

July 2015

ISSN 1650-3414

Volume 26 Number 3

eJIFCC

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: ejifcc@ifcc.org

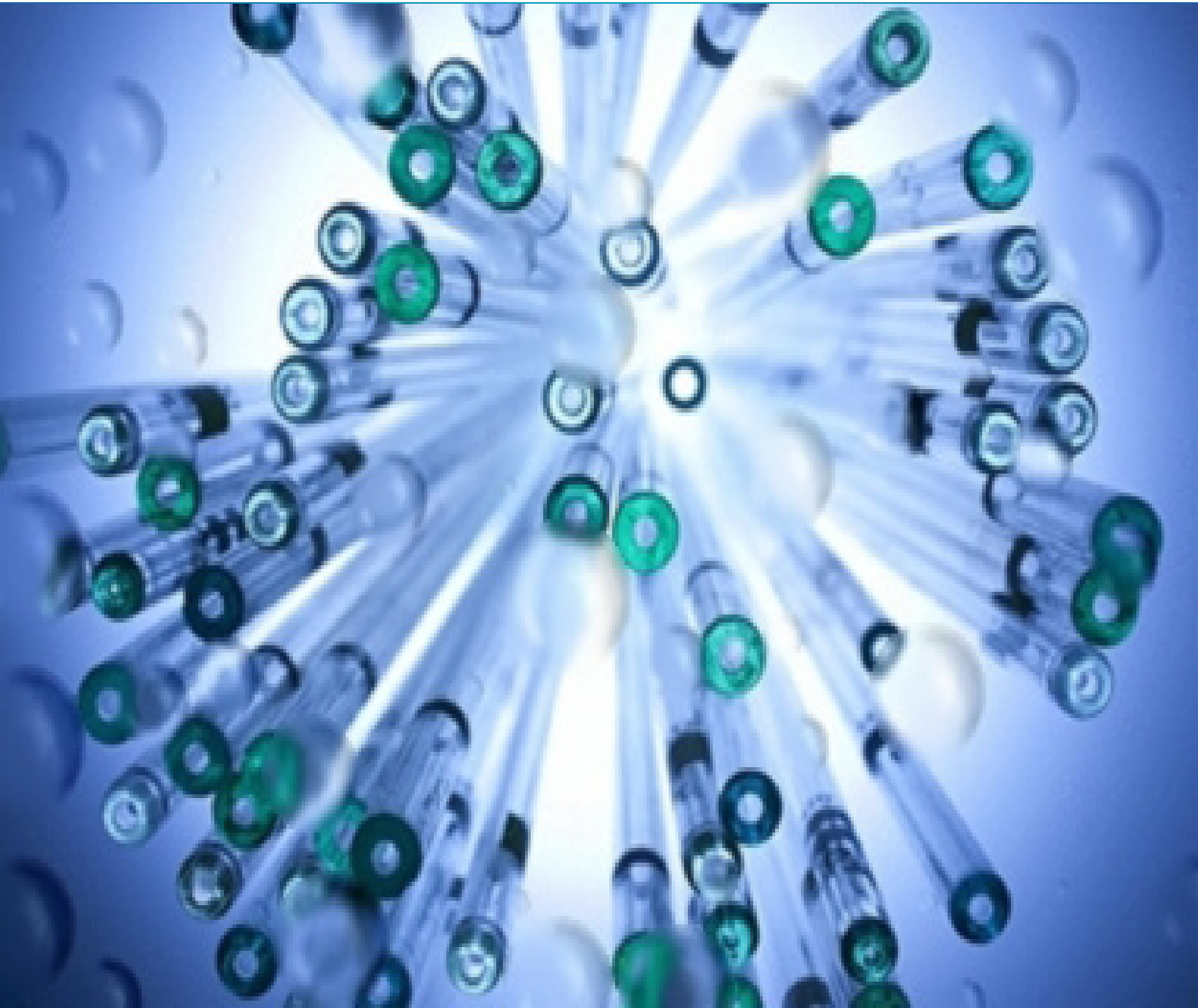
The
Journal of the
International
Federation of
Clinical
Chemistry and
Laboratory
Medicine





IFCC

*International Federation
of Clinical Chemistry
and Laboratory Medicine*



In this issue

Foreword of the editor

Gábor L. Kovács

144

Are guidelines guiding us on how to utilize laboratory tests?

Andrea R. Horvath

146

Defining the path forward: guidance for laboratory medicine guidelines

Stephen E. Kahn, Patricia M. Jones, Alex C. Chin, Robert H. Christenson

158

Grading evidence for laboratory test studies beyond diagnostic accuracy: application to prognostic testing

Andrew C. Don-Wauchope, Pasqualina L. Santaguida

168

Evidence and cost effectiveness requirements for recommending new biomarkers

Paul Collinson

183

How useful are laboratory practice guidelines?

S. Misra, K. Moberg-Aakre, M. Langlois, J. Watine, P.J. Twomey, W.P. Oosterhuis, J.H. Barth **190**

Letter to the Editor — Working terms

Xavier Fuentes-Arderiu

197

Foreword of the editor

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

The current issue of eJIFCC is devoted to the role of guidelines in laboratory medicine. The guest editor of this issue is Professor Andrea Rita Horvath MD, PhD, FRCPath, FRCPA from Australia, an internationally renowned expert of the field. Dr. Horvath has been Clinical Chemistry Network Director of the South Eastern Area Laboratory Services, based at the Prince of Wales Hospital in Sydney since December 2009. She is Conjoint Professor of Medicine at the University of New South Wales and Honorary Professor at the School of Public Health of Sydney University, and at the Faculty of Pharmacology and Biochemistry of Buenos Aires University.

Dr. Horvath obtained her M.D. (1984), her National Board Certification in Clinical Chemistry (1991), and her Ph.D. (1993) from the University of Debrecen in Hungary. She attained Membership by examination (1994) and Fellowships of the Royal College of Pathologists of England (2001) and Australasia (2011). She spent 8 years in Britain: first as postgraduate research fellow in London (1988-1990); later as chemical pathology Registrar in Sheffield (1993-1994); and Senior Registrar and Lecturer in Clinical Biochemistry at Oxford University (1995-1998). She was Professor and Head of Clinical Chemistry at the University of Szeged in Hungary for 11 years (1998-2009).

Rita's key research interest is evidence-based laboratory medicine (EBLM) and evidence-based guideline development. Supported by British Government grants she successfully established an Evidence-Based Medicine Network in Hungary (TUDOR) to teach EBM and advise government in evidence-based health and reimbursement policy and guideline development programs. She currently advises the National Prescribing Service of Australia on test utilization. She has been advising the National Academy of Clinical Biochemistry (NACB) on evidence-based guideline methodology since 2005 and has been involved in NACB's and CLSI's guideline groups. She is currently the Vice Chair of AACC's EBLM Committee.

Her awards include: KoneLab Award, Association of Clinical Biochemists, UK (2003); Per Hyltoft Petersen Award for Distinguished Medical Biopathologist (2006); IFCC Visiting Lecturer Awards: Israel (2004), India, Malaysia and Indonesia (2006), and China, Hong Kong, Taiwan, Thailand and Singapore (2007); Mentor of the Month of AACC (November 2011); Lorand Jendrassik Medal of the Hungarian Society of Laboratory Medicine, the Markusovszky Award for best publication in the Hungarian Medical Journal (2012) and the award of the Croatian Society of Clinical Biochemistry (2013).

Rita has held a number of national and international leadership positions: Chair of the IFCC Committee on EBLM (2003-2008); European Communities Confederation of Clinical Chemistry and Laboratory Medicine Secretary (2005-2007); President of the Hungarian Society (2005-2008) and the Hungarian College of Laboratory Medicine (2008-2009); President-Elect (2007-2009), President (2009-2011) and Past President (2012-2013) of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM). Currently she serves on

AACC's Board of Directors, as member of the board, and she chairs EFLM's Test Evaluation Working Group.

As editor of the eJIFCC, I am personally proud of the fact that while Rita worked in Hungary, we held two of the three chairs in laboratory medicine in the country. We had a lot of professional and scientific interactions. Her involvement in the regulatory affairs of laboratory medicine was absolutely significant for the progress of our discipline. I am very pleased to introduce her as guest editor.

Are guidelines guiding us on how to utilize laboratory tests?

Guest Editor: Andrea R. Horvath

SEALS Department of Clinical Chemistry and Endocrinology, Prince of Wales Hospital and School of Medical Sciences, University of New South Wales, Australia

Screening and Test Evaluation Program, School of Public Health, University of Sydney, Australia

ARTICLE INFO

Corresponding author:

Andrea Rita Horvath
SEALS North, Department of Clinical
Chemistry and Endocrinology
Level 4, Campus Centre
Prince of Wales Hospital
Barker Street, Randwick
NSW 2031, Sydney, Australia
Phone: (+61) 404 027 843
E-mail: rita.horvath@sesiahs.health.nsw.gov.au

Key words:

evidence-based laboratory medicine,
guideline, outcome, test utilisation

Disclaimer:

This article is based on a lecture given
at IFCC Worldlab 2014 in Istanbul.

ABSTRACT

Increasing patient risks and costs associated with the delivery of health care services have been related to inappropriate and uncontrolled use of biomarkers which make evidence-based guideline recommendations for best practice increasingly important. The translation of basic scientific discoveries into clinically meaningful studies and then to evidence-based clinical practice guidelines (CPGs) or health policy is, however, not straightforward. CPGs are potentially the most influential publications as they aim to guide clinical decisions and impact patient outcomes; hence, current approaches to their development often fail scientific publication standards. Critical appraisal of CPGs has revealed that many do not involve laboratory professionals in formulating recommendations on the use of tests; the composition of the panel could influence the scope of guidelines and over-represent certain stakeholders' views; numerous CPGs do not have rigorous evidence-based methodology and miss essential information important for the correct interpretation and application of laboratory results.

Numerous CPGs are released on similar topics worldwide, but their quality and content validity are highly variable and their recommendations may differ even when using the same sources of evidence. This can be due to the limitations of the evidence base, or to the lack of agreed test evaluation methods and easy-to-use evidence rating schemes that could be universally adapted to diagnostic recommendations. Furthermore, value-based judgments on the balance between benefits, harms, risks, patients' preferences and the organizational and financial aspects of care may differ among countries and regions. Addressing these issues requires careful discussions and consensus between relevant multidisciplinary stakeholders involved in the diagnosis and management of health conditions.

INTRODUCTION

Increasing patient risks and costs associated with the delivery of health care services have been related to inappropriate and uncontrolled use of both diagnostic and therapeutic interventions which make evidence-based guideline recommendations for best clinical practice increasingly important. According to the BEACH study in Australia, general practitioners have the greatest difficulty with test ordering and test interpretations for conditions/symptoms that are vague, and/or where there are no guidelines or decision support systems to guide their practice. The least difficulty was reported for conditions such as diabetes, lipids, urinary tract infections where clear management recommendations have been available (1). In response to these needs numerous clinical practice guidelines (CPGs) are released on similar topics worldwide. However, the translation of basic scientific discoveries into clinically meaningful studies and then distillation of study findings into evidence-based

practice recommendations or health policy are not straightforward and pose many methodological and implementation challenges.

WHY DO WE NEED GUIDELINES?

Guidelines are systematically developed statements that assist health care professionals and patients in making decisions about *appropriate* health care in specific clinical circumstances (2). Guidelines aim to:

- disseminate best practice based on scientific evidence;
- decrease practice variation and the potential or frequency of professional misconduct;
- improve patient safety;
- improve the quality and effectiveness of care;
- improve cost-effectiveness of care;
- facilitate training, education and continuous professional development;
- increase explicitness, transparency, patient information and autonomy of choice.

In the context of laboratory medicine, guidelines aim to improve the *appropriateness* of test utilization (i.e. test requesting and interpretation) by (3):

- promoting the use of new tests if evidence proves their **efficacy and effectiveness** – *start starting or stop stopping*
- eliminating poor or useless tests before they become widely available – *stop starting*
- removing old tests with no proven benefit from practice – *start stopping (adapted from 4)*.

Appropriateness in this context refers to care that results in more benefits than harms at

reasonable costs. For example, we have strong evidence from randomised controlled trials that screening with either faecal occult blood testing (FOBT) or sigmoidoscopy decreases the mortality of colorectal cancer (CRC) by 14%-16% and if the cancer is detected at an early localized stage, the 5-year survival rate is 90% (5). As a result, evidence-based recommendations have been issued by a number of guideline organisations and national screening programs have been initiated in many developed countries. For example the United States Preventive Services Task Force (USPSTF) recommends 1) high-sensitivity faecal occult blood testing annually, 2) colonoscopy every 10 years, or 3) sigmoidoscopy every 5 years with FOBT every 3 years for the prevention or early detection of CRC among adults aged 50–75 years (6). Due to these recommendations, the percentage of the U.S. population compliant with recommended CRC screening increased from 54% in 2002 to 65% in 2010 and stayed at the same rate by 2012, primarily through increased use of colonoscopy. To further improve clinical outcomes through the uptake of CRC screening, the CDC introduced more aggressive population-based strategies and set the target for 2014 at 80% (6). The European guidelines issued in 2013 still consider sigmoidoscopy and colonoscopy as a supplement or alternative for CRC screening (5). In the European Union (EU) in 2007 the Council Recommendation for CRC screening targeted approximately 136 million women and men in the age group of 50-74 years primarily by FOBT testing. In 2007 less than 10% of the targeted EU population (approximately 12 million) has taken part in CRC screening and 94% of those were tested by FOBT and the rest by flexible sigmoidoscopy or endoscopy (7).

McDowell et al. published a systematic review of 19 hypertension guidelines, issued between

2001 and 2011 with recommendations for monitoring for adverse drug reactions using biochemical tests in patients taking antihypertensive treatment. They found that guidelines were lacking any evidence behind advice on frequency of biochemical monitoring and both the instructions for monitoring and the extent of advice for subsequent action differed greatly and that such poorly specified recommendations were challenging for clinicians to apply in clinical practice (8).

Clinicians face even more challenges when guideline recommendations are not just vague or diverse but even conflicting. Examples of such confusion are conflicting recommendations for PSA screening from different professional organisations; e.g. in USA the USPSTF recommends against PSA screening to detect prostate cancer, whilst the American Cancer Society and the American Urological Association and many other European cancer societies recommend that patients willing to be screened discuss their options with their physician (9). Another recent area of controversy is related to the screening, diagnosis and management of gestational diabetes including debates about the merits of screening *versus* no screening, universal *versus* selective screening of high risk cases, timing and methods and cut-off glucose values used for defining the condition, and long-term management options for those who have the diagnosis (10,11). These examples illustrate the diversity and complexity of guideline development and implementation even when the same evidence base is available to guide best clinical practice and national policy.

VARIATIONS AND DIVERSITY IN GUIDELINE DEVELOPMENT AND IMPLEMENTATION PRACTICES

Numerous studies have demonstrated that the quality and content validity of guidelines are

highly variable (12-15). This is particularly true for diagnostic recommendations where the evidence base is more limited than in the field of therapeutics. These shortcomings are due to large variations in the analytical and clinical performance of laboratory methods for the same analytes, the lack of agreed test evaluation methods and easy-to-use evidence rating schemes that could be universally adapted to diagnostic recommendations. A recent review identified 12 evidence grading systems that addressed diagnostic testing. Out of these, 5 systems provided varying degree of coverage of the essential items for evidence gathering, review, assessment and linkage to recommendations. However, no single system covers all aspects and supports guideline developers in rating the strength of evidence behind recommendations for the use of laboratory tests (16). To add to the complexity, value-based judgments on the balance between benefits, harms, risks, patients' preferences and the organizational and financial or resource aspects of care may differ among countries and regions and therefore could influence the final recommendation and its grading. Addressing these issues requires a transparent, well-structured and documented process including careful discussions and consensus between relevant multidisciplinary stakeholders involved in the diagnosis and management of health conditions (for more details on grading, see the paper by Don-Wauchope et al., in this issue).

Clinical practice guidelines are potentially the most influential publications as they aim to guide clinical decisions and impact patient outcomes; hence, current approaches to CPG development are often non-systematic, lack clear organisational structure or legislative background and fail the methodological rigour of scientific publication standards. The most widely used critical appraisal tool for assessing the

methodological quality of CPGs is the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument (17). Numerous studies using the AGREE tool on various guideline topics issued by various organisations pointed to significant inconsistencies in terms of best practice recommendations provided to clinicians across the globe which may have an impact on the quality of care provided to patients. Moreover, the findings consistently showed that the least well-addressed domains within the AGREE tool were stakeholder involvement, rigour of development, applicability and editorial independence of the guideline development process (13-15,18). Critical appraisal by the AGREE tool of CPGs primarily addressing laboratory testing in various conditions has similarly revealed that many do not involve laboratory professionals; the composition of the panel could influence the scope of guidelines and over-represent certain stakeholders' views; and miss essential information important for the correct interpretation and application of test results (13,15).

The above factors easily explain why the European Observatory on Health Systems and Policies in its 2013 report also found divergent national guideline development and implementation programs in the EU (18). In addition to the quality of guideline methods, this report investigated the organisational and regulatory framework, the implementation and impact of guidelines developed for chronic non-communicable diseases such as coronary heart disease, chronic obstructive pulmonary disease, asthma, type 2 diabetes mellitus, osteoarthritis, breast cancer, cervical cancer, colorectal cancer and depressive disorders that are responsible for 70-80% of health care costs in the EU. Key findings of this report are listed below (18):

- Regulatory frameworks exist in most EU states for clinical guideline use but relevant laws are not always implemented.

- There is no obvious link between the availability of legislative frameworks and the quality and impact of guideline programs.
- Guidelines are usually developed by government and professional organizations or adopted/adapted from external sources.
- The engagement of multidisciplinary stakeholders in guideline development varies but patients and users of health services are rarely involved in the development of CPGs.
- Few organizations have quality control processes for their guidelines but if they do, they often use the AGREE instrument.

DO WE NEED GUIDELINES FOR MAKING GUIDELINES?

The mentioned shortcomings of guideline development programs are not unique to Europe and call for guidelines for developing guidelines and an assessment of the internal and external validity of recommendations before their implementation is attempted. Numerous government organisations issuing CPGs have guideline development manuals. The so-called GRADE and DECIDE project group systematically reviewed the available guideline development resources and assembled a checklist with 18 topics and 146 items in order to facilitate the standardisation of all stages of the guideline development process. The group provides an interactive webpage (<http://cebgrade.mcmaster.ca/guidecheck.html>) with links to training materials and resources for applying the checklist items (19).

The Institute of Medicine (IOM) has also issued a report entitled “Clinical practice guidelines we can trust”, in order to provide a set of standards for ensuring that guidelines present trustworthy and implementable recommendations (20). Table 1 summarises the IOM standards and their relevance to guideline development on laboratory testing.

In the field of laboratory medicine, the National Academy of Clinical Biochemistry (NACB) of the American Association of Clinical Chemistry (AACC) is a well-recognised source of guidelines. The NACB has recently updated its standard operating procedure for developing laboratory medicine practice guidelines that are based on more systematically gathered evidence (for more details see the paper by Kahn et al., in this issue). Table 2 summarises the main sources of guideline development tools that are relevant to laboratory medicine.

DO GUIDELINES IMPACT CLINICAL PRACTICE AND PATIENT OUTCOMES?

The previously mentioned EU Report also investigated the implementation and impact of guidelines for the management of the most prevalent chronic conditions. They found only two studies that reported effective guideline implementation or impact; five studies showed “partial effectiveness” and three studies did not demonstrate any effectiveness. The BEACH study carried out in Australian general practices investigated pathology test requesting and estimated that 3.1 million tests were reported for Type2 diabetes patients between 2006 and 2008. Seventy two percent of these tests were supported by guideline recommendations, 12.4% were in the grey zone due to unclear guidance and 10.1% were not supported by guidelines (1).

These examples, along with many similar observations published in the literature about guideline implementation, point to the fact that it is not sufficient to develop good evidence-based guidelines and passively disseminate them. Successful translation of the evidence into practice requires a system approach which starts with acknowledging existing gaps in clinical practice and recognising the need for a change, followed by a search for and implementation of

Table 1 Relevance of guideline development standards to laboratory testing-related recommendations

IOM Standard	Explanation	Additional notes relevant to guidelines on laboratory testing
1. Transparency	The guideline development process and its source of funding must be transparent and public	<i>See paper by Kahn et al. in this issue.</i>
2. Conflict of interest	Before guideline panels are established all conflicts of interest must be declared. Chairs and co-chairs should be free from conflicts of interest. Funders of CPGs should not influence the content of the guideline	<i>See paper by Kahn et al. in this issue.</i>
3. Guideline development group composition	Guideline panels should be multidisciplinary involving all key stakeholders targeted by the CPG and methodologists. Patient and consumer involvement should be encouraged.	The involvement of professionals in laboratory medicine should be facilitated in CPG panels where recommendations involve laboratory testing. <i>See paper by Kahn et al. in this issue.</i>
4. Systematic review of the evidence	Guidelines should be based on systematic reviews that meet methodological standards.	Laboratory professionals should be engaged in systemic reviews of diagnostic tests. For recommended tools and checklists see text and the Cochrane DTA and EQUATOR websites.
5. Evidence foundations for and rating the strength of recommendations	<p>Recommendations should have reasoning with clear description of potential benefits and harms and a summary of the evidence behind them.</p> <p>The strength of evidence and the strength of recommendation must be rated.</p> <p>Differences of opinions must be explicitly stated.</p>	<p>The GRADE diagnostic tool is recommended for rating the strength of evidence and the strength of recommendations related to testing.</p> <p><i>See paper by Don-Wauchope et al. in this issue.</i></p>

<p>6. Articulation of recommendation</p>	<p>Recommendations must be clear and unambiguous. Strong recommendations should be worded to allow evaluation of compliance.</p>	<p>Recommendations on laboratory testing should consider covering essential items relevant to the correct use and interpretation of laboratory tests (28)</p> <p><i>See paper by Misra et al. in this issue.</i></p>
<p>7. External review</p>	<p>External review of draft CPGs should be provided by all relevant key stakeholders, including the public.</p> <p>The guideline panel should address all comments and keep a record on how and why those were incorporated or not in the final recommendations.</p>	<p><i>See paper by Kahn et al. in this issue.</i></p>
<p>8. Updating</p>	<p>The CPG publication date, date of systematic evidence review, and proposed time of future update should be documented.</p> <p>The evidence base should be regularly monitored and the CPG updated if significant new evidence emerges that modifies the existing recommendation.</p>	<p><i>See paper by Kahn et al. in this issue.</i></p>

Adapted from Institute of Medicine, Clinical Practice Guidelines We Can Trust, Standards March 2011 (www.iom.edu/cpgstandards).

a solution through raising *awareness* and *acceptance* and leading to *adoption* and *adherence*, i.e. the 4A-pipeline of a behavioural change management process.

Active dissemination of guidelines, using leaflets, electronic alerts and advertisements, outreach visits, lectures by respected senior experts and other tools should be coupled with education to raise *awareness* and facilitate *acceptance* of recommendations. However, even acceptance of the evidence does not guarantee that evidence-based recommendations for best practice are adopted and adhered to. Mickan *et al.* have elegantly demonstrated that there is leakage along the awareness–acceptance–adoption–adherence pipeline. Their

study showed that both adoption and adherence were affected by provider and organisational factors. For example, specialists working in large hospitals with better facilities and resources were more likely to adopt and adhere to recommendations than single-handed general practitioners. Laboratories therefore may need to develop different implementation strategies for their hospital and general practitioner clients. It further emphasizes the importance of joint development of laboratory medicine specific clinical recommendations that this study has also found that national or regional recommendations issued by professional organisations were more likely to be accepted and adopted than global or

Table 2 Checklists and tools for laboratory medicine practice guideline developers

Purpose	Tool	Link
Critical appraisal of guidelines	AGREE	www.agreetrust.org/
Grading the strength of evidence and recommendations	GRADE	www.gradeworkinggroup.org
Implementation of guidelines	GLIA	http://nutmeg.med.yale.edu/glia
Standard operating procedure for laboratory medicine practice guideline development	NACB SOP	https://www.aacc.org/~media/files/nacb/nacb_lmpg_sop_jan_2014.pdf?la=en
Guideline development checklist	GRADE-DECIDE	www.guidelinedevelopment.org http://cebgrade.mcmaster.ca/guidecheck.html
Manual for Diagnostic Test Accuracy systematic reviews	Cochrane DTA	http://srdata.cochrane.org/handbook-dta-reviews
Critical appraisal of diagnostic accuracy studies for systematic reviews	QUADAS-2	http://www.bris.ac.uk/media-library/sites/quadas/migrated/documents/quadas2.pdf
Reporting standards for diagnostic accuracy studies	STARD	http://www.equator-network.org
Reporting standards for multivariable prediction model for individual prognosis or diagnosis	TRIPOD	http://www.equator-network.org
Reporting standards for systematic reviews	PRISMA	http://www.equator-network.org

other international guidelines. This study also showed that informed patients can influence adherence to best practice which highlights the importance of guideline implementation strategies that use patient information and empowerment tools. Clear and consistent laboratory

testing-related guidelines, conceived in collaboration with clinical specialists and which are pilot tested and adapted to local settings and equipped with tools and resources for monitoring, achieve higher success with *adoption* and *adherence* (3,21).

The multidimensional and complex nature of guideline implementation strategies is probably best described by a matrix of multifaceted approaches, including 1/ behavioural and educational, 2/ organisational, 3/ policy, and 4/ professional and other incentives and tools that kick the 4A-cycle into action (3). Various evaluations have also concluded that such multifaceted implementation strategies are more likely to succeed than single interventions. Continuous benchmarking of performance, coupled with feedback and education, seem to be the most successful strategies. For more details on guideline implementation and auditing its impact and effectiveness and cost-effectiveness in practice see the papers by Misra et al. and by Collinson, in this issue.

TOWARDS IMPROVED GUIDELINE DEVELOPMENT

The above mentioned shortcomings of guideline development and implementation raise a couple of questions: 1/ Do we need guidelines at all, or should the laboratory profession focus its efforts and resources on producing more high quality research evidence, and probably less low quality guidelines? 2/ Do we need so many guidelines on the same topic, and 3/ is it necessary to have separate guidelines for covering different aspects of care of a clinical condition?

Considering the first question, one may ask, if guidelines are not implemented or applicable to practice and do not have significant impact on health outcomes, why bother developing them? Would it be better to have high quality trials or systematic reviews or evidence summaries in form of well structured, quality rated evidence-tables that would provide a universal answer to clinically important questions? This might be particularly relevant in laboratory medicine, where systematic reviews,

conducted solely for the purposes of guidelines or economic analyses, are often of poorer quality than single overviews performed by experts trained in evidence-based medicine (22,23). Kahn and Gale also argue whether we need so many guidelines on the same topic and whether we should move away from guidelines that are too generic and directed toward patient populations and replace those by computer generated, individualised guidelines where the evidence is provided as a backbone for local discussions and formulation of local policies on best clinical practice (24).

So, there is an increasing argument for simply providing better evidence and evidence reviews. However, for doing so, laboratory professionals should obtain more skills in systematically reviewing the diagnostic literature which itself has a number of methodological challenges. Various manuals and tools assist in writing systematic reviews related to diagnostic testing (Table 2). For example, a Cochrane Working Group issued a comprehensive handbook for systematic reviews of diagnostic test accuracy (<http://srdta.cochrane.org/handbook-dta-reviews>). The so-called QUADAS tool is a very useful resource for appraising diagnostic accuracy studies for systematic reviews (25). Several reporting standards, such as STARD for diagnostic accuracy studies, TRIPOD for a multivariable prediction model for individual prognosis or diagnosis, and PRISMA for systematic reviews in general can be found on the Equator Network's website (<http://www.equator-network.org>). The Institute of Medicine has also issued methodological standards for producing high quality systematic reviews (http://www.nap.edu/catalog.php?record_id=13059).

Regarding the second question, undoubtedly we have far too many guidelines, often covering the same topic. At the time of writing this article, 2417 CPGs are available in the Agency

for Healthcare Research and Quality's (AHRQ) National Guideline Clearinghouse website (<http://www.guideline.gov/index.aspx>) and 84 CPGs are under development. On diabetes mellitus alone, there are 454 hits for CPGs in the same database. For this phenomenon of multiple guidelines, Kahn and Gale offer some explanations and a solution. If a new guideline is developed for a topic that is already covered by a guideline elsewhere, the organisation should provide a rationale why a new guideline is needed; simply approve the existing guideline if recommendations are the same; or explain how and why the new guideline differs from the previous one (24). The AHRQ also offers guideline synthesis reports that compare the scope, content and the corresponding strength of evidence of various CPGs on the same topic (<http://www.guideline.gov/compare/index.aspx>).

With increasing rigour for development CPGs are becoming too complex and too long. There is an ongoing discussion whether testing-related recommendations should be developed by subspecialty societies, such as AACC's NACB, or whether recommendations on testing should be part of CPGs and developed jointly with clinical societies. On the one hand, guidelines produced by specialty societies are reported to be of lower methodological quality compared to those produced by major guideline organisations that have well-defined processes, rigorous methodologies and adequate resources to hire expertise for evidence-based guideline development (26,27). On the other hand, laboratory testing-related information is not easy to locate and pre- and post-analytical information, important for the appropriate requesting and use of tests, is rarely provided in CPGs. Inappropriate coverage of laboratory testing related information in CPGs has been shown by the Guideline Working Group of the European Federation of Clinical Chemistry and

Laboratory Medicine. For more comprehensive coverage of laboratory-related items in CPGs, this group has suggested a detailed checklist of 33 preanalytical, 37 analytical and 10 post-analytical items and they also provided a reduced list of minimum requirements (28). For more details see the paper by Misra et al., in this issue.

CONCLUSIONS

Whilst clinical practice guidelines aim to close the gap between research and practice, the appearance of so many guidelines seems to have created a new gap between their development and utility in practice. Poor quality and lack of explicitness of recommendations on laboratory testing call for methodological and reporting standards for guidelines. A transparent and explicit evidence-grading scheme and international collaboration of guideline development activities are needed to increase the validity, applicability and cost-effectiveness of recommendations related to the use of laboratory tests in clinical practice.

ACKNOWLEDGEMENTS

Aspects of this review have been discussed in and influenced by the work of the Test Evaluation Working Group and the Guidelines Working Group of the European Federation of Clinical Chemistry and Laboratory Medicine. Roche Diagnostics is thanked for supporting the Test Evaluation Working Group with an independent educational grant to the European Federation of Clinical Chemistry and Laboratory Medicine.

REFERENCES

1. Bayram C, Britt H, Miller G, Valenti L 2009. Evidence-practice gap in GP pathology test ordering: a comparison of BEACH pathology data and recommended testing. Family Medicine Research Centre School of Public Health.

- The University of Sydney, June 2009;pp205. [https://www.health.gov.au/internet/main/publishing.nsf/Content/9C300FE48F876E95CA257BF0001ACE0E/\\$File/Evidence-practice%20gap%20in%20GP%20pathology%20test%20ordering.pdf](https://www.health.gov.au/internet/main/publishing.nsf/Content/9C300FE48F876E95CA257BF0001ACE0E/$File/Evidence-practice%20gap%20in%20GP%20pathology%20test%20ordering.pdf) (Last accessed 2 June 2015)
2. Field MJ, Lohr KN. (eds) Guidelines for Clinical Practice: from development to use, Institute of Medicine, National Academy Press, Washington, DC.1992
 3. Horvath AR. From Evidence to Best Practice in Laboratory Medicine. Clin Biochem Rev 2013;34:47-60.
 4. Evidence-Based Health Care. How to make health policy and management decisions. Ed.: JAM Gray, Churchill Livingstone, 1997;pp270
 5. von Karsa L, Patnick J, Segnan N, Atkin W, Halloran S, et al. European guidelines for quality assurance in CRC screening and diagnosis Overview and introduction to the full Supplement publication. Endoscopy 2013;45:51-59.
 6. Klabunde CN, Joseph DA, MD, King JB, White A, Plescia M. Vital Signs: Colorectal Cancer Screening Test Use — United States, 2012. Morbidity and Mortality Weekly Report 2013; 62(44):881-888. <http://www.cdc.gov/mmwr> (Last accessed 2 June 2015)
 7. von Karsa L, Anttila A, Ronco G, Ponti A et al. Cancer Screening in the European Union Report on the implementation of the Council Recommendation on cancer screening, First Report, European Commission 2008; pp160.
 8. McDowell SE, Thomas SK, Coleman JJ, Aronson JK, Ferner RE. A practical guide to monitoring for adverse drug reactions during antihypertensive drug therapy. J R Soc Med 2013;106:87–95.
 9. Kim EH, Andriole GL. Prostate-specific antigen-based screening: controversy and guidelines. BMC Medicine 2015;13:61. <http://www.biomedcentral.com/1741-7015/13/61> (Last accessed 2 June 2015)
 10. Mpondo BCT, Ernest A, Dee HE. Gestational diabetes mellitus: challenges in diagnosis and management. J Diabetes Metab Disord. 2015;14:42. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4430906> (Last accessed 2 June 2015)
 11. Nankervis A, Conn J. Gestational diabetes mellitus - Negotiating the confusion. Australian Family Physician 2013;42(8):528-531. <http://www.racgp.org.au/afp/2013/august/gestational-diabetes-mellitus/> (Last accessed 2 June 2015)
 12. Watine J, Friedberg B, Nagy E, Onody R, Oosterhuis W, Bunting P, Charet J-C, Horvath AR. Conflict between guideline methodologic quality and recommendation validity: a potential problem for practitioners. Clin Chem 2006;1(52): 65-72.
 13. Nagy E, Watine J, Bunting P, Onody R, Oosterhuis W, Rogic D, Sandberg S, Boda K, Horvath AR. Do guidelines for the diagnosis and monitoring of diabetes mellitus fulfill the criteria of evidence-based guideline development? Clin Chem 2008;54(11):1872-82.
 14. Alonso-Coello P, Irfan A, Solà I, Gich I, Delgado-Noguera M, Rigau D, Tort S, Bonfill X, Burgers J, Schunemann H. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. Quality and Safety in Health Care, 2010;19(6):e58.
 15. Don-Wauchope AC, Sievenpiper JL, Hill SA, Iorio A. Applicability of the AGREE II instrument in evaluating the development process and quality of current National Academy of Clinical Biochemistry guidelines. Clin Chem. 2012;58(10):1426-37.
 16. Gopalakrishna G, Langendam MW, Scholten RJPM, Bossuyt PMM, Leeftang MMG. Guidelines for guideline developers: a systematic review of grading systems for medical tests. Implementation Science 2013;8:78 <http://www.implementationscience.com/content/8/1/78> (Last accessed 2 June 2015)
 17. Brouwers M, Kho ME, Browman GP, Cluzeau F, Feder G, Fervers B, Hanna S, Makarski J on behalf of the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. Can Med Assoc J. 2010;182:E839-842. The AGREE II Instrument is accessible at: <http://www.agreetrust.org/> (Last accessed 2 June 2015)
 18. Legido-Quigley H, Panteli D, Car J, McKee M, Busse R. Clinical guidelines for chronic conditions in the European Union. Observatory Studies Series 30. The European Observatory on Health Systems and Policies. 2013;pp256.http://www.euro.who.int/_data/assets/pdf_file/0009/195876/Clinical-Guidelines-for-Chronic-Conditions-in-the-European-Union.pdf (Last accessed 2 June 2015)
 19. Schünemann HJ, Wiercioch W, Etzeandia I, Falavigna M, Santesso N, Mustafa R, Ventresca M et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ 2014;186(3):E123-142.
 20. Institute of Medicine, Clinical Practice Guidelines We Can Trust. Accessible at: www.iom.edu/cpgstandards (Last accessed 2 June 2015)
 21. Mickan S, Burls A, Glasziou P. Patterns of 'leakage' in the utilisation of clinical guidelines: a systematic review. Postgrad Med J 2011;87:670-9.
 22. Vigna-Taglianti F, Vineis P, Liberati A, Faggiano F. Quality of systematic reviews used in guidelines for oncology practice Ann Oncol 2006;17:691-701,

23. Mallett S, Deeks JJ, Halligan S, Hopewell S, Cornelius V, Altman DG. Systematic reviews of diagnostic tests in cancer: review of methods and reporting. *BMJ* 2006;333:413-19.
24. Kahn R, Gale EAM, Gridlocked guidelines for diabetes. *Lancet* 2010;375:2203-4.
25. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, the QUADAS-2 Group. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med.* 2011;155:529-36.
26. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R: Characteristics of high quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. *Int J Technol Assess Healthcare* 2003;19(1):148–157.
27. Grilli R, Magrini N, Penna A, Mura G, Liberati A: Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000;355:103–105.
28. Aakre KM, Langlois MR, Watine J, Barth JH, Baum H, Collinson P, Laitinen P, Oosterhuis WP. Critical review of laboratory investigations in clinical practice guidelines: proposals for the description of investigations. *Clin Chem Lab Med* 2013;51:1217-26.

Defining the path forward: guidance for laboratory medicine guidelines

Stephen E. Kahn¹, Patricia M. Jones², Alex C. Chin³, Robert H. Christenson⁴

¹ Department of Pathology, Loyola University Health System, A Member of Trinity Health, Maywood, IL, USA

² Department of Pathology, University of Texas Southwestern Medical Center and Children's Medical Center, Dallas, TX, USA

³ Calgary Laboratory Services; and Department of Pathology and Laboratory Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

⁴ Department of Pathology, University of Maryland School of Medicine, Baltimore, MD, USA

ARTICLE INFO

Corresponding author:

Stephen E. Kahn, Ph.D., DABCC, FACB
Professor and Vice Chair
Clinical Services, Pathology
Loyola University Health System –
A Member of Trinity Health
2160 South First Avenue
Maywood, IL, USA 60153
Phone: +017082164725
Fax: +017082161639
E-mail: skahn@lumc.edu

Key words:

evidence-based medicine, evidence-based laboratory medicine, laboratory management practice guideline, standard operating procedure, AGREE II

ABSTRACT

The National Academy of Clinical Biochemistry (NACB) has developed consensus-based guidelines for the laboratory evaluation and monitoring of patients with specified disorders for two decades. In 1997, the NACB recognized the need to standardize the process of guideline development and promulgated its first Standard Operating Procedure (SOP) for this purpose. In 2010, the American Association of Clinical Chemistry (AACC) and NACB created the Evidence-Based Laboratory Medicine Committee (EBLMC). Among other roles, this group was given responsibility to provide oversight of clinical practice guideline development in accordance with SOP guidance and using currently accepted good practices. In 2011, the U.S. Institute of Medicine (IOM) published two reports of relevance: 'Clinical Practice Guidelines We Can Trust' and 'Finding What Works in Health Care – Standards for Systematic Reviews.' These reports were created as part of a response to a legislative mandate from the U.S. Congress requesting that steps be taken to implement recommendations from IOM's report on 'Knowing What Works in Health Care' (2008). The latest revision of the laboratory medicine practice guidelines (LMPG) SOP was in part driven by these reports. NACB continues to

develop LMPGs at a rate of roughly one per year through standard processes detailed in its 2014 revision of the SOP.

This article describes the NACB and EBLMC experience in developing LMPGs with a focus on the evolution and use of the latest SOP. AACC and NACB have established a solid track record in collaboratively working with many clinical societies and professional organizations on clinical practice guideline development. Presently, three LMPG's are in various stages of development and all with the collaboration of other clinical/professional groups. The practices and tools being used for current LMPGs in progress are also highlighted in the context of the challenges that presently exist for effective clinical practice guideline development in the U.S.

INTRODUCTION

Over the past decade, a transformation has swept across the U.S. healthcare system. Delivering patient-centered care and improving resource utilization have become 'mission critical' goals for healthcare providers. After promising during his election campaign to make U.S. health care reform a top priority, Barack Obama became the 44th U.S. President in 2009. The following year, President Obama signed the *Patient Protection and Affordable Care Act* into law. Often referred to as '*Obamacare*,' this act set the stage for an even greater transformation of the U.S. healthcare landscape by creating a new paradigm for providers' delivery of healthcare with a focus shift from volume to value. As a result of these factors, interest in the practice of evidence-based medicine (EBM) in the U.S. has never been stronger.

With this increased interest in EBM and associated evidence-based laboratory medicine (EBLM) efforts, clinical societies, professional organizations and governmental groups have developed a greater awareness on the importance

of clinical practice guidelines as well as the methods used for their development. In 2011, the Institute of Medicine (IOM) published two relevant reports: '*Clinical Practice Guidelines We Can Trust*' (1) and '*Finding What Works in Health Care – Standards for Systematic Reviews*' (2). Promulgating these reports was part of the IOM's response to a legislative mandate from Congress requesting that steps be taken to implement recommendations from an earlier IOM report on '*Knowing What Works in Health Care (2008)*' (3). As a result, the Department of Health and Human Services was commissioned to develop evidence-based, methodological standards for systematic reviews (SRs) and clinical practice guidelines (CPGs) (1). These events also provided new resources for the National Academy of Clinical Biochemistry (NACB) at a time when it had become the 'Academy of AACC' and was reassessing their processes for development of laboratory medicine practice guidelines (LMPGs).

A brief history of NACB and a key program of the Academy – Laboratory Medicine Practice Guidelines

The NACB was founded by a group of members from the Chicago Section of the American Association for Clinical Chemistry (AACC) in 1976. This core group of clinical chemists envisioned a learned professional society of doctoral level scientists employed in academic, research and/or hospital-based settings. Throughout its history, the scope, visibility and impact of the Academy's programs have grown steadily. Early on, two key Academy programs were a specific NACB Annual Meeting and the *Journal of Clinical Biochemistry*. In the 1990's, the overlap of individuals who held leadership positions in both AACC and NACB began to increase. Additionally, recognition of the benefits in synergistic collaboration across multiple programs and venues led to formal agreements between NACB and

AACC, and were established in the spirit of working together more closely. The mission of NACB is to ‘*advance clinical practice and research and to promote education and professional development in clinical laboratory medicine*’. In 2006, AACC and NACB leaders signed an agreement to merge, expanding NACB’s mission to include ‘*serving as the Academy of AACC*’.

One of the NACB’s most visible programmatic initiatives continues to be the staging of conferences and symposia focusing on important topics in the disciplines of clinical biochemistry and laboratory medicine. In the mid-1990’s, NACB leaders decided to replace the scientific symposia at their annual meetings with conferences aimed at *Standards of Laboratory*

Practice (SOLPs). The model for this new format was a small meeting, often a satellite of the larger AACC conference, for which the proceedings and issues discussed would be published in the form of a monograph. These NACB monographs were early versions of clinical laboratory practice guidelines. Once published, they allowed for broader dissemination of conference findings and education of laboratory professionals. In 1999, NACB leaders decided to use a new name for future SOLPs, *Laboratory Medicine Practice Guidelines* (LMPGs). Since 1994, the NACB has developed, or is currently developing, nearly 20 SOLPs and LMPGs. A list of these documents is provided in Table 1.

Table 1 The year, topic and status of the SOLPs and LMPGs of the AACC Academy (the NACB)

Year	Topic	Status
1994	Nutritional Status	(Out of Print)
1996	Diagnosis of Thyroid Disease	-
1998	Evaluation and Management of Newborns	(Out of Print)
1999	Therapeutic drug Monitoring	(Out of Print)
1999	Cardiac Markers	(Archived)
2000	Hepatic Injury	(Archived)
2000	Electronic Medical Records	-
2002	Thyroid Disease	(Archived)
2002	Diabetes Mellitus	(Archived)
2003	Tumor Markers in the Clinic	(Archived)
2005	Emergency Toxicology	(Archived)
2006	Maternal-Fetal Risk Assessment	(Archived)

2007	Point of Care Testing	(Archived)
2007	Biomarkers of Acute Coronary Syndrome	(Published)
2008	Expanded Newborn Screening	(Published)
2009	Emerging CV Risk Factors	(Published)
2009	Tumor Markers in Testicular, Prostate, Colorectal, Breast, and Ovarian Cancers	(Published)
2009	Use of Tumor Markers in Clinical Practice: Quality Requirements	(Published)
2010	Tumor Markers in Liver, Bladder, Cervical, and Gastric Cancers	(Published)
2010	Laboratory Analysis and Application of Pharmacogenetics to Clinical Practice	(Published)
2011	Diagnosis and Management of Diabetes Mellitus	(Published)
(In development)	Pain Management	(Final title to be determined)
(In development)	Biomarkers of Cardiac Disease	(Final title to be determined)
(In development)	Guidelines and Recommendations for Laboratory Analysis of Human Chorionic Gonadotropin (hCG) in Clinical Practice	(Final title to be determined)

Earlier SOLP's and LMPG's of the Academy are now either out of print or archived.

Today, LMPGs are documented practice recommendations resulting from evidence-based approaches to addressing questions regarding appropriate use of diagnostic laboratory testing in a specific scientific and/or clinical discipline. LMPGs are intended to improve the use of diagnostic laboratory tests in a manner that optimizes patient care outcomes and are based on practice recommendations informed by systematic review of the evidence. LMPGs include recommendations based on weighting and grading the relevant evidence. LMPGs also address the benefits and harms of alternative laboratory testing strategies. A key component of SOLPs and LMPGs has always been development in col-

laboration with other relevant clinical societies, stakeholders and/or professional organizations.

Sustaining guideline quality through standard operating procedures

Not long after the first SOLP was published, NACB leaders recognized that a long-term approach for ensuring the quality and impact of their guidelines would best be served by the development of policies or procedures for guideline development. This recognition led to a decision made by the NACB's Board of Director's (BOD) to include in their own manual a policy on LMPGs that also required the creation, use, and periodic revision of a Standard Operating Procedures (SOP) instrument for NACB Guideline Development Groups (GDGs) (4). The first NACB SOLP SOP was created in 1997. Prior to the 2005 approval of

the SOP by NACB's BOD, it had already been revised twice during the 8 years since the initial SOP was created.

Initially, responsibility for oversight of the development of LMPGs rested with the NACB's Education and Scientific Affairs Committee (ESAC). In 2009, AACC and NACB leaders decided that activities of AACC's EBM Committee should integrate more closely with programs of the Academy. By that time, the Committee had capably demonstrated a strong track record in offering programs and products as well as establishing a solid working relationship with the Agency for Healthcare Research and Quality (AHRQ) in the U.S. AHRQ is a government agency, part of the Department of Health and Human Services, that functions to support research to improve the quality of health care (5).

Key members of AACC's EBM Committee continue their contributions to the field that contribute to maintaining EBM and EBLM at the forefront of numerous AACC and NACB initiatives (6,7). In 2010, a new Evidence-Based Laboratory Medicine Committee (EBLMC) was formed combining the activities of AACC's EBM Committee and NACB's ESAC.

The EBLMC was charged with several responsibilities including oversight of LMPG development. The EBLMC is also charged with promoting and/or overseeing the collaborative efforts required in review and approval of other society or organizational guidelines for potential AACC endorsement. In fact, all NACB guideline development groups must have a member of the EBLMC who is selected through collaborative discussion between the LMPG committee chair and the EBLMC chair.

Given these roles, it made sense that the EBLMC would also take on the responsibility for ensuring that revisions of the LMPG SOP remained consistent with current best practices in clinical practice guideline development. Through

an extended process that began in 2011 that involved multiple stages of review by key stakeholder groups in AACC and its Academy, the 2014 revision of the LMPG SOP was approved and is available to AACC and/or NACB members on NACB's webpage on the AACC website (8). Before final AACC and NACB BOD approval, a draft of the 2014 LMPG SOP was posted allowing for and inviting open public comment on the proposed content in order to achieve openness and transparency of EBLMC's efforts to arrive at a final revision that could be widely utilized.

Content in the 2014 SOP was influenced significantly by the 2011 IOM report as well as by other available guideline development resources (9,10). AACC and NACB leaders as well as members of the EBLMC recognized, acknowledged, and underscored the importance of developing LMPGs in a process consistent with the below key principles articulated in the 2011 IOM report on developing trustworthy clinical practice guidelines:

- Establishing transparency
- Management of conflict of interest (COI)
- Guideline development group composition
- Clinical practice guideline-systematic review intersection
- Establishing evidence foundations for and rating strength of recommendations
- Articulation of recommendations
- External review
- Updating

LMPG committees are strongly encouraged to keep all elements of these standards in mind during the guideline development process and incorporate specifics, where applicable, in the final LMPG.

Components in the AACC organizational structure associated with guideline development, review and approval are numerous and varied.

As a result, the nature of interactions and responsibilities between these components as well as with external groups, when applicable, are complex. Consequently, the importance and benefits of using the SOP as a mandatory guide by LMPG committees in LMPG development should be apparent. Organizational elements that can potentially be involved with the development of LMPGs and the use of the LMPG SOP as well as the activities related to review/approval of other external society or organizational guidelines are shown in Figure 1.

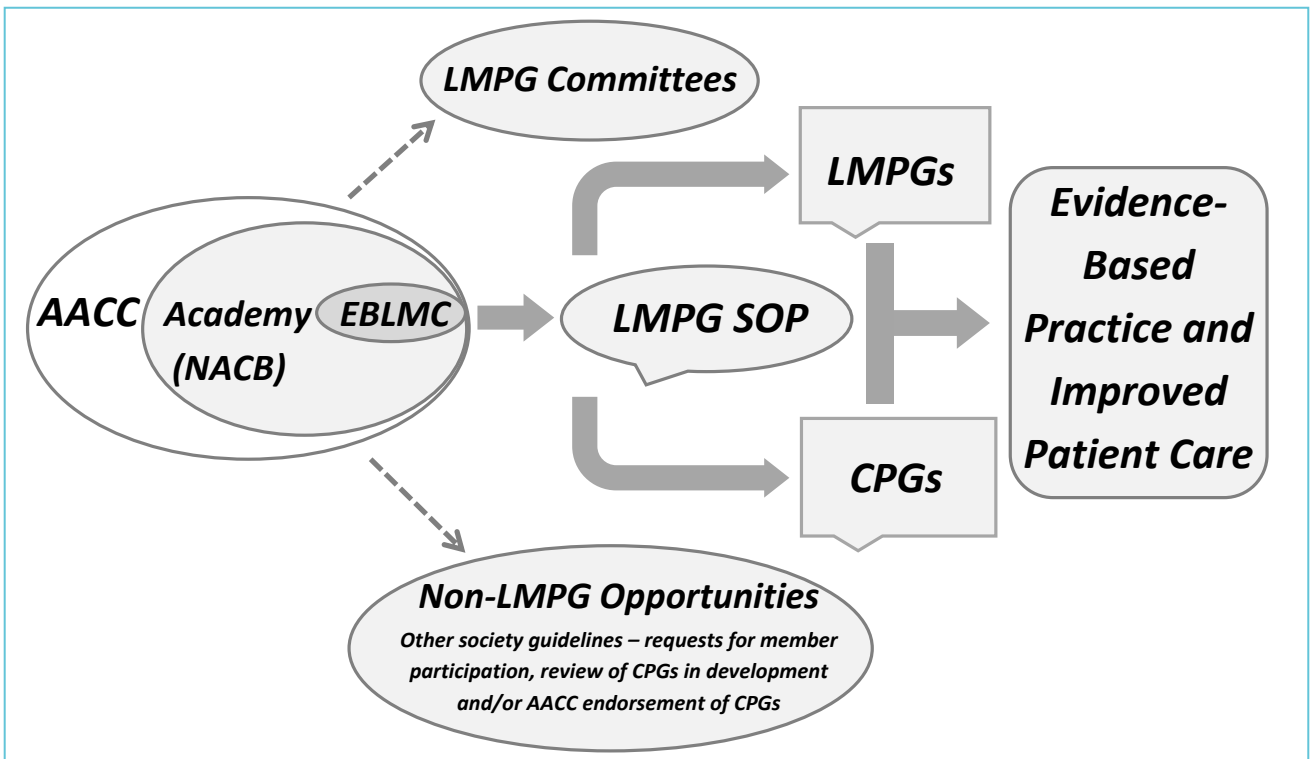
Shouldering the bulk of responsibility for the tremendous amount of work required for LMPG development is the LMPG committee itself. Several other groups have typically played a key role in the overall process including the

EBLMC, AACC and NACB BODs as well as conference or meeting organizing groups such as the AACC's annual meeting organizing committees. LMPG committees are expected to be multidisciplinary and typically have members from other relevant clinical societies or professional organizations.

With the understanding that LMPGs are more likely to be utilized fully by both laboratorians and clinicians with the endorsement and support of the appropriate clinical society, LMPG Committees are strongly encouraged by the SOP to include clinical society members and to sign a collaborative Co-Sponsorship Agreement with the clinical society(s) involved.

The most important role on the LMPG committee is that of the LMPG chair. This individual,

Figure 1 Promoting development of LMPGs and CPGs by the AACC Academy and EBLMC



How the AACC Academy and the Evidence-Based Laboratory Medicine Committee (EBLMC) promote evidence-based practice and improved patient care through use of Standard Operating Procedures (SOP) for the development of laboratory medicine practice guidelines (LMPGs) and clinical practice guidelines (CPGs).

or individuals if there are co-chairs, may often be a key and active participant in one or more of AACC's Divisions that are unique groups or 'communities' that AACC members may join that focus on their specific area(s) of interest or expertise within the field of laboratory medicine. Presently, there are 18 scientific divisions within AACC.

Other LMPG development issues addressed in the 2014 SOP are the roles and responsibilities of all key stakeholder groups or individuals, how LMPG topics are selected, how to conduct the systematic review of the evidence (including selected examples of past and current data abstraction forms for this review) and how to evaluate the strength as well as grading of the final evidence-based recommendations. Significant ancillary activities required for LMPG development are also addressed. This category of information includes public presentation of LMPG information in selected program categories or venues, public posting of LMPGs including digital media, processes for guideline finalization and approval, requirements for LMPG publication, expected LMPG development timelines and requirements of a plan for future updating of the LMPG. This last item has not been well addressed in past versions of the SOP. In turn, being able to update key LMPGs when the 5 year active period has expired has, in the past, often been a challenge for NACB, ESAC and now, also the EBLMC.

Being able to assess the effectiveness of an LMPG is another area that has been lacking previously and is in keeping with the current SOP. Any initial proposal of a LMPG topic now takes into account such issues as target audience, guideline promotion and optimal utility and priority gaps that should be addressed. In addition, LMPG Committee selection focuses on bringing the appropriate partners to the table to facilitate the production of effective guidelines.

Critical issues to address for achieving best practices in guideline development

Working collaboratively with other clinical societies and organizations remains a top priority in LMPG development. For more than two decades, this collaboration has involved close to 100 other clinical societies and/or professional organizations. Frequent partners include the College of American Pathologists, the Endocrine Society, the U.S. Centers for Disease Control and Prevention as well as the AHRQ. Now, with the emphasis being placed on clinical society collaboration on LMPGs, partnership organizations are increasing in number and variety. Current LMPGs in development involve collaboration with the American Academy of Pain Management, the American Congress of Obstetrics and Gynecology and other clinical groups.

Grading the quality of the evidence and the strength of recommendations also presents a challenge given the lack of systems effectively designed for use with diagnostic tests (11). For grading the evidence and assigning the strength of recommendations LMPG committees, especially if the LMPG is developed in collaboration with leading clinical societies, use those systems that are routinely employed by the relevant clinical societies in their guideline development process. When this has not been the case, the system that has often been used by LMPG committees was an adapted and modified version from the U.S. Preventive Services Task Force Recommendations for Preventive Services (12).

LMPGs have been typically posted on AHRQ's National Guideline Clearinghouse website for the five year active period per AHRQ policy (13). As of early 2015, three LMPGs remain actively listed. One of these LMPGs on '*Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus*' also reported development of a new

system designed by the LMPG committee for grading evidence and assigning the strength of recommendations (14,15). This system also incorporates a new and specific expert-based consensus recommendation known as 'Best Practice Points.' The overall system reported and used by this LMPG committee will be an option for future guideline development groups, including LMPG committees, to consider since it was specifically designed for a guideline focusing on diagnostic testing.

Continuing to strive towards achieving best practices in guideline development and being able to sustain a greater level of consistency in LMPG development are emphasized in the 2014 SOP. That this would be a significant opportunity for improvement is not surprising considering the length of time since NACB groups first began developing SOLPs and, presently, AACC's Academy developing LMPGs in collaboration with an extensive number of partner societies.

The EBLMC and NACB leaders have underscored the importance of addressing and resolving these issues (16). In 2011, the EBLMC decided that the guideline evaluation tool to be used by LMPG committees for this purpose should be the second edition of the AGREE (Appraisal of Guidelines for Research and Evaluation) instrument (17). Use of the AGREE II instrument to evaluate the methodological quality of clinical practice guidelines has been reported (18).

Another group reported using the AGREE II instrument to evaluate eleven NACB LMPGs (most now archived). This group found that five of eleven LMPGs had overall scores $\geq 50\%$.

However, while all provided useful information seen as applicable to clinical practice by the evaluators, there was still a wide variability in AGREE II domain scores (19). Notably, the one guideline published (15) after the development

of the AGREE II instrument achieved a very high score (19). To further advance the necessary support by EBLMC in LMPG or external society guideline review, the 2014 SOP describes a significant change made by EBLMC compared to previous methods of 'linking' the LMPG developing groups and those helping to oversee the development process. Historically, NACB required guideline chairs to be members of the Education and Scientific Affairs Committee. As noted previously, the practice now required in the 2014 SOP is for at least one EBLMC member, preferably with relevant content expertise and experience, to also serve on each new LMPG committee. In this manner, representative EBLMC members will be able to provide updates to the EBLMC on the progress and challenges experienced by the respective LMPG committees. Reciprocally, given that the EBLMC includes members with experience in guideline development and methodology, efforts are under way to make this expertise more available to LMPG committees.

CONCLUSION

For all clinical guideline development groups, effective application of evidence-based laboratory medicine will continue to require openness and transparency as well as adaptability in their procedures and activities in the future.

For the EBLMC and future LMPG committees, significant opportunities remain for identifying ways that can increase the effectiveness of LMPGs, provide measurable indicators of their impact and document related changes in clinical practice associated with the new evidence-based recommendations regarding use of diagnostic tests. Indeed, employment of new communication strategies including digital media may prove useful to promote LMPG activities and evidence-based laboratory medicine.

Presently, there are three LMPG committees two focusing on new LMPG topics and one being an update of the widely recognized 2007 LMPG on Biomarkers of Cardiac Disease (20,21). The LMPG committee for the latter is being formed, in part, from members of AACC's Biomarkers of Acute Cardiac Disease Division. This committee will undoubtedly include key members from other clinical societies in the cardiac disease and/or cardiology disciplines. It is anticipated that the next LMPG to be finalized from these three LMPG committees will be on the laboratory aspects of Pain Management and another one is under development on the clinical use of hCG testing.

All three of the current LMPGs in some stage of development include working with, and involving individuals from other clinical societies under the auspices of the EBLMC using the procedures contained in the 2014 SOP. The efforts will include the monitoring of the methodological quality of LMPGs by application of the AGREE II instrument.

Within the EBLMC as well as the leaders of NACB and the AACC, there is a strong, sustained commitment to ensure that the LMPG development process will continue to evolve and improve over time. This commitment must include the EBLMC and other groups remaining open to making future revisions to the 2014 SOP when necessary. For as so many individuals have stated in a quote, known widely: "if you're not getting better, you're getting worse."

REFERENCES

1. Clinical Practice Guidelines We Can Trust, Institute of Medicine, The National Academies Press, Washington, D.C., 2011.
2. Finding What Works in Health Care – Standards for Systematic Reviews, Institute of Medicine, The National Academies Press, Washington, D.C., 2011.
3. Knowing What Works in Health Care: A Roadmap for the Nation, Institute of Medicine, The National Academies Press, Washington, D.C. 2008.
4. NACB Policy on Laboratory Medicine Practice Guidelines, Page 25, NACB Board of Directors Manual, April 2007 Revision.
5. Agency for Healthcare Research and Quality website at <http://www.ahrq.gov>
6. Price CP, Glen JL, Christenson RH. Applying Evidence-Based Laboratory Medicine: A Step-by-Step Guide., Washington DC, AACC Press, 2006.
7. Price CP and Christenson RH (Eds.). Evidence-Based Laboratory Medicine; Principles, Practice and Outcomes. Second Edition. Washington DC, AACC Press, 2007.
8. NACB Standard Operating Procedures for: Preparing, Publishing and Revising NACB LMPGs (Including Review and Approval of External Society/organization Guidelines for Endorsement and Support by AACC/NACB). January 2014 Version. Available at (<https://www.aacc.org/community/national-academy-of-clinical-biochemistry>).
9. American College of Chest Physician's Guideline Development Workshop Materials, ACCP, 2012.
10. American Society of Clinical Oncology Guidelines Procedures Manual, Expert Panel Version 4.0, available at www.asco.org, ASCO, 2011.
11. Horvath AR. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. Clin Chem 2009;56(5); 853–5.
12. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al. Current methods of the U.S. Preventive Services Task Force: a review of the process. Am J Prev Med 2001;20:21–35.
13. Agency for Healthcare Research and Quality National Guidelines Clearinghouse at <http://www.guideline.gov>
14. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Kirkman MS et al. Executive summary: Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem 2011;57(6):793-8.
15. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Kirkman MS et al. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem 2011;57(6):e1–47.
16. Kahn SE, Astles JR, Lo SF, Bennett, MJ. The AGREE II instrument is helpful for creation of NACB LMPGs. A letter to the editor. Clin Chem 2013;59(2);446–7.

17. The AGREE Enterprise Website. Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. www.agreetrust.org

18. Brouwers MC *et al.* AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010;182(18):e839-42.

19. Don-Wauchope AC, Sievenpiper JL, Hill SA, Iorio A. Applicability of the AGREE II instrument in evaluating

the development process and quality of current NACB guidelines. *Clin Chem* 2012;58(10):1426–37.

20. Morrow DA, Cannon CP, Jesse RL, Newby LK *et al.* NACB LMPGs: Clinical characteristics and utilization of biochemical markers in acute coronary syndromes. *Circ* 2007;115:e356–e375.

21. Christenson RH. NACB LMPGs for utilization of biochemical markers in acute coronary syndromes and heart failure. *Clin Chem* 2007;53(4):545–6.

Grading evidence for laboratory test studies beyond diagnostic accuracy: application to prognostic testing

Andrew C. Don-Wauchope^{1,2}, Pasqualina L. Santaguida³

¹ Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada

² Hamilton Regional Laboratory Medicine Program, Hamilton, Ontario, Canada

³ Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Corresponding author:

Andrew C. Don-Wauchope
Core Laboratory
Juravinski Hospital and Cancer Centre
711 Concession Street
Hamilton, L8V 1C3
Ontario, Canada
Phone: +1 905 521 2100 x76306
Fax: +1 905 381 7066
E-mail: donwauc@mcmaster.ca

ABSTRACT

Background: Evidence-based guideline development requires transparent methodology for gathering, synthesizing and grading the quality and strength of evidence behind recommendations. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) project has addressed diagnostic test use in many of their publications. Most of the work has been directed at diagnostic tests and no consensus has been reached for prognostic biomarkers.

Aim of this paper: The GRADE system for rating the quality of evidence and the strength of a recommendation is described. The application of GRADE to diagnostic testing is discussed and a description of application to prognostic testing is detailed. Some strengths and limitations of the GRADE process in relation to clinical laboratory testing are presented.

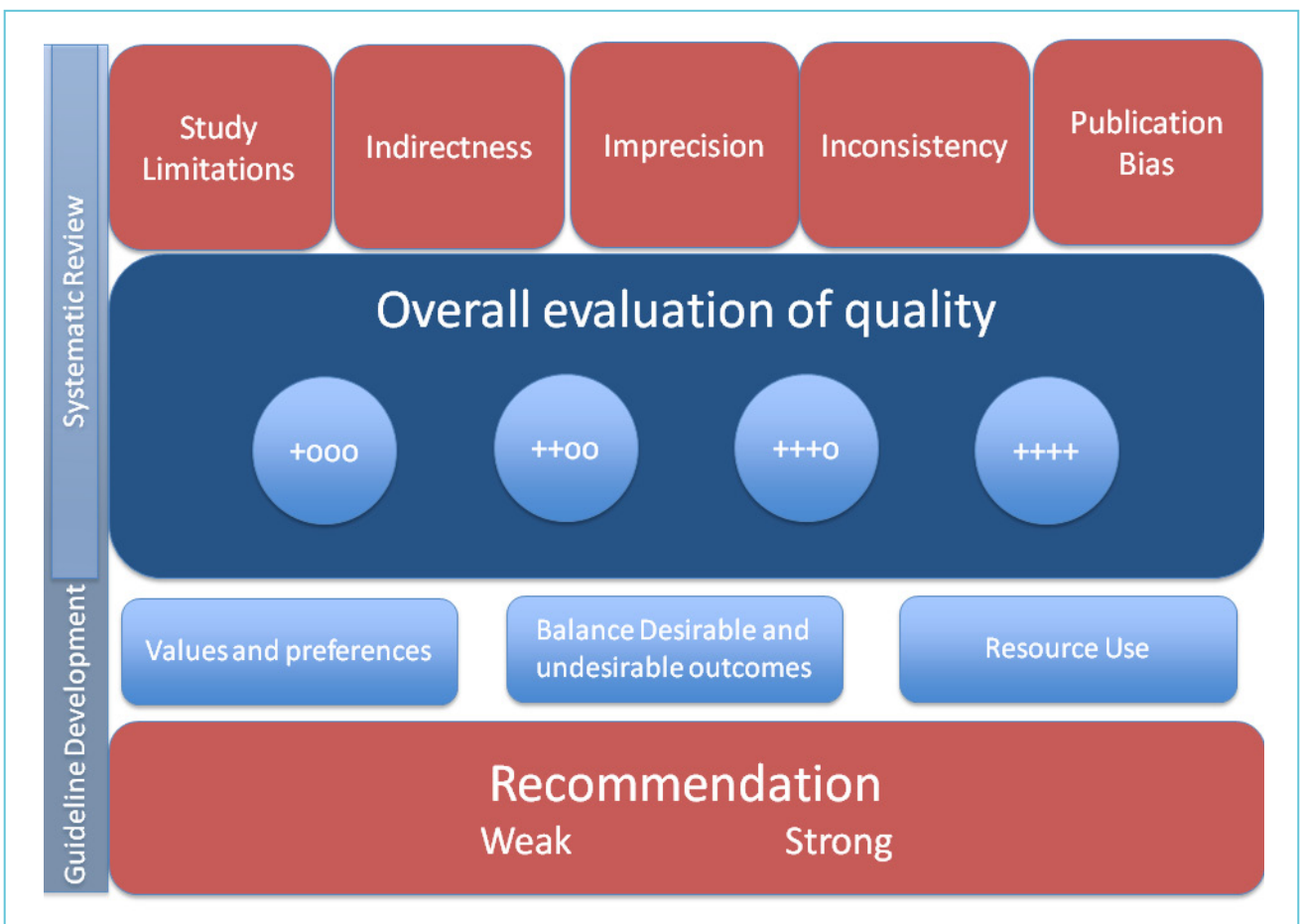
Conclusions: The GRADE system is applicable to clinical laboratory testing and if correctly applied should improve the reporting of recommendations for clinical laboratory tests by standardising the style of recommendation and by encouraging transparent reporting of the actual guideline process.

INTRODUCTION

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) project was initiated to standardise the grading of guideline recommendations (1). The GRADE system addresses both the quality of evidence as well as the level of recommendation (2). Numerous systems exist for grading the evidence and recommendations, generated by a range of organisations representing professional societies and national/provincial/international bodies amongst others (3). The GRADE

project has published two sets of papers with the most recent series still appearing in the literature (4). These provide a combination of general guidance and examples of specific application to a range of areas in medicine. This article will briefly describe the GRADE approach to evaluating the quality of evidence for diagnostic testing with a focus on laboratory tests. Figure 1 gives an overview of how this fits into the overall GRADE process that includes a number of other factors in the formation of a recommendation classified as strong or weak. Subsequently, we will describe how this can be

Figure 1 The GRADE domains – the basis for the evaluation of quality of evidence



This information is obtained as part of a systematic review that allows for full evaluation of the evidence for each individual paper and then a collation of this into an overall summary of the quality of evidence. The guideline developers then need to consider the quality of evidence in context of a number of other important factors to judge a final recommendation.

applied to prognostic testing using our previous work on natriuretic peptides as the example. Finally, the strengths and limitations of the GRADE approach will be considered in the context of laboratory medicine.

OVERVIEW OF THE GRADE SYSTEM OF RATING THE QUALITY OF THE EVIDENCE

The GRADE system uses four major domains to evaluate the quality of the evidence for a research question (Figure 1). Typically research questions would be expected to follow the Population-Intervention-Comparator-Outcome (PICO) format (5). There are four major domains and several minor domains that can be considered as modifiers of the final quality of evidence (6).

The first major domain investigates the *risk of bias* or limitations of primary papers that are considered for answering the specific PICO research question behind the guideline recommendations (7). This is based on evaluation of the study design (i.e. cohort or randomized trials), the application of the study design (identification of any threats to internal validity), the reporting and analysis of the results and the conclusions presented. There are a range of validated tools available to assist researchers and guideline teams to evaluate the risk of bias in the primary papers. Systematic reviewers should include their GRADE assessment and the supporting data in the results of the systematic review.

The second major domain investigates the *inconsistency* of the evidence (8). This domain considers all the primary papers related to each outcome (defined in the PICO) and evaluates the direction of the effect for consistency. The presence of inconsistency in the direction or magnitude of the effect (i.e. specificity) would result in a downward grading of the evidence

for the outcome. It is evaluated by considering the range of point estimates, the confidence interval around each point estimate and the statistical testing for heterogeneity. When several outcomes are considered, inconsistency is evaluated separately for each outcome.

The third major domain investigates the *indirectness* of evidence in relation to outcomes (9). This domain considers the plausible or proved link between the factor (e.g. the diagnostic intervention) being considered and the outcome being evaluated. This requires consideration of the potential differences in population, type of intervention, outcome measures and the comparisons made. The overall indirectness needs to be judged based on the PICO and if present would downgrade the quality of evidence. Similar to inconsistency, each outcome is evaluated for indirectness.

The fourth major domain is about the *imprecision* of the evidence (10). Ideally, this domain evaluates outcomes for which a summary pooled estimate is calculated in a meta-analysis to provide a measure of overall effect across different studies. The width of the 95% CI in this context would give an estimate of the imprecision of the summarised data. If an intervention is being compared to a control then the 95% CI of the individual point estimates for each included study would be precise if there was no overlap, and imprecise if there was overlap. When the study effects cannot be meta-analyzed a number of factors (such as sample size) are considered across the literature being evaluated and graded for imprecision.

There are several minor domains that can also be considered when grading evidence and recommendations. One minor domain is publication bias (11). This domain is generally evaluated using statistical techniques to assess the probability of publication bias. There must be sufficient number of studies included so that

the statistical test has validity. In the case where there are too few studies, one may likely assume that publication bias is likely present. Other aspects to consider when assessing publication bias are small numbers of studies with small populations and predominate funding from industry sponsors whose role within the study is not specified. Other minor domains include any evidence for dose response, the magnitude of the effect size and plausible residual confounding (12).

Using the GRADE approach, the quality of evidence is reported as one of 4 levels: High (++++); Moderate (+++o); Low (oo); or Very Low (oooo) (13). The use of symbols to convey the strength of evidence is becoming more apparent in clinical practice guidelines and assists readers in quickly assessing the quality upon which the recommendations are based. The definitions of these categories have been well described for therapeutic interventions (13) and we have suggested some additional descriptions applicable to diagnostic accuracy and prognostic studies. Table 1 (on the next page) is an adaptation of the practical interpretation of the quality of the evidence when considering intervention (13), diagnostic accuracy (14), and prognostic studies (15).

GRADE FOR DIAGNOSTIC TESTING USING LABORATORY TESTS

Diagnostic testing was considered a separate category when the GRADE project published the first set of articles describing the process for evaluating quality of the evidence and recommendations (16). This was received with some scepticism from the laboratory community but has been successfully applied in some situations with a number of limitations. The challenge to diagnostic testing is often in the nature of the study design providing data to support the PICO question. The Oxford Centre for Evidence-Based

Medicine (CEBM) has articulated this well in their table for levels of evidence in diagnostic accuracy testing (17). Within this hierarchy, the highest order (i.e. most rigorous and valid) of study are cohort and case-control studies and thus quite different from therapeutic interventions where randomised controlled trials are considered the highest order of study design. This is noted in the GRADE description for diagnostic test strategies, where exception is made for diagnostic accuracy studies that would include cross-sectional or cohort designs as an acceptable study type with no downgrading based on for the domain of study limitations. However, the evidence is quickly down-ranked when considering the indirectness and imprecision often associated with these study design types. As more experience with the use of GRADE was gained, the approach to evaluating diagnostic accuracy studies was further developed (18, 19).

The same general principles and categories apply and it remains essential to set the question well with consideration of the PICO elements. There is some evidence to suggest that many clinical questions posed in diagnostic test studies do not distinguish between the population being tested and the problem (disease) of interest (20).

The PICO format for interventions typically combines the problem with population while for diagnosis it may be important to separately define these two components. For diagnostic accuracy studies the outcomes are typically the classification of the results into the proportion of true positive, true negative, false positive and false negative (21). This assumes that the patient-relevant clinical outcome is the correct diagnosis, and this encourages focus on diagnostic accuracy data. However, there is debate about what is considered the most appropriate clinical outcome of testing and that more emphasis should be placed on the role of testing

Table 1 Interpretation of the quality of evidence for GRADE			
Quality	Interventions (13)	Diagnostic test for diagnostic accuracy (14)	Prognostic use of diagnostic test (15)
High Quality	We are confident that the true effect lies close to the estimate of the effect	We are confident that the diagnostic accuracy estimates are accurate.	We are confident that the test makes an important contribution to the determination of outcome (predictive strength).
Moderate Quality	We are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.	We are moderately confident in the estimates of accuracy. The true accuracy estimate is likely to be close to the observed accuracy, but there is a possibility that it is substantially different.	We are moderately confident that the test makes an important contribution to the determination of the outcome. The estimate of the observed predictive strength is likely close to the true effect, but there is a possibility that it is substantially different.
Low Quality	Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect	Our confidence in the accuracy estimate is limited; the true accuracy may be substantially different from the accuracy observed.	Our confidence in the predictive strength is limited; the true predictive strength may be substantially different from the estimate of predictive strength observed.
Very Low Quality	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.	We have very little confidence in the accuracy. The true accuracy is likely to be substantially different from the observed accuracy.	We have very little confidence in the predictive estimate of the test. The true predictive strength is likely to be substantially different from the estimate of predictive strength.

in clinical pathways, and that the purpose of the test (diagnosis, monitoring, screening, prognosis, risk stratification and guiding therapy) and the clinical effectiveness of testing should be considered in the wider context of health care and the

role for diagnostic testing (22). If the clinically important outcome includes appropriate management and improvement in patient health, then there is great difficulty in linking the diagnostic test to the health outcome directly and the

assessment of imprecision requires that multiple other factors are considered (22, 23). There are a number of outcome options that could be considered for diagnostic testing and the most appropriate of these should be defined as part of the PICO (22, 24).

Thus far most of the published literature has focused on diagnostic accuracy studies. The STARD document has helped improve the reporting of diagnostic accuracy studies (25). The comparator could be a “gold” standard test but this may not be available and other options are mentioned in the STARD document. This concept has been explored further by the Agency for Health Care Research and Quality (AHRQ) in their methods guide for medical test reviews (26). Other parts of the extended PICO question definition may include the timing and setting for the question (i.e. PICOTS) (27). Timing is one aspect that is often considered critical for diagnostic testing as the time between the test being investigated and the comparator test is essential. Timing plays an important role, particularly if the investigators are not blinded to the index and reference test results are not masked. It is also important if the two tests are carried out at different time points in the disease process. For index tests and reference tests, that require samples or procedures other than blood (for example tissue or diagnostic imaging), then the two tests must be conducted in a time frame in which change in the disease process would not impact the interpretation of the test result. For laboratory testing based on blood samples the ideal situation is collection of all samples at the same point in time. The setting often helps define the population more clearly. When the prevalence of the diagnosis is changed because of the setting (e.g. primary care versus specialist clinic), it becomes an important component as consideration of prevalence will impact the diagnostic accuracy data. This can be illustrated by two of the questions asked in the AHRQ

comparative effectiveness review on the use of Natriuretic peptides in Heart Failure (28, 29). Two diagnostic settings were considered and this allowed for the primary papers to be grouped correctly and evaluated in the appropriate context (Table 2).

Assessing risk of bias for diagnostic accuracy studies is discussed extensively in the GRADE papers as this is seen as particularly challenging (18, 30). The AHRQ Methods Guide describes the challenges of assessing risk of bias in more detail (31). Validated tools such as the QUADAS II(32) tool or its predecessor the QUADAS(33) can be helpful to carefully consider a range of important factors that impact on the evaluation of risk of bias. For any new systematic reviews or clinical practice guidelines the use of QUADAS II would be recommended as it has improved from the earlier version. QUADAS II focuses on 4 aspects of risk of bias (patient selection, conduct or interpretation of the index test, conduct or interpretation of the reference test, flow and timing of the tests) and four aspects of applicability (whether the study is applicable to the population and settings of interest). In the AHRQ Methods Guide, the domain of indirectness, which is the link between diagnostic accuracy and clinical outcome, and the domain of imprecision were identified as challenging to assess (34).

This section provides an overview of the theoretical framework to identify ways in which the domains of risk of bias/study limitations, inconsistency, indirectness, imprecision and publication bias can be considered for evaluating the evidence for diagnostic tests. This has been successfully applied to diagnostic applications of laboratory tests and Table 2 provides an example of how GRADE was applied in the recent AHRQ systematic review for Natriuretic peptides in the diagnosis of heart failure (28, 29).

Table 2 Grading of evidence for the diagnostic use of B-type Natriuretic peptides

PICO	Diagnostic measure	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Strength of evidence
Use of B-type natriuretic peptides for the diagnosis of heart failure in the emergency department (28)	Sensitivity	low	Consistent for BNP Inconsistent for NT-proBNP	Direct	Imprecise	n/a	For both BNP and NT-proBNP High or ++++
	Specificity	low	Consistent for BNP Inconsistent for NT-proBNP	Direct	Imprecise	n/a	BNP High or ++++ NT-proBNP Moderate or +++o
Diagnostic performance of B-type natriuretic peptide for the diagnosis of heart failure in primary care (27)	Sensitivity	low	Consistent	Direct	Imprecise	No evidence	High or ++++
	Specificity	low	Inconsistent	Direct	Imprecise	No evidence	Moderate or +++o

APPLICATION OF GRADE TO PROGNOSTIC TESTING

Although the GRADE has been widely adopted for assessing the quality of the evidence in both studies of interventions and diagnostic accuracy, it has not yet been applied to studies evaluating prognosis. In large part, this is because GRADE has not reached consensus on how to apply the criteria in the four major domains and in the minor domains specific to prognosis research.

Prognosis is defined as the probable course and outcome of a health condition over time.

A prognostic factor is any measure in people with a health condition that from a specific start point is associated with subsequent clinical outcome (endpoint) (35). Prognostic factors, if well established, function to stratify individuals with the health condition into categories of risk or probability for the outcomes of interest. Research into prognostic factors aims to establish which factors are modifiable, which should be included in more complex models predicting outcome, monitor disease progression, or show differential responses to treatment.

We had the opportunity to explore the application of the GRADE approach in a systematic review in which 3 prognostic questions were addressed (36). In the diagnostic examples (Table 2), we considered the use of natriuretic peptides with respect to diagnosing heart failure. In addition, our systematic review considered natriuretic peptides as potential markers predicting mortality and morbidity in both acutely ill and chronic heart failure patients (37-40), as well as in the general population (41). Our review showed that both BNP and NT-proBNP generally functioned as an independent predictor of subsequent mortality and morbidity at different time frames.

Huguet et al. (2013) have recently proposed some guidance for adapting GRADE for prognostic studies based on their work in identifying factors associated with chronic pain (15). The main differences from GRADE applied to intervention studies, occur with respect to study limitations and to factors that may increase overall quality. With regards to study limitations, there is consideration of the phases of prognostic research. This differs from evaluating evidence from intervention and diagnostic accuracy studies, where the type of specific design (e.g. RCT or cohort study) is given specific weighting. In the context of prognostic studies, there is no consensus on the taxonomy for phases of prognosis research (Table 3). The simplest approach considers three phases of prognostic research. At the lowest level of prediction (PHASE 1), prognosis studies are designed to identify potential associations of the factors of interest and are termed "exploration" (42) or "predictor finding" (43) or "developmental studies" (44). PHASE 2 explanatory studies typically establish or confirm independent association between prognostic factors and outcomes, and are also labelled as "validation" studies (44). The highest level of evidence is from PHASE 3 studies where the prognosis study attempts to

evaluate the underlying processes that link the prognostic factor with the outcome. High quality evidence is likely found in PHASE 3 studies (15); conversely, moderate to very low quality evidence is based on PHASE 1 and 2 studies.

In prognostic research, setting the clinical question is still the most important aspect as patient important outcomes need to be addressed in the appropriate context. Using the PICOTS format is central to this process to adequately define the population, the intervention, the timing and the setting. The comparator and the outcome are also critical but often challenging to define. The comparator test could be a wide range of items when it comes to delineating probable course and outcome. In our examples we included a full range of reported comparators in the form of any type of diagnosis of heart failure.

This could prove to be challenging if one form of confirmation is clearly better than another or if the different confirmatory tests include different sub-populations. For the heart failure populations we did not attempt to divide these out, apart from the division between acute decompensated and chronic stable heart failure. However, we could have tried to use different diagnostic criteria such as echocardiography findings to delineate severity and diastolic from systolic dysfunction.

As discussed in the diagnostic accuracy section the range of clinically relevant outcomes can be quite diverse. For prognostic outcomes the use of clinical pathways and clinical effectiveness should be considered in addition to the more traditional mortality and morbidity outcomes. The length of time from the test to the evaluation of the outcome status may be an important consideration as this may change with differing lengths of time. Bearing all these concepts in mind is important when defining the outcome as the applicability of the findings will be dependent on patient important outcomes.

Table 3 Frameworks for sequential development of prediction models that assess the contribution of potential prognostic factors

Framework of an explanatory approach to studying prognosis (42)	Consecutive phases of multivariable prognostic research (44)	Types of multivariable prediction research (43)
PHASE 1: Identifying associations		Predictor Finding Studies
PHASE 2: Testing independent associations	Developmental Studies (new prognostic model is designed)	Model Development studies without external validation
PHASE 3: Understanding Prognostic Pathways	Validation Studies (External replication of the model)	Model Development studies with external validation
		External validation with or without model updating
	Impact Studies (prognostic models are technologies which require assessment of their impact on health outcomes)	Model Impact Studies

Risk of bias for the prognostic studies in the natriuretic peptide systematic review was evaluated using the underlying principles of the Quality in Prognosis Studies (QUIPS) tool (45). The elements of the QUIPS tool had been previously published and we adapted these very slightly for the prognostic questions in our study (46). This considers 6 domains that may impact bias of a prognostic study: participation; attrition; prognostic factor measurement; confounding measurement and control; outcome measurement; and analysis and reporting (45). The type of study design for prognostic evaluation is largely cohort studies and these are primarily prospective in nature. However, in many reports the original study was a prospective or randomised controlled trial and the analysis

of the prognostic factor was done as an afterthought and hence the study design should be classified as retrospective cohort. There are randomised controlled trials that could be considered as true evaluations of prognostic testing but these are rare.

One additional advantage of using the QUIPS is that there is a thorough assessment of the potential for confounding bias. When applying the GRADE to intervention studies, where the presence of plausible confounding in cohort studies can be expected to reduce the effect size observed, the study limitations can be upgraded. However, this assumption may not be applicable to prognostic studies which are predominately observational in design; residual confounding can effect predictions in either direction (over or

under estimation of the predictive strength) or have no effect at all (15). Our systematic review for natriuretic peptides and heart failure showed that most studies had many plausible confounders (biases) that were not accounted for in the adjusted analysis (i.e. residual confounders) (38, 40). The methods used in our comparative effectiveness review attempted to establish a minimum of three critical confounders; age, renal function, BMI (or other measure of height and weight) considered in the study design or in the analysis. As an example to evaluate confounding from renal function we considered multiple terms to identify the tests and conditions (Table 4). Our findings showed consistent problems with studies measuring these three plausible confounders, not considering several other

potential confounders. However, it was not clear which if any of these affected our estimates of prediction or the direction of impact. The domain of confounder measurement and control is essential in prognostic studies because the link between the prognostic test and the outcome is most often not direct and thus consideration of all other known factors that influence the outcome need to be taken into account. This evaluation of primary papers allowed us to judge the overall bias for the papers included for each sub-question that we addressed as well as obtain some insight into the other relevant domains of GRADE. Hugué et al (2013) have also made use of the QUIPS tool in their experience with chronic pain systematic reviews (15).

Table 4 Example of the range of terms used to identify renal dysfunction in the prognostic evaluation of natriuretic peptides

Terms used for renal function	Test used for renal function
renal failure	urea or BUN
acute renal failure	blood (serum or plasma) creatinine
ARF	creatinine clearance
primary acute renal failure	urine creatinine
chronic renal failure	
CRF	
acute interstitial nephritis	
acute tubular necrosis	
azotemia	
dialysis	
glomerulonephritis	
hemodialysis	
obstructive renal failure	
renal insufficiency	
kidneys	
acute kidney failure	
diabetes	

Inconsistency can be estimated from the summary tables with the point estimates and 95% CI from odds ratio (OR), hazards ratio (HR) and relative risk (RR). This follows the description from the GRADE group and application of this category does not differ from tests of intervention or diagnostic tests (8).

The proposed adaptation of the GRADE to prognostic studies for indirectness asks raters to consider this domain in the context of the population, the prognostic factor, and the outcome. The less generalizable the results for each of these contexts, the higher the likelihood of down-rating this category increases. Indirectness is typically present when one considers prognostic use of a test as there is very seldom a direct link between the test and the outcome of interest. There are typically numerous steps in the process and many of these are completely independent of the test being evaluated. If the factors described by the GRADE group (population; intervention, outcome and comparator) are well described in the PICOTS then it may be possible to find a group of primary studies that match all factors in the same way. If such a group of studies could be found then indirectness may not be present. In the natriuretic peptide systematic review primary studies differed in outcome and comparators that clearly made the evidence-to-outcomes link indirect (38, 40).

Imprecision has some interesting difference between application in guidelines and systematic reviews (10). For systematic reviews the goal is estimating the effect size while for guidelines the goal is to support a recommendation. Thus in a systematic review the precision will be interpreted on the width of the 95% CI while in guidelines it would be interpreted on the ability to separate from the comparator. When possible the pooled effect size and confidence limit would be the ideal tool to evaluate imprecision. Consideration should also be given to the

sample size of studies (10). However meta-analysis is not always available as the appropriate application of meta-analysis requires that the studies being included match the PICOTS closely. When meta-analysis is not possible the range of effect size and the spread of 95% CI need to be considered.

Publication bias will follow the same principles described in the GRADE papers (11). Although the issue has been noted in recent literature, in the context of prognostic studies (47), there is currently no registry of studies, or studies related to laboratory testing. Thus it is difficult to make informed judgements about the likelihood of publication bias.

Careful consideration and description of all the GRADE domains need to be made by the guideline developers or systematic reviewers. This should be documented and written up as an appendix to allow users of the guideline to consider the details used by the guideline writers and to allow methodologists the opportunity to further develop the concepts around evaluation of diagnostic tests.

STRENGTHS AND LIMITATIONS OF GRADE FOR LABORATORY TESTS

The major strengths when using the GRADE approach for the evaluation of the strength of evidence and recommendations is the explicitness and reproducibility of the process (48). An advantage is the requirement to define a useful and appropriate clinical question that includes the necessary components of PICOTS. The GRADE system takes into account key domains to assess quality and strength of evidence. The process of GRADE allows for transparency when users of the guideline review the evidence behind the recommendations (49).

Limitations can be grouped in a number of areas. Firstly guideline writers often do not fully understand the GRADE system. Methodological

experts are most often aware of the system but many of them invited to participate in the guideline team will not have had sufficient exposure to GRADE or training to incorporate the GRADE assessment of the strength of evidence strength or to the process for making recommendations. The GRADE system has been available for a number of years but as it continues to develop it can be difficult for non-methodologists to keep pace with the changes. The application of GRADE requires judgment of the evidence in the domains as well as judgement of the factors that help form the recommendation. This judgment is often construed as expert opinion and this has formed the core of clinical practice guidelines in many instances. The GRADE process is designed to move away from expert opinion alone to one that includes an evidence-formed judgement. If the team is well versed in the GRADE literature and suitably trained then the judgement aspect will be a strength; however, it could be a limitation if the team is not able to sufficiently consider the evidence and be unduly influenced by their own expert opinion.

The second group of limitations relates to the challenges guideline teams face in meeting the explicit criteria required for developing structured clinical questions and for the evaluation of the evidence as described in the GRADE process. Although the domains of GRADE and how to apply these are well defined, the heterogeneity of evidence presents practical challenges to guideline development teams. For example, defining the appropriate type of study design for the highest rank of evidence can be challenging. As noted previously, the designs that are considered to have greater rigour (i.e. higher form of evidence) will depend on the actual purpose of the study. For diagnostic testing and prognostic testing these will be different and these nuances require careful reflection from the

guideline developers. Initially the researchers may consider using the currently published models (for example CEBM tables and Table 3) and use these if seen as appropriate (17, 42-44). If an alternative system is used it should be justified in the method description. The aspects of PICOTS require careful consideration to make the question applicable to the target audience. This is reasonably straightforward for diagnostic testing (19), but definitions may be more challenging in prognostic questions as the distinction between population and disease become even more important. Often more than a single outcome should be considered in order to capture the complexity of the contribution of diagnostic testing in relation to patient important outcomes. There are practical challenges when judgements are based on patient-relevant *versus* a test accuracy perspective (19). Similarly, there are some challenges to adequately judge imprecision as statistical approaches are somewhat limited for assessing heterogeneity in diagnostic tests. The complexity and diversity of clinical care pathways may complicate the assessment of indirectness. Here the factors that may impact the clinical care pathway need to be accounted for when the directness or indirectness of the evidence is rated. The choice of outcome measures will further influence the considered judgement process of the GRADE approach.

CONCLUSIONS

The GRADE system can be used to rate the evidence for diagnostic and prognostic use of laboratory testing. There are numerous challenges and the results may not always be seen as consistent between different guideline groups. However, the GRADE evidence rating system allows users of the guideline to compare and contrast guidelines covering the same or similar content. The transparency of the approach also

allows better-informed adaptation and implementation of guideline recommendations to local practice.

REFERENCES

- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. Grade: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924-6.
- Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. Grade guidelines: 14. Going from evidence to recommendations: The significance and presentation of recommendations. *Journal of clinical epidemiology*. 2013;66:719-25.
- Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of recommendations i: Critical appraisal of existing approaches the grade working group. *BMC health services research*. 2004;4:38.
- List of grade working group publications and grants. GRADE Working Group. <http://www.gradeworkinggroup.org/publications/index.htm> (accessed 20 February 2015)
- Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. Grade guidelines: 2. Framing the question and deciding on important outcomes. *Journal of clinical epidemiology*. 2011;64:395-400.
- Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. Grade guidelines: 1. Introduction-grade evidence profiles and summary of findings tables. *Journal of clinical epidemiology*. 2011;64:383-94.
- Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. Grade guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *Journal of clinical epidemiology*. 2011;64:407-15.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. Grade guidelines: 7. Rating the quality of evidence--inconsistency. *Journal of clinical epidemiology*. 2011;64:1294-302.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. Grade guidelines: 8. Rating the quality of evidence--indirectness. *Journal of clinical epidemiology*. 2011;64:1303-10.
- Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. Grade guidelines 6. Rating the quality of evidence-imprecision. *Journal of clinical epidemiology*. 2011;64:1283-93.
- Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. Grade guidelines: 5. Rating the quality of evidence-publication bias. *Journal of clinical epidemiology*. 2011;64:1277-82.
- Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. Grade guidelines: 9. Rating up the quality of evidence. *Journal of clinical epidemiology*. 2011;64:1311-6.
- Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. Grade guidelines: 3. Rating the quality of evidence. *Journal of clinical epidemiology*. 2011;64:401-6.
- Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the grade approach and grading quality of evidence about interventions. *Allergy*. 2009;64:669-77.
- Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: Adapting the grade framework. *Systematic Reviews*. 2013;2:1-.
- Schünemann HJ, Schünemann AHJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336:1106-10.
- Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. The oxford levels of evidence 2. Oxford Centre for Evidence-Based Medicine. <http://www.cebm.net/index.aspx?o=5653> (accessed 20 February 2015)
- Brozek JL, Akl EA, Compalati E, Kreis J, Terracciano L, Fiocchi A, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines part 3 of 3. The grade approach to developing recommendations. *Allergy*. 2011;66:588-95.
- Gopalakrishna G, Mustafa RA, Davenport C, Scholten RJP, Hyde C, Brozek J, et al. Applying grading of recommendations assessment, development and evaluation (grade) to diagnostic tests was challenging but doable. *Journal of clinical epidemiology*. 2014;67:760-8.
- Huang X, Lin J, Demner-Fushman D. Evaluation of pico as a knowledge representation for clinical questions. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium*. 2006:359-63.
- Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, et al. Challenges in systematic reviews of diagnostic technologies. *Annals of internal medicine*. 2005;142:1048-55.
- 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. *Clinica chi-*

mica acta; international journal of clinical chemistry. 2014;427:49-57.

23. Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? The European Journal of Health Economics. 2003;4:143-50.

24. Matchar DB. Chapter 1: Introduction to the methods guide for medical test reviews. Journal of General Internal Medicine. 2012;27:4-10.

25. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The star statement for reporting studies of diagnostic accuracy: Explanation and elaboration. Clinical Chemistry. 2003;49:7-18.

26. Trikalinos TA and Balion CM. Chapter 9: Options for summarizing medical test performance in the absence of a "gold standard". Journal of General Internal Medicine. 2012;27:S67-75.

27. Samson D, Schoelles KM. Chapter 2: Medical tests guidance (2) developing the topic and structuring systematic reviews of medical tests: Utility of picots, analytic frameworks, decision trees, and other frameworks. Journal of General Internal Medicine. 2012;27 Suppl 1:S11-9.

28. Booth RA, Hill SA, Don-Wauchope A, Santaguida PL, Oremus M, McKelvie R, et al. Performance of bnp and nt-probnp for diagnosis of heart failure in primary care patients: A systematic review. Heart failure reviews. 2014;19:439-51.

29. Hill SA, Booth RA, Santaguida PL, Don-Wauchope A, Brown JA, Oremus M, et al. Use of bnp and nt-probnp for the diagnosis of heart failure in the emergency department: A systematic review of the evidence. Heart failure reviews. 2014;19:421-38.

30. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, et al. Grade: Assessing the quality of evidence for diagnostic recommendations. Evidence-based medicine. 2008;13:162-3.

31. Santaguida PL, Riley CM, Matchar DB. Chapter 5: Assessing risk of bias as a domain of quality in medical test studies. Journal of General Internal Medicine. 2012;27 Suppl 1:S33-8.

32. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155:529-36.

33. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of quadas: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC medical research methodology. 2003;3:25.

34. Singh S, Chang SM, Matchar DB, Bass EB. Chapter 7: Grading a body of evidence on diagnostic tests. Journal of General Internal Medicine. 2012;27 Suppl 1:S47-55.

35. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, et al. Prognosis research strategy (progress) 2: Prognostic factor research. PLoS medicine. 2013;10:e1001380.

36. Oremus M, McKelvie R, Don-Wauchope A, Santaguida PL, Ali U, Balion C, et al. A systematic review of bnp and nt-probnp in the management of heart failure: Overview and methods. Heart failure reviews. 2014;19:413-9.

37. Don-Wauchope AC, Santaguida PL, Oremus M, McKelvie R, Ali U, Brown JA, et al. Incremental predictive value of natriuretic peptides for prognosis in the chronic stable heart failure population: A systematic review. Heart failure reviews. 2014;19:521-40.

38. Oremus M, Don-Wauchope A, McKelvie R, Santaguida PL, Hill S, Balion C, et al. Bnp and nt-probnp as prognostic markers in persons with chronic stable heart failure. Heart failure reviews. 2014;19:471-505.

39. Santaguida PL, Don-Wauchope AC, Ali U, Oremus M, Brown JA, Bustamam A, et al. Incremental value of natriuretic peptide measurement in acute decompensated heart failure: A systematic review. Heart failure reviews. 2014;19:507-19.

40. Santaguida PL, Don-Wauchope AC, Oremus M, McKelvie R, Ali U, Hill SA, et al. Bnp and nt-probnp as prognostic markers in persons with acute decompensated heart failure: A systematic review. Heart failure reviews. 2014;19:453-70.

41. Don-Wauchope AC, Santaguida PL, McKelvie R, Brown JA, Oremus M, Ali U, et al. Prediction of clinical outcomes using b-type natriuretic peptides in the general population: A systematic review. Heart failure reviews. 2014;19:541-51.

42. Hayden JA, Côté P, Steenstra IA, Bombardier C, et al. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. Journal of clinical epidemiology. 2008;61:552-60.

43. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: A systematic review. PLoS medicine. 2012;9:e1001221.

44. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? BMJ (Clinical research ed). 2009;338:b375-b.

45. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Annals of internal medicine. 2013;158:280-6.

46. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Annals of internal medicine*. 2006;144:427-37.

47. Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KGM, et al. Improving the transparency of prognosis research: The role of reporting, data sharing, registration, and protocols. *PLoS medicine*. 2014;11:e1001671.

48. Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, et al. The grade approach is reproducible

in assessing the quality of evidence of quantitative evidence syntheses. *Journal of clinical epidemiology*. 2013;66:736-42.

49. Hsu J, Brožek JL, Terracciano L, Kreis J, Compalati E, Stein AT, et al. Application of grade: Making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implementation science*. 2011;6:62.

Evidence and cost effectiveness requirements for recommending new biomarkers

Paul Collinson

Departments of Clinical Blood Sciences and Cardiology, St George's Hospital and Medical School, London, UK

ARTICLE INFO

Corresponding author:

Paul Collinson
Departments of Clinical Blood Sciences
and Cardiology
St George's Hospital and Medical School
London SW17 0QT, UK

Key words:

biomarkers, cost effectiveness, evidence-based medicine, test evaluation

Disclaimer:

This article is based on a lecture given at IFCC Worldlab 2014 in Istanbul.

IS THIS TEST 'APT'?

The literature is full of new biomarkers which are claimed to add to the laboratory repertoire in a range of conditions. The literature is often confusing and may be contradictory. The past 20 years is littered with publications claiming the next big thing in a biomarker, some of which have been implemented on high throughput laboratory platforms. The number of novel biomarkers which have reached widespread clinical acceptance and implementation is relatively small. How can the laboratory community realistically assess claims for new markers? There is, to date, no completely defined set of criteria which should be used. However, there are some common themes in biomarker assessment. The two major areas which need to be considered are evidence required to assess test performance and cost effectiveness.

Assessment of test performance can be broadly considered under three categories, **Analytical suitability**, **Plausibility** and **Treatment effectiveness**; is the test APT. *Analytical suitability* means an assessment of the evidence-based analytical performance of the assay. This will include at least the following. Pre-analytical factors that will affect the test must be well understood before a test can be put into routine clinical

practice. This will include the collection conditions required, anticoagulant requirements, pre-analytical sample handling factors and stability in storage. A marker needs to be measurable in the routine clinical laboratory without the need for special handling conditions if it is to form part of the routine work-up of the patient. Tests requiring complex pre-analytical steps are tolerated by the laboratory, rather than embraced. Often there is no alternative; the test is confined to special circumstances and particular patient types which are usually rare. A test in the clinical routine which will be ordered in large numbers requires simplicity of laboratory handling. A recent example is the measurement of soluble CD40 ligand (sCD40l), a marker of platelet activation. Measurement of sCD40l was shown to be a powerful predictor of mortality in patients with unstable angina. In addition, it was shown to be a predictor of a successful therapeutic response to the anti-glycoprotein IIb/IIIa antagonist abciximab (1). These studies were done using serum as matrix. It was subsequently found that clotting releases significant but variable amounts of sCD40l. Studies demonstrated that the release of sCD40l was critically affected by sample handling and the assay utilised for measurement (2). Only EDTA plasma could be used and values were significantly affected by delay in sample processing (3,4). Finally, it was shown that sCD40l was primarily produced by *in vitro* platelet activation (5) and the first use of a commercial assay failed to confirm the promise of the initial publication (6). Analytical performance of the test needs to be also appropriate for clinical use. Bodies such as the Clinical Laboratory Standards Institute produce protocols for the routine assessment of limit of blank, limit of detection and imprecision profile. It is also important that these analytical performance measures are independently assessed and that laboratories do not rely on the manufacturers' datasheets as the

sole source of this information. Assay imprecision has a profound influence on the ability to define the 99th percentile and the value of the relative change required between two consecutive measurements to be reliably different. It is an interesting observation that the redefinition of myocardial infarction (7-9) considers a 10% imprecision to be adequate at the 99th percentile but also recommended a 20% change in values. Unfortunately, if the data is modelled it is apparent that an imprecision rather less than 10% is required to reliably detect a 20% change (<http://www.westgard.com/troponin-interpretations.htm>). In addition to the ability to measure the biomarker with precision and accuracy, the analysis must be simple and have a rapid turnaround time. Ideally it should be implemented on existing laboratory equipment rather than requiring additional apparatus. In practice this means that a colorimetric or more likely an immunoassay for the marker is available. Population aspects of the test need to be understood in particular the influence of age, gender, ethnicity and comorbid conditions on the reference interval need to be considered. These can be quite subtle. Occult comorbid conditions profoundly influence the reference interval for cardiac troponin but can only be unmasked by the use of rigorous patient selection including cardiac imaging (10,11). The need for appropriate patient selection for troponin reference intervals has been the subject of discussion and recommendations made (12,13).

The *plausibility* of the biomarker for the putative clinical role needs also to be established. The pathobiology of the biomarker needs to be understood. This means an understanding of the genesis of the biomarker and of the relationship of the biomarker to the medical condition of interest. A good example of this is ischaemia modified albumin (IMA). The concept of a biomarker of ischaemia is very attractive.

Ischaemia would be detected prior to necrosis (we have excellent markers for this in the cardiac troponins) allowing intervention to abort the pathophysiology before irreversible cardiac injury occurs. The background concept of IMA was that the N terminus of albumin was altered during an ischaemic event resulting in the loss of the ability to bind transition metals. This was detectable by loss of the ability to bind cobalt, which could be determined by a simple colorimetric reaction (14). Preliminary studies using angioplasty as a model of human myocardial ischaemia showed that IMA increased after balloon inflation then returned rapidly to baseline levels, supporting the role as a biomarker of ischaemia (15,16). Subsequently, sequencing of the N terminus of IMA positive albumin showed that the N-terminal amino acid sequence was not removed (17). Physicochemical studies suggested that it was the binding of free fatty acids to albumin that induced a conformational change that reduced transition metal binding (18). A lack of fundamental understanding of the biomarker was therefore apparent and contributed to the lack of any clinical application (19). Plausibility also includes the clinical plausibility for the putative clinical role. This means that the biomarker must have appropriate sensitivity and specificity to detect the medical condition of interest in clinically appropriate populations where the test will actually

be used in routine clinical practice. Many studies on biomarkers have evaluated them in clinical trial sample banks or alternatively in highly selected patient groups. This does not constitute an appropriate environment to evaluate test performance as disease prevalence is inappropriately high, often close to 100%. Such studies allow proof of concept that needs to be followed up by prospective evaluation in clinically representative populations. Comparison of a sensitive with a less sensitive troponin assay clearly shows earlier diagnostic sensitivity (20), as would be expected. Early studies of the new high sensitivity assays showed excellent analytical performance but compared them with the conventional assays and included patients with ST segment elevation in the evaluation (21,22), overstating the diagnostic performance of the assays.

Treatment effectiveness is the final and most important strand to assessment. This may be summed up as the “so what” factor. This is short for the question that should be asked by any clinician of a test “so what do I need to do differently with the result of this test”. A new biomarker must offer either a significant proven diagnostic efficiency or result in a change in treatment. Ideally it should do both. The change in treatment may be a decision to give or withhold drug or other therapeutic

Table 1 Key questions for evaluating the evidence base for clinical use

Has this marker been measured with an appropriate method and been shown to be additive to or replace a contemporary test?

Have there been independent studies?

Has there been a multicentre study?

Is there meta-analysis of evidence?

Has there been an RCT?

Can I measure it in the routine lab without additional equipment and staff?

intervention or to change the management pathway such as more prompt hospital discharge or admission to an appropriate level of clinical care. The questions which should pass through the laboratory practitioners' mind are shown in Table 1 below.

An example of a randomised controlled trial of the diagnostic test is the Randomised Assessment of Treatment using Panel Assays of Cardiac markers (RATPAC) (23). This was a pragmatic randomised controlled trial which compared two treatment strategies, conventional management with measurement on admission and at 90 minutes of a panel of cardiac troponin I, creatine kinase MB and myoglobin by point of care testing. The outcome measure was a proportion of patients discharged or a decision to discharge within four hours of attendance with no adverse events during the following three months. Randomisation to the point of care arm of the study was reflected in increased successful discharge and no change in the frequency of adverse events. There was increased use of coronary care in the point of care arm. One of the most interesting aspects of this study was the significant differences between the six different sites with only two showing very large differences in length of stay in those randomised to the point of care arm (24). It highlights the importance of process within the utilisation of test results. Simple

provision of rapid results will be ineffective unless it is accompanied by treatment decision.

IS THIS TEST COST EFFECTIVE?

Cost effectiveness considers the impact on health care resources utilisation and how we assess it. Cost effectiveness can be considered under four categories as shown in Table 2 below. It should be noted however that the terminology is often mixed.

Cost minimisation analysis is the most straightforward. It assumes that the consequences of the two interventions being compared are identical so the analysis reduces to the comparison of costs alone. An example would be the diagnosis of acute myocardial infarction using cardiac troponin (cTn) compared to the measurement of creatine kinase MB isoenzyme (CK-MB). If the assumption is that CK-MB costs 20 currency units (CU) and cTn 30 CU then a protocol involving three hourly CK-MB measurements for 12 hours (total cost 80 CU) will be more expensive than a protocol measuring cTn on admission and 12 hours from admission (total cost 60 CU). In cost effectiveness analysis differences can be expressed in terms of changes in one main parameter. The differences in costs are related to the main differences in events. An example of this type of analysis is the use of measurement of B type natriuretic peptide (BNP) in patients with suspected chronic heart failure. The basic

Table 2 Cost effectiveness categories

Type	Measurement and valuation of consequences
Cost minimisation analysis	No measurement. Consequences assumed or shown to be equivalent.
Cost effectiveness analysis	Natural units (Life years gained)
Cost utility analysis	Health state preference values (quality adjusted life years gained)
Cost benefit analysis	Monetary gains

premise is that two pathways are compared: direct referral for hospital assessment of patients with suspected heart failure and referral only of those with an elevated BNP. A simple analysis compares costs at the pathway level where the costs of echocardiography on all patients is compared with the combined cost of BNP measurement followed by echocardiography only in the those with BNP levels above a certain designated threshold. This is effectively a cost minimisation analysis and shows that the BNP based pathway is cheaper (25). A more sophisticated approach utilising a sequential testing strategy modelled on individual patient data meta-analysis was performed as part of a health technology assessment informing the National Institute of Clinical and health Excellence (NICE) guidelines on BNP testing. This modelling produced very similar results to the cost minimisation model. Cost effectiveness was driven by the prior probability of disease and favoured BNP measurement as the first test (as in strategy discussed above) unless the probability of heart failure was very high (26). Cost utility analysis typically utilises the quality adjusted life year (QALY). A QALY takes into account longevity and quality-of-life. The number of QALYs accrued by a patient is estimated by multiplying the years of survival by quality-of-life measured on a scale from zero (equivalent to death) to 1 (perfect health). States of health below zero are possible for a health state considered worse than death. QALYs have the advantage of allowing comparison between any healthcare intervention that can influence survival or quality-of-life. Analysis is based on willingness to pay (cost per QALY) with a typical threshold of £20,000 in the UK. An example would be comparison of the cost effectiveness of measurement of high sensitivity troponin on admission versus conventional troponin management at 10 hours (27). Such a study shows that high sensitivity troponin measurement on admission is superior to

conventional troponin measurement and that measurement on admission and at three hours is the most sensitive approach. Measurement of conventional troponin at 10 hours is only cost effective if an immediate decision to discharge is made, highlighting again the importance of process in the application of laboratory testing. One problem with cost effectiveness analysis in diagnostics is that the data is often inadequate or even non-existent. Modelling approaches are typically used but the accuracy of the cost modelling is often challenging though mitigated by sensitivity analysis (changing the model parameter and looking at the impact, a large change suggests that the modelling is not robust). Very small differences in QALY's may be present.

A systematic attempt to evaluate the evidence for diagnostics including laboratory testing is used by the Diagnostics Assessment Committee of NICE. They utilise a systematic evidence-based review followed by cost economic modelling. The recommendations and their evidence base can be found on the NICE website (www.nice.org.uk) and in the publications of the UK health technology assessment programme. These are all available online. Examples are the recent recommendations for the use of faecal calprotectin (www.nice.org.uk/guidance/dg11) and the accompanying evidence report (28).

CONCLUSIONS

In conclusion, assessment of test suitability is a combination of the traditional laboratory attributes of the analytical performance of the test but combined with other features. The underlying scientific validity of the test needs to be understood and the diagnostic utility demonstrated in appropriate populations, to show the test is plausible. Finally, the test result must produce a treatment change. All of these, Analytical, Plausibility, Treatment will make a test APT. But an APT test clinically also

needs to be cost effective. Conversely, unless a test has been shown to be APT, the probability of demonstrating cost effectiveness is small. The challenge for the laboratory is to work together with clinicians to develop test evaluation strategies that will allow demonstration of all the attributes to show that the test is both APT and cost effective.

REFERENCES

1. Heeschen C, Dimmeler S, Hamm CW, van den Brand MJ, Boersma E, Zeiher AM, Simoons ML. Soluble CD40 ligand in acute coronary syndromes. *N Engl J Med* 2003;348:1104-11.
2. Halldorsdottir AM, Stoker J, Porche-Sorbet R, Eby CS. Soluble CD40 ligand measurement inaccuracies attributable to specimen type, processing time, and ELISA method. *Clin Chem* 2005;51:1054-7.
3. Weber M, Rabenau B, Stanisch M, Elsaesser A, Mitrovic V, Heeschen C, Hamm C. Influence of sample type and storage conditions on soluble CD40 ligand assessment. *Clin Chem* 2006;52:888-91.
4. Weber M, Rabenau B, Stanisch M, Nef HM, Mollmann H, Elsaesser A et al. Influence of sample type on soluble CD40 ligand assessment in patients with acute coronary syndromes. *Thromb Res* 2007;120:811-4.
5. Ivandic BT, Spanuth E, Haase D, Lestin HG, Katus HA. Increased plasma concentrations of soluble CD40 ligand in acute coronary syndrome depend on in vitro platelet activation. *Clin Chem* 2007;53:1231-4.
6. Plaikner M, Peer A, Falkensammer G, Schmidauer C, Pechlaner C, Griesmacher A et al. Lack of association of soluble CD40 ligand with the presence of acute myocardial infarction or ischemic stroke in the emergency department. *Clin Chem* 2009;55:175-8.
7. Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. *Eur Heart J* 2000;21:1502-13.
8. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD et al. Third universal definition of myocardial infarction. *Eur Heart J* 2012;33:2551-67.
9. Thygesen K, Alpert JS, White HD, Jaffe AS, Apple FS, Galvani M et al. Universal definition of myocardial infarction. *Circulation* 2007;116:2634-53.
10. Collinson PO, Heung YM, Gaze D, Boa F, Senior R, Christenson R, Apple FS. Influence of population selection on the 99th percentile reference value for cardiac troponin assays. *Clin Chem* 2012;58:219-25.
11. Koerbin G, Abhayaratna WP, Potter JM, Apple FS, Jaffe AS, Ravalico TH, Hickman PE. Effect of population selection on 99th percentile values for a high sensitivity cardiac troponin I and T assays. *Clin Biochem* 2013;46:1636-43.
12. Sandoval Y, Apple FS. The global need to define normality: the 99th percentile value of cardiac troponin. *Clin Chem* 2014;60:455-62.
13. Apple FS, Jaffe AS, Collinson P, Mockel M, Ordonez-Llanos J, Lindahl B et al. IFCC educational materials on selected analytical and clinical applications of high sensitivity cardiac troponin assays. *Clin Biochem* 2015;48(4-5):201-3.
14. Bar-Or D, Lau E, Winkler JV. A novel assay for cobalt-albumin binding and its potential as a marker for myocardial ischemia--a preliminary report. *J Emerg Med* 2000;19:311-5.
15. Bar-Or D, Winkler JV, Vanbenthuyzen K, Harris L, Lau E, Hetzel FW. Reduced albumin-cobalt binding with transient myocardial ischemia after elective percutaneous transluminal coronary angioplasty: a preliminary comparison to creatine kinase-MB, myoglobin, and troponin I. *Am Heart J* 2001;141:985-91.
16. Sinha MK, Gaze DC, Tippins JR, Collinson PO, Kaski JC. Ischemia modified albumin is a sensitive marker of myocardial ischemia after percutaneous coronary intervention. *Circulation* 2003;107:2403-5.
17. Bhagavan NV, Lai EM, Rios PA, Yang J, Ortega-Lopez AM, Shinoda H et al. Evaluation of human serum albumin cobalt binding assay for the assessment of myocardial ischemia and myocardial infarction. *Clin Chem* 2003;49:581-5.
18. Mothes E, Faller P. Evidence that the principal CoII-binding site in human serum albumin is not at the N-terminus: implication on the albumin cobalt binding test for detecting myocardial ischemia. *Biochemistry* 2007;46:2267-74.
19. Collinson PO, Gaze DC. Ischaemia-modified albumin: clinical utility and pitfalls in measurement. *J Clin Pathol* 2008;61:1025-8.
20. Melanson SE, Morrow DA, Jarolim P. Earlier detection of myocardial injury in a preliminary evaluation using a new troponin I assay with improved sensitivity. *Am J Clin Pathol* 2007;128:282-6.
21. Keller T, Zeller T, Peetz D, Tzikas S, Roth A, Czyz E et al. Sensitive troponin I assay in early diagnosis of acute myocardial infarction. *N Engl J Med* 2009;361:868-77.
22. Reichlin T, Hochholzer W, Bassetti S, Steuer S, Stelzig C, Hartwiger S et al. Early diagnosis of myocardial

infarction with sensitive cardiac troponin assays. *N Engl J Med* 2009;361:858-67.

23. Goodacre SW, Bradburn M, Cross E, Collinson P, Gray A, Hall AS. The Randomised Assessment of Treatment using Panel Assay of Cardiac Markers (RATPAC) trial: a randomised controlled trial of point-of-care cardiac markers in the emergency department. *Heart* 2011;97:190-6.

24. Bradburn M, Goodacre SW, Fitzgerald P, Coats T, Gray A, Hassan T et al. Interhospital variation in the RATPAC Trial (Randomised Assessment of Treatment using Panel Assay of Cardiac markers). *Emerg Med J* 2012;29:233-8.

25. Collinson PO. The cost effectiveness of B-Type natriuretic peptide measurement in the primary care setting-a UK perspective. *Congest Heart Fail* 2006;12:103-7.

26. Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P et al. Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. *Health Technol Assess* 2009;13:1-207, iii.

27. Thokala P, Goodacre SW, Collinson PO, Stevens JW, Mills NL, Newby DE et al. Cost-effectiveness of presentation versus delayed troponin testing for acute myocardial infarction. *Heart* 2012;98:1498-503.

28. Waugh N, Cummins E, Royle P, Kandala NB, Shyangdan D, Arasaradnam R et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health Technol Assess* 2013;17:xv-211.

How useful are laboratory practice guidelines?

S. Misra¹, K. Moberg-Aakre², M. Langlois³, J. Watine⁴, P.J. Twomey⁵,
W.P. Oosterhuis⁶, J.H. Barth⁷

1 Diabetes, Endocrinology & Metabolism, Imperial College London, St Mary's Campus, London, UK

2 Laboratory of Clinical Biochemistry, Haukeland University Hospital, Bergen, Norway

3 Dept. Laboratory Medicine, AZ St Jan Hospital, Bruges, and Ghent University, Belgium

4 Laboratoire de Biologie Polyvalente, Hôpital de la Chartreuse, Villefranche-de-Rouergue, France

5 Clinical Chemistry, St. Vincent's University Hospital, Elm Park, Dublin, Ireland

6 Atrium-Orbis. Department of Clinical Chemistry and Haematology, Heerlen, The Netherlands

7 Blood Sciences, Old Medical School, Leeds General infirmary, Leeds, UK

ARTICLE INFO

Authorship on behalf of:

The Guidelines Working Group
of the European Federation of Clinical
Chemistry and Laboratory Medicine
and UEMS Section of Laboratory
Medicine/Medical Biopathology

ABSTRACT

Clinical practice guidelines (CPGs) relating to laboratory diagnostic testing are increasingly produced with the aim of standardizing practice and improving patient care based on the best available evidence. However, the production of a CPG is merely the first step in the process of getting evidence into practice, to be undertaken by laboratories and other stakeholders. This process should evaluate the information provided in the guidelines on laboratory tests, devise a strategy for implementing the CPG or the laboratory aspects of the CPG and finally, once implemented, assess the impact of the CPG on clinical practice, patient outcomes and costs of care.

The purpose of CPG evaluation by the laboratory is to determine whether sufficient information is provided on the particular test recommended. CPGs may not always be written with the involvement of a laboratory specialist and this underlies the paucity of relevant information in some national guidelines. When laboratory specialists are involved, CPGs can provide practical information which supports local laboratories as

well as clinicians in the implementation and appropriate use of recommendations.

Implementation of CPGs is an often neglected area that needs attention and thought. There are many barriers to successful implementation, which may vary at local level. These need to be identified early if CPGs are to be successfully adhered to. The effectiveness of CPGs also needs to be audited using process and health outcome indicators. Clinical audit is an effective tool for assessing adherence to recommendations and for measuring the impact and success of the CPG.

INTRODUCTION

Clinical practice guidelines (CPGs) relating to laboratory diagnostic testing are increasingly produced with the aim of standardizing practice and improving patient care, based on the best available evidence. There are relatively few guidelines dealing purely with the laboratory medicine aspects of patient care and the majority of laboratory related recommendations are embedded in CPGs. Unfortunately, these are often inadequately detailed to be useful for laboratories (1). Therefore laboratory medicine specialists should be more actively involved in the production of clinical guidelines to ensure that advice is given about the appropriate utilization of laboratory tests.

It is important to remember that the production of a CPG is merely the first step in a larger process that needs to be undertaken by laboratories and other stakeholders. The next step is implementation. This process, firstly, should evaluate and assess the quality of the guideline in order to ensure that a sufficient level of information is provided. Secondly, the process for guideline implementation needs to be planned with due consideration given to local barriers that may prevent guideline adoption. CPG production is a lengthy process, but largely futile

if efforts are not made to ensure adoption, dissemination and implementation. Finally once implemented, the effectiveness of the CPG needs to be evaluated. This may be through clinical audit, which is an essential tool to evaluate uptake, impact on practice, patient outcome and resource utilization.

An understanding of this process is important as it underpins good laboratory practice and forms the basis of practicing evidence-based laboratory medicine. Below we highlight key aspects of this three-step process of evaluation, implementation and audit of CPGs.

HOW SHOULD LABORATORY TEST ADVICE BE INCLUDED IN CPGs?

Guidelines are typically produced by specialist groups, often national or international societies, frequently involving only single clinical specialties. Whilst the classification of the hierarchy of evidence is well described, there appears to be no standardized approach to reporting guidelines (see Kahn *et al*, this issue). Both these factors hinder the development of good laboratory based guidance as laboratory medicine specialists are rarely included in writing committees and the evidence base for diagnostic tests is largely observational with few randomized trials assessing the impact of the diagnostic test on clinical pathways. Only observational evidence supports, for example, the use of glycated haemoglobin in the diagnosis of diabetes and that of troponin in acute coronary syndrome.

There is clearly a need to ensure that good diagnostic test guidance is included in CPGs. This has been achieved for a few disorders that are managed by a number of different disciplines and where guidelines have been written by multidisciplinary teams. Successful examples of this co-operative approach include work by the European Atherosclerosis Society (EAS) in

association with the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) which has resulted in the development of two consensus papers (2, 3) and more recently, a joint Consensus Panel which has written guidelines for lipid testing in the management of dyslipidemia and cardiovascular risk (4). Another example would be the recent British Thyroid Cancer guidelines, written by a range of clinicians and laboratory medicine specialists, and providing detailed information on the appropriate use of thyroglobulin and calcitonin assays (5). There are also inter-society collaborations whereby a practice guideline primarily developed by a laboratory medicine organization is adopted by a clinical society. For example, the Diabetes Mellitus guideline of the National Academy of Clinical Biochemistry (NACB) (6) has been adopted and published as an officially endorsed recommendation by the American Diabetes Association (7). Joint development and endorsement of CPGs by clinical and laboratory medicine societies is complementary and safeguards that the most appropriate and relevant advice is provided for the use and interpretation of laboratory results.

Similar to any other types of guidelines, the actual guideline development process requires searching for and critically appraising the current evidence for diagnostic tests. Guideline panel members need to meet face-to-face several times during the process to achieve consensus on various key issues. Firstly, they need to agree on the scope and specific key questions to be addressed in the guideline, including the pre-analytical, analytical, and post-analytical aspects of testing and related candidate biomarker(s). Secondly, panel members need to critically review the best available evidence published in the literature. These may come from analytical and clinical performance studies, randomized controlled clinical trials or meta-analyses that

assess the impact of biomarker-targeted strategies on patient outcomes. Thirdly, members need to review additional literature and formulate recommendations based on the body of evidence and considered judgment of the guideline panel. The process of writing guidelines is expensive and it is essential that all sources of funding and other conflicts of interest are clearly identified so that these factors are not used to disparage the value of the guidelines.

HOW CAN THE ADVICE ON LABORATORY TESTS IN CPGs BE OPTIMISED?

CPGs are usually produced around a clinical scenario in which a laboratory investigation plays only a small, but often critical part within the overall management of that situation. When the CPG writing committee involves no laboratory specialist, the appropriate description of the testing modality and the laboratory issues surrounding it could easily be omitted. Even when the utility of a test is thoroughly evaluated within a clearly defined clinical scenario, there is a risk that the test may then be employed in a different clinical scenario for which the diagnostic utility has not been tested. It is also important to consider whether the guideline provides appropriate methodological information about the actual test recommended, particularly when a test result or clinical decision limit is highly dependent on the assay methodology. The transferability of the evidence from one scenario to the other therefore, must be critically assessed. Arguably this should be to a level of detail above and beyond that required for the clinical aspect, given the test may be used for other purposes.

Laboratory-oriented CPGs often provide detailed and appropriate methodological information on the laboratory tests outlined; e.g. NACB guidance on laboratory testing in diabetes and on tumor markers (8, 9). However, more

commonly, guidelines produced by clinical groups without laboratory professional input, often lack sufficient information. For example in the NICE guidance on chest pain, troponin elevation is discussed, however there is no mention of non-ischaemic causes of a raised troponin, which may be of particular relevance when considering the patient groups in whom troponin is commonly requested. Nor is there any discussion regarding differences between the analytical and clinical performance of assays available on the market (10).

Strategies to improve reporting of analyte-specific laboratory information include a checklist of criteria to consider when interpreting laboratory information in CPGs. A comprehensive list was published in 2012 and suggested 33 pre-analytical, 37 analytical and 10 post-analytical items that should be addressed in a guideline process including laboratory testing (1). Twelve CPGs covering common diseases and conditions were evaluated during the development of the checklist and the mean percentage of topics dealt with by the guidelines was 33%. Information about patient status, biological and analytical interferences and sample handling were scarce in most guidelines even if the inclusion of a laboratory medicine specialist in the guideline production led to increased focus on some typical laboratory related items (e.g., sample type, sample handling and analytical variation).

The checklist has further been used to evaluate the major international CPGs that give advice on using troponins for diagnosing acute coronary syndrome (11). Of the nine CPGs studied, most of the laboratory related checklist items were not considered or needed to be updated. For example, the suggested analytical quality goals were not applicable for the high sensitivity troponin assays and important interferences that may lead to false positive or

negative diagnoses were not commonly mentioned. Recently, another group has appraised the checklist and proposed additional items and modifications (12).

The effectiveness of a CPG needs to be evaluated by assessing the potential improvement in outcome of patients who are managed by the process described in the guideline. This will firstly require an assessment of whether the guidance has been successfully implemented. Secondly, whether its advice has been adhered to and thirdly, that some tangible and measurable quality indicators have been benchmarked against other users of the guideline. It should be recognised that adherence to CPGs is a real issue to be overcome.

WHAT ARE THE BARRIERS TO GUIDELINE IMPLEMENTATION?

How should a laboratory implement a guideline and what are the barriers to implementation? There are many reasons why CPGs are not implemented and this varies with both the condition under scrutiny and the different clinical practitioners. Moreover, since a single CPG can have a number of recommendations, there will be a variation in the overall compliance with the guidelines. Finally, there may be an element of self-deception. In the early days of CPGs, Lomas *et al.* reported that obstetricians were aware of and agreed with CPG recommendations in regards to Cesarean sections but their actual practice did not reflect recommended care (13).

In general, the barriers to implementation can be classified into three domains – knowledge, attitudes and behavior (14). A study of Dutch general practitioners explored the reasons for non-compliance and the key barriers identified were lack of agreement with the recommendations, environmental factors and lack of knowledge of the guidance. The

environmental factors included time pressures, lack of resources, organizational restraints and lack of reimbursements (15). We have recently surveyed laboratory medicine specialists in England regarding two guidelines, one on diabetes and one on chronic kidney disease and found that only 41% and 12% were compliant, respectively (Barth *et al*, unpublished data 2015). Since the barriers of using CPGs in practice can be manifold, it is very important that guideline panels consider these potential issues before they start the actual development process.

HOW DO WE KNOW IF THE GUIDELINE IS EFFECTIVE?

CPGs are written after a distillation of the clinical evidence available for that condition. In the ideal case, the evidence will be of high quality and based on studies examining clinical outcome. However, when there are no outcome data or the evidence is poor, clinical audit of the guideline becomes a means of not only evaluating the adherence and the clinical effectiveness of the CPG, but also providing primary evidence for effectiveness. Meanwhile where guidelines are underpinned by high quality evidence, audit can provide a useful tool for laboratories to assist with demand management, working practices and to aid decision support.

Clinical audit is therefore an essential tool and recommendations for measurable key quality indicators should be included in all CPGs in order to aid the process of monitoring and evaluation of the guidelines' effectiveness. A systematic review suggested that evaluation through audit or other means may improve the effectiveness of the CPG on outcomes overall (16). This would indeed make audit or assessment of guideline effectiveness a

key part of the success of CPGs in changing outcomes.

At present, routine clinical audit to evaluate CPGs following their introduction is not a mandated activity. It is unclear who would be responsible for auditing the diagnostic testing in a CPG. However, since it is well recognized that audits that are not supported by the group being audited have little impact, audits of test usage should be performed by the clinicians ordering the tests. Despite this, laboratories have much to gain by auditing laboratory test utilization and the clinical and cost-effectiveness of testing. In fact, laboratory testing can be used as a surrogate marker of adherence to clinical guidelines e.g. Hb1Ac in diabetes (17), and to support laboratory-level decision making as outlined above.

National schemes for auditing laboratory practices are undertaken in the UK through the activities of the Association for Clinical Biochemistry. These audits have been used successfully to evaluate adherence and practices following CPG introduction (18, 19, 20, 21), however it is not known how many other countries have similar national audit programmes. Clinical Pathology Accreditation (CPA, UK) or other accreditation bodies stipulate the requirement of laboratory practices to be audited regularly (22). Despite the value of this activity in assessing the uptake and wider implementation of best laboratory practice, there is no formal obligation for auditing CPG compliance at present.

CONCLUSIONS

CPGs are widespread and being increasingly produced. Other articles in this journal have focussed on the role of laboratories in synthesizing the evidence-base underpinning guidelines and in ensuring the quality of guideline production. However, the production of a CPG

is merely the first step of a complex process that ultimately puts the best available evidence into daily clinical practice. This process, firstly, involves an evaluation of the laboratory information contained within the CPG to determine if any relevant information is missing. Secondly, attempts should be made to encourage that laboratory professionals are included in CPG development. Thirdly, strategies need to be developed to enhance compliance with national and international CPGs and some form of evaluation, through audit or other means, is developed after the guideline is published and disseminated. The laboratory should rightly be involved in each of these steps, if it is to subscribe to evidence-based best practice.

REFERENCES

1. Aakre KM, Langlois MR, Watine J, Barth JH, Baum H, Collinson P, Laitinen P, Oosterhuis WP. Critical review of laboratory investigations in clinical practice guidelines: proposals for the description of investigations. *Clin Chem Lab Med* 2013;51:1217-26.
2. Nordestgaard BG, Chapman MJ, Ray K, et al. Lipoprotein(a) as a cardiovascular risk factor: current status. *Eur Heart J* 2010;31:2844-53.
3. Nordestgaard BG, Chapman MJ, Humphries SE et al. Familial hypercholesterolemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease. *Eur Heart J* 2013;34:3478-90.
4. Langlois MR, Descamps OS, van der Laarse A, et al. Clinical impact of direct HDLc and LDLc method bias in hypertriglyceridemia. A simulation study of the EAS-EFLM Collaborative Project Group. *Atherosclerosis* 2014;233:83-90.
5. Perros P, Colley S, Boelaert K, Evans C, Evans RM, Gerard GE, Gilbert JA, Harrison B, Johnson SJ, Giles TE, Moss L, Lewington V, Newbold KL, Taylor J, Thakker RV, Watkinson J, Williams GR. Guidelines for the management of thyroid cancer. *Clin Endo* 2014;81 suppl 1:1-122.
6. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Kirkman MS, Lernmark A, Metzger BE, Nathan DM. Special Report. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Clin Chem* 2011; 57(6): e1-e47
7. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Kirkman MS, Lernmark A, Metzger BE, Nathan DM. Position Statement Executive Summary: Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus. *Diabetes Care* 2011; 34:1419-1423.
8. Sturgeon CM, Duffy MJ, Stenman UH, Lilja H, Br nner N, Chan DW, et al. Babaian R, Bast RC Jr, Dowell B, Esteva FJ, Haglund C, Harbeck N, Hayes DF, Holten-Andersen M, Klee GG, Lamerz R, Looijenga LH, Molina R, Nielsen HJ, Rittenhouse H, Semjonow A, Shih leM, Sibley P, S l toromos G, Stephan C, Sokoll L, Hoffman BR, Diamandis EP. National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast and ovarian cancers. *Clin Chem* 2008; 54: e11-e79.
9. Sacks DB, Arnold M, Bakris GL, et al. Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus. *Diabetes Care* 2011;34:e61-e99.
10. National Institute of Clinical Excellence. Chest pain of recent onset, NICE Clinical Guideline 95. 2010.
11. Aakre KM, Langlois MR, Barth JH, Misra S, Watine J, Oosterhuis WP. [The quality of laboratory aspects of troponin testing in clinical practice guidelines and consensus documents needs to be improved](#). *Clin Chim Acta*, 2014: 437, 58-61.
12. Wils J, Fonfr de M, Augereau C and Watine J. Further comments on 'Critical review of laboratory investigations in clinical practice guidelines: proposals for the description of investigation'. *Clin Chem Lab Med* 2014;52:e155
13. Lomas J, Andreson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice. *New Eng J Med* 1989;321:1306-11.
14. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, Rubian HR. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999;282:1458-65.
15. Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. *Implement Sci* 2009;4:54
16. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic Review of rigorous evaluations. *Lancet*. 1993; 342: 1317-22.
17. Norman C, Zarrinkoub R, Hasselstrom J, Godman B, Granath F & Wettermark B. Potential savings without compromising the quality of care. *Int J Clin Pract*. 2009;63:1320-26.
18. <http://www.acb.org.uk/whatwedo/science/audit/audits-by-topic>

19. Kilpatrick ES, Verrill H; National Clinical Biochemistry Audit Group. A national audit of estimated glomerular filtration rate and proteinuria reporting in the UK. *Ann Clin Biochem* 2011;48:558-61.

20. Barlow IM, Mussa H, Hyde P, Jagger C. An audit of cerebrospinal fluid bilirubin analysis against national audit guidelines/standards within a UK pathology network. *Ann Clin Biochem* 2011;48:479-80.

21. Hadfield SG, Horara S, Starr BJ, Yazdgerdi S, Marks D, Bhatnagar D et al?, Cramb R, Egan S, Everdell R, Ferns

G, Jones A, Marenah CB, Marples J, Prinsloo P, Sneyd A, Stewart MF, Sandle L, Wang T, Watson MS, Humphries SE; Steering Group for the Department of Health Familial Hypercholesterolaemia Cascade Testing Audit Project. Family tracing to identify patients with familial hypercholesterolaemia: the second audit of the Department of Health Familial Hypercholesterolaemia Cascade Testing Project. *Ann Clin Biochem* 2009;46:24-32.

22. Misra S, Barth JH. [How good is the evidence base for test selection in clinical guidelines?](#) *Clin Chim Acta* 2014; 432, 27-32.

Working terms

Xavier Fuentes-Arderiu

Clinical Laboratory Sciences Consulting, Barcelona, Catalonia, Spain

INFO

Corresponding author:

Xavier Fuentes-Arderiu
Clinical Laboratory Sciences Consulting
Barcelona, Catalonia
Spain
E-mail: 2461xfa@gmail.com

Key words:

terminology, standardization,
translation, management

LETTER TO THE EDITOR

Many international bodies have recommended systematic terms to describe quantities and other properties in clinical laboratory sciences (1, 2).

These systematic terms are essential to understanding the foundations of clinical laboratory sciences; however, it has been demonstrated that their implementation in the actual environment of a clinical laboratory is very difficult.

The few clinical laboratories that have adopted these terms in their day-to-day usage is a demonstration of such an implementation being very difficult. Thus, an easier alternative (herein called *working terms*) to systematic terms can be more acceptable, in the same way as enzyme nomenclature, which has systematic names (too long to be convenient for practical use) and working names (more convenient for practical use). In addition, in the day-to-day practice of a clinical laboratory (and probably in other kinds of laboratories), the proposed working terms have the advantage over the systematic ones of being more easily translated from English to other languages.

Starting with the concept (and term) of **property**, defined as “that which when possessed by an object

contributes to it being as it is" (e.g. mass concentration is 50 mg/L; colour is yellow) (3), this concept can be divided into four related concepts (and terms) having different levels of abstraction regarding the object involved:

1) Working term for concept: **generic property**

- Concept definition: "property that refers neither to any system, nor to any component" (4,5)
- Corresponding systematic term: *kind-of-property*
- EXAMPLES: Mass concentration, form.

2) Working term for concept: **subgeneric property**

- Concept definition: "property that does not refer to a system, but refers to a given component of a system, although considered abstractly" (4,5)
- Corresponding systematic term: None proposed
- EXAMPLES: Mass concentration of protein; form of bacteria.

3) Working term for concept: **specific property**

- Concept definition: "property that refers to a given system, or to a given system and some of its components, although considered abstractly" (4,5)
- Corresponding systematic term: *dedicated kind-of-property*.
- EXAMPLES: Mass concentration of protein in blood plasma; form of bacteria in tap water.

4) Working term for concept: **individual property**

- Concept definition: "property that refers to a given system, or to a given system and some of its components, spatiotemporally defined" (4,5)
- Corresponding systematic term: None proposed, but described as an instance of a *dedicated kind-of-property*
- EXAMPLES: Mass concentration of protein in the blood plasma of the patient YZ, on day D, at time T; form of bacteria in Barcelona tap water, on day D, at time T.

In all cases, the same applies to **quantity**, changing the terms, definitions, and examples accordingly.

REFERENCES

1. International Union of Pure and Applied Chemistry, Clinical Chemistry, International Federation of Clinical Chemistry and Laboratory Medicine. Prepared for publication by U. Magdal Petersen, R. Dybkær, and H. Olesen. Properties and units in the clinical laboratory sciences. Part XXIII. The NPU terminology, principles, and implementation: A user's guide (IUPAC Technical Report). Pure Appl Chem 2012;84:137–65.
2. European Committee for Standardization. Healthcare informatics – Structure for nomenclature, classification, and coding of properties in clinical laboratory sciences. EN1614:2006. Brussels: CEN; 2006.
3. Fuentes-Arderiu X. Vocabulary of terms in protometrology. Accred Qual Assur 2006; 11:640–3.
4. Fuentes-Arderiu X. Ordinal quantities in clinical laboratory sciences. Accred Qual Assur 2012;17:349–51
5. Fuentes-Arderiu X. Taxonomy of quantities. Biochem Med 2012;22:274–5.



Editor-in-chief

Gábor L. Kovács

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

Assistant Editor

Harjit Pal Bhattoa

Department of Laboratory Medicine, University of Debrecen, 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

Mike Hallworth, Shrewsbury, United Kingdom

Ellis Jacobs, Alere Inc., 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

Peter Vervaart, LabMed Consulting, Australia

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:


IFCC
International Federation
of Clinical Chemistry
and Laboratory Medicine

www.ifcc.org