

Procjena tumačenja rezultata pretraga u laboratorijskoj medicini

The assessment of interpretation of test results in laboratory medicine

Gordon S. Challand¹, Ping Li²

¹Zavod za kliničku biokemiju, Bolnica Royal Berkshire, Reading, Berkshire, Velika Britanija

¹Department of Clinical Biochemistry, Royal Berkshire Hospital, Reading, Berkshire, UK

²Zavod za laboratorijsku medicinu, Medicinski fakultet i bolnica Zapadne Kine, Sveučilište u Sichuanu, Chengdu, Sichuan 610041, NR Kina

²Department of Laboratory Medicine, West China Hospital and School of Medicine, Sichuan University, Chengdu, Sichuan 610041, PR China

Sažetak

Sve je veća potreba za laboratorijskim stručnjacima koji mogu pružiti savjete o tumačenju rezultata, što obično čine dodavanjem kratkih napomena na nalazu. Iznenadjuće je koliko je malo podataka temeljenih na dokazima dostupno za tumačenje rezultata laboratorijskih pretraga. Kratke napomene koje prate laboratorijske nalaze su kompleksne te se obično sastoje od nekoliko sastavnica kojima se mogu predložiti moguće dijagnoze i dodatne pretrage. Svaka je takva napomena različita i procjena tumačenja nije jednostavna. U ovom ćemo članku govoriti o različitim mogućim pristupima: procjeni cijele napomene, samo nekih njenih sastavnica ili ključnih riječi, te sudjelovanju nezavisnog procjenitelja ili skupine stručnjaka. Niti jedan se pristup ne može smatrati zlatnim standardom, budući da procjena predstavlja nesavršenu znanost te je kao takva samo smjernica, a ne definicija jedinstvenih rješenja. Iako programi vanjske procjene kvalitete (engl. *External Quality Assurance Schemes*, EQAS) ispituju tumačenje rezultata te na taj način pružaju sudionicima programa informaciju o tome kakve su njihove napomene u usporedbi s drugima, primarna je svrha tih programa ipak obrazovna i, što je još bitnije, ti nam programi pomažu u prikupljanju spoznaja o tome kako poboljšati i razviti tumačenje.

Ključne riječi: tumačenje rezultata; program vanjske procjene kvalitete

Abstract

There is a growing need for laboratory professionals to offer advice on the interpretation of results, typically by adding a brief interpretative comment to a report. Surprisingly little evidence-based data exists to support interpretation of laboratory test results, and interpretative comments accompanying laboratory reports are complex, usually consisting of several components which may suggest possible diagnoses and additional investigations. Every comment is different, and assessment of interpretation is difficult. We report different approaches that can be used: assessing whole comments or comment components or key phrases; and using independent assessors or a pooled panel of experts. No approach can be considered to be a 'gold standard' since assessment is an imperfect science and is a guide to, not a definition of unequivocal solutions. Although External Quality Assurance Schemes examining interpretation provide information to individual participants on how their comments compare with others, their primary purpose is educational, and even more importantly, these Schemes enable us to gather knowledge on how to better establish and develop interpretation.

Key words: interpretation of results; External Quality Assurance Schemes

Pristiglo: 25. veljače 2009.

Prihvaćeno: 21. travnja 2009.

Received: February 25, 2009

Accepted: April 21, 2009

Uvod

Cjelokupni ciklus laboratorijskog testiranja, poznat kao Lundbergerova petlja (engl. *Lundberger's 'brain-to-brain loop'*), opisuje sve korake od naručivanja adekvatne pretrage do ispravnog postupka i tumačenja rezultata za učinkovito donošenje kliničke odluke (1,2). Prije četrde-

Introduction

The total laboratory testing cycle, formerly known as Lundberger's 'brain-to-brain loop', describes all steps ranging from ordering the right test to correctly reacting and interpreting test results for effective clinical decision-making (1,2). Forty years ago, the majority of laboratory

set godina većina je zahtjeva za laboratorijskim pretragama dolazila od bolničkoga kliničkog osoblja, koje je u sklopu svoga visokoga medicinskog obrazovanja završilo školovanje iz područja patologije i laboratorijske medicine. Osim u kompliciranim slučajevima i onima koji su zahtijevali specijalističko stručno mišljenje, nije bilo velike potrebe da osoblje laboratorija daje savjete o tumačenju rezultata. U mnogim je zemljama došlo do smanjenja obujma osnovnih znanstvenih disciplina koje obuhvaća dodiplomsko medicinsko obrazovanje (3), dramatično je porasla kompleksnost laboratorijske dijagnostike, te većina kliničara traži pretrage s kojima sami nisu dovoljno upoznati. Stoga je sve veća potreba za savjetima iz laboratorija, koji bi pomogli u prikladnosti i učinkovitom tumačenju rezultata pretraga (1,4). Idealno bi bilo da se takvi savjeti daju telefonskim putem ili razgovorom liječnika na odjelima, no to je onemogućeno postojećim radnim opterećenjem laboratorija i činjenicom da većina uputnica za laboratorijske pretrage dolazi od liječnika primarne zdravstvene zaštite. U Ujedinjenom Kraljevstvu većina laboratorija ima zaposlenog biokemičara starije službe (engl. *Duty Biochemist*), koji tumači rezultate biokemijskih pretraga prilažući nalazu kratku napomenu. Iako bi se o toj praksi moglo raspravljati (5), prema anketama kliničko je osoblje uvelike odobrava i cijeni (6) te smatra da utječe na obradu bolesnika (7). Čak će i laboratoriji koji inače ne pružaju tu uslugu na izravan zahtjev kliničara dati napomenu s tumačenjem rezultata (8). Akreditacijski standardi kliničke patologije Ujedinjenog Kraljevstva (engl. *Clinical Pathology Accreditation (UK) Standards*) navode da je tumačenje rezultata bitna sastavnica usluge koju pružaju klinički laboratoriji (9). Međutim, ne slažu se svi oko toga koliko se često takve napomene zahtijevaju i upotrebljavaju u laboratorijskoj medicini, te oko prikladnosti ili nekih drugih značajki određenih napomena u danoj situaciji. Prije desetak godina biokemičar starije službe često je radio u izolaciji s vrlo malo ili čak bez povratnih informacija od korisnika usluga ili svojih kolega te je iznenađujuće koliko je malo podataka temeljenih na dokazima dostupno čak i za tumačenje uobičajenih laboratorijskih patoloških rezultata.

Godine 1997. započelo se s jednom interpretativnom vježbom za tumačenje rezultata (engl. *'Cases for Comment'*; Slučajevi za tumačenje rezultata), koja se distribuirala preko baze elektronskih adresa za opće rasprave Društva kliničkih biokemičara (engl. *Association of Clinical Biochemists, ABC*) (acb-clin-chem-gen@jiscmail.ac.uk) (10). Godine 2001. ta je vježba prerasla u formalni program vanjske procjene kvalitete (engl. *External Quality Assurance Scheme, EQAS*) tumačenja rezultata kliničke biokemije (11). Godine 2000. pokrenut je sličan pilot-program u Australiji, no bez procjene prikladnosti napomena, jer se smatralo da će sudionici shvatiti što je primjereno kada se njihove napomene usporede s napomenama referentne skupine

requests were from hospital-based clinical staff who had received formal training in Pathology and Laboratory Medicine while attending Medical Schools. Except in difficult or in highly specialised Cases, there was little need for laboratory staff to offer advice on the interpretation of results. In many countries there has been a decline in the amount of basic sciences taught at medical undergraduate level (3), the complexity of laboratory diagnostics has dramatically increased, and many clinical staff ask for tests which they themselves are unfamiliar with. There is therefore a growing need for interpretative advice from the laboratory both to help appropriateness and effective interpretation of test results (1,4). Ideally, such advice would be provided through telephone conversations or visits to wards, but this is hampered by the size of present-day workloads, and by the fact that the majority of laboratory requests may come from primary care physicians. In the UK, the majority of laboratories use a Duty Biochemist (a senior member of medical or scientific staff holding a professional qualification) to offer interpretative advice on Clinical Biochemistry results, usually by attaching a brief comment to a report. Although this is itself controversial (5), surveys of clinical staff indicate both that this is very much welcomed and appreciated (6) and that it influences patient management (7). Even in those laboratories where such a service is not provided, an interpretative comment is usually available on direct enquiry by a clinician (8). Clinical Pathology Accreditation (UK) Standards state that interpretation of results is an important component of the service provided by clinical laboratories (9). However, there is little consensus on the required frequency or the user needs for such comments in laboratory medicine, let alone the appropriateness or otherwise of specific comments in a given situation. Ten years ago the Duty Biochemist often worked in isolation with little or no feedback from users of the service or from colleagues, and it is surprising how little evidence based data there is for interpretation even of common laboratory abnormalities.

In 1997 an interpretative exercise 'Cases for Comment' began to be distributed through the general discussion mailbase of the ACB (acb-clin-chem-gen@jiscmail.ac.uk) (10); in 2001 this was developed into a formal EQAS for interpretation in Clinical Biochemistry (11). In 2000 a similar scheme was piloted in Australia but without any assessment of the appropriateness of comments as it was felt that participants would get a sense of what was an appropriate comment by peer group comparison (12). This turned out not to be the case: at the end of the pilot survey, both participants and organisers felt the need for formal assessment and feedback on the appropriateness of the comments (13). The form of assessment used by each scheme has been different and each has advantages and disadvantages (14).

(12). To se, međutim, nije dogodilo; na kraju tog pilot-istraživanja i sudionici i organizatori osjetili su potrebu za formalnom procjenom i povratnom informacijom o prikladnosti napomena (13). Svaki program ima drugačiji oblik procjene sa svojim prednostima i nedostacima (14).

Metode procjene

Slučajevi za tumačenje rezultata (10)

Organizator je napomene raščlanio u sastavnice od kojih je referentna skupina (engl. *peer review group*) svakoj dodijelila brojčanu ocjenu. Svi su članovi te skupine bili stručno kvalificirani te su radili odvojeno jedan od drugoga. Izračunom srednjih vrijednosti ocjene svake sastavnice, dobila se procjena njene prikladnosti. Tablica 1 pruža prikaz jednoga takvog slučaja. Budući da je distribuiran elektronskom poštom, to je najneformalniji od svih pristupa. Tijekom 4 godine trajanja tog programa distribuirano je više od stotinu primjera.

Nacionalni program vanjske procjene kvalitete Ujedinjenog Kraljevstva (engl. *United Kingdom National External Quality Assessment Service, UK NEQAS*) za kratke napomene (11,16)

Ovim se pristupom primjeri i prikazi distribuiraju preko mrežnih stranica te se procjenjuju cjelovite napomene, a ne pojedine sastavnice napomena, čime su onemogućeni subjektivni elementi prisutni pri raščlanjivanju napomena u sastavnice ili ključne riječi. I kod ovog pristupa nezavisna referentna skupina dodjeljuje brojčanu ocjenu svakoj napomeni. Srednja vrijednost ocjena omogućuje stupnjevanje svih napomena prema njihovoj prikladnosti, a u istraživanje su uključeni i primjeri s niskom, srednjom i visokom prikladnošću. Tablica 2. prikazuje primjer ovog pristupa. Svake je godine bilo distribuirano barem 20 primjera.

Australski program procjene kvalitete (engl. *Royal College of Pathologists of Australasia-Quality Assurance Program, RCPA-QAP*) (12,13)

Organizator raščlanjuje napomene na ključne fraze koje procjenjuje skupina (engl. *expert panel*) sastavljena od stručnjaka iz određenih specijalističkih i stručnih područja koji surađujući klasificiraju te fraze kao poželjne, manje vrijedne ili neprikladne. Skupina stručnjaka također predlaže adekvatnu napomenu (12). Budući da se distribuira običnom poštom, ovaj je pristup najformalniji od svih. U svojoj je biti sličan izvornom pristupu „Slučajeva za napomene“ zbog raščlambe napomene na sastavnice ili ključne fraze, no razlikuje se u tome što se usuglašeni usmeni sud donosi kroz raspravu unutar skupine stručnjaka, a ne kao srednja brojčana ocjena koju šalju procjenitelji radeći nezavisno jedan od drugoga. Prednost ovog pristupa leži u činjenici da će sudionici radije prihvatiti usmenu ocje-

Methods of the survey

Cases for Comment (10)

The organiser broke down comments into components, each of which was then allotted a numerical score by a peer review group all group members holding professional qualifications and working in isolation from each other. The mean score for each component then gave an assessment of its comparative appropriateness. An example of a case summary produced using this approach is shown in Table 1. Distributed by e-mail, this was the least formal of the approaches. One hundred Cases were distributed over the four years this Scheme operated.

The UK NEQAS for Interpretative Comments (11,16)

This distributes Cases and Summaries through a web page, and assesses whole comments rather than individual components which prevents the subjective element inherent in breaking down comments into components or key phrases, but again uses independent peer review to allot a numerical score to each comment. The mean score enables ranking of all comments in terms of their appropriateness, and examples of low, median and high scoring comments are included in the summary. An example of this approach is shown in Table 2. At least 20 Cases are distributed each year.

The Australian RCPA-QAP for Interpretative Comments (12,13)

The organiser breaks down comments into key phrases which are assessed by an expert panel consisting of peers holding specified offices in professional bodies who work together to classify them as preferred, of lesser value, or inappropriate. A suggested comment is also made by the expert panel (12). Distributed by conventional mail, it is the most formal of the approaches. In its essence, this is similar to the original 'Cases for Comment' in its breakdown of comments into components or key phrases, but it differs in that a consensus verbal view is reached through discussion within the expert panel group, rather than a mean numerical view reached by assessors working independently. An advantage is that the expert panel consensus view is more likely to be accepted by participants than the mean view obtained from independent assessors. A limitation is that the necessity for expert group discussion and the distribution by conventional mail limit the practicable case distribution to around 10 each year.

Results of the survey

Although in general all approaches to assessment give broadly similar results, there is not a particularly good correlation between the assessment schemes (Vasikaran S D, unpublished observation) and on occasions there is a profound difference both between assessors and

TABLICA 1. Pregled 'Slučajevi za tumačenje rezultata': nezavisnan procjenitelj ocjenjuje sastavnice napomene

Dvadesetdvogodišnjakinja na pregledu kod svog obiteljskog liječnika. Anamneza glasi: Prekid trudnoće krajem kolovoza. Nezaštićeni odnos 20. rujna, uslijedila hitna kontracepcija („pilula za dan poslije“) i menstruacija. Nezaštićeni odnos 26. listopada. Kućni test za trudnoću pozitivan. Status?’

Koncentracija humanoga korionskog gonadotropina (HCG) u serumu je 240 U/L

U ovom je slučaju sudjelovalo 46 ispitanika od kojih je 8 telefonskim putem raspravljalo o njemu.

Deset ispitanika je komentiralo ‘trudna’ [0,0]; dva su komentirala ‘rana trudnoća’ [0,8]; pet ih je komentiralo ‘vrlo rana trudnoća’ [1,0]; jedan je rekao ‘trudna 1-2 tjedna’ [-0,3]; jedan je rekao ‘trudna 3–4 tjedana’ [0,5]; pet ih je reklo ‘trudna 4-5 tjedana’ [1,0].

Dva ispitanika su rekla ‘trudnoća rezultat događaja od 26. listopada’ [0,3]; dva su rekla ‘moguće rezultat događaja od 20. rujna’ [0,3]; šest ih je reklo da je ‘rezultat previsok kako bi se mogao odnositi na događaj od 26. listopada’ [1,0]; šest ih je reklo da je ‘trudnoća uslijedila nakon događaja od 20. rujna’ [1,3].

Pet ispitanika je reklo da bi se rezultat mogao povezati s nepotpunim prekidom trudnoće [0,3]; dva su rekla da bi to mogla biti ostatna koncentracija HCG od ranijeg prekida trudnoće [0,8]; dva ispitanika su spomenula mogućnost korionkarcinoma [0,3]; dva su spomenula mogućnost izvanmaternične trudnoće [0,8].

Dvanaest ispitanika bi predložilo ponovljeno određivanje koncentracije HCG nakon 2 dana [1,5]; osmero bi predložilo ponovljeno određivanje nakon 1 tjedna [1,5]; jedan nakon 2 tjedna [0,0]; jedan nakon 3 tjedana [-0,5].

Dva ispitanika bi zatražila uzorak mokraće za test trudnoće [-0,5]; jedan bi predložio određivanje koncentracije progesterona s ponovljenim određivanjem koncentracije HCG [-1,0]; devetero bi uputilo bolesnicu na ultrazvuk [-0,5].

Komentar organizatora programa bio je ‘Određena koncentracija HCG je prosječna koncentracija HCG za 30. dan trudnoće, preniska za začeće kod prvog događaja, a vjerojatno je previsoka za začeće kod drugog. Mogla bi biti ostatna koncentracija HCG od prošlog prekida trudnoće, no trudnoća kao rezultat drugog događaja ne može se isključiti. Predlaže se ponovljena pretraga za tjedan dana’.

nu skupine stručnjaka nego brojčanu ocjenu nezavisnih procjenitelja. Ograničenje ovog pristupa je u nužnosti održavanja sastanaka zbog rasprave skupine stručnjaka, te poštanskoj distribuciji koja ograničava distribuiranje slučajeva na 10 slučajeva godišnje.

TABLE 1. A ‘Cases for Comment’ summary: comment component marking by independent assessors

A 22 year old girl seeing her family doctor. Clinical information is ‘Termination of pregnancy at end of August. Unprotected intercourse September 20th, followed by ‘morning-after’ pill and a period. Unprotected intercourse October 26th. Positive home pregnancy test. Status?’

Serum HCG was 240 U/L

This Case attracted 46 participants, 8 of whom would phone to discuss the Case.

10 participants commented ‘pregnant’ [0.0]; 2 commented ‘early pregnancy’ [0.8]; 5 commented ‘very early pregnancy’ [1.0]; 1 said ‘pregnant 1 – 2 weeks’ [-0.3]; 1 said ‘pregnant 3–4 weeks’ [0.5]; 5 said ‘pregnant 4 – 5 weeks’ [1.0].

2 said ‘consistent with 26 October event’ [0.3]; 2 said ‘possibly consistent with 20 September event’ [0.3]; 6 said the result is too high to be related to the October 26 event [1.0]; 6 said ‘pregnant subsequent to the September 20 event’ [1.3].

5 said the result could be consistent with incomplete termination of pregnancy [0.3]; 2 said that this could be residual HCG from the earlier termination [0.8]; 2 mentioned the possibility of choriocarcinoma [0.3]; 2 mentioned the possibility of an ectopic pregnancy [0.8].

12 would suggest a repeat HCG after 2 days [1.5]; 8 would suggest a repeat after 1 week [1.5]; 1 after 2 weeks [0.0]; 1 after 3 weeks [-0.5].

2 would ask for a urine sample for a pregnancy test [-0.5]; 1 would suggest a progesterone with the repeat HCG [-1.0]; 9 would suggest referral for ultrasound [-0.5].

The Scheme Organiser’s comment was ‘Average for day 30 pregnancy, too low for first event and likely to be too high for second. Could be residual HCG from earlier termination, but pregnancy arising from second event cannot be excluded. Suggest repeat after 1 week’.

between members of an expert panel on the value of a comment, whether it is assessed in its entirety or broken down into components (12). There is some evidence from the UK NEQAS that assessors find it easier to agree on what might be called good or bad comments than on intermediate comments; and there may also be differences between what could be called a Teaching Hospital (or specialised centre) approach and a General Hospital (or non-specialised centre) approach, particularly in the extent to which follow-up tests are suggested or carried out (Challand G S, unpublished observations).

TABLICA 2. Prikaz slučaja iz programa UK NEQAS: nezavisni procjenitelji ocjenjuju cijeli komentar**TABLE 2.** A case summary from the UK NEQAS scheme: whole comment marking by independent assessors

Dvadesetogodišnjakinja je primljena na odjel bolničke hitne kirurške službe s bolovima u trbuhu. Klinički detalji s uputnice su bol u desnom gornjem kvadrantu (engl. *right upper quadrant pain*, RUQ pain). Aktivnost amilaze u serumu bila je unutar referentnih granica.

Sljedeći dan je premještena na odjel kirurgije te je odatle stigao uzorak s anamnezom „Bol u trbuhu. Izvanmaternična trudnoća?“.

Koncentracija ureje i elektrolita u serumu je bila unutar referentnih granica. Koncentracija HCG u serumu bila je 345.000 U/L.

U ovom je slučaju sudjelovalo 143 ispitanika od kojih je 140 dobilo pozitivnu ocjenu (medijan (interkvartilni raspon) = 1,22 (0,82-1,50)). Gotovo su svi ispitanici napomenuli kako visoka koncentracija HCG ukazuje da nije vjerojatno da se radi o izvanmaterničnoj trudnoći. Mnogi su se ispitanici pitali je li ona uzrokovana višeplođnom trudnoćom ili tumorom koji luči HCG; u polovini napomena spomenuta je potreba za ultrazvukom.

Zanimljivo je da je nekoliko sudionika koji rade na analizatoru DPC Immulite smatralo da je koncentracija HCG normalna za trudnoću. Iako analizator DPC za trudnoću od 7-11 tjedana navodi medijan vrijednosti koncentracije HCG 118.000 U/L, njihov raspon je do 291.000 U/L kod jednoplođnih trudnoća. Nama koji radimo na uređaju Bayer Centaur neobično je vidjeti koncentracije HCG koje su više od 180.000 U/L kod jednoplođnih trudnoća.

Bolesnica je premještena na odjel porodništva. Bol u trbuhu se stišala, standardne pretrage funkcije jetre dale su normalne rezultate. Ultrazvuk napravljen nešto kasnije tog dana pokazao je 10. tjedan blizanačke trudnoće. Sljedeći je dan bolesnica otpuštena te se osjećala bolje.

Komentar za nisku ocjenu bio je:

Rezultati upućuju na trudnoću. Ponoviti određivanje aktivnosti amilaze za kontrolu bolesnika – očekuje se povišenje aktivnosti amilaze ukoliko je došlo do rupture jajovoda.

Komentar za srednju ocjenu bio je:

Koncentracija HCG previsoka za izvanmaterničnu trudnoću, iako je ona moguća. Treba isključiti trofoblastični tumor.

Komentar za visoku ocjenu bio je:

Koncentracija HCG previsoka za izvanmaterničnu trudnoću; u stvari je vrlo visoka i za normalnu trudnoću. Blizanačku trudnoću i trudnoću s mrtvim plodom treba isključiti. Predlaže se ultrazvuk.

A 23 year old woman was admitted to the Emergency Surgical Unit of the hospital with abdominal pain. The clinical details on the initial request form were 'RUQ pain'. Serum amylase was within reference limits.

The following day she was transferred to a surgical ward, and a further sample was received, with clinical information 'Abdominal pain. Tubal ectopic?'.

Serum urea and electrolytes were normal. Serum HCG was 345,000 U/L.

This Case attracted 143 participants, 140 achieving a positive score (median and interquartile range = 1.22 (0.82-1.50)). Nearly all the participants commented that this high HCG level suggested that an ectopic pregnancy was unlikely. Most participants queried if this was caused by a multiple pregnancy or an HCG secreting tumour; half of these comments mentioned the need for an ultrasound scan.

It is interesting that a few DPC Immulite users felt that this HCG was normal for pregnancy. Although DPC quote a median HCG of 118,000 U/L for gestation of 7-11 weeks, their quoted range is up to 291,000 U/L in singleton pregnancies. It is unusual from our experience with the Bayer Centaur to see HCG levels of more than 180,000 U/L in singleton pregnancies.

This lady was transferred to the care of the Obstetricians. Her abdominal pain had resolved and liver function tests were normal. An ultrasound scan later that day showed a twin pregnancy of 10 weeks gestation. She was discharged feeling well the following day.

A low scoring comment was

Results consistent with pregnancy. Repeat amylase to follow on this sample – would expect an increase if a tubal pregnancy has ruptured.

A median scoring comment was

HCG level too high for ectopic, although possible. Trophoblastic tumour should be excluded.

A high scoring comment was

HCG is too high for ectopic pregnancy; in fact it is very high for a normal pregnancy & a twin/ molar pregnancy should be excluded. Suggest scan.

Rezultati procjene

Iako, općenito gledano, svi pristupi procjeni nude uglavnom slične rezultate, ne postoji osobito dobra korelacija između različitih programa procjene (Vasikaran SD, neobjavljeni podaci) te ponekad dolazi do bitne razlike u ocjeni vrijednosti napomene s obzirom na to je li napomenu procijenio nezavisni procjenitelj ili članovi stručne skupine, odnosno procjenjuje li se napomena u cijelosti ili je raščlanjena na sastavnice (12). Pristup UK NEQAS pokazao je da se procjenitelji lakše slože oko toga što znači dobra ili loša napomena, a teže oko toga što bi bila napomena srednje vrijednosti; također se mogu pojaviti razlike između pristupa kliničke bolnice (engl. *Teaching hospital approach*) (ili specijaliziranog centra) i pristupa opće bolnice (engl. *General Hospital approach*) (ili nespecijaliziranog centra), posebno u obujmu u kojem se predlažu i vrše pretrage u sklopu praćenja (Challand GS, neobjavljeni podaci).

Tumačenje rezultata

Ne postoji idealno rješenje vezano za metode koje se mogu rabiti u procjeni interpretativnih napomena. U njima se ne bi smjelo govoriti o ishodu nekog stanja, budući da je ishod u trenutku pisanja napomene nepoznat, možda nikada neće biti poznat, a može biti da je neuobičajen uzrok poznatih patologija. Usuglašene se ocjene (povezane s metodom srednje vrijednosti kod analitičke procjene kvalitete) ne mogu rabiti jer se suglasnost rijetko postiže (15,16), a čak i kada se postigne može doći do širokog razilaženja u mišljenjima o prikladnom praćenju stanja. Usprkos činjenici da su na prvi pogled jednostavne, kratke su napomene kompleksne i tipično sadrže barem tri različite stavke (npr. predlažu vjerojatnu dijagnozu; predlažu koja se dijagnoza može isključiti; te predlažu dodatne pretrage) od kojih se svaka može smatrati prikladnom ili neprikladnom. Štoviše, napomena zahtijeva određenu vještinu u prenošenju informacija. Procijeniti se može cijela napomena ili njene pojedinačne sastavnice. Procjenitelji mogu biti organizator programa vanjske procjene kvalitete, skupina stručnjaka ili referentna skupina; mogu raditi zajedno, pa donijeti opisnu ocjenu ili pojedinačno (nezavisni procjenjivači), pa dati brojčanu ocjenu u obliku aritmetičkog prosjeka. Prednost procjenjivanja sastavnica leži u tome da su mnoge od njih poznate većini sudionika te da gotovo svi sudionici mogu vidjeti da je dio njihovih napomena bio dobar. Nedostatak jest da raščlamba napomene na njene sastavnice može biti prilično subjektivna. Prednost procjenjivanja cijele napomene je u tome da se na taj način izbjegava subjektivan pristup (osobna pristranost) te se ocjenjuju komunikacijske vještine kao i interpretativne ideje. Nedostatak je da taj pristup nedvojbeno predstavlja puno veći izazov i za sudionika i za procjeni-

Interpretation issues

There is no ideal solution to the methods that can be used to assess interpretative comments. The outcome of a Case should not be used, since at the time advice is offered the outcome is unknown, may never be known, and may be an unusual cause of a common set of abnormalities. The consensus views (related to the method mean in analytical QA) cannot be used since a consensus seldom occurs (15,16), and even when it does there may be wide differences of opinion in appropriate follow-up investigations.

Despite being superficially simple, interpretative comments are complex and typically contain at least three distinct ideas (for example suggesting a probable diagnosis; suggesting which diagnoses which can be excluded; and suggesting additional investigations), each of which may be regarded as appropriate or inappropriate. Moreover, a comment requires communication skills. Either the whole comment or its individual components can be assessed, and the assessment can be carried out by an EQAS Organiser, by a group of experts, or by a group of peers, the groups working either together to arrive at a consensus view, or independently to provide an arithmetically average view. An advantage to assessing components is that many of these are common to most participants, and that almost all participants can feel that part of their comment was good. A disadvantage is that the breakdown of each comment into components is rather subjective. An advantage to assessing whole comments is that this avoids a subjective (personally biased) approach, and that it tests communication skills as well as interpretative ideas. A disadvantage is that it is undoubtedly more challenging both for the participants and the assessors. We also have to accept that any assessment process inevitably includes an element of ranking each single comment against the others: there are seldom definitive guidelines which can be used as a basis for assessment.

ISO 15189 clause 5.6.4 states "External quality assessment programmes should, as far as possible, provide clinically relevant challenges that mimic patient samples and that check the entire examination process including pre- and post-examination procedures" (17). Results interpretation offered by laboratories, a post-examination procedure, is thus included. Assessment is widely used in many academic and professional activities, but others have reported on the lack of reproducibility of assessment (18). A recent article identified both intra-assessor and inter-assessor differences, and suggested that a minimum of 16 assessors was needed to arrive at a consistent (but not necessarily correct) verdict in 95% of cases (19). However this extrapolation depends on an implicit assumption that the sole reason for intra- and inter-assessor differences is random variation, akin to intra- and inter-assay impre-

telja. Također moramo prihvatiti činjenicu da svaki proces procjene neizbježno uključuje stupnjevanje svake napomene u odnosu na druge: rijetko će se naći gotove smjernice koje bi se mogle koristiti kao osnova za procjenu.

Norma ISO 15189, stavak 5.6.4. navodi da „Programi vanjske procjene kvalitete trebaju, koliko je to moguće ponuditi, klinički važne izazove koji imitiraju uzorke bolesnika i koji provjeravaju čitav proces testiranja uključujući i prijeanalitičku i poslijeanalitičku fazu“ (17). To, dakle, uključuje tumačenje rezultata koje nudi laboratorij kao poslijeanalitički postupak. Procjenjivanje je u širokoj uporabi kod mnogih akademskih i stručnih aktivnosti, no neki autori izvještavaju o nedostatnoj ponovljivosti procjena (18). Nedavno objavljen članak istaknuo je razlike u sukladnosti pojedinog te između više procjenitelja (engl. *intra- and inter-assessor differences*) i ukazao da je potrebno najmanje 16 procjenitelja kako bi se postigla dosljedna odluka (koja ne mora nužno biti točna) u 95% slučajeva (19).

Međutim, ta ekstrapolacija ovisi o implicitnoj pretpostavki da je jedini razlog razlika u sukladnosti između procjenitelja nasumična varijacija srodna nepreciznosti u seriji i iz dana u dan kod konvencionalne analitičke EQA. Očigledno je da je to nije točno, budući da sustavne razlike između procjenjivača zaista postoje te vode različitim ishodišta. Prva razlika jest da različiti procjenitelji pridaju različiti stupanj važnosti pojedinoj neprikladnoj sastavnici inače prikladnog komentara ili vještini prenošenja informacije te mogu rabiti različite i neusuglašene smjernice za procjenu prikladnosti komentara (srodno razlikama u umjerenju među različitim testovima). Druga je nešto temeljnija: različiti procjenitelji mogu donijeti različite zaključke zasnovane na njihovom osobnom prijašnjem iskustvu koje se može jako razlikovati. To je kao da se različitim imunokemijskim metodama pokušavaju kvantificirati složene smjese fragmenata i intaktnih molekula pomoću različitih protutijela koja prepoznaju različite epitope i kombinacije epitopa. U situacijama kada sustavne varijable (engl. *non-random variables*) utječu na različite procjenitelje nemoguće je definirati koliko je procjenitelja potrebno kako bi dali ocjenu (koja ne mora biti nužno točna) istom većinom glasova u 95% slučajeva. Kao i ostale okolnosti i ova ovisi o relativnom omjeru procjenitelja koji zastupaju jedno mišljenje i onih koji zastupaju drugo. Primjerice, ako se postavi pitanje na koje polovica procjenitelja odgovori potvrdno, a druga polovica negativno (plus-minus slučajna varijacija), nikada neće biti donesena odluka većinom glasova. Odgovor koji se vrednuje srednjom ocjenom je „možda“, što ne zadovoljava nijednu polovinu, ali može biti točna odluka.

Navedeno bi predstavljalo veliki problem kad bi se procjena odlukom većine koristila kao zlatni standard (što je čest slučaj na mnogim područjima akademske i stručne aktivnosti). Što se tumačenja tiče, odluka o ocjeni može se smatrati jedino smjernicom u pitanju koja se napome-

cision in conventional analytical EQA. This is obviously invalid since non-random differences between different assessors do exist and lead to different outcomes. The first is that different assessors may put different weight on a single inappropriate component in an otherwise appropriate comment, or on communication skills, and may use different and contradictory guidelines for assessing the appropriateness of a comment (akin to calibration differences between different assays). The second is more fundamental: different assessors may come to different conclusions based upon their individual past experiences which may differ widely. This is akin to different immunoassays attempting to quantify a complex mixture of fragments and intact molecules using different antibodies recognising different epitopes or combinations of epitopes. In the situation of non-random variables influencing different assessors, it is impossible to define how many assessors are needed to achieve the same majority (but not necessarily the correct) verdict in 95% of cases. Among other circumstances, this depends on the relative proportions of assessors holding one view as opposed to another. For example, if a question is asked which half the assessors believe the answer is 'yes' and the other half 'no' there is (give or take random variation) never a majority verdict. The mean answer is 'maybe', which satisfies neither half, but may be the correct verdict.

This would be of major concern if the majority assessment verdict were used as the 'gold standard' (which it often is in many areas of academic and professional activity). For interpretation, an assessment verdict can only be regarded as a guide as to which comments may be viewed as more or less appropriate. In the current schemes, there are several 'over-rides' to prevent the highest scoring comment or key phrase being necessarily regarded as the gold standard. The first is to use the actual outcome of the Case when this is known. The second is to use the opinion of the scheme organiser, reinforced if necessary by seeking expert opinion. Neither of these over-rides is in itself a gold standard: the outcome of a Case cannot be regarded as such since it may have been a rare cause of a common set of abnormalities; and experience from the UK NEQAS is that experts sometimes do not realise that they hold extreme minority (but not necessarily incorrect) opinions. We have to accept that assessment of any sort is imperfect; that the primary purpose of interpretative EQAS is educational (20); and that assessment of any kind is a guide to participants, not a definition of exact solutions which seldom if ever exist. Although assessment is imperfect, it is an essential component of any EQAS assessing interpretation, and there is evidence that assessment procedures incorporated in an EQA scheme improve the standard of interpretation (20).

na može smatrati više, odnosno manje prikladnom. U sadašnjim programima postoji nekoliko ograničenja (engl. *over-rides*), koja sprječavaju da se komentari i ključne fraze s najvišim rezultatom postave kao zlatni standard. Prvo je da se rabi istinski ishod nekog slučaja kada je on poznat. Drugo je da se kao referentno mišljenje uzima mišljenje organizatora programa, koji u slučaju potrebe može zatražiti stručno mišljenje. Niti jedno od ovih ograničenja ne može se samo po sebi smatrati zlatnim standardom: ne može se znati kakav će biti ishod nekog stanja, budući da može imati neuobičajen uzrok poznatih patologija. Iskustvo EQAS kaže da stručnjaci ponekad ne shvaćaju da je njihovo mišljenje (koje ne mora nužno biti neispravno) u velikoj manjini. Moramo prihvatiti činjenicu da je procjenjivanje bilo kakve vrste nesavršeno, da je primarna svrha interpretativnog EQAS obrazovna (20) te da procjena bilo koje vrste sudionicima predstavlja smjernicu, a ne definiciju ispravnog rješenja koje se pojavljuje jako rijetko, ako uopće i postoji. Iako je procjenjivanje nesavršen proces, ono je ključna sastavnica svakog tumačenja programa vanjske procjene kvalitete te postoje jasni pokazatelji koji potvrđuju da postupci procjenjivanja unutar programa vanjske procjene kvalitete poboljšavaju standard tumačenja (20).

Adresa za dopisivanje:

Gordon S. Challand
Department of Clinical Biochemistry
Royal Berkshire Hospital
London Road
Reading
Berkshire RG1 5AN
UK
e-pošta: gqn62@dial.pipex.com

Literatura/References:

- Lippi G, Fostini R, Guidi GC. Quality improvement in laboratory medicine: extra-analytical issues. *Clin Lab Med* 2008;28:285-94.
- Plebani M. Laboratory errors: How to improve pre- and post-analytical phases? *Biochimica Medica* 2007;17:5-9.
- Freedman DB. Is the medical undergraduate curriculum 'fit for purpose'? *Ann Clin Biochem* 2008;45:1-2.
- Simundic AM, Topic E. Quality indicators. *Biochimica Medica* 2008;18:311-9.
- Kilpatrick ES, Barth JH. Whither clinical validation? *Ann Clin Biochem* 2006;43:171-2.
- Barlow IM. Are biochemistry interpretative comments helpful? Results of a general practitioner and nurse practitioner survey. *Ann Clin Biochem* 2008;45:88-90.
- Barlow IM. Do interpretative comments influence patient management and do our users approve of the laboratory 'adding on' requests? A follow-up on?of? General Practitioner and Nurse Practitioner survey. *Ann Clin Biochem* 2009;46:65-86.
- Sciavolli L, Zardo L, Secchiero S, Zaninotti M, Plebani M. Interpretative comments and reference ranges in EQA programs as a tool for improving laboratory appropriateness and effectiveness. *Clin Chim Acta* 2003;333:209-19.
- Clinical Pathology Accreditation (UK). Standards for the medical laboratory.
- Challand G. Cases for comment, education and audit. *JIFCC* 1998;10:53-5.
- Challand GS, Osypiw JC, MacKenzie F, Middle JG, Fister D, Stezhka L. Web-based external quality assessment of individual skills – the UK NEQAS for interpretative comments in Clinical Chemistry. *EuroMed-Lab 2001, Prague, abstract S259*
- Vasikaran SD, Penberthy L, Gill J, Scott S, Sikaris KA. Review of a pilot Quality Assessment Program for interpretative commenting. *Ann Clin Biochem* 2002;39:261-72.
- Lim EM, Sikaris KA, Gill J, Calleja J, Hickman PE, Beilby J, Vasikaran SD. Quality assessment of interpretative commenting in Clinical Chemistry. *Clin Chem* 2004;50:632-37.
- Challand GS, Vasikaran SD. The assessment of interpretation in Clinical Biochemistry: a personal view. *Ann Clin Biochem* 2007;44:101-5.
- Challand GS. Assessing the quality of comments on reports: a retrospective study. *Ann Clin Biochem* 1999;36:316-32.
- Li P, Challand GS. Experience with assessing the quality of comments on Clinical Biochemistry reports. *Ann Clin Biochem* 1999;36:759-65.
- <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26301&COMMID=&scopelist=>
- Goldman RL. The reliability of peer assessments: a meta-analysis. *Eval Health Prof* 1994;17(1):3-21.
- Bindels R, Hasman A, van Wersch JWW, Pop P, Winkens RAG. The reliability of assessing the appropriateness of requested diagnostic tests. *Med Decis Making* 2003;23:31-7.
- Osypiw JC, Challand GS. 'Cases for Comment' on the ACB mailbox: were they educational? *Proceedings of the ACB National Meeting 2002: A105: p85.*