When laboratory tests can mislead even when they appear plausible

Author: Adel AA Ismail

ABSTRACT

A laboratory test has three phases, pre-analytical, analytical and post-analytical. The purpose of this review is to highlight an issue concerning the analytical phase of one of the most widely deployed groups of *in vitro* diagnostic tests using a common technology – namely immunoassay.

Immunoassay entails an inherently high error rate and, therefore, has the potential for inaccurate and misleading results susceptible to misinterpretation and/or diagnostic misapplication by clinicians. An approach based on Bayesian inference (without mathematics or equations) – illustrated by examples – is presented; this may help clinicians in discerning potentially erroneous results even when they appear plausible and not unreasonable.

Essentially, false positive results are most likely to occur when the disease prevalence/incidence is low. False negative results become more prominent when the prevalence/ incidence of disease increases. When concern is raised, available follow-up laboratory tests should be initiated to establish with confidence the diagnostic reliability or unreliability of such results.

KEYWORDS: Bayesian principle, false negative, false positive, immunoassays, misdiagnosis, reliability

Introduction

Laboratory tests are commonly used in the diagnosis and management of disease, and millions of tests on a wide range of analytes are requested annually by clinicians in the UK. Clinical summaries on request forms may be scant or absent; conversely, analytical results received by clinicians are generally regarded as accurate and reliable and, if plausible, are accepted as diagnostic. The purpose of this review is to highlight the element of risk in this rationale for certain tests with limited accuracy. Unquestioning acceptance at face value may lead to misdiagnosis and mistreatment or unnecessary surgery for unconfirmed disease with consequent iatrogenic morbidity.[1–9]

A probabilistic rationale based on Bayes' theorem[1] – in which key information, such as (a) the rate of inherent analytical error in a test and (b) the disease prevalence (number of cases per head of

Author: retired consultant in clinical biochemistry and chemical endocrinology, Wakefield, UK

population) or incidence (rate of new cases over a period of time), is taken into account – can be used to evaluate the likelihood of an erroneous result occurring and its nature, ie false positive or false negative. Essentially, in disorders with low prevalence/incidence and/or when a test is used as a screen with vague or insufficient clinical correlates, erroneous results are likely to be false positives. On the other hand, when the disease prevalence/incidence is high and the index of suspicion is significant with good clinical correlates, erroneous results are more likely to be false negative.

Immunoassays

Immunoassays (the largest *in vitro* diagnostic technology worldwide (Box 1)) have an analytical error rate of 0.4–4%, which is considerably higher than other routine tests such as renal and liver function tests or full blood count.[1–9] Abnormal results, including false positives, may routinely be interpreted as 'pathological' or diagnostic by clinicians,

---

Box 1. Common tests performed by immunoassays

> All endocrine tests (eg pituitary, thyroid, adrenal, parathyroid, pancreatic, gonadal, hCG, GI hormones)

> Tumour markers, eg PSA, β-hCG, thyroglobulin, CEA

> Cardiac biomarkers such as troponins and myoglobin

> Rheumatoid factor

> Allergy and allergen-specific IgE

> Vitamin B12/folic acid

> Specific serum proteins such as Ferritin, α-fetoprotein, sex hormone binding globulin

> Specific antibodies against bacteria/viruses, eg hepatitis, HIV, CMV, rubella, syphilis

> Anti-endocrine gland antibodies, eg thyroid peroxidase, thyroglobulin, intrinsic factor, adrenal

> Therapeutic drug monitoring such as digoxin, gentamicin, cyclosporine

> Drugs of abuse, eg cannabis and opiates

Although this list does not include all immunoassay tests, it highlights the wide range of analyses predominantly performed by this technology

CEA = carcinembryonic antigen, CMV = cytomegalovirus; GI = gastrointestinal hormones; hCG = human chorionic gonadotrophin; PSA = prostate-specific antigen

---

leading to corresponding intervention. Conversely (and possibly more seriously), false-negative results may wrongly exclude pathology, thus denying patients necessary treatment. Greater awareness of potential wide variation in accuracy between different laboratory tests is therefore important for clinicians.

Examples of misleading results

It would be realistic in practice for clinicians to exercise extra care before accepting and acting on potentially misleading results carried out by immunoassay, even when they appear plausible and not unreasonable. This can be achieved by grasping the concept underpinning probabilistic reasoning. This is exemplified here (avoiding equations and unnecessary mathematics) using two illustrative scenarios, namely (1) thyroid-stimulating hormone (TSH) in the diagnosis of subclinical hypothyroidism and (2) troponin in patients presenting with chest pain suspicious of acute coronary syndrome.

Prevalence of subclinical hypothyroidism and utility of TSH in diagnosis

Subclinical hypothyroidism has a low prevalence of 1% or less in young adults and children but much higher incidence of approximately 17% in older women.[10,11]

The immunoassay used to measure TSH may be assumed to have an excellent diagnostic accuracy of 99.6% and an error rate of ±0.4% (considered among the lowest reported).[1-9] This means four false positives or negatives are predicted per thousand in any population sample.

In a 1,000 sample of the low (1%) prevalence group (young adults and children), the number of individuals with genuine subclinical hypothyroidism (ie truly raised TSH) would be anticipated to be 10. However, including the 4 false positives, there would be a total of 14 cases in which TSH is raised, and potentially interpreted as consistent with the diagnosis. This equates in this cohort to a test sensitivity of about 70% and an error rate of about 30% (4/14).

Conversely, in 1,000 samples of the higher (17%) prevalence group (older women), the corresponding figures would be 170 and 174 – ie a test sensitivity of about 98% and an error rate of about 2% (4/174).

Needless to say, these figures are approximations, but would not be radically different from a precise statistical calculation:

$$P(a \mid b) = P(b \mid a) P(a) \div \Sigma (P(b \mid a) P(a) + P(b \mid \text{not } a) P(\text{not } a))$$

With a representing the condition in which the patient has the disease and b representing a positive test result.

Or simply

True positives ÷ the sum of 'true positives + false positives'

The number of false negatives (also 4) on the other hand would be proportionately very small in both groups. The numbers of individuals without the disease in young adults/children and in older women would be 990 and 830, respectively, making the rate of false negative results 4/994 and 4/834 – ie an error rate of <0.5% and a specificity of >99.5% for exclusion of subclinical hypothyroidism in both cohorts.

A decision-making rationale based on probabilistic Bayesian inference thus supports the clinician in adopting different clinical strategies in the interpretation of TSH results in these two cohorts of patients. For example, initiating replacement therapy for subclinical hypothyroidism might be controversial in some older patients with other comorbidities,[12] while lifelong treatment of younger individuals would warrant compelling diagnostic confirmation before embarking on that course of action.

Use of troponin in patients with chest pain suspicious of acute coronary syndrome

Acute coronary syndrome is considered to be the likely diagnosis in about 80% of patients who present with a history of typical cardiac chest pain in the presence of multiple risk factors, such as smoking, age, diabetes mellitus, hypertension and dyslipidemia.[13] Cardiac troponin immunoassay also has an inaccuracy of ±0.4% – ie in any 1,000 cohort, four falsely raised and four falsely low results.

As in the previous example, taking these two key factors into account (ie disease incidence and test inaccuracy), the rate of false negatives is likely to be four times greater than that of false positives. Serum troponin would be correctly elevated in 800 patients with the disease (true positives) with four false positives caused by the inaccuracy of immunoassay – a false positive rate of about 0.5% (4/804) and sensitivity of about 99.5%. In the 200 without acute coronary syndrome, the corresponding false negative rate will be about 2% (4/204), with a specificity of about 98%.

Again, these figures are approximations, but would not differ radically from a more precise calculation using the equation present earlier.

A falsely elevated result may lead to hospitalisation and expensive investigation (such as coronary angiography), while a falsely low result may deny the patient necessary investigations and treatment with potentially more serious sequelae. Of interest is that false positive case reports have been widely reported in recent literature (being more identifiable) while the number of false negative case reports is lower.[1,14-16] Importantly, these analytical errors, either negative or positive, may persist for some time. Protocols such as repeat analysis after 6–12 hours or 6–18 hours from onset may therefore be unhelpful or even misleading if the initial reading is false and other corroborative follow-up investigations are not carried out.

Discussion

This brief outline is intended to promote increased awareness among clinicians about the specific nature and significant incidence of inaccurate test results from immunoassays, and to propose practical approaches to address the issues raised.

'False positive' is a term generally used to indicate a test result suggesting a disease presence when there is none, and 'false negative' the reverse. Clinicians are generally aware that this may arise from the common/accepted practice of establishing a 'statistical normal/reference' range for an individual analyte. The quoted normal range for each test (including immunoassays) is obtained by using cut-off points (about 95%) of a continuum obtained from 'normal' individuals. This statistical truncation delineates some false positive or

false negative calculable data, expressed as predictive false positive and false negative rates using the well-established 2×2 contingency table.

The terms false positive and false negative described in this review and illustrated by the two examples are, however, different and should not be confused with results falling outside a normal established reference range produced by statistical truncation as above. The error rate in immunoassay tests is variable, random, insidious and unpredictable and may occur in any immunoassay test irrespective of its nature and/or format.

Interference in immunoassays leading to inaccurate results is generally unique to individuals who may fortuitously have endogenous immunoglobulin antibodies capable of interfering with those used as biological-reagents in immunoassay analysis. Each immunoassay 'host' presents his/her own antibody(s) as interacting 'analytical reagents'. Because of the huge array and diversity of immunoglobulin antibodies that may be endogenously produced (around 10 billion), this form of potential interference is impossible to predict *a priori* and it would be time consuming, expensive and clinically unhelpful to subsequently establish its specific nature. The potential interference and error rate of any individual test (eg Box 1), though inevitable, cannot be known to the 'host' or clinician, or predicted by the laboratory. The presence of interfering endogenous antibodies may be transient (ie for months following infection, immunisation, allergic reactions, blood transfusion, therapy with monoclonal antibodies etc) or permanent (eg in autoimmune disease).

Research over the last two decades has shown immunoassay error rates ranging from 0.4–4%. A more realistic general 'guestimate' in practice would be 0.4–1%, which is still high; hence the need for awareness and caution.[1–9]

Probabilistic reasoning based on Bayesian inference, which takes into account key relevant information such as the rate of analytical inherent error and disease prevalence/incidence, is a concept that can help in identifying potentially false results and quantifying the likelihood of false positive and negative data.

What practical approaches may help in addressing these issues? At what point in the process should alarms be raised and action be triggered and by whom – laboratory scientists or clinicians?

Laboratory scientists have a role. Although they often receive scant clinical information, they are nevertheless able to detect gross aberrations, eg gross inconsistences in multiple parameter profiles or sudden unexplained changes in serial measurements of an analyte, and to respond by reporting their concern.

Clinicians may independently encounter such aberrations too, including more subtle/plausible ones, particularly in two main scenarios:

1 when an immunoassay test is used for screening and/or as a diagnostic test for diseases with low prevalence/incidence
2 when inconsistences occur between an immunoassay test and other clinical or radiological correlates.

In such cases, potentially false analytical data in the immunoassay tests must be on the list of differential diagnoses before embarking on invasive, expensive investigations or making a definitive diagnosis.

Good communication and dialogue between clinicians and laboratory scientists is clearly paramount in avoiding misuse of potentially misleading results.

Conclusion

Immunoassay was one of the most important 20th century innovations and enabled the measurement of numerous analytes with exquisite specificity, high sensitivity and remarkable precision. It is here to stay, being highly practicable and cost effective. Nevertheless, tests performed by immunoassays continue to be reported as fallible[17,18] because of inherent high analytical error rates. Therefore, if in doubt, the integrity and reliability of any given analysis should be verified by relevant corroborative laboratory tests. If analytical interference is confirmed and accuracy remains in doubt,[19] such results must be discounted for diagnostic and management purposes, even if plausible. ■

Conflicts of interest

The author has no conflicts of interest to declare.

Acknowledgement

References

1 Ismail AAA. Identifying and reducing potentially wrong immuno-assay results even when plausible and "not-unreasonable". *Adv Clin Chem* 2014;66:241–93.

2 Bjerner J, BØmer OP, Nustad K. The war on heterophilic antibody interference. *Clin Chem* 2005;51:9–11.

3 Rotmensch S, Cole LA. False diagnosis and needless therapy of presumed malignant disease in women with false-positive human chorionic gonadotropin concentrations. *Lancet* 2000;355:712–5.

4 Mills F, Jeffery J, Mackenzie A, Cranfield A, Ayling RM. An immuno-globulin G complexed form of thyroid stimulating hormone (macro thyroid-stimulating hormone) is a cause of elevated serum thyroid-stimulating hormone concentration. *Ann Clin Biochem* 2013;50:416–20.

5 Saiegh L, Odeh M, Elias L *et al*. A possible analytical and clinical role of endogenous antibodies causing discrepant adrenocortico-tropic hormone measurement in a case of ectopic Cushing's syndrome. *Ann Clin Biochem* 2014;51:490–4.

6 Plebani M, Mion M, Altinier *et al*. False positive troponin I attributed to a macrocomplex. *Clin Chem* 2002;48:677–8.

7 Norden AGW, Jackson RA, Norden LE *et al*. Misleading results from immunoassays of serum free thyroxine in the presence of rheumatoid factor. *Clin Chem* 1997;43:957–62.

8 Bolland MJ, Chiu WW, Davidson JS, Croxson MS. Heterophilic antibodies may cause falsely lowered serum cortisol values. *J Endocrinol Invest* 2005;28:643–5.

9 Ross HA, Menheere PPCA, Thomas CMG *et al*. Interference from heterophilic antibodies in seven current TSH assays. *Ann Clin Biochem* 2008;45:616–8.

10 Surks MI, Ortiz E, Daniels GH *et al*. Subclinical thyroid disease: Scientific review and guidelines for diagnosis and management. *JAMA* 2004;291:228–38.

11 Weetman AP. Fortnightly review. Hypothyroidism: screening and subclinical disease. *BMJ* 1997;314:1175–7.

12 Monzani F, Dardano A, Caraccio N. Does treating subclinical hypothyroidism improve markers of cardiovascular risk? *Treat Endocrinol* 2006;5:65–81.

13 Carlton E, Greenslade J, Cullen L *et al*. Evaluation of high-sensitivity cardiac troponin I levels in patients with suspected acute coronary syndrome. *JAMA Cardiol* 2016;1:405–12.

Adel AA Ismail

14 Legendre-Bazydlo LA, Haverstick DM, Kennedy JLW, Dent JM, Burns DE. Persistent increase in cardiac troponin I in plasma without evidence of cardiac injury. *Clin Chem* 2010;56:702–5.

15 Bohner J, von Pape KW, Hannes W, Stegmann T. False negative immunoassay results from cardiac troponin I probably due to circulating troponin I autoantibodies. *Clin Chem* 1996;42:2046–7.

16 Eriksson S, Halenius H, Pulkki K, Hellman J, Petterson K. Negative interference in cardiac troponin I immunoassays by circulating troponin autoantibodies. *Clin Chem* 2005;51:839–17.

17 Wang H, Bi X, Xu L, Li Y. Negative interference by rheumatoid factor in alpha-fetoprotein chemiluminescent microparticle immunoassay. *Ann Clin Biochem* 2017;54:55–9.

18 Choy KW, Teng J, Wijeratne N, Tan CY, Doery JCG. Immunoassay interference complicating management of Cushing's disease: the onus is on the clinician and the laboratory. *Ann Clin Biochem* 2017;54:183–4.

19 Ismail AAA, Walker PL, Barth JH *et al*. Wrong biochemistry results: two case reports and observational study in 5310 patients on potentially misleading thyroid stimulating hormone and gonadotropins immunoassay results. *Clin Chem* 2002;48:2023–29.

Address for correspondence: Dr Adel Ismail, Tanglewood, Chevet Lane, Wakefield WF2 6HL, UK.
Email: adelaaismail@aol.com

The UK's multidisciplinary response to an Ebola epidemic

Authors: Sian Reece,[A,*] Colin S Brown,[B,*] Jake Dunning,[C] Meera A Chand,[D] Maria C Zambon[E] and Michael Jacobs[F]

ABSTRACT

The West African Ebola virus disease (EVD) epidemic was the largest and most devastating outbreak of EVD the world has ever seen. Its impact was felt far from the shores of Guinea, Liberia and Sierra Leone, with public health systems and clinicians across the globe confronted with an international response both in the affected region and within their own borders. The UK had a prominent role in response efforts, particularly in Sierra Leone. This article highlights how UK academic, health service, military, commercial and public health professionals all played a significant role both at home and abroad.

KEYWORDS: Ebola, outbreak, planning, resilience, response, Sierra Leone, UK

Introduction

The West Africa Ebola virus disease (EVD) epidemic of 2013–16 was an unprecedented, complex and grave humanitarian crisis. It was the largest known outbreak of EVD in history, with over

Authors: [A]NHS medical director's clinical fellow, Public Health England, London, UK; [B]locum consultant in infectious diseases and medical microbiology, Public Health England, London, UK and ex-specialty registrar, University College London Hospitals NHS Foundation Trust, London, UK; [C]consultant in infectious diseases, Public Health England, London, UK and honorary visiting research fellow, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK; [D]consultant in medical microbiology, Public Health England, London, UK and Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, London, UK; [E]deputy director, National Infectious Service, Public Health England, London, UK; [F]consultant in infectious diseases, Royal Free London NHS Foundation Trust, London, UK

28,500 clinically compatible cases and 11,000 deaths reported in Guinea, Liberia and Sierra Leone.[1] The impact on the three most affected countries was immense and had the potential to undo important progress and efforts to rebuild following periods of conflict. Some neighbouring countries were also affected, and many countries outside Africa were required to diagnose and manage EVD cases for the first time.

The first World Health Organization (WHO) public announcement of the EVD outbreak was on 23 March 2014.[2] This announcement was shortly followed by a call to action from Médecins Sans Frontières, who announced their concerns regarding the lack of international response to the rapidly evolving outbreak.[3] It has since been widely acknowledged that the international response to this outbreak was too little, too late.[4] On 8 August 2014, the WHO declared the Ebola outbreak a public health emergency of international concern (PHEIC).[5]

The case fatality rate (CFR) for this epidemic was estimated to be between 31% and 74% in the three most affected countries in West Africa, with an average CFR of 48% for hospitalised patients; this is largely consistent with the range of CFRs reported for previous Ebola Zaire (EBOV) outbreaks.[6–11] Cases were also diagnosed in Senegal, Nigeria, Mali, USA, UK, Italy and Spain. Evacuated cases were treated in France, Italy, the Netherlands, Switzerland, Norway, Germany, the USA and the UK. Additionally, a more typical, smaller and unrelated EVD outbreak was confirmed in the Democratic Republic of Congo in August 2014.

On 29 March 2016, the WHO declared that the PHEIC had ended.[12] In the three most affected countries in West Africa, sporadic cases remained a concern despite all countries having been declared free of human-to-human EVD transmission. Small hotspots of EVD continued to occur, with new cases reported as late as April 2016. The complex nature of this outbreak posed significant challenges moving forward, for these countries and for the wider public health community, given