Med* 2014; 370: 589-92.
25. Elshaug AG, Watt AM, Mundy L, Wills CD. Over 150 potentially low value healthcare practices: an Australian study. *Med J Aust* 2012; 197: 556-60.
26. The "Top 5" lists in primary care: meeting the responsibility of professionalism. *Arch Intern Med* 2011; 171: 1385-90.
27. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med* 2006; 145: 488-96.
28. Laposata M. Putting the patient first - using the expertise of laboratory professionals to produce rapid and accurate diagnoses. *LabMedicine* 2014; 45: 4-5.
29. Moynihan R, Doust J, Henry D. Preventing overdiagnosis - how to stop harming the healthy. *BMJ* 2012; 344: e3502.

30. Van Walraven C, Naylor D. Do we know what inappropriate laboratory utilization is? JAMA 1998; 280: 550-8.
31. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The landscape of inappropriate laboratory testing: A 15-year meta-analysis. PLoS ONE 2013; 8(11): e78962.
32. Plebani M. The detection and prevention of errors in laboratory medicine. Ann Clin Biochem 2010; 47: 101-10.
33. Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. BMJ Qual Saf 2013; 22: ii6-ii10. doi 10.1136/bmjqs-2012-001621.
34. Protecting resources, promoting value: a doctors' guide to cutting waste in clinical care. Academy of Medical Royal Colleges, 2014. <http://www.aomrc.org.uk/dmdocuments/Promoting%20value%20FINAL.pdf> Accessed November 2014.
35. Fu C, Ji L, Wang W, Luan R, Chen W, Zhan S, Xu B. Frequency of glycated haemoglobin monitoring was inversely associated with glycemic control of patients with type 2 diabetes mellitus. J Endocrinol Invest 2012; 35: 269-73.
36. Using labs best. Bandolier 1999 61: 4. <http://www.medicines.ox.ac.uk/bandolier/band61/b61-4.html>. Accessed November 2014.
37. Astion M. Interventions that improve laboratory utilization: from gentle guidance to strong restrictions. Lab Errors Patient Saf 2006; 2: 2.
38. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. Lancet 2006; 367: 1990-6.
39. Khromova V, Gray T. Learning needs in clinical biochemistry for doctors in foundation years. Ann Clin Biochem 2008; 45: 33-8.
40. Stanfliet JC, Macauley J, Pillay TS. Quality of teaching in chemical pathology: ability of interns to order and interpret laboratory tests. J Clin Pathol 2009; 62: 664-6.
41. Clarke R, Littlewood T. 'Haematophobia?' Attitudes of Oxford medical students towards haematology. Bull Royal Coll Pathol 2010; 150: 118-22.
42. Freedman DB. Is the medical undergraduate curriculum fit for purpose? Ann Clin Biochem 2008; 45: 1-2.
43. List of available tests. Brigham and Women's Hospital Department of Pathology, Boston, MA, USA. <http://bwhpathology.partners.org/labmanual.aspx>. Accessed November 2014
44. National laboratory medicine catalogue. Royal College of Pathologists. <http://www.rcpath.org/clinical-effectiveness/pathology-catalogues/national-laboratory-medicine-catalogue/national-laboratory-medicine-catalogue.htm>. Accessed November 2014.
45. Keele University Benchmarking Service. www.keele.ac.uk/benchmarking Accessed November 2014.
46. Consultation on standardized test order sets. Association for Clinical Biochemistry and Laboratory Medicine. www.acb.org.uk/whatwedo/science/best_practice/test_profiles.aspx. Accessed November 2014.
47. Neilsen EG, Johnson KB, Rosenbloom ST, Dupont WD, Talbert D, Giuse DA et al. The impact of peer management on testing ordering behaviour. Ann Intern Med 2004; 141: 196-204.
48. Bairstow PJ, Mendelson R, Dhillon R, Valton F. Diagnostic imaging pathways: development, dissemination, implementation, and evaluation. Int J Qual Health Care 2006; 18: 51-7.
49. Georgiou A, Williamson M, Westbrook JI, Ray S. The impact of computerised physician order entry systems on pathology services: a systematic review. Int J Med Inform 2007; 76: 514-29.
50. Collin S, Reeves BC, Hendy J, Fulop N, Hutchings A, Priedaner E. Implementation of computerised physician order entry (CPOE) and picture archiving and communication systems (PACS) in the NHS: quantitative before and after study. BMJ 2008; 337: a939.
51. Turner HE, Deans KA, Kite A, Croal BL. The effect of electronic ordering on pre-analytical errors in primary care. Ann Clin Biochem 2013; 50: 485-8.
52. Final report of the national minimum retesting intervals project. Association for Clinical Biochemistry and Laboratory Medicine. <http://www.acb.org.uk/docs/default-source/guidelines/acb-mri-recommendations-a4-computer.pdf?sfvrsn=2> Accessed November 2014.
53. Epner P, Astion M. Focusing on test ordering practices to cut diagnostic errors. Clin Lab News 2012; 38 (No 7) 17-8.
54. Fryer AA, Smellie WS. Managing demand for laboratory tests: a laboratory toolkit. J Clin Pathol 2013; 66: 62-72.
55. Hutton HD, Drummond HS, Fryer AA. The rise and fall of C-reactive protein: managing demand within clinical biochemistry. Ann Clin Biochem 2009; 46: 155-8.
56. Srivastava R, Bartlett WA, Kennedy IM, Hiney A, Fletcher C, Murphy MJ. Reflex and reflective testing: efficiency and effectiveness of adding on laboratory tests. Ann Clin Biochem 2010; 47: 223-7.

57. Report of the second phase of the review of NHS pathology services in England. Chaired by Lord Carter of Coles. December 2008. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_091985. Accessed November 2014.

58. Feldman LS, Shihab HM, Thiemann D, Yeh HC, Ardolino M, Mandell S, Brotman DJ. Impact of providing fee

data on laboratory test ordering: a controlled clinical trial. JAMA Intern Med. 2013; 173: 903-8.

59. Horn DM, Koplan KE, Senese MD, Orav EJ, Sequist TD. The impact of cost displays on primary care physician laboratory test ordering. J Gen Intern Med 2014; 29: 708-14.

60. Goetz T. The decision tree: taking control of your health in the new era of personalized medicine. Emmaus, PA: Rodale Books, 2010.

Collaborating with international clinical organizations

Howard A. Morris

*School of Pharmacy and Medical Science, University of South Australia
Chemical Pathology Directorate, SA Pathology, Adelaide, South Australia 5000*

ARTICLE INFO

Corresponding author:

Prof. H. A. Morris
Chemical Pathology Directorate
SA Pathology
Box 14 Rundle Mall Post Office
Adelaide, SA 5000
Australia
E-mail: howard.morris@health.sa.gov.au

Key words:

clinical collaboration, chronic
kidney disease, KDIGO, IOF, IFCC

ABSTRACT

The provision of quality laboratory services for patient care to improve healthcare outcomes is at the centre of the work of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC). However the day to day work of laboratory medicine practitioners largely does not involve direct contact with patients. The IFCC Executive Board has therefore included in its strategic plan activities to highlight collaboration with clinical organizations.

A review of IFCC activities demonstrates a wide range of such collaborations with international health and clinical organizations at all levels. The IFCC Executive Board leads such collaborations with leading international bodies including the World Health Organization and World Association of Societies of Pathology and Laboratory Medicine (WASPaLM). The work of the Scientific Division has involved collaborations with 16 clinical organizations at the level of the Executive Committee as well as with specific Committees and Working Groups.

Furthermore in recent years the Executive Board has established a number of Task Forces with strong interaction with clinicians and clinical organizations. The harmonization of the assay for haemoglobin A1c is just one example of technological improvement to not only improve the performance of the test for monitoring disease but increase its utility for diagnosis which currently involves collaboration with clinicians.

The IFCC is continuing to expand its relations with international clinical organizations to enhance both the translation of developments in laboratory medicine to improve patient care and clinical outcomes and their adoption into clinical practice via inclusion in clinical guidelines.

INTRODUCTION

The provision of quality clinical laboratory services to improve patient care and healthcare outcomes is at the centre of the work of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC). Included in the Mission of the IFCC are statements highlighting this goal including working “to enhance the scientific level and the quality of diagnosis and therapy for patients throughout the world” as well as to “build on the professionalism of our members to provide quality services to patients.” All of us working in Laboratory Medicine are conscious that direct contact with patients in our day to day work is unusual. Consequently the Executive Board of the IFCC has considered as a priority activities to enhance the Laboratory Medicine – Clinical interface, aiming to improve the efficacy of our practice and to facilitate the translation of developments in our field into patient care. Within the strategic plan of the 2012-2014 IFCC Executive Board were two goals: “Develop a plan to increase collaboration between IFCC and international clinical organizations” and “Establish at least one new collaboration each year with an international clinical organization”.

In recent years the IFCC has advocated for laboratory medicine to adopt a strategy focussed on the patient to improve clinical effectiveness and outcomes through activities such as clinical interpretation and provision of advice on laboratory results. Furthermore, patients

are increasingly taking more responsibility for their own health and are requiring such information to influence decisions on their healthcare. Models of healthcare delivery are changing with the integration of imaging and other sources of data into clinical guidelines improving knowledge to speed up healthcare and improve patient outcomes. One consequence of integrated diagnostics is the erosion of traditional boundaries within laboratory medicine and between diagnostic modalities.

THE RANGE OF IFCC CLINICAL COLLABORATIONS

The IFCC collaborates at the level of the Executive Board with international clinical organizations including the World Health Organization (WHO) and the World Association of Societies of Pathology and Laboratory Medicine (WASP-aLM). Collaboration with WHO has increased particularly since this organization has recognised the increasing importance of non-communicable diseases for health throughout the world. Diabetes mellitus is one such disease and is becoming much more common in resource-limited settings. There has been an increased understanding of the improvement of clinical outcomes associated with effective monitoring of blood glucose and haemoglobin A1c (HbA1c) as well as significant changes to the diagnostic criteria for diabetes mellitus during the past decade. IFCC has pioneered the standardization of HbA1c measurement with considerable benefit to method comparability and the effective monitoring of diabetic patients. IFCC has also been working with WHO to revise a WHO booklet entitled ‘Laboratory diagnosis and monitoring of diabetes mellitus’ for the use of clinical and laboratory medicine specialists, especially in resource-limited settings.

Table 1 Clinical organizations collaborating with the IFCC Scientific Division

World Gastroenterology Organization
The International Association of Therapeutic Drug Monitoring and Clinical Toxicology
International Society of Endocrinology
American Thyroid Association
International Diabetes Federation
International Growth Hormone Society
European Association of Allergy and Clinical Immunology
International Osteoporosis Foundation

The IFCC Scientific Division has supervised a variety of clinical collaborations undertaken by their Committees and Working Groups over many years. These have included a wide range of clinical disciplines working on specific projects (Table 1). Current projects include the Working Group – Standardization of Albumin Assay in Urine (WG-SAU) working in collaboration with the US National Kidney Disease Education Program (NKDEP); Working Group – Standardization of Insulin Assays (WG-SIA) working in collaboration with the American Diabetes Association and the European Association of Diabetes Societies; Working Group – Standardization of Bone Marker Assays (WG-SBMA) working in collaboration with the International Osteoporosis Foundation; and the Task Force on Chronic Kidney Disease working in collaboration with WASPaLM, Kidney Disease Improving Global Outcomes (KDIGO) and Asia Forum for CKD Initiative. The IFCC has seven Task Forces working on integrated projects across the Scientific and Education and Management Divisions of the IFCC to which numerous clinicians have been appointed to provide expertise on the clinical translation of developments in laboratory medicine to improve healthcare outcomes.

CASE STUDIES ON IFCC CLINICAL COLLABORATIONS

1. Chronic kidney disease

Kidney Disease–Improving Global Outcomes (KDIGO) is a collaboration between the International Society of Nephrology, the Transplantation Society, a number of national nephrology societies, Canadian national organizations for health research and provision of clinical services and some pharmaceutical companies. In 2009 it held a conference to discuss the definition, classification and prognosis of chronic kidney disease where data from some 1,558,332 participants ranging across healthy subjects to high-risk subjects for kidney disease and patients suffering from kidney disease were subject to meta-analyses. Summaries of the data and the decisions arrived at from the congress were published (1).

The data indicated that both all-cause mortality and cardiovascular mortality were strongly inversely related to estimated-glomerular filtration rate (eGFR) and independently directly related to urine albumin excretion expressed as the albumin to creatinine ration (ACR). These data allowed for the derivation of clinically critical decision limits for these parameters.

For eGFR a value lower than 60 mL/min/1.73 m² was accepted because the hazard ratio for mortality increases sharply below this level. For urine albumin, cut-offs were adopted at less than 30 mg albumin per g creatinine, 30 to 300 mg/g and greater than 300 mg/g. These values were used to develop clinical guidelines for primary care medical practitioners as well as medical specialists.

Collaboration between IFCC WG-Glomerular Filtration Rate Assessment and NKDEP had ensured the development and adoption of a reference measurement procedure for creatinine in serum or urine utilizing isotope dilution-mass spectroscopy technology and standard reference materials for serum creatinine (2). By 2012 KIDIGO was able to recommend adoption of reporting eGFR for the assessment of kidney function for all patients calculated with the CKD-EPI formula (3) using serum creatinine assays aligned to the reference measurement procedure. Thus with the availability of appropriate clinical data, in collaboration with clinical laboratory professionals and the *in vitro* diagnostics industry, the highest level of clinical practice for the diagnosis and monitoring of chronic kidney disease can be made available to all patients through their primary care physicians.

The next step is to ensure the adoption of these practices internationally. In a number of countries or regions the leading nephrology and laboratory medicine organizations have come together to implement these recommendations. For example, in Australia and New Zealand the Australasian Creatinine Consensus Working group was established for such an implementation program. This involved nephrologists (through Kidney Australia) and laboratory medicine practitioners (through the Royal College of Pathologists of Australasia and the Australasian Association of Clinical Biochemists, the member society representing

the region within the IFCC). Manuscripts were published in the Medical Journal of Australia, the official publication of the Australian Medical Association, reporting on automatic reporting of estimated glomerular filtration rate and clinical interpretation of results (4) as well as chronic kidney disease and measurement of albuminuria or proteinuria (5). Consequently Australasian clinical laboratories have adopted the following practices to ensure optimal treatment for diagnosis and monitoring of chronic kidney disease at all levels of medical practice but especially at the level of the primary care practitioner: standardized creatinine assays; common units for reporting creatinine and eGFR; universal reporting of eGFR; standardized interpretation of results; universal definition of CKD according to eGFR and urine albumin:creatinine ratio values; a link between laboratory results and interpretation and clinical management of the patient; and finally, communication of nephrology specialist advice to primary care practitioners via the clinical laboratory result report, allowing patients to receive the highest quality of medical treatment for CKD from their primary care physician close to their home.

While adoption of these recommendations is widespread across North America, Europe, Australasia and parts of Asia, in other regions such clinical collaborations have not yet occurred. Therefore the IFCC, in collaboration with WASPaLM, has established a Joint Task Force to help national organizations to implement best clinical practice guidelines for the management of CKD. At this level, a major obstacle is effective communication between all stakeholders. Ensuring traceability of creatinine assays to recognized international reference materials using reference measurement procedures has been adopted by the established major international *in vitro* diagnostic (IVD) equipment and reagent providers.

However, there are currently approximately 100 IVD companies who provide reagents and instruments for creatinine measurements in clinical laboratories and therefore there is considerable work to be done to ensure all these assays meet the quality requirements for optimal clinical care.

2. Postmenopausal osteoporosis

Osteoporosis is highly prevalent with some 50% of women and 30% of men over the age of 60 years expected to experience a fracture, which imposes considerable morbidity and premature mortality on our aging populations. Treatment of these fractures and their consequences, such as loss of the ability to live independently, results in the largest costs on healthcare budgets of any medical condition. The ability to identify patients at increased risk of fracture and those who are not responding to treatments would be a great benefit. Bone turnover markers have been in use in research settings and by medical specialists for over 50 years with the aim of identifying appropriate biomarkers. While many markers show promise, no consistent data have been generated from clinical research either to allow interpretation of bone turnover measurements for the individual patient at the level of the primary care physician or to permit bone turnover markers to be incorporated in clinical guidelines for osteoporosis. Recognizing the clinical importance of improving osteoporosis care, the International Osteoporosis Foundation (IOF) and the IFCC agreed to collaborate on investigations of the role of bone turnover markers, if any, in clinical management and a Joint Working Group between IOF and IFCC (WG-BMSO) was established.

The WG undertook the evaluation of current data and review of publications between 2000 and 2010 (6), aiming to recommend particular bone turnover marker assays for standard

clinical practice. Some 7 bone formation marker assays are available for clinical laboratories as reagent kits or available on automated clinical chemistry platforms and 8 bone resorption marker assays are similarly available. Some clinical data suggested that these assays were useful for both prediction of fracture risk and for response to treatment. However there is significant variation between studies, and the data were largely of low quality and restricted to European populations. Therefore current data are inconclusive with regard to interpretation of assay results for the individual patient.

The WG concluded that it was necessary to enlarge the experience with a limited number of designated bone turnover marker assays for fracture risk assessment in population-based studies in subjects of a variety of ethnicities as well as for monitoring response to osteoporosis treatments. Furthermore, consensus was reached that there is no evidence identifying a perfect bone turnover marker. Criteria for an ideal marker have been delineated, including the following: adequately characterized biologically and chemically; anatomical specificity for bone; high performance in fracture risk prediction and in monitoring osteoporosis treatments among women and men; widely available on automated platforms and not the monopoly of a single supplier; assays to demonstrate suitable biological and analytical variability, sample handling, stability and ease of analysis; and finally, assays that are available for analysis from blood specimens. Two bone turnover markers, C-telopeptide fragments of collagen type 1 α 1 chains, also known as serum Crosslaps (CTX-1), assayed in plasma or serum as an assessment of bone resorption, and N-terminal Propeptide of Type 1 Procollagen (P1NP) assayed in serum as an assessment of bone formation best met these criteria and were recommended to be assessed in all future clinical trials.

Between 2010 and 2011, the National Bone Health Alliance (a North American collaboration between clinical organizations involved with the clinical management of osteoporosis), pharmaceutical companies and organizations representing laboratory medicine (American Association of Clinical Chemistry) also conducted a review of the scientific literature (7). Their conclusions were similar, recommending collecting further clinical research data for the bone turnover markers CTX-1 and PINP.

Currently these assays are available on automated platforms from two manufacturers: Immunodiagnosics Systems Ltd (IDS) iSYS® and Roche Diagnostics Cobas®. If data from clinical trials are to be combined to conduct meta-analyses for assessment of the efficacy of bone turnover marker levels, the values generated by these platforms must be comparable. Currently for CTX-1 there are conflicting preliminary data, while data for PINP suggest that these values are comparable from the two systems. A second Joint Working Group between the IOF and the IFCC has been established to define the comparability of assays for CTX-1 and PINP (WG-SBMA) and to harmonize or standardize as feasible if they are not comparable. This project is currently underway in collaboration with the NBHA.

CONCLUSIONS

The IFCC is participating in a range of collaborations with clinical organizations based on specific projects. Such projects arise from developing requirements for optimal healthcare delivery and improved patient outcomes. For some well-established biomarkers clinical utility is being improved by, for example, standardization of the various assays used in clinical laboratories to ensure the optimal application of clinical guidelines based on critical levels for these biomarkers. The examples of serum

creatinine and urine albumin for the diagnosis and assessment for the clinical management of chronic kidney disease are examples. For new biomarkers, the goal of these collaborations is the appropriate clinical application and development of clinical guidelines incorporating biomarker values. The use of cardiac troponins for the diagnosis of acute cardiac syndrome is an example. Other projects are working to establish the appropriate status of biomarker assays such that their clinical efficacy can be investigated. Bone turnover markers in the diagnosis and management of osteoporosis are an example of this activity.

The IFCC is actively promoting collaborations with clinical organizations to enhance the contribution of laboratory medicine in healthcare delivery, clinical effectiveness and patient outcomes. These goals are achieved through a patient-focussed strategy by increasing the quality of laboratory medicine practice to improve patient safety and clinical usefulness; ensuring the timely presentation of clinical laboratory results and provision of appropriate clinical interpretation and advice; and maintaining the financial sustainability of laboratory medicine by improving cost effectiveness, ensuring the appropriate use of the laboratory and providing value for money.

REFERENCES

1. Levey AS, de Jong PE, Coresh J, El Nahas M, Astor BC, Matsushita K, et al. The definition, diagnosis and prognosis of chronic kidney disease: A KDIGO Controversies Conference Report. *Kidney Int* 2011; 80: 7-28.
2. International Federation of Clinical Chemistry and Laboratory Medicine: Working Group on Standardization of Glomerular Filtration Rate Assessment (WG-GFRA); Panteghini M, Myers GL, Miller WG, Greenberg N. The importance of metrological traceability on the validity of creatinine measurement as an index of renal function. *Clin Chem Lab Med* 2006; 44: 1287-1292.
3. Levey AS, Inker LA, Coresh J. GFR estimation: from physiology to public health. *Am J Kidney Dis* 2014; 63: 820-834.

4. Johnson DW, Jones GR, Mathew TH, Ludlow MJ, Doogue MP, Jose MD et al. Chronic kidney disease and automatic reporting of estimated glomerular filtration rate: new developments and revised recommendations. *Med J Aust.* 2012; 197: 1-5.

5. Johnson DW, Jones GR, Mathew TH, Ludlow MJ, Chadban SJ, Usherwood T et al. Chronic kidney disease and measurement of albuminuria or proteinuria: a position statement. *Med J Aust* 2012; 197: 224-225.

6. Vasikaran S, Eastell R, Bruyere O, Foldes AJ, Garnero P, Griesmacher A, et al. Markers of bone turnover for the prediction of fracture risk and monitoring of osteoporosis treatment: A need for international reference standards. *Osteoporos Int* 2011; 22: 391-420.

7. Bauer D, Krege J, Lane N, Leary E, Libanati C, Miller P, et al. National Bone Health Alliance Bone Turnover Marker Project: current practices and the need for US harmonization, standardization, and common reference ranges. *Osteoporos Int* 2012; 23: 2425-2433.

The impact for patient outcomes of failure to follow up on test results. How can we do better?

Joanne Callen, Andrew Georgiou, Julie Li, Johanna I Westbrook

*Centre for Health Systems and Safety Research, Australian Institute of Health Innovation,
Macquarie University, Sydney, Australia*

ARTICLE INFO

Corresponding author:

Joanne Callen
Associate Professor, Centre for Health
Systems and Safety Research
Australian Institute of Health Innovation
Macquarie University
Level 6, 75 Talavera Road
Sydney, 2109
Australia
E-mail: Joanne.Callen@mq.edu.au

Key words:

patient safety, test result follow-up,
quality improvement, diagnostic tests,
patient access to records,
health information systems,
laboratory information systems,
medical order entry systems

ABSTRACT

Background: The World Health Organization–World Alliance for Patient Safety has identified test result management as a priority area. Poor test result follow-up can have major consequences for the quality of care, including missed diagnoses and suboptimal patient outcomes. Over the last three decades there has been considerable growth in the number of requests for pathology and radiology services which has added to the complexity of how patient care is delivered and test results are managed. This can contribute to a lack of clarity about where and with whom responsibility for test follow-up should reside: a problem that is compounded by a lack of clear definitions about what are critical, unexpected or significantly abnormal results.

Aim of this paper: This paper will present a narrative review highlighting key issues related to the problem of failure to follow up laboratory test results, and outline potential solutions.

Conclusions: Information technology (IT) has the potential to enhance the performance and safety of test result management processes. Effective solutions must engage all stakeholders, including consumers, in arriving at decisions about who needs to receive results, how and when they are communicated, and how they are acknowledged and acted upon and the documentation of these actions.

Meeting these challenges requires the establishment and maintenance of resilient governance approaches and a culture dedicated to ensuring the reliability and safety of patient care.

INTRODUCTION

The World Health Organization–World Alliance for Patient Safety has identified poor test result management as a priority patient safety area (1). Poor test result follow-up can have major consequences for the quality of care, including missed diagnoses and suboptimal patient outcomes. A root cause analysis of aggregated information from a national Australian incident management information system showed that 11% (3/27) of clinical incidents resulting in a serious outcome (e.g., patient death), and 32% (24/75) of clinical incidents with major patient-related consequences, were related to problems with test follow-up (2). Clinicians themselves acknowledge that their test management practices are inefficient (3). The urgency of the problem was underscored by the US Emergency Care Research Institute's (ECRI) 2014 report on patient safety concerns for health care organizations (4). The report listed data integrity failures associated with health information systems, poor care coordination across levels of care and test result reporting problems as the leading three items of their top 10 patient safety concerns (4). Each of these problems is intrinsically connected to the issue of poor test result follow-up.

A systematic review published in 2011 (5) identified 12 studies over a 20 year period which investigated the extent of failure to follow up laboratory and radiology results for hospital patients. The review reported the lack of follow-up of test results for hospitalized inpatients ranged from 20.04% to 61.6%, and 1.0% to 75% for patients treated in the Emergency Department (ED), when calculated

as a proportion of tests. The consequences of missed test results for patient care included delayed diagnoses such as malignancies, hypothyroidism, hyperthyroidism, and osteoporosis, reinforcing the urgent need to address the problem. In situations involving missed microbiological results consequences included failure to commence or change antibiotic therapy. The review also highlighted that there were cases of missed positive serological test results for *Helicobacter pylori* and *Chlamydia* and in the latter the patient subsequently developed pelvic inflammatory disease (5). Another systematic review which quantified the extent of failure to follow up test results in ambulatory care settings (6) identified 19 studies and reported wide variation in the proportions of tests not followed up: 6.8% to 62% for laboratory test results and 1.0% to 35.7% for radiology (6). These failings included missed cancer diagnoses in four of the seven studies reporting on the impact on patient outcomes (6). Increased hospital presentations resulting from hyperkalaemia related to missed abnormal serum potassium levels and adverse drug events related to insufficient supplementation with levothyroxine due to missed follow-up of abnormal TSH results were examples of other reported negative patient outcomes.

Results pending at discharge was identified as an area of particular concern for hospitalized patients (5). A 2012 study of test orders in an Australian hospital revealed that 47% of missed results stemmed from tests ordered on the day of discharge, which raises concerns about the appropriateness of those tests where results are not followed up (7). The systematic review of missed test results for hospital patients also flagged follow-up of critical results as a problem area (5). Despite practice guidelines requiring critical values to be telephoned to the clinical team, compliance remains an issue and information may

not always reach the clinician involved in the patient's care. Several studies report the absence of guidelines regarding responsibility for patient notification, and documentation of actions related to test follow-up (8-12). The management of test results involves communication between many individuals, including physicians, nurses, clerical and laboratory staff and patients, across a variety of settings using a range of manual and electronic systems. The systematic reviews (5, 6) identified varying test management practices between care settings and the information systems used in the process included paper-based, electronic, and a combination of paper and electronic systems. Information technology (IT) has a key role to play in supporting the management of test results in terms of ordering, reporting, accessing and tracking follow-up with documentation of actions. However, evidence of the effectiveness of IT is limited although studies show a general trend towards improved test follow-up when electronic systems are used (5, 6). New models of test management supported by IT can only succeed when a systems approach is adopted which recognises the complex clinical governance challenges associated with safe test management (13). The two systematic reviews on test follow-up for hospitalized and ambulatory patients identified evidence that failure to follow up test results is a substantial problem, but with only 31 studies conducted across a span of 20 years (5, 6) the evidence base is not substantial. What is particularly lacking is evidence of potential interventions to support clinicians and patients to reduce the rates and thus risks associated with failure to follow up test results. Further studies are urgently needed to evaluate solutions such as on-line endorsement/acknowledgment of test results and particular attention must be paid to the integration of solutions with work

practices of clinicians and laboratories and the needs of consumers.

ENHANCING THE QUALITY AND SAFETY OF TEST RESULT MANAGEMENT

Harmonization of test result management

Pathology and medical imaging services perform a major role in the delivery of patient care by ensuring that reliable and accurate results are delivered in a timely fashion to inform clinical management decisions (1). One of the main errors associated with delayed follow-up of pathology and medical imaging results originates in the post-analytic phase of the testing process, or once a report or test result has been issued to the requesting (or referring) doctor. Failures in this phase are linked to a lack of clarity about where and with whom responsibility for test result follow-up should reside (14), and clear definitions about what are critical, unexpected or significantly abnormal results. There is also no consensus regarding the reporting timeframe for these abnormal results between laboratories, medical imaging departments, hospitals and other health care settings (15). A 2012 survey of test result management in Australasian laboratories, conducted by Campbell and Horvath (16), revealed large variations in how critical results are managed and the failure of laboratories to uniformly follow internationally-recognised guidelines. Out of a total of 58 participating laboratories across Australia, New Zealand, and Hong Kong, 97% included critical results and 81% incorporated significantly abnormal results in their critical limit list. Only 41% of laboratories stated that they compiled their list in consultation with doctors, even though this is an accreditation requirement specified by the ISO 15189 quality management system standard for medical laboratories. In this paper the authors also stated that there was a

subjective element in the compilation of critical limit lists and this was a factor in the substantial variation in range of values between institutions(16). Inconsistent policies also existed between laboratories regarding critical result notification procedures, including the identification of critical results, timeliness of reporting critical results, how critical results are notified, to whom the result is notified, and the acknowledgement of results receipt (16).

Evidence-based recommendations in this area (15-18) emphasise the importance of clear definitions of key terms and the need for agreed alert thresholds and timeframes and specified procedures for fail-safe communication of test results that pose critical or significant risk to patient safety. Many doctors describe existing test result management systems as inefficient and chaotic (3, 19). It is an important issue faced by pathology and medical imaging departments world-wide (15), and requires establishment of standardized pathology information structures and terminologies to improve recording, decision support and communication of laboratory information (18).

Information technology initiatives

IT offers solutions to enhance the performance and safety of test result management processes. The process of identifying missing test results can involve time-consuming and cumbersome audits involving paper (and electronic) records (20). In such cases, the identification of missed test results may be too late to have any positive effect on patient safety (5). IT systems can be used to track pending test results at hospital discharge (21), deliver result alerts and document test result acknowledgement and subsequent clinical actions (22). An on-line test result endorsement function provides an auditable trail of test follow-up actions and as such provides a continuous quality audit

capability which can be used by clinicians and management (5).

The existence of hospital data silos and poor integration of electronic systems remains a well-documented problem and major patient safety hazard in Australia and internationally (23). The establishment of integrated electronic data sources is a key component for safely monitoring, identifying and acting upon any instances of failure to follow up test results, to ensure that appropriate treatment is delivered (24). The use of hybrid medical records, that is paper and electronic systems, has been shown to be associated with errors and duplications compared to complete electronic systems (25, 26). In relation to test follow-up, the use of a partial electronic medical record (eMR; paper based progress notes and electronic test results or vice versa) was shown to be associated with higher rates of failure to inform patients of clinically significant results compared to using a complete manual or electronic system (8).

Successful implementation of IT must recognise the dynamic between the technology and the complex social environment in which healthcare is delivered (27). Management of test results needs to ensure that the requirements of clinicians in different clinical settings need to be taken into account. Sittig and Singh (28) have made recommendations which aim to reconcile the social (personal, workflow, organizational) and technical (hardware/software, clinical content, user interface) elements of test result follow-up in the clinical environment to facilitate correct use of eMR-based IT initiatives and realization of potential benefits. These recommendations include: the provision of standardized clear definitions of test result categories to facilitate prioritization (flagging of significantly abnormal test results) and electronic reporting; that physicians be trained to process test result notifications in

a timely manner and consistently document all follow-up actions in the eMR, as multiple sources of documentation may lead to a breakdown in communication of test results and follow-up failure; and that responsibility for test result follow-up and communication under all clinical circumstances should be clear, formally documented and regularly reviewed, and understood by all professional parties concerned (28). The Safer Self-Assessment Guides (29) also recommend that automated result alerts should be limited to those that are clinically relevant to avoid information overload or “alert fatigue”, and all test ordering should be completed using Computerized Provider Order Entry systems to allow access to tests electronically and avoid the creation of hybrid information environments.

Establishment of a safety and quality governance structure and culture

Tackling the issue of test result follow-up requires the establishment and maintenance of integrated governance systems and a culture dedicated to ensuring the reliability and safety of patient care. Effective clinical governance systems require integration across all parts of an organization. This involves the clear delineation of responsibilities and workforce accountability, along with systems to monitor progress and deal with any risks or impediments (30). The US Joint Commission Journal on Quality and Patient Safety, “Safe Practice Recommendations for Communicating Critical Test Results,” outlines this process as starting with the identification of the ordering or responsible provider as the person who should receive results, followed by the person the result is directed to if the ordering provider is not available, to ensure that patients receive timely clinical attention (17).

A 2014 study investigated the successful implementation of an electronic test management

system at a major Australian hospital (24). The system provided an electronic safety net based upon a test management governance model. This system ensures that, if the responsible medical officer who ordered a test does not acknowledge the receipt of a test result within 3 days, a notification-escalation process is set in motion so that as each day passes, email or pager alerts are sent to increasingly senior members of the hospital staff. This process begins with the clinical unit’s designated medical officer (day 4), and then escalates to the clinical unit support supervisor (administration or medical) (day 5), clinical unit director (day 7) and division director (day 10). This process enabled the ongoing monitoring of test results and allowed delays in test result follow-up to be identified and remedied in a timely fashion. Evaluation of the system identified that over a period of one year *all* test results had been acknowledged, with 60% of laboratory and 44% of medical imaging results acknowledged within 24 hours of result availability (24).

Enhancing the role of consumers in test result access

The engagement of consumers in their health care is an important trend in Australia and internationally. It is increasingly acknowledged that the benefits of increased consumer engagement encompass better quality and safer health care practice (31, 32). Consumer involvement is particularly relevant to test result management, where failure to inform patients of test results has been described as legally indefensible in malpractice claims (33). Hospital eMRs can be used to provide consumers with access to information on-line using a secure electronic patient portal, which in addition to allowing access to appointment and personal clinical information, including test results, also facilitates communication with health professionals (34, 35).

Patients have regularly expressed interest in being involved in medical decision making and in being notified of their test results, both abnormal and normal (36). It has also been argued that sharing information and engaging patients to take responsibility for follow-up lead to improvements in the efficiency and effectiveness of the laboratory test process (e.g. decrease in test redundancy) (37). However, there are major obstacles which hinder the active involvement of consumers in test follow-up, including a lack of access to both clinical information and tools and checklists that help consumers understand and engage in their own care (33). Clinicians do not agree on the level and timing for consumers to have access to their test results (38); clinical unease may also be related to the impact that direct patient access to test results has on the traditional physician role and authority as the information gatekeeper (36). Concerns about patient anxiety, confusion and lack of expertise to appropriately interpret their test results have also divided physicians in their attitudes toward direct patient notification of test results (39, 40). This contrasts with the findings of a quasi-experimental pilot of a patient portal (41) in primary care practices across three regions in the United States which found that only a very small proportion of patients (1% to 8%) experienced confusion or worry when directly accessing their electronic notes, and 77% to 87% across the three sites reported that *Open Notes* helped them feel more in control of their care. All participating physicians also expressed a willingness to continue use of the portal. However, generalizability of the study was limited by sampling bias as all participants were volunteers who responded positively in their attitudes and expectations of the patient portal administered pre-study (41).

Evidence of patient portal use and impact has been, in general, limited and inconclusive (34, 35, 42, 43), as patient portals are relatively new technology and the health care community has only begun to understand how they can engage with this innovation to optimize care delivery, outcomes and patient engagement (44). A recent systematic review examining the effect of patient portals on clinical care concluded that there was insufficient evidence to determine whether patient portals had a positive, negative or neutral impact, although patient outcomes and satisfaction appeared positive when portals were integrated within a larger case management program (42). The review highlighted important gaps in the literature, advocating studies that look at context and implementation factors. Patient race and ethnicity, education level or literacy, and degree of comorbid conditions may influence portal use. The review identified disparities between patients who access portals and those who do not, and described instances of suboptimal patient attitudes of their worth. It suggests that increased acceptance will require attention to overcoming these disparities and addressing usability and patient-perceived value to engage certain populations that are not readily embracing personal health record systems (42).

CONCLUSION

Failure to follow up laboratory test results is a significant concern and a priority patient safety area. The issue of missed test results is multi-dimensional, and involves a number of interconnected issues encompassing both test result management practices and the systems involved in the process. An examination of existing research has revealed a lack of consistency in how test results are managed in the post-analytic laboratory testing phase,

including variations and ambiguity in policies regarding result notification procedures, identification of critical results, timeliness of results reporting, and acknowledgement of result receipt. Evidence of the impact of IT on improving the safety of the test result management process has also been inconsistent with few published evaluations to date. Electronic systems have yet to overcome issues with integration and hospital information silos, whilst partial uptake of the eMR has resulted in hybrid paper and electronic systems which may add to the risk of missed test results. Improving the safety of test result management through IT initiatives involves the establishment of a fully integrated electronic system that is implemented as a component of the solution alongside appropriate clinical and organizational governance elements. The success of IT interventions is intrinsically linked to resilient management arrangements, attention to clinical governance and commitment to robust evaluation practices which address issues with laboratory test management work practices and guidelines at the post-analytic testing phase. The empowerment and engagement of consumers in the management of their own healthcare data will further the move towards a culture which delivers reliable and safe patient care (32).

REFERENCES

1. The Research Priority Setting Working Group of the World Alliance for Patient Safety . Summary of the evidence on patient safety: implications for research. Geneva, Switzerland: World Health Organization, 2008. http://www.who.int/patientsafety/information-centre/20080523_Summary_of_the_evidence_on_patient_safety.pdf. Accessed December 2014.
2. Clinical Excellence Commission. Clinical Focus Report: Diagnostic tests - How access and follow-up affect patient outcomes. Sydney, Australia: Clinical Excellence Commission, 2011. http://www.cec.health.nsw.gov.au/_documents/programs/patient-safety/patient-safety-report-diagnostic-tests.pdf. Accessed December 2014.
3. Gandhi TK. Fumbled handoffs: one dropped ball after another. *Ann Intern Med* 2005; 142: 352-8.
4. ECRI Institute. Top 10 patient safety concerns for healthcare organizations. Plymouth Meeting, PA: ECRI Institute, 2014. <https://www.ecri.org/PatientSafety-Top10> Accessed December 2014
5. Callen J, Georgiou A, Li J, Westbrook JI. The safety implications of missed test results for hospitalised patients: a systematic review. *BMJ Qual Saf* 2011; 20: 194-9.
6. Callen JL, Westbrook JI, Georgiou A, Li J. Failure to follow-up test results for ambulatory patients: a systematic review. *J Gen Intern Med* 2012; 27: 1334-48.
7. Ong MS, Magrabi F, Jones G, Coiera E. Last orders: follow-up of tests ordered on the day of hospital discharge. *Arch Intern Med* 2012; 172: 1347-9.
8. Casalino LP, Dunham D, Chin MH, Bielang R, Kistner EO, Karrison TG, et al. Frequency of failure to inform patients of clinically significant outpatient test results. *Arch Intern Med* 2009; 169: 1123-9.
9. Chen ZJ, Kammer D, Bond JH, Ho SB. Evaluating follow-up of positive fecal occult blood test results: lessons learned. *J Healthc Qual* 2007; 29: 16-34.
10. Elder NC, McEwen TR, Flach JM, Gallimore JJ. Management of test results in family medicine offices. *Ann Fam Med* 2009; 7: 343-51.
11. Elder NC, McEwen TR, Flach J, Gallimore J, Pallerla H. The management of test results in primary care: does an electronic medical record make a difference? *Fam Med* 2010; 42: 327-33.
12. Singh H, Arora HS, Vij MS, Rao R, Khan MM, Petersen LA. Communication outcomes of critical imaging results in a computerized notification system. *J Am Med Inform Assoc* 2007; 14: 459-66.
13. Singh H, Wilson L, Reis B, Sawhney MK, Espadas D, Sittig DF. Ten strategies to improve management of abnormal test result alerts in the electronic health record. *J Patient Saf* 2010; 6: 121-3.
14. Sherman H, Castro G, Fletcher M, Hatlie M, Hibbert P, Jakob R, et al. Towards an international classification for patient safety: the conceptual framework. *Int J Qual Health Care* 2009; 21: 2-8.
15. Campbell CA, Horvath AR. Harmonization of critical result management in laboratory medicine. *Clin Chim Acta* 2014; 432: 135-47.
16. Campbell CA, Horvath AR. Towards harmonisation of critical laboratory result management-review of the literature and survey of Australasian practices. *Clin Biochem Rev* 2012; 33: 149-60.

17. Hanna D, Griswold P, Leape LL, Bates DW. Communicating critical test results: safe practice recommendations. *Jt Comm J Qual Patient Saf* 2005; 31: 68-80.
18. Roy CL, Rothschild JM, Dighe AS, Schiff GD, Graydon-Baker E, Lenoci-Edwards J, et al. An initiative to improve the management of clinically significant test results in a large health care network. *Jt Comm J Qual Patient Saf* 2013; 39: 517-27.
19. Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care. *Arch Intern Med* 2004; 164: 2223-8.
20. Georgiou A, Vecellio E, Toouli G, Eigenstetter A, Li L, Wilson R, Westbrook JI. The impact of the implementation of electronic ordering on hospital pathology services. Report to the Commonwealth of Australia, Department of Health and Ageing, Quality Use of Pathology Committee. Sydney, Australia: Australian Institute of Health Innovation 2012. <http://aihi.mq.edu.au/resource/impact-implementation-electronic-ordering-hospital-pathology-services> Accessed December 2014.
21. Dalal AK, Poon EG, Karson AS, Gandhi TK, Roy CL. Lessons learned from implementation of a computerized application for pending tests at hospital discharge. *J Hosp Med* 2011; 6: 16-21.
22. Schiff GD, Bates DW. Can electronic clinical documentation help prevent diagnostic errors? *N Engl J Med* 2010; 362: 1066-9.
23. Georgiou A, Prgomet M, Paoloni R, Creswick N, Hordern A, Walter S, Westbrook J. The effect of computerized provider order entry systems on clinical care and work processes in emergency departments: a systematic review of the quantitative literature. *Ann Emerg Med* 2013; 61: 644-53. e16.
24. Georgiou A, Lymer S, Forster M, Strachan M, Graham S, Hirst G, et al. Lessons learned from the introduction of an electronic safety net to enhance test result management in an Australian mothers' hospital. *J Am Med Inform Assoc* 2014; doi:10.1136/amiainl-2013-002466.
25. Callen J, Westbrook JI, Braithwaite J. The effect of physicians' long-term use of CPOE on their test management work practices. *J Am Med Inform Assoc* 2006; 13: 643-52.
26. Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing paper-based with electronic patient records: lessons learned during a study on diagnosis and procedure codes. *J Am Med Inform Assoc* 2003; 10: 470-7.
27. Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Smith MW, Murphy DR, et al. Primary care practitioners' views on test result management in EHR-enabled health systems: a national survey. *J Am Med Inform Assoc* 2013; 20: 727-35.
28. Sittig DF, Singh H. Improving test result follow-up through electronic health records requires more than just an alert. *J Gen Intern Med* 2012; 27: 1235-7.
29. Ash J, Singh H, Sittig D. SAFER Self Assessment - Test Results Reporting and Follow-Up. Washington, DC: The Office of the National Coordinator for Health Information Technology, 2013. http://www.healthit.gov/sites/safer/files/guides/safer_testresultsreporting_sg008_form.pdf Accessed December 2014
30. Greenfield D, Nugus P, Fairbrother G, Milne J, Debono D. Applying and developing health service theory: an empirical study into clinical governance. *Clin Govern Int J* 2011; 16: 8-19.
31. The National Patient Safety Foundation. Lucian Leape Institute. Safety is personal; Partnering with patients and families for the safest care. Boston, MA: National Patient Safety Foundation, 2014. http://c.ymcdn.com/sites/www.npsf.org/resource/resmgr/LLI/Safety_Is_Personal_Executive.pdf Accessed December 2014
32. Australian Commission on Safety and Quality in Health Care (ACSQHC). National Safety and Quality Health Service Standards. Sydney, Australia: ACSQHC, 2011. <http://www.safetyandquality.gov.au/wp-content/uploads/2011/09/NSQHS-Standards-Sept-2012.pdf>. Accessed December 2014.
33. Bolton P. A doctor's duty to follow up preventable conditions: Young v Central Australian Aboriginal Congress - a bridge too far? *Northern Territory Law J* 2012; 2 (3).
34. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. *J Med Internet Res* 2012; 14(6): e162.
35. Forster M, Dennison K, Callen J, Georgiou A, Westbrook JI. Maternity patients' access to their electronic medical records: use and perspectives of a patient portal. *Health Inf Manag J* 2014; <http://dx.doi.org/10.12826/18333575.2014.0011.Forster>
36. Giardina TD, Singh H. Should patients get direct access to their laboratory test results?: An answer with many questions. *JAMA* 2011; 306: 2502-3.
37. Feeley TW, Shine KI. Access to the medical record for patients and involved providers: transparency through electronic tools. *Ann Intern Med* 2011; 155: 853-4.
38. Callen J, Georgiou A, Westbrook J, editors. Evaluation of solutions aimed at reducing the incidence of missed test results: can technology assist? [abstract] Sydney, Australia: 14th Asian Oceanian Congress of Radiology (AOCR); 2012.

39. Callen J, Singh H, Giardina TD, Li L, Paoloni R, Georgiou A, et al. Access to patient test results using the internet: a multi-site survey of emergency physicians' views of direct notification of laboratory and radiology test results. *J Med Internet Res* 2015; (in press).
40. Giardina TD, Callen J, Georgiou A, Westbrook JI, Greisinger A, Esquivel A, et al. Releasing test results directly to patients: a multisite survey of physician perspectives. *Patient Educ Couns* 2015; (in press).
41. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. *Ann Intern Med* 2012; 157: 461-70.
42. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Mlake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. *Ann Intern Med* 2013; 159: 677-87.
43. Giardina TD, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. *J Am Med Inform Assoc* 2013; doi:10.1136/amiajnl-2013-002239.
44. Nettleton S, Hanlon G. Pathways to the doctor in the information age: The role of ICTs in contemporary lay referral systems. In: *Innovative health technologies:*

The clinical and health economic value of clinical laboratory diagnostics

Bruce Jordan, Cheryl Mitchell, Andy Anderson, Norbert Farkas,
Richard Batrla

Roche Diagnostics Ltd, Rotkreuz, Switzerland

ARTICLE INFO

Corresponding author:

Bruce Jordan
Roche Diagnostics
CH-6343
Rotkreuz
Switzerland
e-mail: bruce.jordan@roche.com

Key words:

in vitro diagnostics, guided therapy,
clinical benefit, health economic benefit

Disclosure:

All authors of this paper are full-time employees of Roche Diagnostics Ltd. Medical writing assistance was provided by Julia Heagerty of Elements Communications, Westerham, UK and was funded by Roche Diagnostics Ltd.

ABSTRACT

The ultimate goal of diagnostic testing is to guide disease management in order to improve patient outcomes and patient well-being. Patient populations are rarely homogenous and accurate diagnostic tests can dissect the patient population and identify those patients with similar symptoms but very different underlying pathophysiology that will respond differently to different treatments. This stratification of patients can direct patients to appropriate treatment and is likely to result in clinical benefits for patients and economic benefits for the healthcare system. In this article we look at the clinical and economic benefits afforded by clinical laboratory diagnostics in three disease areas that represent substantial clinical and healthcare burdens to society; heart failure, Alzheimer's disease and asthma.

The relative spend on diagnostics compared with pharmaceuticals indicates that diagnostic tests are underappreciated in relation to the medical and economic value that they deliver. Clinical laboratory diagnostics should be viewed as a pivotal part of the healthcare system and valued accordingly. The skills available in clinical laboratories around the world should be harnessed to ensure the continued development of accurate tests that inform the healthcare community with respect to the pathophysiology of disease and facilitate the screening, diagnosis, appropriate treatment and monitoring of patients

INTRODUCTION

Clinical laboratory diagnostics are central to the integrated management of many different diseases. Without accurate diagnosis, appropriate treatment is not possible. However, the central role of clinical diagnostics is often underappreciated because the impact on patient care is not as readily apparent as medical intervention.

European expenditure on diagnostic procedures represents just 0.8% (€10.8 billion) of total healthcare expenditure (~€1,350 billion). Moreover, patients across Europe have unequal access to *in vitro* diagnostics because resources spent on these tests vary from €3.6 (Romania) to €43.5 (Switzerland) per capita per annum [1]. This expenditure belies the importance of clinical diagnostics, which is said to influence more than 60% of clinical decision making. Accurate diagnosis, based on detection of biomarkers and other tests, with subsequent guided therapy can result in clinical benefits for patients and economic benefits for the healthcare system [2,3]. As the population expands and ages, clinical laboratory diagnostics can help to reduce the associated healthcare costs by directing care and resources to those who are most likely to benefit.

Although automated platforms have accelerated the testing procedure and reduced the necessary labour intensity, many tests still require highly trained, skilled clinical scientists to interpret the results and relay these effectively to the clinician responsible for a patient's care. These clinical scientists should be viewed as highly valued members of the broader healthcare team.

The purpose of this article is to highlight the value of the diagnostic work conducted by clinical laboratories from two key perspectives; clinical and economic. The impact of clinical laboratory diagnostics is examined in three key areas; heart failure (HF), Alzheimer's disease (AD) and asthma.

IMPACT OF DIAGNOSIS ON UNDERSTANDING MEDICINE: HOW LESSONS FROM THE PAST ENABLE TREATMENT IN THE FUTURE

An understanding of diseases has always been fostered by a better understanding of underlying causes. In one of the earliest examples, diabetes mellitus (meaning 'honey-like'), was able to be separated from diabetes 'insipidus' (meaning 'tasteless') based on the observation that ants are attracted to the urine from a patient with diabetes mellitus [4,5]. Such an individual and definitive diagnosis is fundamental in separating patients with similar symptoms into subgroups with very different underlying pathophysiology.

Understanding how these diseases develop is key to appropriate patient management. It improves understanding of clinical symptoms and in turn improves early and accurate diagnosis of disease through the identification of at-risk groups. This is a progressive, iterative process with individual developments continually refining the initial wider spectrum diagnosis.

Asthma is an example of progressive refinement of diagnosis. Traditionally treatment of asthma has largely been symptomatic with increases in symptoms leading to escalation of therapy, with no knowledge or understanding of the different pathological causes responsible for symptoms in different patient groups. As a consequence, the cause of symptoms was not addressed and treatment response was suboptimal. Subsequent recognition that asthma patients can be eosinophilic or non-eosinophilic based on the presence or absence of sputum eosinophilia is leading to a better understanding of response to treatment in these patients [6]. However, this necessitates sputum testing for eosinophilic status becoming more widely accepted so that patients more likely to respond to therapy can be identified.

More recently further dissection of the patient population based on observed heterogeneity of interleukin-13 (IL-13) expression has identified a group of patients with high levels of periostin who are more likely to respond to therapy with lebrikizumab, an anti-IL-13 medicine currently in Phase III clinical development [7].

AD, the most frequent cause of dementia [8,9], may be a further example of such refinement. One characteristic of AD is the presence of amyloid-beta plaques. In the past a definitive diagnosis could only be made through identification of these plaques at autopsy, although more recently there has been a shift towards *in vivo* diagnosis based on amyloid-binding positron emission tomography (PET) tracers and cerebrospinal fluid (CSF) biomarkers. However, studies comparing clinical diagnosis and autopsy findings have shown that an incorrect diagnosis is made in as many as 12–23% of cases [10], and up to 32% of patients with clinically probable AD have shown no amyloid pathology on PET [11–13]. The potential impact of this was observed in the EXPEDITION 1 and 2 studies, which investigated the use of the humanized analogue of the murine antibody, solanezumab, in patients with mild-to-moderate AD [14]. In this study, there was no significant improvement in cognition or functional ability. However, 22% of the patient population did not meet the cut-off for being amyloid positive [15] and probably did not have AD. This may have diluted the efficacy. Using a biomarker like amyloid-beta it is possible to identify a purer population of the specific disease and gain an understanding of their disease progression and ability to be targeted with specific therapies, such as anti-amyloid therapy, that may be effective in this selected population. In fact, a subanalysis of these patients has demonstrated a trend to respond in amyloid-beta-enriched patients and the ongoing EXPEDITION 3 study is looking into this further [NCT01900665].

The understanding of the role of the specific Tau proteins in disease progression may further aid the understanding of the pathophysiological causes of AD. Stronger investment into biomarker research and provision of these biomarkers to physicians in the form of reliable and accessible diagnostic tools may be an effective route to developing a better understanding of the disease and ultimately help to develop more specific and effective therapies. For this reason the imbalance of expenditure on diagnostics and interventional drugs needs to be reduced. Diagnostics needs to play a more prominent role in medicine and these innovations should receive greater recognition by the healthcare community.

HEALTH ECONOMIC IMPACT: HOW THIS IS MEASURED

Problems central to the provision of healthcare include the scarcity of resources and the need to contain costs within healthcare systems against a background of increasing demand as a result of an ageing population, poor diet, increasing rates of obesity and other healthcare megatrends. Since the 1960s, expenditure on healthcare has risen faster than the general rate of inflation [16].

Health economic evaluations help decision makers to allocate scarce resources based on cost vs benefit. This mainly involves undertaking prospective and retrospective comparative studies and/or economic modelling [17]. Economic modelling falls into four major categories: cost minimization, cost-effectiveness, cost utility and cost-benefit analysis. Analysis can be performed from different perspectives; societal/economic perspective, healthcare system perspective, social insurance perspective or from the perspective of specific providers, such as hospitals. In general, choice of comparator must be appropriate for the specific analysis.

Costs are usually described in monetary units, while associated benefits are described in terms of quality-adjusted life years (QALYs) gained or lost [17]. The relationship between the two is the incremental cost-effectiveness ratio (ICER). Threshold values for 'willingness to pay' (e.g. approximately £20–30 k/QALY gained in the UK) could inform decision makers as to whether the technology in question is 'good value for money', keeping in mind the budgetary implications on the healthcare system.

Health economic evaluation of diagnostic technologies is complex, involving combined modelling of diagnostics and treatment, timing of tests and different test cut-off points, and is further complicated by the lack of universally accepted general guidelines and methodologies.

IMPACT OF CLINICAL LABORATORY DIAGNOSTICS: CLINICAL AND ECONOMIC PERSPECTIVES

Without reliable diagnostic tests appropriate clinical decisions cannot be made. Point of care tests allow these decisions to be made within hours, if not minutes. A single test can identify the need for additional tests, indicate that further tests are futile, or be sufficient to rule-out a disease and discharge a patient. They can be used to monitor treatment progress and to indicate when or whether treatment should be initiated or stopped as well as informing the optimal dose or treatment frequency needed to achieve a desired therapeutic effect in an individual patient.

A diagnosis based solely on clinical symptoms, as described above, can lead to the wrong conclusion. Laboratory diagnostics provide an objective measure. This is particularly important in areas where key symptoms are non-specific, such as dyspnoea or headache, and where diagnosis is problematic based on clinical history alone. Dyspnoea is one of the most common

symptoms. It is also one of the most non-specific; the online diagnostic tool, DiagnosticPro, lists close to 500 causes of dyspnoea, which can be challenging to distinguish between. Laboratory diagnostics, together with the clinical assessment, can give a definitive answer, or at least narrow down the options. For example, although acute coronary syndrome usually presents as dyspnoea associated with chest discomfort, it may typically present as dyspnoea alone. In this circumstance, cardiac markers are important for diagnosis and directing treatment. Nowadays, diagnostic tests can be performed at a centralized laboratory, in hospital, in the clinic, and at work or home, offering flexibility around clinical decision making.

Diagnostic tests have the ability to safeguard public health as well the health of an individual by providing rapid information during public health emergencies to confirm the presence of infectious disease, triage and treat accordingly. Evidence-based clinical practice guidelines are increasingly recommending the use of specific diagnostic tests because of their role in informing healthcare decision making.

Clinical diagnostics allow for the stratification of patients with heterogeneous diseases to enable targeted therapy for patients most likely to respond. Not only can diagnostic tests in some cases predict therapeutic efficacy, but they may also predict those who are more likely to experience adverse events. Thus, they inform the risk: benefit trade-off that is central to healthcare.

The real health economic benefit of clinical laboratory diagnostics is evident when the impact on tertiary care is examined. In particular, clinical laboratory diagnostics can be used effectively to triage patients to the appropriate level of care with a related reduction in costs associated with hospitalisation [3]. Additional cost benefits of clinical laboratory diagnostics may be realized through a reduction in the number needed

to treat, a reduction in drug costs associated with identifiable non-responders, avoided costs from predictable side effects, improved compliance and persistence and improved health outcomes [18]. Thus, clinical laboratory diagnostics play a key role by influencing the quality of patient care, health outcomes and downstream resource requirements. These considerations will become more and more important as the global population expands and ages. Using the example of AD, with an estimated projected worldwide patient population of 115 million by 2050 [8], employing a diagnostic test to exclude the proportion of patients unlikely to respond to therapy alone has the potential to drastically reduce associated healthcare costs.

The clinical benefit of an accurate diagnosis is apparent for all diseases. An associated health economic impact is most relevant in diseases that are highly prevalent or resource-intensive to manage. Three examples are HF, AD and asthma.

CLINICAL AND HEALTH ECONOMIC IMPACT OF LABORATORY DIAGNOSTICS IN HEART FAILURE

HF is one of the most costly medical conditions to manage, due to high prevalence and frequent and prolonged periods of hospitalization; in the US, in patients aged 18–64, each hospitalization due to HF costs an estimated \$23,077 [19]. Although HF-related hospitalization rates are declining [20], HF remains one of the leading causes of hospitalization among people aged >60 years [21], with patients staying on average 4 days longer in hospital than for other diseases [21,22]. In addition, over one-quarter of patients are readmitted within 30 days of initial discharge [23].

The prevalence of HF increases with age. In the UK, analysis from the British Heart Foundation estimates that 0.9% of men and 0.7% of women

suffer from HF, rising to 13.1% of men and 11.9% of women aged over 75 years [24]. Thus, as is the case in the US, HF constitutes a substantial burden on the National Health Service (NHS), accounting for one million inpatient bed-days (2% of the NHS total) and 5% of all emergency hospital admissions [25]. Given the age-related prevalence of HF, as well as age-related increases in recognised risk factors, such as hypertension, coronary heart disease, obesity, diabetes and hyperlipoproteinaemia, associated costs can be expected to increase. Indeed, the American Heart Association predicts that by 2030 the prevalence of HF will be 3.5%, equating to \$77.7 billion in direct costs [26].

A cardinal symptom of HF is dyspnoea. As noted above, this symptom is non-specific and subjective and patients presenting with dyspnoea may have multiple comorbidities that complicate diagnosis. This means that patients with HF may be missed or that patients may be misdiagnosed or hospitalized unnecessarily. Each of these consequences has clinical and economic implications. In a study of 592 dyspnoeic patients, clinical uncertainty (a diagnostic certainty estimate between 21% and 79%) for acutely destabilized HF was associated with increased morbidity and mortality. Significantly more patients in the clinical uncertainty group were admitted to hospital (86% vs 71%; $P<0.001$) and median length of stay in hospital was also longer (6.6 days vs 5.4 days; $P=0.02$). In addition, in the clinical uncertainty group >90% of patients were discharged within 14 days compared with 9 days in the clinical certainty group [2]. Clinical uncertainty was found to be an independent predictor of death (hazard ratio [HR] 1.88 [95% confidence interval {CI}: 1.02–2.25; $P=0.05$]) as well as death or hospitalization within one year (HR 2.18 [95% CI: 1.71–2.49; $P=0.01$]) [2]. Although not evaluated directly in this study, the observed increased hospitalization of patients

in the clinical uncertainty group is highly likely to be associated with increased healthcare spend.

Value of measuring N-terminal prohormone of brain natriuretic peptide

The data in the study by Green and colleagues [2] suggest that reducing diagnostic uncertainty has the potential to improve patient outcomes as well as reducing costs associated with hospitalization. This can be achieved by including other tests to inform diagnosis and not relying on non-specific clinical symptoms, such as dyspnoea, alone.

Echocardiography is the most reliable method for assessing cardiac pathology. However, echocardiographic assessment of all dyspnoeic patients is likely not to be cost-effective, with many patients referred for evaluation showing no evidence of significant heart disease [27,28]. Tests that can accurately and rapidly confirm or rule-out a diagnosis of HF have potential to improve subsequent patient management and

significantly reduce the costs associated with clinical uncertainty. A number of biomarkers have been identified as being associated with HF. Among these, the natriuretic peptides are of proven diagnostic/prognostic value, based on the observation that levels increase following atrial or ventricular dilatation [29].

Brain natriuretic peptide (BNP) is derived from pre-prohormone of brain natriuretic peptide, which is cleaved to remove the 26 amino acid signal protein and then subsequently to produce active BNP and its inactive N-terminal portion, NT-proBNP [29]. Both BNP and NT-proBNP have been shown to be of considerable utility for the clinical evaluation and risk prediction of HF [30]. NT-proBNP, however, does have a number of advantages over BNP, including a substantially longer half-life [30], higher circulating concentrations [30], greater stability [31], lower vulnerability to circadian variation [32] and more flexible sampling [30]. Unlike BNP, the

Table 1 Optimal NT-proBNP cut-points for the diagnosis/exclusion of acute HF among dyspnoeic patients [34]

Category	Optimal cut-point	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Accuracy (%)
Confirmatory (rule-in) cut-points						
<50 years (n=184)	450 pg/mL	97	93	76	99	94
50–75 years (n=537)	900 pg/mL	90	82	83	88	85
>75 years (n=535)	1,800 pg/mL	85	73	92	55	83
Rule-in overall		90	84	88	66	85
Exclusionary (rule-out) cut-points						
All patients (n=1,256)	300 pg/mL	99	60	77	98	83

k

\

y

h

-

o

#

available NT-proBNP assays are standardized and show relatively reproducible results [33].

In the first large-scale international analysis of NT-proBNP testing in the evaluation of patients with suspected HF, NT-proBNP was a sensitive and specific indicator of HF [34; Table 1]. Measuring NT-proBNP levels can reduce the uncertainty associated with HF diagnosis based on clinical symptoms alone [2] and thereby ensure appropriate care [35]. In the study described above [2], among the 185 patients in the clinical uncertainty group, 103 (56%) had acutely destabilized HF. In this group, the value of clinical judgement alone, determined by the area under the receiver operating characteristic curve (ROC AUC) was found to be 0.76 compared with 0.88 in the clinical certainty group ($P<0.001$). In the same population, measurement of NT-proBNP had an overall sensitivity of 90% (95% CI: 81%–94%), 84% specificity (95% CI: 72%–88%) and a positive predictive value of 86% for the diagnosis of acutely destabilised HF [2]. ROC AUC for NT-proBNP was 0.91 and 0.96 in the clinical uncertainty and clinical certainty groups, respectively (Table 2). Combining NT-proBNP with clinical judgement improved diagnostic accuracy in both the clinical certainty (ROC AUC

0.98) and clinical uncertainty groups (ROC AUC 0.94; Table 2) [2].

The IMPROVE CHF (Improved Management of Patients with CHF) trial evaluated the clinical and economic impact of NT-proBNP testing in addition to usual care compared to usual care alone on the management of 500 patients presenting to the emergency department with dyspnoea. This study also demonstrated increased diagnostic accuracy when combining NT-proBNP measurement with clinical judgement (ROC AUC of 0.90 [95% CI: 0.90–0.93] vs 0.83 [95% CI: 0.80–0.84]; $P=0.00001$) [35]. Overall, the median duration of the initial visit to the emergency department was significantly shorter in the NT-proBNP group compared with usual care (6.3 vs 5.6 hours; $P=0.0309$). There were no significant differences in initial hospitalizations, length of hospital stay, time in intensive care or initial and 60-day mortality. However, a significant reduction in the number of patients readmitted within 60 days was observed (13% vs 20%; $P=0.0463$). In addition NTpro-BNP-guided therapy resulted in a 15% reduction in total direct medical costs to 60 days follow up (\$6,129 vs \$5,180; $P=0.0232$) [35].

The studies above describe how the addition of NT-proBNP testing to clinical judgement based

Table 2 Accuracy of clinical judgement and NT-proBNP-guided judgement in dyspnoeic patients according to clinical certainty of a diagnosis of acutely destabilised HF [2]

Judgement	ROC (95% CI)	
	Clinical certainty (n=407)	Clinical uncertainty (n=185)
Clinical	0.88 (0.83–0.92)	0.76 (0.69–0.83)
NT-proBNP-guided	0.96 (0.94–0.97) ^a	0.91 (0.87–0.96) ^a
Clinical plus NT-proBNP	0.98 ^b	0.94 ^b

^a $P<0.001$ compared with clinical judgement;

^b $P<0.05$ for comparison with each of clinical and NT-proBNP-guided judgement alone

on symptoms and other evaluations improves the accuracy of diagnosis and can reduce direct medical costs. Other studies have demonstrated the value of NT-proBNP measurements in the stratification of patient care, also with the accompanying benefit of reducing associated healthcare costs. The PROMPT study resulted in improved stratification of patient care, with knowledge of elevated NT-proBNP levels resulting in early and more aggressive patient management. More patients with high levels of NT-proBNP ($>1,800$ pg/mL) were likely to be admitted to a higher level of care if the physician was aware of the NT-proBNP level than if they were not (21.9% vs 12.9%; $P=0.037$). Patients with a low NT-proBNP level (<150 pg/mL) were less likely to be admitted (4.6% vs 13.8%; $P=0.036$). There was no difference in admission rates in those patients with intermediate values of NT-proBNP [3]. In addition, compared with low levels, high levels of NT-proBNP were associated with higher rates of hospital admission (odds ratio [OR] 2.9), longer hospital stays (8.5 days vs 3.5 days, $P<0.01$), higher rates of in-hospital death (3.9% vs 0%, $P<0.01$), greater likelihood of re-hospitalization within 6 months (OR 5.1, $P<0.001$), and greater likelihood of death or re-hospitalization within 6 months (OR 5.7). Overall, NT-proBNP levels were associated with better stratification of patient care and were strongly correlated with subsequent utilization of hospital resources and prognosis [3]. In agreement with these observations, a cost-utility analysis of NT-proBNP-guided therapy in Canada found that NT-proBNP-guided intensive HF patient management, in addition to multidisciplinary care, not only reduced death and hospitalisation but was cost effective compared with multidisciplinary care alone or usual care, without adverse effects on safety [36]. NT-proBNP-guided intensive management cost less per patient compared with usual care and

multidisciplinary care (CAN\$55,946 vs \$57,729 and \$61,500, respectively). Quality-adjusted life-years were also greater (3.20 vs 2.36 and 3.04 for usual care and multidisciplinary care, respectively).

Taken together, these studies clearly demonstrate the considerable value of NT-proBNP testing from both a clinical and health economic perspective.

CLINICAL AND HEALTH ECONOMIC IMPACT OF LABORATORY DIAGNOSTICS IN ALZHEIMER'S DISEASE

According to the World Alzheimer's Report (2010), the global economic burden of dementia - which affects 36 million people around the world - has been estimated at \$604 billion [37]. The strongest risk factor for the development of AD is advancing age [8]. Therefore, increasing life expectancy will result in more and more people becoming affected by the disease; the number of people suffering from AD is estimated to reach 65.7 million by 2030 and 115.4 million by 2050 [37]. This same report predicts a rise of 85% in costs associated with dementia by 2030. As the most common cause of dementia, responsible for 60–80% of cases [9], AD is the largest contributor to this clinical and economic burden.

In Europe, annual costs per person with dementia vary widely. Based on Eurocodes estimates for dementia prevalence, a cost model based on published European cost of illness papers determined that the total cost of illness in the European Union in 2008 was €160 billion, which equates to €22,000 per person with dementia per year [38]. This annual burden varied from €4,473 in Eastern Europe to €35,987 in Northern Europe.

In the US, Medicare costs for beneficiaries with AD were \$91 billion in 2005 and reached a staggering \$160 billion in 2010. While direct medical

costs are substantial, the costs from lost wages of patients and families and the costs for non-nursing home patients is \$120 billion annually in the US. In high-income countries, informal care (45%) and formal social care (40%) account for the majority of costs, while the proportionate contribution of direct medical costs (15%) is much lower.

In the US, development of an intervention found to delay onset of AD by 5 years is estimated to result in a 57% reduction in the number of people affected and to almost halve projected annual Medicare costs from ~\$630 to ~\$340 billion [8]. Currently, however, there are no effective disease-modifying drugs that will prevent the disease, slow its progression or delay its onset [8]. In the absence of such drugs, early symptomatic treatment is the optimal strategy. Studies have shown that a patient's level of function will be preserved for longer if managed earlier and that community-dwelling patients with AD incur less societal cost than those who require long-term institutionalisation [39]. Early intervention, however, requires early diagnosis. As discussed earlier, diagnosis based on clinical signs and symptoms alone is incorrect in a substantial proportion of patients [10–13].

Biomarkers have diagnostic value in AD. Although several have been studied, evidence for three is strongest [8,40]; the 42 amino acid species of amyloid-beta (amyloid β 42 [$A\beta$ 42]), which is the principal constituent of amyloid plaques, and total Tau (t-Tau) and phosphorylated Tau (p-Tau), which aggregate to form intraneuronal neurofibrillary tangles and are associated with neuronal degeneration or injury. Both are measured in CSF. $A\beta$ 42 has been shown to have an inverse correlation with plaque load at autopsy, and whereas t-Tau and p-Tau are generally highly correlated and typically elevated in individuals with Alzheimer's disease, p-Tau may be more specific for AD as, unlike t-Tau, elevations are not observed in traumatic brain injury,

stroke or Creutzfeldt–Jakob disease [8]. Indeed, low circulating $A\beta$ 42 and high levels of Tau have been shown to have diagnostic and prognostic value in AD and are able to predict which individuals with mild cognitive impairment (MCI) and asymptomatic/preclinical AD are likely to progress to AD [8].

The ability to identify individuals whose disease is likely to progress using clinical laboratory assessment of biomarkers is important. Even in the absence of effective disease-modifying therapies, the timely detection of AD can be cost effective because treatments that are available can improve symptoms sufficiently to reduce healthcare costs by keeping patients living in the community for longer [41]. Because few treatments are available, this study modelled the effects of two hypothetical interventions; one modestly effective symptomatic treatment, and another that halted cognitive decline for a short period. Although hypothetical, the study demonstrates that early intervention is necessary for current symptomatic treatments to maximise cost-effectiveness. For disease-modifying drugs, maximal cost-effectiveness is achieved by intervening early enough to anticipate the period of rapid cognitive decline [41]. A diagnostic and economic evaluation of new biomarkers for AD is ongoing, which aims to assess the diagnostic test accuracy of current clinical diagnostic work-up and emerging biomarkers, perform a cost-consequence analysis and assess long-term cost-effectiveness using an economic model [42].

Recently, the use of AD pathology biomarkers has been included in the new consensus research diagnostic criteria for AD, MCI, and preclinical AD, proposed by the National Institute on Aging and the Alzheimer's Association. These new criteria take into account that AD dementia is part of a continuum of clinical and biological phenomena [43–45]. The new International Working Group (IWG) criteria, IWG-2,

recommend the use of either CSF biomarkers or PET imaging for the evaluation of AD patients [46]. In Europe, the Committee for Medicinal Products for Human Use published a number of qualification opinions on the use of biomarkers in the context of AD for enrichment of clinical trials in pre-dementia and mild-to-moderate AD [47]. The use of AD biomarkers for clinical trial enrichment is also supported by the recent FDA draft guidance for treatment of early AD; at this point the role of clinical laboratory diagnostics can be expected to be central in the effective clinical and cost-effective management of patients with AD.

CLINICAL AND HEALTH ECONOMIC IMPACT OF LABORATORY DIAGNOSTICS IN ASTHMA

Asthma is a highly heterogeneous disease. It is a global public health problem and the prevalence is increasing in most countries [48]. According to the Global Asthma Report, as many as 334 million people may be affected and the burden of disability is high [49]. Asthma was once considered a disease of high income societies, but this is no longer the case and rates of asthma are increasing fastest in low to middle income societies [49]. It is responsible for an estimated 1% of the worldwide disability-adjusted life years lost [50] and ranks 22nd worldwide, similar to other chronic diseases, such as diabetes [48]. In Western Europe one in four patients requires either an emergency room or unscheduled urgent care visit, and in North America this figure reached 40% [50]. In the US, patients with asthma exacerbations had significantly higher total healthcare costs compared with those who did not (\$9,223 vs \$5,011; $P < 0.0001$). Asthma-related costs were also significantly higher (\$1,740 vs \$847; $P < 0.0001$), and they tend to have co-morbidities such as sinusitis, pneumonia, and mental disorders [51].

In the UK, the NHS spends around £1 billion a year for the treatment of patients with asthma. In the year 2008/2009 up to 1.1 million working days were lost due to lung problems [52,53]. Asthma exacerbations led to over 50,000 hospital admissions with an annual spend of £800 million on pharmaceutical therapy alone [54]. In Germany, the direct and indirect medical costs reached €2.74 billion during 1999. Age-specific hospital costs per admission ranged from €564 (in those <5 years of age) to €2,800 (in those ≥75 years of age) [55]. Moreover, despite the availability of effective preventive therapy, costs associated with asthma appear to be increasing [56].

The heterogeneity of the disease makes it a challenge to manage. Patients present with different clinical, inflammatory and immunological phenotypes, the identification of which is key to providing effective treatment. Traditional diagnostic techniques rely on clinical judgement and pulmonary function tests, despite the limitations of both [57]. Associated exacerbations, defined as the need for courses of high-dose oral corticosteroids or hospitalization, are a major cause of morbidity as a result of an accelerated decline in lung function [58,59] and are associated with high healthcare costs comparable to diabetes and hypertension [59,60]. Approximately 5–14% of the total asthma population have severe asthma [61,62] (Table 3) and this population is associated with disproportionate healthcare use and costs [62,63], both in terms of direct and indirect costs [64,65] (Figure 1). Disease exacerbations, in particular hospitalizations, account for 55% of direct costs in the EU.

It is not possible to predict the risk of exacerbation based on asthma phenotype without the use of biomarkers. However, along with a patient's clinical history, biomarkers may help identify individuals at risk of exacerbations, which may in turn improve patient care and reduce associated healthcare costs. Currently

Table 3 Severe asthma is associated with disproportionate healthcare use and costs [32]

Asthma severity	% Asthma population	Mean direct costs* (€)
Mild	13.7	263
Moderate	33.3	686
Moderate–severe	38.9	1,196
Severe	14.1	2,782

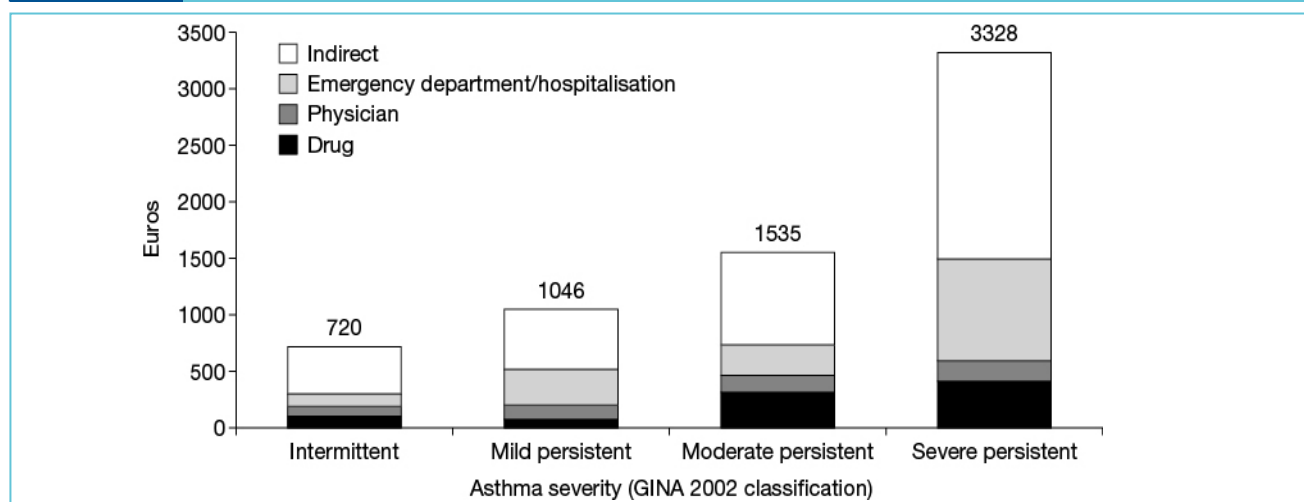
*Direct costs of asthma: mean costs of goods and services except hospitalization.

available biomarkers for clinical practice, such as those in bronchial lavage, bronchial biopsies, sputum or fraction of exhaled nitric oxide (FeNO) are limited due to invasiveness or lack of specificity [66], and there is a need for easily interpreted biomarkers that can be exploited in clinical laboratory diagnostic tests to assess the nature and severity of disease.

Serum total IgE and allergen specific IgE are biomarkers to define phenotype in asthmatic patients [67]. Serum periostin, a systemic marker

of T2-derived asthma, is upregulated by IL-13 and may be the marker with a highest accuracy for identifying eosinophilic airway inflammation in asthma [68–70]. Lebrikizumab, a monoclonal antibody to IL-13, has been shown to have a more pronounced anti-asthmatic effect in patients with elevated periostin [7]. Thus, diagnostic tests for periostin have the potential to identify a subgroup of asthma patients who will benefit from treatment with lebrikizumab. IL-5 has also been proposed as a potential therapeutic target in eosinophilic asthma. FeNO may

Figure 1 Asthma-related direct (drug; physician visits; emergency department/hospitalization) and indirect annual costs per patient by severity of asthma



Reproduced by permission of the European Respiratory Society: *Eur Respir J* May 2004; 23: 723–29; doi:10.1183/09031936.04.00004904.

help predict exacerbations and may identify patients most likely to respond to inhaled corticosteroids [71], although results are conflicting.

Treatment directed by serial sputum eosinophil count measurements has been shown to prevent exacerbations in patients with severe asthma, resulting in fewer hospital admissions [72]. In this study, compared with treatment based on symptoms and spirometry, sputum count-directed corticosteroid therapy resulted in fewer exacerbations (47 vs 79; $P=0.04$), a longer period until first exacerbation (607 days vs 394 days) and fewer exacerbations requiring prednisolone (78% occurred in the symptoms and spirometry group). Since exacerbations are responsible for a substantial proportion of asthma-related costs, these observations may be expected to reduce healthcare expenditure.

As well as identifying those most likely to respond to certain therapies, eosinophil counts can similarly be used to identify patients likely to have a poor response to corticosteroids [73]. Identifying subpopulations of patients with improved clinical response to specific drugs allows targeted therapy and is likely to reduce costs. Individualized management plans have been shown to improve asthma control and reduce hospitalization (relative risk [RR] 0.64 [95% CI: 0.50–0.82]) and emergency room attendance (RR 0.82 [95% CI: 0.73–0.94]) as a result of exacerbations [74] as well as reducing the number of days off work (RR 0.79 [95% CI: 0.67–0.93]).

Clinical laboratory diagnostics clearly have a central role to play in the appropriate, cost-effective management of patients with asthma. The heterogeneity of the asthma phenotype requires clinical laboratory diagnostic tests for a biomarker panel to improve disease diagnosis [67].

DISCUSSION

The literature reviewed in this paper is not exhaustive. However, in the three therapeutic areas discussed there appear to be clear clinical and/or economic benefits to guided therapy facilitated by accurate clinical laboratory diagnostics. NT-proBNP-guided therapy has the potential to triage patients to the appropriate level of care [3], and to reduce costs associated with hospitalization [2,35]. Earlier intervention with symptomatic treatments in AD based on diagnosis with A β 42 and Tau has the potential to reduce associated costs by keeping patients functioning in the community for longer [41]. When disease modifying drugs do become available, they have the potential for substantially reducing the financial impact of AD [8,41]. In asthma, emerging biomarkers, such as periostin, have the potential to dissect the heterogeneous asthma population and to direct care to those most likely to respond to therapy [7]. However, these apparent benefits of individualized healthcare need to be balanced against costs associated with this approach. These include additional costs associated with the true and false positive patients, the costs associated with expanding patient populations through screening and prevention, which will need potentially costly therapeutic intervention, and increased spending on diagnostics [18].

Diagnosis is a vital part of medical innovation and novel diagnostic tools enable the identification of patients with a specific pathophysiological cause within a group of patients with similar symptoms. This, in turn, fosters better understanding of the disease and perpetuates the cycle of medical innovation; provision of innovative and reliable/reproducible diagnostic tools to physicians is crucial for reliable outcomes in this process.

The clinical laboratory is thus central to the provision of effective patient care, identifying

disease, guiding treatment, and monitoring response. The skills within clinical diagnostic laboratories must be used to further refine diagnostic processes and realize the promise to identify patients more or less likely to respond to a particular therapy and to ensure appropriate, targeted therapy for all. This in turn should help to control healthcare costs associated with an expanding ageing population. To achieve this, manufacturers will need to focus on developing diagnostic tests that better predict clinical outcomes and deliver savings in healthcare costs and improve patient management. They will also need to collaborate more systematically to demonstrate the significant contribution of diagnostics to improving delivery of healthcare to patients.

The relative spend on diagnostics compared with pharmaceuticals underlines the fact that currently diagnostic tests are in general underappreciated in relation to the medical and economic value that they deliver. Unlike the 'value-based' reimbursement of innovative pharmaceuticals, in many markets *in vitro* diagnostics have been treated as low-margin commodities with low reimbursement rates that are based solely on the method of testing and not according to value brought to the patient [75]. In addition, in most healthcare systems, codings are non-specific, covering procedures, rather than technologies or brands, and new tests are linked to existing Diagnosis-Related Group codes [75].

The ultimate goal of diagnostic testing is to guide disease management in order to improve patient outcomes and patient well-being. Clinical laboratory diagnostics should be viewed as a pivotal part of the healthcare system and valued accordingly. The skills available in clinical laboratories around the world should be harnessed to ensure the continued development of accurate tests that inform the healthcare community with respect to the pathophysiology of

disease and facilitate the diagnosis, appropriate treatment and monitoring of patients. Laboratory medicine will need to form alliances with clinicians, healthcare managers and insurers, as well as the general public, and gain these stakeholders as advocates for valuing laboratory medicine according to the information it delivers to facilitate optimum clinical care.

REFERENCES

1. European IVD Market Statistics Report. Available at: <http://www.edma-ivd.eu/index.php?mact=EuropeanIVDMarketStatistics,me8bc7,view,1&me8bc7details=26&me8bc7returnid=104&page=104>. [Accessed November 2014].
2. Green SM, Martinez-Rumayor A, Gregory SA, Baggish AL, O'Donoghue ML, Green JA, et al. Uncertainty, diagnostic accuracy, and outcomes in emergency department patients presenting with dyspnea. *Arch Intern Med* 2008; 168: 741–8.
3. Luchner A, Möckel M, Spanuth E, Möcks J, Peetz D, Baum H, et al. N-terminal pro brain natriuretic peptide in the management of patients in the medical emergency department (PROMPT): correlation with disease severity, utilization of hospital resources, and prognosis in a large, prospective, randomized multicentre trial. *Eur J Heart Fail* 2012; 14: 259–67.
4. Lindholm J. Diabetes insipidus: historical aspects. *Pituitary* 2004; 7: 33–8.
5. Lakhtakia R. The history of diabetes mellitus. *Sultan Qaboos Univ Med J* 2013; 13: 368–70.
6. McGrath KW, Icitovic N, Boushey HA, Lazarus SC, Sutherland ER, Chinchilli VM, Fahy JV. Asthma Clinical Research Network of the National Heart, Lung, and Blood Institute. A large subgroup of mild-to-moderate asthma is persistently noneosinophilic. *Am J Respir Crit Care Med* 2012; 185: 612–9.
7. Corren J, Lemanske RF, Hanania NA, Korenblat PE, Parsey MV, Arron JR, et al. Lebrikizumab treatment in adults with asthma. *N Engl J Med* 2011; 365: 1088–98.
8. Fagan AM. CSF Biomarkers of Alzheimer's disease: Impact on disease concept, diagnosis, and clinical trial design. *Adv Geriatr* 2014; Article ID 302712: doi:10.1155/2014/302712.
9. Alzheimer's Association. 2013 Alzheimer's disease facts and figures. *Alzheimers Dement* 2013; 9: 208–45.
10. Gaugler JE, Ascher-Svanum H, Roth DL, Fafowora T, Siderowf A, Beach TG. Characteristics of patients misdiagnosed with Alzheimer's disease and their medication

use: an analysis of the NACC-UDS database. *BMC Geriatr* 2013; 13: 137.

11. Doraiswamy PM, Sperling RA, Coleman RE, Johnson KA, Reiman EM, Davis MD, et al. (AV45-A11 Study Group). Amyloid- β assessed by florbetapir F 18 PET and 18-month cognitive decline: a multicenter study. *Neurology* 2012; 79: 1636–44.

12. Devanand DP, Mikhno A, Pelton GH, Cuasay K, Pradhaban G, Dileep Kumar JS, et al. Pittsburgh compound B (11C-PIB) and fluorodeoxyglucose (18 F-FDG) PET in patients with Alzheimer disease, mild cognitive impairment, and healthy controls. *J Geriatr Psychiatry Neurol* 2010; 23: 185–98.

13. Nelissen N, Van Laere K, Thurfjell L, Owenius R, Vandenbulcke M, Koole M, et al. Phase 1 study of the Pittsburgh compound B derivative 18F-flutemetamol in healthy volunteers and patients with probable Alzheimer disease. *J Nucl Med* 2009; 50: 1251–9.

14. Doody RS, Thomas RG, Farlow M, Iwatsubo T, Vellas B, Joffe S, et al. Alzheimer's Disease Cooperative Study Steering Committee; Solanezumab Study Group. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med* 2014; 370: 311–21.

15. Dean R, Shaw LM, Waligorska TW, Korecka M, Figurski M, Trojanowski JQ, et al. Inclusion of patients with Alzheimer's disease pathology in solanezumab expedition 3 using florbetapir PET imaging or inno-bia AlzBio3 CSF $\text{a}\beta$ 1-42. *Alzheimers Dement* 2014; 10: 811.

16. Meltzer MI. Introduction to health economics for physicians. *Lancet* 2001; 358(9286): 993–8.

17. National Institute of Clinical Excellence. Guidance for Manufacturers and Sponsors. London: NICE, 2001. Available at: <http://www.gserve.nice.org.uk/niceMedia/pdf/technicalguidanceformanufacturersandsponsors.pdf>. [Accessed December 2014].

18. Faulkner E, Annemans L, Garrison L, Helfand M, Holtorf A-P, Hornberger J, et al, for Personalized Medicine Development and Reimbursement Working Group. Challenges in the development and reimbursement of personalized medicine—payer and manufacturer perspectives and implications for health economics and outcomes research: A report of the ISPOR Personalized Medicine Special Interest Group. *Value Health* 2012; 15: 1162–71.

19. Wang G, Zhang Z, Ayala C, Wall HK, Fang J. Costs of heart failure-related hospitalizations in patients aged 18 to 64 years. *Am J Manag Care* 2010; 16: 769–76.

20. Chen J, Normand SL, Wang Y, Krumholz HM. National and regional trends in heart failure hospitalization and mortality rates for Medicare beneficiaries, 1998–2008. *JAMA* 2011; 306: 1669–78.

21. Cleland JG, Swedberg K, Follath F, Komajda M, Cohen-Solal A, Aguilar JC, et al, Study Group on Diagnosis of the

Working Group on Heart Failure of the European Society of Cardiology. The EuroHeart Failure survey programme—a survey on the quality of care among patients with heart failure in Europe. Part 1: patient characteristics and diagnosis. *Eur Heart J* 2003; 24: 442–63.

22. Organisation for Economic Co-operation and Development (OECD). Average length of stay in hospitals. In: *Health at a glance 2011: OECD indicators*. OECD, 2011. pp88–89.

23. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med* 2009; 360: 1418–28.

24. British Heart Foundation Health Promotion Research Group. Coronary heart disease statistics: 2012 edition, London: British Heart Foundation, 2012. p.81. Available at: <https://www.bhf.org.uk/publications/statistics/coronary-heart-disease-statistics-2012>. [Accessed December 2014].

25. National Institute for Health and Care Excellence. NICE Guideline CG108 Chronic heart failure: Management of chronic heart failure in adults in primary and secondary care. London: NICE, 2010. Available at: <http://www.nice.org.uk/guidance/cg108>. [Accessed December 2014].

26. Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation* 2011; 123: 933–44.

27. Hobbs FD, Jones MI, Allan TF, Wilson S, Tobias R. European survey of primary care physician perceptions on heart failure diagnosis and management (Euro-HF). *Eur Heart J* 2000; 21: 1877–87.

28. Khunti K. Systematic review of open access echocardiography for primary care. *Eur J Heart Fail* 2004; 6: 79–83.

29. Januzzi JL Jr. Natriuretic peptides as biomarkers in heart failure. *J Investig Med* 2013; 61: 950–5.

30. Steiner J, Guglin M. BNP or NTproBNP? A clinician's perspective. *Int J Cardiol* 2008; 129: 5–14.

31. Yeo KT, Wu AH, Apple FS, Kroll MH, Christenson RH, Lewandowski KB, et al. Multicenter evaluation of the Roche NT-proBNP assay and comparison to the Biosite Triage BNP assay. *Clin Chim Acta* 2003; 338: 107–15.

32. Bruins S, Fokkema MR, Römer JW, Dejongste MJ, van der Dijs FP, van den Ouweland JM, Muskiet FA. High intraindividual variation of B-type natriuretic peptide (BNP) and amino-terminal proBNP in patients with stable chronic heart failure. *Clin Chem* 2004; 50: 2052–8.

33. Clerico A, Zaninotto M, Prontera C, Giovannini S, Ndreu R, Franzini M, et al. (Study Group on Cardiovascular Risk Biomarkers of the Italian Society of Clinical Biochemistry). State of the art of BNP and NT-proBNP im-

munoassays: the CardioOrmoCheck study. Clin Chim Acta 2012; 414: 112–9.

34. Januzzi JL, van Kimmenade R, Lainchbury J, Bayes-Genis A, Ordóñez-Llanos J, Santalo-Bel M, et al. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the International Collaborative of NT-proBNP Study. Eur Heart J 2006; 27: 330–7.

35. Moe GW, Howlett J, Januzzi JL, Zowall H. (Canadian Multicenter Improved Management of Patients with Congestive Heart Failure (IMPROVE-CHF) Study Investigators). N-terminal pro-B-type natriuretic peptide testing improves the management of patients with suspected acute heart failure: primary results of the Canadian prospective randomized multicenter IMPROVE-CHF study. Circulation 2007; 115: 3103–10.

36. Moertl D, Steiner S, Coyle D, Berger R. Cost-utility analysis of nt-proBNP-guided multidisciplinary care in chronic heart failure. Int J Technol Assess Health Care 2013; 29: 3–11.

37. Alzheimer's Disease International. World Alzheimer Report 2010. Available at: <http://www.alz.co.uk/research/files/WorldAlzheimerReport2010.pdf>. [Accessed December 2014].

38. Wimo A, Jönsson L, Gustavsson A, McDaid D, Ersek K, Georges J, et al. The economic impact of dementia in Europe in 2008 - cost estimates from the Eurocode project. Int J Geriatr Psychiatry 2011; 26: 825–32.

39. Leifer BP. Early diagnosis of Alzheimer's disease: clinical and economic benefits. J Am Geriatr Soc 2003;51:S281–8.

40. Humpel C. Identifying and validating biomarkers for Alzheimer's disease. Trends Biotechnol 2011; 29: 26–32.

41. Barnett JH, Lewis L, Blackwell AD, Taylor M. Early intervention in Alzheimer's disease: a health economic study of the effects of diagnostic timing. BMC Neurol 2014; 14: 101.

42. Handels RL, Aalten P, Wolfs CA, OldeRikkert M, Scheltens P, Visser PJ, et al. Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease: the research protocol of a prospective cohort study. BMC Neurol 2012; 12: 72.

43. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging - Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7: 263–9.

44. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging - Alzheimer's Association

workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7: 270–9.

45. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7: 280–92.

46. Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol 2014; 13: 614–29.

47. Isaac M, Vamvakas S, Abadie E, Jonsson B, Gispen C, Pani L. Qualification opinion of novel methodologies in the predementia stage of Alzheimer's disease: cerebro-spinal-fluid related biomarkers for drugs affecting amyloid burden--regulatory considerations by European Medicines Agency focusing in improving benefit/risk in regulatory trials. Eur Neuropsychopharmacol 2011; 21: 781–8.

48. Bousquet J, Khaltaev NG. Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach. World Health Organization, Geneva, 2007. Available at: <http://www.who.int/gard/publications/GARD%20Book%202007.pdf>. [Accessed November 2014].

49. Global Asthma Network. Global Asthma Report 2014. Available at: <http://www.globalasthmareport.org/>. [Accessed December 2014].

50. Global Initiative for Asthma. Global Burden of Asthma, 2004. Available at: <http://www.ginasthma.org/documents/8>. [Accessed December 2014].

51. Ivanova JI, Bergman R, Birnbaum HG, Colice GL, Silverman RA, McLaurin K. Effect of asthma exacerbations on health care costs among asthmatic patients with moderate and severe persistent asthma. J Allergy Clin Immunol 2012; 129: 1229–35.

52. Asthma UK. Asthma facts and FAQs. Available at: <http://www.asthma.org.uk/asthma-facts-and-statistics>. [Accessed December 2014].

53. National Health Service. An Outcomes Strategy for COPD and Asthma: NHS Companion Document. Department of Health, 2012. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216531/dh_134001.pdf. [Accessed December 2014]

54. National Health Service. 2013/14 standard contract for respiratory:severe asthma (adult) Available at: <http://www.england.nhs.uk/wp-content/uploads/2013/06/a14-respiratory-sev-asthma.pdf>. [Accessed December 2014]

55. Stock S, Redaelli M, Luengen M, Wendland G, Civello D, Lauterbach KW. Asthma: prevalence and cost of illness. Eur Respir J 2005; 25: 47–53.

56. Bahadori K, Doyle-Waters MM, Marra C, Lynd L, Alasaly K, Swiston J, FitzGerald JM. Economic burden of asthma: a systematic review. *BMC Pulm Med* 2009; 9: 24.
57. Taylor DR. Biomarkers of inflammation in asthma: a clinical perspective. *Semin Respir Crit Care Med* 2012; 33: 620–9.
58. Bai TR, Vonk JM, Postma DS, Boezen HM. Severe exacerbations predict excess lung function decline in asthma. *Eur Resp J* 2007; 30: 452–6.
59. Thomson NC, Chaudhuri R. Identification and management of adults with asthma prone to exacerbations: can we do better? *BMC Pulm Med* 2008; 8: 27.
60. Lane S, Molina J, Plusa T. An international observational prospective study to determine the Cost Of Asthma exacerbations (COAX). *Resp Med* 2006; 100: 434–50.
61. Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J* 2014; 43: 343–73.
62. Braman SS. The global burden of asthma. *Chest* 2006; 130(1 Suppl): 4S–12S.
63. Godard P, Chanez P, Siraudin L, Nicoloyannis N, Duru G. Costs of asthma are correlated with severity: a 1-yr prospective study. *Eur Respir J* 2002; 19: 61–7.
64. Peters SP, Ferguson G, Deniz Y, Reisner C. Uncontrolled asthma: a review of the prevalence, disease burden and options for treatment. *Respir Med* 2006; 100: 1139–51.
65. Antonicelli L, Bucca C, Neri M, De Benedetto F, Sabbatani P, Bonifazi F, et al. Asthma severity and medical resource utilisation. *Eur Respir J* 2004; 23: 723–9.
66. Vijverberg SJ, Hilvering B, Raaijmakers JA, Lammers JW, Maitland-van der Zee AH, Koenderman L. Clinical utility of asthma biomarkers: from bench to bedside. *Biologics* 2013; 7: 199–210.
67. Fatemi F, Sadroddiny E, Gheibi A, Mohammadi Farsani T, Kardar GA. Biomolecular markers in assessment and treatment of asthma. *Respirology* 2014; 19: 514–23.
68. Jia G, Erickson RW, Choy DF, Mosesova S, Wu LC, Solberg OD, et al. (Bronchoscopic Exploratory Research Study of Biomarkers in Corticosteroid-refractory Asthma (BOBCAT) Study Group.) Periostin is a systemic biomarker of eosinophilic airway inflammation in asthmatic patients. *J Allergy Clin Immunol* 2012; 130: 647–54.
69. Nair P, Kraft M. Serum periostin as a marker of T(H)2-dependent eosinophilic airway inflammation. *J Allergy Clin Immunol* 2012; 130: 655–6.
70. Parulekar AD, Atik MA, Hanania NA. Periostin, a novel biomarker of TH2-driven asthma. *Curr Opin Pulm Med* 2014; 20: 60–5.
71. Dweik RA, Sorkness RL, Wenzel S, Hammel J, Curran-Everett D, Comhair SA, et al. Use of exhaled nitric oxide measurement to identify a reactive, at-risk phenotype among patients with asthma. *Am J Respir Crit Care Med* 2010; 181: 1033–41.
72. Jayaram L, Pizzichini MM, Cook RJ, Boulet LP, Lemièrre C, Pizzichini E, et al. Determining asthma treatment by monitoring sputum cell counts: effect on exacerbations. *Eur Respir J* 2006; 27: 483–94.
73. Green RH, Brightling CE, Woltmann G, Parker D, Wardlaw AJ, Pavord ID. Analysis of induced sputum in adults with asthma: identification of subgroup with isolated sputum neutrophilia and poor response to inhaled corticosteroids. *Thorax* 2002; 57: 875–9.
74. Gibson P, Powell H, Coughlan J, Wilson A, Abramson M, Haywood P, et al. Self-management education and regular practitioner review for adults with asthma. *Cochrane Database Syst Rev* 2003; 1: CD001117.
75. Fugel HJ, Nuijten M, Postma M. Stratified medicine and reimbursement issues. *Front Pharmacol* 2012; 3: 181. doi: 10.3389/fphar.2012.00181.

Evaluating biomarkers for guiding treatment decisions

Patrick M. Bossuyt, Parvin Tajik

*Department of Clinical Epidemiology, Biostatistics & Bioinformatics,
Academic Medical Center, University of Amsterdam, The Netherlands*

ARTICLE INFO

Corresponding author:

Patrick M. Bossuyt
Dept. Clinical Epidemiology,
Biostatistics & Bioinformatics
Academic Medical Center
University of Amsterdam
Room J2-127 PO Box 22700
1100 DE Amsterdam
The Netherlands
E-mail: p.m.bossuyt@amc.uva.nl

Key words:

biomarker, medical test evaluation,
evidence-based medicine,
clinical effectiveness

ABSTRACT

The genetic revolution is expected to lead to improved targeting of new and existing forms of treatment. Rather than a one-size-fits-all blockbuster strategy in battling disease with drugs and other interventions, a more precise approach is becoming available, one in which treatment is only offered to those likely to benefit. The identification of those likely to benefit from treatment could be based on one or more biomarkers, but in an era where medical decisions aim to be evidence-based, the use of treatment selection markers should not just be based on hope and optimism, but on solid data from sound research. The performance of the treatment selection marker should be expressed in quantitative terms, similar to the way we express the clinical performance of diagnostic markers, or the performance of prognostic markers.

We describe recent research on this issue. First we present in intuitive terms a general, decision-theoretical framework for making treatment decisions. We then describe some measures for expressing the performance of treatment selection markers, showing that conventional measures of clinical performance, such as clinical sensitivity and specificity, are not decisive or helpful. In the last part of the paper, we provide a brief summary of study designs for evaluating treatment selection markers. Like all other forms of medical testing, potential treatment selection markers should be properly evaluated before they are implemented in routine clinical practice.

INTRODUCTION

The unraveling of the human genome has fuelled high hopes for the advancement of clinical medicine. Many believed that our improved understanding of the role of genes, the function of proteins, and the characterization of small-molecule metabolite profiles would strengthen our understanding of the origins of disease, and would help to clarify disease mechanisms. This would eventually lead to new and better forms of treatment, enabling clinicians to sustain and restore health for their patients, and to prevent premature death.

The benefits from the genetic revolution would not just come from new forms of treatment. The advances in knowledge were also expected to lead to improved targeting of new and existing forms of treatment. Rather than a one-size-fits-all blockbuster strategy in battling disease with drugs and other interventions, a more precise approach would become available, one in which treatment is only offered to those likely to benefit. The identification of those likely to benefit from treatment would be made based on one or more biomarkers. We will refer to such biomarkers as “treatment selection markers”.

In an era where medical decisions aim to be evidence-based, the use of treatment selection markers would not just be based on hope and optimism, but on solid data from sound research. It is not sufficient to expect a benefit from using a biomarker to guide treatment decisions, one should also have convincing evidence that the marker is actually able to do so. The performance of the treatment selection marker should be expressed in quantitative terms, similar to the way we express the clinical performance of diagnostic markers, or the performance of prognostic markers.

These new ambitions pose a challenge for laboratory professionals, and for researchers and

methodologists in general. How does one know that a marker is fit to serve as a guide for treatment decisions? How can one express the performance of a treatment selection marker?

This paper summarizes some recent research on this issue. First we present in intuitive terms a general, decision-theoretical framework for making treatment decisions. We then present some measures for expressing the performance of treatment selection markers, showing that conventional measures of clinical performance, such as (clinical) sensitivity and specificity, are not decisive or helpful. In the last part of the paper, we provide a brief summary of study designs for evaluating treatment selection markers.

THE ANATOMY OF TREATMENT DECISIONS

In general, a treatment decision is based on balancing the positive, hoped-for effects against the negative, feared effects. The latter could be a combination of the side-effects of treatment, the burden of treatment (going to the hospital at regular intervals, or taking pills daily), and the societal costs: the resources used to develop, build and administer treatment. The positive effects are the health gains expected from treatment: restoration of health, or the prevention of worsening.

If we assume the negative effects are all known, we can re-express the treatment decision as a threshold issue. Are the positive effects large enough to offset the negative ones? Assume, for example, that the positive, hoped-for effect is an increase in 5-year survival from adjuvant chemotherapy for a cancer patient. Assume, additionally, that we have a reliable estimate of the 5-year survival for that patient. We then can present the negative effects of adjuvant chemotherapy to the patients and ask the patient how large the gain in 5-year survival have to be to justify treatment for that patient. Assume

then that a new, large RCT comes out that has estimated the survival benefit of this form of adjuvant chemotherapy for patients similar to the one facing the decision. That patient then can compare the gain in survival – in absolute terms – with the personal threshold. If the gain is larger than the threshold, adjuvant chemotherapy seems justified. Otherwise, if the gain is smaller than the threshold, this is not the case.

In this case we base the recommendation about treatment not on the statistical significance of the treatment effect, as estimated in the randomized trial. As is well known, such a significance test only evaluates whether the difference in survival is zero. In case of a significant result, we have rejected the null hypothesis of equality. With a two-sided test, this implies that the alternative hypothesis specifies that the survival difference is either negative or positive; with a one-sided test, the alternative hypothesis typically specifies that there is some survival gain. So conventional statistical significance tests typically do not indicate whether the health gains are large enough. We must add that, in principle, it would be perfectly possible to formulate an alternative statistical hypothesis test, one in which we test whether the treatment effect exceeds a pre-specified threshold, but this is not typically done in randomized trials.

The recommendation about treatment is also not based on the target difference, as used in the sample size calculations. This target difference helps to calculate the desired precision of a study, which is typically driven by the number of included study participants. The target difference can provide reassurance that the study will be informative, in the sense that a relevant difference, if one exists, is likely to be detected with the required statistical precision (1).

Asking for a threshold for the treatment effect sounds like a complicated question to ask a patient. It is probably not an easy task to define a personal threshold, but existing research has shown that the question is indeed answerable. For adjuvant chemotherapy, for example, the actual question “what makes it worthwhile” has been asked to patients with non-small-cell lung cancer (2), to patients with early colon cancer (3), and patients with early breast cancer (4).

The threshold does not have to be same for every individual patient: for some the required gain may be fairly large, while for others extending survival is extremely important, and their threshold for accepting treatment is close to zero. This is definitely an area for personalized medicine: not in the abundant use of next-generation sequencing, but in the recognition that personal values and trade-offs differ. Despite this recognition, we will assume for now that there is one common threshold, to ease the exposition.

In itself, the threshold approach is as old as decision theory. It was introduced, or re-introduced, into medicine in the 1970s, through impressive articles written by Steve Pauker and Jerome Kassirer, which formed the start of clinical decision analysis and helped to launch economic evaluations in health care (5, 6).

Note also that the question about a large benefit is usually phrased in terms of the absolute benefit: the survival gain in percentage points at five years, for example. Although treatment effects in trials are typically expressed in relative terms, answering the question about the threshold in such relative terms is much more challenging and complicated.

TREATMENT SELECTION MARKERS

So, when can a marker act as a treatment selection marker, to guide decisions about treatment?

The threshold approach to decision-making, as just introduced, allows us a simple rule to arrive at a conclusion when evaluating a biomarker's potential to guide treatment. We assume for now that the marker is present or absent, or takes values in a well-known range. To be sufficiently general, we suppose that the marker is quantitative, be it on a dichotomous (1/0), ordinal, or interval scale.

One condition for a marker to act as a treatment selection marker is the existence of heterogeneity in the treatment effect. Keeping to the example of survival gain from adjuvant chemotherapy, this means that not everybody in the trial population is expected to benefit to the same degree from the treatment: for some the benefit is larger, for others smaller, and there may be subgroups who do not benefit from chemotherapy, but are even harmed by it: their 5-year survival is lower after treatment.

A second condition is then the existence of a reliable association between the putative treatment selection marker and treatment benefit. We can further specify this condition in terms of a classification, relative to the (common) threshold: the marker is able to identify a subgroup for which the survival gain is equal to or larger than the threshold, separating it from another subgroup where the survival gain is smaller, or even nonexistent: patients are not helped or even harmed by the treatment. The first group benefits from treatment – the gains exceed the threshold – while the second group does not.

A marker can then act as a treatment selection marker if there is a value, or a range of values, that corresponds to a group who benefits, and the remaining values correspond to a group that does not benefit.

What then if the personal treatment thresholds vary? In that case we have to generalize the second condition, over the distribution of values for the treatment threshold. The

marker may be able to act as a treatment selection marker for some, but not for all. If it can act as a marker for at least one (group of) patients, in the sense we just described, then it can be qualified as a (potential) treatment selection marker.

To further facilitate presentation of concepts and performance measures we describe a clinical decision scenario with a potential treatment selection marker and discuss which measures do and which ones do not measure the performance of the marker for guiding treatment decision.

As an example, we consider using vaginal culture in women with preterm premature rupture of membranes to guide the decision for immediate delivery. In pregnant women in whom rupture of membranes occurs prematurely and before the onset of labour, a decision dilemma is whether to follow a strategy of wait-and-see or to perform immediate delivery to prevent infection and sepsis in the foetus. Bacterial infection causing neonatal sepsis is most commonly associated with the Group B *streptococcus* (GBS) from the mother's vagina. Therefore testing the vaginal GBS colonisation in mothers could potentially identify foetuses at higher risk of infection and may be a good candidate marker for guiding the decision for immediate delivery.

In a trial, about 700 women with premature rupture of membranes were randomly assigned to immediate delivery or wait-and-see strategy (7, 8). Among the women studied, 14% had GBS-colonization and were marker-positive. Table 1 shows the association between the GBS-colonization and the outcome in the trial participants (9).

It may seem that we could quantify the performance of a treatment selection marker with the usual measures of clinical performance: why not use sensitivity or specificity here? Indeed,

Table 1 GBS - colonization and outcomes (9)

Strategy	Patients with neonatal sepsis	% of total	Patients without neonatal sepsis	Total patients
Wait-and-see				
GBS Colonization	7	15.2%	39	46
No GBS Colonization	8	2.6%	305	313
Total	15	4.2%	344	359
Immediate delivery				
GBS Colonization	1	1.8%	56	57
No GBS Colonization	9	2.9%	297	306
Total	10	2.8%	353	363

we could do so, but only if there was a straight-forward clinical reference standard to identify with sufficient certainty those who benefited (sufficiently) from treatment, separating those from the rest, who did not. In that case the sensitivity of the treatment selection marker would be the proportion of those who benefited, correctly identified as such by the marker, and the specificity would be the proportion of those who did not benefit, correctly identified as such by the marker.

Unfortunately, this distinction is less easy to make on an individual basis in most treatment studies, where only the outcome under treatment is observed, or the outcome under the absence of treatment. It requires a counterfactual approach then to specify what would have happened with an alternative course of action. Below we will describe how we can use the information from the group of trial participants to evaluate the performance of a putative treatment selection marker.

PERFORMANCE OF TREATMENT SELECTION MARKERS

We have just described the necessary conditions for a marker to act as a treatment selection marker. These are absolute conditions: a marker either is or is not a (potential) treatment selection marker. Yet to make decisions about the actual use of the marker, a more quantitative estimate of its performance is required.

Janes and colleagues have explored a number of statistics to express biomarker performance, with descriptive and inferential methods to evaluate individual markers and to compare candidate markers (10, 11). They proposed useful measures for analyzing marker performance. By combining them they calculate the population benefit from using the marker as a treatment selection marker, compared to a strategy of not using the marker to decide about treatment in subgroups of patients.

We use our clinical example in Table 1 to present these measures.

Proportion of marker-positives

First we turn to the subgroup of patients who are marker-positive, in our example women with GBS colonization. GBS-positive women comprised 14% of women participating in the trial: 103 out of 722 (Table 1). So the proportion of patients in whom treatment recommendations could change following marker measurement is 0.14.

Average benefit of treatment among marker-positives

If marker-positive women receive a wait-and-see strategy, 15.2% of their neonates will develop neonatal sepsis. In contrast, when undergoing immediate delivery only 1.8% of their neonates will develop sepsis. Immediate delivery will therefore result in a reduction of 13.5% in the neonatal sepsis rate in this group: this is the average benefit of intervention in this subgroup.

Change in population event rate with marker-based treatment

This is the main composite measure of marker performance for treatment selection. It is based on the difference in overall outcome between not using the marker and using the marker for treatment decisions, aggregated over all members of the target population. Based on marker status, we will only treat marker positives, so the expected change can be calculated by multiplying the proportion of marker-positives (0.14) with the average benefit of treatment in marker-positives (13.5%): $(0.14 \times 13.5\%) = 1.9\%$.

In other words, a strategy in which immediate delivery is only considered for marker positives will lead to an absolute decrease of 1.9% in the neonatal sepsis rate, compared to a wait-and-see strategy for all.

The impressive reduction in the neonatal sepsis rate in the GSB positives (minus 13.5%) may

look like an adequate expression of marker performance, but it is quite clear that the prevalence of the marker positives should also be included in the evaluation.

The result is a clinically interpretable measure of performance of GBS testing for treatment selection. It evaluates the treatment selection marker in terms of its clinical effectiveness: its ability to lower the number of adverse events in the study population (12). With the same approach one can calculate the impact of application of GBS-based strategy on other outcomes such as cost of care or rate of premature birth to complete an evaluation of the costs and consequences of the marker-based strategy.

In our example we did not discuss chance variability. Janes and colleagues have described methods for statistical inference and hypothesis testing (11). They suggest that the performance measures are only estimated if a null hypothesis corresponding to no marker performance is rejected.

This approach assumed that the marker only acts as a selection mechanism, and that, in itself, it does not lead to the event one tries to prevent. It only does so by guiding treatment. We also assume that the effectiveness of the treatment itself is not affected by knowing the marker status. This could happen with some strategies, for example, through better adherence or a different way of handling side-effects. If these assumptions do not hold, the only way to evaluate the effectiveness of a marker-based strategy would be a randomized trial, allocating eligible participants, to this marker-based strategy or to an alternative: no treatment in all.

By further extending this approach, Huang and colleagues define an extension of the net benefit measure: expected benefit. This measure expresses the reduction in the sum of disease and treatment cost by using the marker, based

on the comparison between a marker-based treatment-selection rule and the optimal treatment strategy without the marker information (13).

PREDICTIVE AND PROGNOSTIC MARKERS

In the oncology literature, the terms predictive and prognostic markers have increasingly been used within the context of stratified or personalized medicine, but their use has been somewhat confusing. Some have stipulated, for example, that predictive markers are associated with drug response, in contrast with prognostic markers, which are associated with disease outcome (14). We have shown that is not so much the association with outcome or drug response that counts, but the ability to separate groups who benefit – with difference in outcome compared to the threshold – from those who do not.

In this relatively young field, several other metrics and statistics have been proposed to express the performance of treatment selection markers. Some of these can be severely misleading, since they cannot provide evidence that a marker is helpful in guiding treatment decisions.

These questionable measures include expressions of the strength of the association between marker status and outcome, not benefit. In Table 1, for example, one can see that marker positives have a six-fold higher risk of neonatal sepsis under a wait-and-see strategy. With a strategy of immediate delivery, the relative risk is 0.6.

Both relative risks give information about the association between GBS colonization and outcome, but in themselves they do not reflect marker performance. Treatment decisions should not be guided by outcome in itself, but by benefit: the expected *change* in outcome produced by the treatment.

CONCLUDING COMMENTS

It is exciting to see the developments resulting from rapid progress in our understanding of molecular processes. Biomarkers and other forms of medical tests are not only used for making a diagnosis or staging a disease, but for many other purposes, including decisions about treatment. To express the performance of such treatment selection markers, and to see whether they can actually be used for this purpose, we need a different set of measures. The classical clinical performance measures, such as clinical sensitivity and specificity, can only be used in rare circumstances. Relying on familiar statistics, such as relative risks, or simple significance tests, may actually be misleading. Like all other forms of medical testing, potential treatment selection markers should be properly evaluated before they can be implemented in daily clinical practice.

REFERENCES

1. Hislop J, Adewuyi TE, Vale LD, Harrild K, Fraser C, Gurning T, et al. Methods for specifying the target difference in a randomised controlled trial: the Difference ELicitation in TriAls (DELTA) systematic review. *PLoS Med* 2014; 11: e1001645.
2. Blinman P, McLachlan SA, Nowak AK, Duric VM, Brown C, Wright G, et al. Lung cancer clinicians' preferences for adjuvant chemotherapy in non-small-cell lung cancer: what makes it worthwhile? *Lung Cancer* 2011; 72: 213-8.
3. Blinman P, Duric V, Nowak AK, Beale P, Clarke S, Briscoe K, et al. Adjuvant chemotherapy for early colon cancer: what survival benefits make it worthwhile? *Eur J Cancer* 2010; 46: 1800-7.
4. Duric VM, Butow PN, Sharpe L, Heritier S, Boyle F, Beith J, et al. Comparing patients' and their partners' preferences for adjuvant chemotherapy in early breast cancer. *Patient Educ Couns* 2008; 72: 239-45.
5. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980; 302: 1109-17.
6. Pauker SG, Kassirer JP. Clinical application of decision analysis: a detailed illustration. *Semin Nucl Med* 1978; 8: 324-35.

7. Van der Ham DP, van der Heyden JL, Opmeer BC, Mulder AL, Moonen RM, van Beek JH, et al. Management of late-preterm premature rupture of membranes: the PPROMEXIL-2 trial. *Am J Obstet Gynecol* 2012; 207: 276 e1-10.
8. Van der Ham DP, Vijgen SM, Nijhuis JG, van Beek JJ, Opmeer BC, Mulder AL, et al. Induction of labor versus expectant management in women with preterm prelabor rupture of membranes between 34 and 37 weeks: a randomized controlled trial. *PLoS Med* 2012; 9: e1001208.
9. Tajik P, van der Ham DP, Zafarmand MH, Hof MH, Morris J, Franssen MT, et al. Using vaginal Group B Streptococcus colonisation in women with preterm premature rupture of membranes to guide the decision for immediate delivery: a secondary analysis of the PPROMEXIL trials. *BJOG* 2014; 121: 1263-72.
10. Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. *Ann Intern Med* 2011; 154: 253-9.
11. Janes H, Brown MD, Huang Y, Pepe MS. An approach to evaluating and comparing biomarkers for patient treatment selection. *Int J Biostatistics* 2014; 10: 99-121.
12. Horvath AR, Lord SJ, StJohn A, Sandberg S, Cobbaert CM, Lorenz S, et al. From biomarkers to medical tests: the changing landscape of test evaluation. *Clin Chim Acta* 2014; 427: 49-57.
13. Huang Y, Laber EB, Janes H. Characterizing expected benefits of biomarkers in treatment selection. *Biostatistics* 2014 doi: 10.1093/biostatistics/kxu039
14. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* 2005; 23: 2020-7.

Book review: “A Practical Guide to ISO 15189 in Laboratory Medicine”

Joseph B. Lopez

MAHSA University, Kuala Lumpur, Malaysia

REVIEWED BOOK

“A Practical Guide to ISO 15189 in Laboratory Medicine”

by David Burnett

Publisher: ACB Venture Publications, 2013

ISBN 978-0-902429-49-9

Corresponding author:

Joseph B. Lopez

MAHSA University

Kuala Lumpur, Malaysia

RECESSION

Laboratory accreditation is still optional in many countries but it is increasingly the path that many have volunteered to take. The ultimate beneficiary of accreditation is the patient who is the ultimate end-user of the laboratory, since the reliability of results is assured with it. Reliable results in turn contribute to patient safety and effective treatment.

This book is yet another contribution from that well known authority on laboratory accreditation, David Burnett and is a successor to his previous publications, “Understanding Accreditation in Laboratory Medicine” in 1996 and, “A Practical Guide to Accreditation in Laboratory Medicine” in 2002. The author has indicated that this will be his “third and final book”. It is primarily intended as a guide for laboratory professionals seeking to implement or renew accreditation that uses the ISO 15189:2012 standard. It should also be useful to auditors working for accreditation bodies. While other standards for the clinical laboratory, exist the ISO 15189 has been recommended by the IFCC.

This easily readable book is more than a manual for accreditation. The opening chapter is a description of the ISO 15189 standard and this is followed by one on defining and managing quality in the medical laboratory. In subsequent chapters due attention is given to organisation and management, personnel, accommodation, environmental conditions and safety.

The actual testing process is covered in three chapters, beginning with a chapter on equipment, reagents, consumables and external services. This is followed by chapters on pre- and post-examination processes and the quality of examination results. The book is replete with diagrams and tables which makes it easy to read. However, though pre-analytical errors may constitute up to 75% of laboratory errors (1), they receive a scant 18 pages of attention.

There are three useful appendices. The first is on “The ‘Ideal Standard’ and ISO 15189:2012” which is an index to the clauses of these standards and the chapters in which they are discussed. The next appendix provides some sample pages from the Quality Manual of a fictional hospital while the last appendix is a bibliography of for each chapter of the book and includes several online references.

The book is comprehensive in its coverage of the subject. Yet one is left to wonder if more could have been done. Though it discusses hazards and risk management in terms of the immediate environmental and personnel of the laboratory, the book does not address the wider impact to the environment of the laboratory’s

activities. While this topic may not be within the remit of the ISO 15189 standard, Burnett would have done us a great service had he devoted a small section to the environmental impact of laboratories and how it could be incorporated into overall quality management. This is but a suggestion and is not meant to be a criticism of what is otherwise a very useful textbook on accreditation.

A check with Google reveals that there are not many books on the ISO 15189 standard. It is usually a challenge for those new to accreditation to understand and interpret the technical jargon of manuals (the more familiar word “test”, for example, is replaced with the all-encompassing “examination” which is usually used for histopathology) which require subtlety and appreciation of nuance. Hence, there is a need for clearly written guides for the implementation of this standard. This book is an answer to that need.

REFERENCE

1. Plebani M. Quality Indicators to Detect Pre-Analytical Errors in Laboratory Testing. Clin Biochem Rev. 2012; 33: 85–88.

Book review

“Practical Clinical Chemistry: Core Concepts”

Barna Vasarhelyi

Department of Laboratory Medicine, Semmelweis University, Budapest, Hungary

REVIEWED BOOK

***“Practical Clinical Chemistry:
Core Concepts”***

by Tahir Pillay, MBChB, PhD, FRCPath

Corresponding author:

Prof. Barna Vasarhelyi, MD, PhD, DSc
Chair of Laboratory Medicine
Department of Laboratory Medicine
Semmelweis University
Budapest, Hungary
vasarhelyi.barna@med.semmelweis-univ.hu

RECENSION

Nowadays, the majority of clinical laboratories operate as high throughput factories. Financial reasons and quality control issues both justify the widespread application of automated systems and ready-to-use reagents. While technical evolution lead to an exceptional improvement in the productivity of clinical labs and the generation of high quality data, on the other hand, however, we have to pay the price of this development. This price is particularly not one of an economic nature; it is rather some loss of the miracles of the *in vitro* world that attracted older generations – assistants, chemists, laboratory doctors and other professionals – to the lab.

Once these miracles had included a quite range of colorful reactions reflecting the successful compilation of reagents; the dramatically increasing absorbance values as the result of the functioning of a vivid serum enzyme in the tube; or the appearance of stripes on an electrophoretic gel after a fierce overnight work to isolate proteins. Unfortunately for many, these miracles were switched by such simple processes as ‘push the button’, ‘load the system’, ‘save’ or ‘print’ the results etc. and all the lab work was getting to resemble to a black box that generates somehow results from samples. And, while all the results may completely fit to external and internal QC programs, the personal contribution of professionals gets to become minimal. (In line with this trend I also noticed in a personal pilot survey that an almost negligible minority of labs working in the field of

clinical chemistry use routinely scales, pH-measurement device, spectrophotometers in a university environment.) Finally, there is a new generation already working in clinical chemistry 'factories' that has less and less personal experience with in vitro world.

The book "Practical Clinical Chemistry: Core Concepts" may be a tool to bring back some of the essence of the good old times for students and even for younger colleagues. Its structure, problem-based approach and well-designed practical demonstrations in each chapter are efficient tools to introduce the beauties and also the challenges of clinical chemistry methods for the interested readers. Although the size of this book is limited to 126 pages, being presented in 14 chapters gives a very impressive overview of clinical chemistry lab work.

The first page of each chapter defines clearly the objectives of the knowledge transfer. Then, the chapter is divided on two major parts. Yellow color-coded pages indicate the practical tasks for students to perform, while light blue pages contain instructor's guide along with the explanation of observation and, also, a condensed explanation of clinical usefulness of results. Clear and high-quality figures and photos are also incorporated in a justified amount that supports the understanding of the described material. This design and structure largely support the professional consistency of the book and help the readers to orientate themselves easily.

This fascinating book provides a well-designed and didactic way that covers almost all the major fields that are 70-100 per cent automated in our days. Chapters let a short insight into basic laboratory practices; present the difference between end-point and kinetic assays; indicate basic principles of protein assays and immunoassays etc. The readers can learn basic ideas behind immunoassays, electrophoresis of proteins. A specific chapter is also devoted to introduce some very basic molecular biology methods. After the successful completion of presented practical tasks and learning the attached explanations the students will be understand basic mechanisms leading to the generation of laboratory results. With this knowledge they may be able to look critically at data generated with the use of automated systems and, possibly, to intervene in case of urgency.

I am convinced that this novel book will largely support the graduate and post-graduate education in clinical chemistry. One minor limitation with this educational tool is that currently iPhone/iPad/Mac is the only medium where it is available; therefore, a significant part of the clinical chemistry community having no iPhone still has no chance to obtain it. I hope that observed positive experience will foster producers to extend this book for PC or other systems and make it more widely available.



Editor-in-chief

Gábor L. Kovács

Institute of Laboratory Medicine, Faculty of Medicine,
University of Pécs, Hungary

Editorial Board

Khosrow Adeli, The Hospital for Sick Children, University of Toronto, Canada

Borut Božič, University Medical Center, Ljubljana, Slovenia

Rajiv Erasmus, Dept. of Chemical Pathology, Tygerberg, South Africa

Nilda E. Fink, Universidad Nacional de La Plata, La Plata, Argentina

Ellis Jacobs, New York University School of Medicine, New York, USA

Bruce Jordan, Roche Diagnostics, Rotkreuz, Switzerland

Evelyn Koay, National University, Singapore

Maria D. Pasic, Laboratory Medicine and Pathobiology, University of Toronto, Canada

Oliver Racz, University of Kosice, Slovakia

Rosa Sierra Amor, Laboratorio Laquims, Veracruz, Mexico

Sanja Stankovic, Institute of Medical Biochemistry, Clinical Center of Serbia, Belgrade, Serbia

Danyal Syed, Ryancenter, New York, USA

Grazyna Sypniewska, Collegium Medicum, NC University, Bydgoszcz, Poland

Istvan Vermes, University of Twente, The Netherlands

Stacy E. Walz, Arkansas State University, USA



Publisher: IFCC Communications and Publications Division (IFCC-CPD)

Copyright © 2015 IFCC. All rights reserved.

The eJIFCC is a member of the **Committee on Publication Ethics (COPE)**.

The eJIFCC (Journal of the International Federation of Clinical Chemistry) is an electronic journal with frequent updates on its home page. Our articles, debates, reviews and editorials are addressed to clinical laboratorians. Besides offering original scientific thought in our featured columns, we provide pointers to quality resources on the World Wide Web.

Contents may not be reproduced without the prior permission of the Communications and Publications Division (CPD) of the IFCC.

Produced by:

 **Insoft Digital**
Web Solutions

www.insoftdigital.com

Published by:


International Federation
of Clinical Chemistry
and Laboratory Medicine

www.ifcc.org