The Standardisation and Control of Testing for Infectious Diseases

By

Wayne Dimech

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy

College of Medicine and Public Health

20th March 2023

Table of Contents

Tables and Figures

Tables

Figures

Significant Abbreviations

ANOVA	Analysis of variance
CIA	Commercial immunoassays
CLSI	Clinical and Laboratory Standards Institute
c/mL	Copies per millilitre
CMV	Cytomegalovirus
COI	Cut-off index
CRM	Certified reference materials
DV	Delta Value; Δ Value
EBV	Epstein-Barr virus infection
EIA	Enzyme-based colour reactions
EQA	External quality assessment
EQAS	External quality assessment scheme
HAI	Haemagglutination inhibition assays
HBsAb	Anti-hepatitis B surface antigen
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human Immunodeficiency Virus
HPLC	High performance liquid chromatography
IAs	Immunoassays
IB	Immunoblot
IFU	Instructions for use
IgG	Immunoglobulin G
IH-EIA	In-house enzyme immunoassay
ISO	International Organization for Standardization
IU/ml	international units per millilitre
IVDs	*In-vitro* diagnostic devices
MTP	Microtiter plate
MU	Measurement uncertainty
NAT	Nucleic acid test (testing)
NIBSC	National Institute of Biological Standards and Controls
NRL	National Reference Laboratory, Australia
PCR	Polymerase chain reaction
PNT	Plaque neutralisation assays
PoCT	Point of care testing
QC	Quality control
qPCR	Quantitative polymerase chain reaction
RDT	Rapid lateral flow test device
RiliBÄK	German Richtlinien der Bundesärztekammer
ROC	Receiver operating characteristic
S/Co	Signal to cut-off ratio
SD	Standard deviations
TEa	Allowable total error
TGA	Therapeutics Goods Administration
VCA	Viral capsid antigen
VL	Viral load
WB	Western blot
WHO	World Health Organization

Acknowledgements

I would like to acknowledge my colleagues at the National Serology Reference Laboratory, Australia (NRL) for collaborating with my work over the past decade or more. NRL is dedicated to the promotion of quality of infectious disease testing. Collectively and individually many colleagues have supported this journey of understanding and documenting the standardisation and control of testing for infectious diseases. Their scientific contribution is acknowledged through co-authorship of peer-reviewed papers, but their day-to-day support and interest in the topic is also recognised here. Notable contributions are from Joe (Giuseppe) Vincini who has been a sounding board for the past decade, and Marina Karakaltsas, another colleague who has performed background data cleaning, review and analysis. I would also like to acknowledge previous NRL Directors, Susan Best and Dr. Elizabeth M. Dax. Both have supported this work during their time at NRL, as well as acted as mentors. I also would like to acknowledge and thank my current Director, Dr Philippa (Pip) Hetzel, who continues to support this research.

Of particular note, I would like to thank my wife Margaret Dimech, who has always been supportive of my career in medical science and understands, and tolerates, my passion for this subject. Her support of my endeavours has been invaluable even thorough difficult medical episodes.

Finally, a huge thanks to my supervisors Dr. Dusan Matusica and A/Prof Jill Carr from Finders University who helped me through the process of enrolment and for their invaluable advice in the development and completion of this thesis. They have both been open to my questions and supportive of the project. I thank them both personally.

Declaration

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university

2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and

3. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Signed: 

Date: 20th March 2023

Contribution to Submitted Papers

Chapter 2

Dimech W, Panagiotopoulos L, Francis B, Laven N, Marler J, Dickeson D, Panayotou T, Wilson K, Wootten R, Dax EM. Evaluation of Eight Anti-rubella Virus Immunoglobulin G Immunoassays that Report Results in International Units per Milliliter. J Clin Microbiol. 2008 Jun; 46(6):1955-60.

As lead author I coordinated the study, enrolled collaborating laboratories, and acquired the samples used in the study. I wrote the protocol, collected and analysed the results, and was the main person drafting the manuscript. All other authors were engaged in testing samples and reporting results excepting EMD who was the Director of NRL and who reviewed the manuscript. All authors reviewed the manuscript prior to publication.

SCImago Journal Rank Impact Factor 4.06

Dimech W, Arachchi N, Cai J, Sahin T, Wilson K. Investigation into Low-level Anti-rubella Virus IgG Results Reported by Commercial Immunoassays. Clin Vaccine Immunol. 2013 Feb;20(2):255-6.

As with the paper above, I coordinated the study, acquired sample and developed the protocol. KW developed the Western blot used in the study. The other authors performed the testing of the samples. I wrote the manuscript and was the corresponded author. All authors reviewed the manuscript prior to publication.

SCImago Journal Rank Impact Factor: 2.83

Chapter Three

Dimech W, Cabuang LM, Grunert H-P, Lindig V, James V, Senechal B, Vincini G, Zeichhardt H. Results of Cytomegalovirus DNA Viral Loads Expressed in Copies per Millilitre and International Units per Millilitre are Equivalent. J Virol Methods. 2017 Nov 7;252:15-23.

I initiated the study. The co-authors each had equal contribution to this study, supporting the acquisition of the sample and the sending of samples to laboratories participating in the EQA conducted. HPG and VL produced the Youden plots and other graphs included in the paper. I drafted the manuscript and was the corresponding author on publication. All authors reviewed the manuscript prior to publication.

SCImago Journal Rank Impact Factor: 0.86

Chapter Four

Dimech W, Vincini G, Karakaltsas M. Determination of quality control limits for serological infectious disease testing using historical data. Clin Chem Lab Med. 2015 Feb;53(2):329-36.

I initiated this study and collaborated with a biostatistician to develop the statistical algorithm used in QConnect™. GV and MK extracted the data for analysis from EDCNet™ and conducted the data analysis according to the algorithm developed. I wrote the manuscript and submitted in for publication as lead author. All authors reviewed the manuscript prior to publication.

SCImago Journal Rank Impact Factor: 0.88

Dimech W, Karakaltsas M, Vincini GA. Comparison of four methods of establishing control limits for monitoring quality controls in infectious disease serology testing. Clin Chem Lab Med. 2018 Oct 25;56(11):1970-1978.

I designed the study and was lead investigator. GV developed a spreadsheet to perform the data analysis and, along with MK, conducted much of data analysis. Along with GV, I interpreted the results. I drafted the manuscript and submitted it for publication, as well as being the corresponding author. All authors reviewed the manuscript prior to publication.

SCImago Journal Rank Impact Factor: 1.19

Chapter Five

Dimech W, Giuseppe A Vincini, Liza M Cabuang, Megan Wieringa. Does a Change in Quality Control Results Influence the Sensitivity of an anti-HCV Test? Clin Chem Lab Med. Mar 2020.

I initiated and developed the protocol for this study, as well as identifying and acquiring the samples used. MW conducted testing of the samples on different reagent lots. LC and GV supported the study through provision of external quality assessment and quality control data, respectively. I wrote the manuscript and submitted it for publication as the corresponding author. All authors reviewed the manuscript prior to publication.

SCImago Journal Rank Impact Factor: 0.98

Abstract

Testing for infectious diseases underpins diagnosis, treatment, surveillance, and epidemiology of disease as well as maintaining safe transplantation of blood and tissue. Over the past decades, infectious disease testing has changed from manual, qualitative functional biological methodologies such as haemagglutination inhibition and complement fixation, to highly automated immunoassay and molecular platforms. In developed countries, much of this testing has moved from specialised microbiology laboratories to high throughput "core" laboratories testing a range of clinical chemistry and other medical pathology analytes on the same test platforms. As these test systems are generally managed and controlled by clinical chemists, it is not surprising that methods used to standardise and control testing usually employed in clinical chemistry are being introduced for infectious disease testing. However, there are significant differences between inert, homogeneous clinical chemistry analytes, such as glucose, potassium and urea, and the highly variable and heterogeneous biological testing used for infectious diseases. These differences in testing have been identified as the reason why standardisation and control processes have been largely unsuccessful when applied to infectious disease testing.

This thesis details my original and significant body of work identifying the deficiencies of traditional approaches to standardisation and control when applied to infectious disease testing. Papers presented in the thesis demonstrate the lack of standardisation of rubella IgG tests across two decades and investigates the clinical impact caused by poor standardisation. A further publication demonstrates that the traditional standardisation approach is appropriate for molecular testing using the quantification of CMV DNA as the example. My work in developing a novel approach to understanding and interpreting external run control results has had significant impact, being the only scientifically validated method for controlling infectious disease testing and is now licensed of use by two large quality control manufacturers. One paper presented demonstrates the QConnect™ concept and associated software EDCNet™. Additional papers demonstrate that QConnect™ is more fit-for-purpose than traditional methods such as Westgard rules when applied to infectious disease serology. Finally, a significant paper investigated whether variation detected by quality control had an impact on the clinical sensitivity and specificity of a test system.

The concepts presented in this thesis have been developed over time in a systematic manner, building upon each study to develop an understanding of the standardisation of infectious disease testing. This concept remains relevant and topical today with the emergence of SARS-CoV-2 infections and the release of an international standard for anti-SARS-CoV-2. The thesis highlights the differences between clinical chemistry and infectious disease analytes; reviews the utility of traditional methods for standardisation and control of medical testing when applied to infectious disease testing and describes and validates novel, alternative approaches. The clinical impact of the proposed alternatives was investigated. Together, these peer-reviewed publications form a significant and on-going impact on knowledge in the areas of standardisation and controls of infectious disease testing and continue to inform scientific discussion and influence national and international policy in this area.

Chapter One: Introduction to standardisation and control of infectious disease testing

This introduction is a modification of my review entitled "The Standardization and Control of Serology and Nucleic Acid Testing for Infectious Diseases" (1), first published in 2021 in the Journal of Clinical Microbiology Reviews, a journal with an impact factor of 26.13 in 2020-21. The modifications were made to update the thesis with recent, relevant information, in particular relating to SARS-CoV-2, reflect additional areas of relevant knowledge required of a comprehensive PhD thesis, and to maintain the flow of the thesis.

Introduction

Testing for infectious diseases detects infection, monitors efficacy of treatment, identifies stages of disease and provides evidence of immunity and disease prevalence. Since the 1960s, infectious disease testing has experienced dramatic changes as a result of increased knowledge of the immune system; the advent of significant technologies such as immunoassays (IAs) and nucleic acid testing (NAT); biomedical engineering and robotics and the introduction of stringent, national regulatory systems, each having a major impact on the delivery of more accurate and timely results. More recently, near-patient or point of care testing (PoCT), as well as self-testing, has extended the provision of infectious disease testing into remote and regional, community-based facilities. A recent example is molecular testing for SARS-CoV-2, now performed in remote regions such as first nations communities in Australia and Canada, and SARS-CoV-2 antigen testing in non-heath facilities such as schools, aged care facilities, mines and airports (2, 3). Irrespective of the mode of testing, infectious disease science has faced the challenge of managing the standardisation and control of these tests. It is important that patients receive comparable results when tested in different testing facilities, using different tests (standardised) and that the results from each facility are reproducible over time (controlled).

Disease Transmission and Human Immune Response to Infection

The routes of transmission of infectious agents to humans are varied. Although outside the scope of this paper a brief, representative summary of the ways pathogens breaches the human immune system is presented (4). Skin damage can be a primary route of transmission. Damage can be intentional, such as through medical procedures, resulting in transfusion transmitted infections with HIV and hepatitis B and C. Accidental tissue damage, including cuts and abrasions, punctures (e.g. hypodermic needles, trauma, insect or animal bites), burns or necrosis due to lack of blood supply (e.g. diabetes, tumour, crush injuries) allow the introduction of pathogens. Parasites such as *Strongyloides spp*. and *Schistosoma spp*. actively penetrate the skin on contact. Organisms can bypass mucosal barriers though insertion (e.g. catheter insertion, intrauterine devices, foreign bodies, sexual activity). These breaches can allow the introduction of pathogens of various sources, such as environmental, zoonotic, commensal or nosocomial. Pathogens are also introduced through inhalation of viruses in respiratory droplets (e.g. rubella, SARS-CoV-2 and other respiratory viruses) (5), spores (e.g. anthrax, histoplasma), or fomites (6). Other sources of infection include ingestion of meat from infected animals, food and water or fomites contaminated by faecal material, dried respiratory droplets or animal excretions (e.g. bat droppings) (4). A commensal organism that usually does not cause infection may become a pathogen in certain circumstances. Overgrowth of organisms, such as in bacterial vaginosis or candida vaginosis can occur due to changes in localised environmental conditions, such as pH. Infections caused by commensal organisms can also occur during immunodeficiency (e.g. Human herpes virus 8 causing Kaposi's sarcoma), or poor health (*Streptococcus pneumoniae* pneumonia of

alcoholics). Commensal or environmental organisms may cause infections when introduced into normally sterile anatomical organs such as the bloodstream, brain, lower respiratory system or peritoneum (6).

The immune response is designed to protect humans from potential infection. There are multiple layers to the immune response, including physical and chemical barriers (7, 8). The non-specific host-defence barriers of the innate immune system includes the physical barrier to infection created by skin and mucous membranes of the mouth, genitourinary system and external gastrointestinal tract. Potential pathogens contend with normal flora, entrapment in secreted mucous, low pH and chemical inhibitors, while cilia remove organisms trapped in mucous. If these barriers are breached, local immune cells produce cytokines such as tumour necrosis factor, interleukin 1 and interleukin 6 (8). Numerous other cell types are involved in the innate immune response, including neutrophils, macrophages, eosinophils and natural killer cells. These cells have a range of activities. Neutrophils and macrophages actively phagocytose bacteria and fungi. Eosinophils degranulate releasing enzymes, growth factors and cytokines, while natural killer cells destroy infected cells through the release of perforins and granzymes (8).

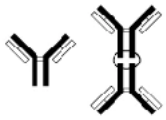
The adaptive immune response has two main cell types – T cell response and the B cell response (7, 8). When exposed to antigen-presenting cells that have digested an antigen and is displaying a specific peptide, T cells receptors are activated to differentiate into either mature cytotoxic T cells (CD8+ cells) that destroys cells containing foreign antigens, or become mature helper T cells (CD4+) that secrete cytokines that promote B cell maturity and mediate the immune response by directing other cells (9). B cells are produced in the bone marrow and circulate in blood and lymphatics. They can recognise antigens without the aid of antigen-producing cells. Once in contact with a recognised antigen, B cells proliferate and differentiate into either antibody-secreting plasma cells or memory cells (8, 9). Plasma cells have a relatively short life span, often undergoing apoptosis when the antigen is eliminated. However, plasma cells can produce large amounts of antibodies during its lifetime. In contrast, memory cells persist for long periods and, on subsequent contact with the same antigen, can release a rapid antibody response at re-exposure. Immunoglobulins produced by B cells can inactivate or inhibit viral attachment and reproduction but cannot eliminate infection. B cells produce five different immunoglobulins, IgA, IgD, IgE, IgG and IgM (**Table 1**), each having different functions (8). IgA is associated with mucosal protection, preventing pathogens binding to mucosal sites. The function of IgD is poorly understood but appears to be involved in homeostasis. It is also thought to play a central role in the regulation of tolerogenic and protective B cell responses (10). IgE antibodies are involved in allergic reactions and are also associated with parasitic infection (8). The detection of IgM and IgG is the primary method of laboratory diagnosis of infection, as described below.

When an immunocompetent human is exposed to an organism, they elicit an immune response. Typically, the response will be a primary or secondary immune response. In a primary response, the B lymphocytes circulating in a person who is naïve i.e. has never been in contract with the antigen, will produce antibodies to the antigen(s) present on the organism. This response is delayed and initially produces antibodies of low affinity (9). The initial antibody response is usually the production of antigen-specific IgM, followed by IgG antibodies. Initially the antibody reaction is immature and not necessarily specific to the antigen and sometimes target only certain immunogenic sites of the pathogen. Over time, the elicited antibodies increase in affinity (the strength of the bond between the antibody and the specific antigen) and avidity (the overall strength of the bonds between all antibodies and the antigen) and develop reactivity to the full range of that organism's immunogenic sites. At the same time a cell mediated response occurs and memory B cells are also created. On re-exposure to the antigen, the memory cells are primed to develop high avidity IgG, enabling a more rapid and specific secondary immune response. The difference between high and low avidity antibodies can be detected using an avidity assay. The changes in the affinity and avidity of the antibodies over time can directly impact on the binding of these antibodies to antigens used in assays.

Therefore, the stage of disease is an important consideration when understanding an assay's ability to detect specific antibodies.

IgG can be divided into four sub-classes, (IgG1, IgG2, IgG3, and IgG4) each having different functions, and unique profiles with respect to antigen binding, immune complex formation, complement activation and triggering of effector cells (11). Along with other immunoglobulins, IgG antibodies are neutralising antibodies, inhibiting the ability of viruses, in particular, to bind to receptor sites on human cells, limiting the viruses' ability to replicate (5, 8, 12).

Table 1. Structure and functions of immunoglobulins.

Immunoglobulin	Structure	Immune function	Activity
 IgA	Monomer or dimer in plasma and/or secretions.	Critical for protecting mucosal surfaces.	Moderate serum levels; Predominant immunoglobulin in secretions (saliva and breast milk).
 IgD	"Y-shaped" monomer.	Unclear function; Possible role in homeostasis.	Low levels in plasma; Short lived response.
 IgE	"Y-shaped" monomer.	associated with hypersensitivity and allergic reactions; Involved in parasitic immune response.	Lowest serum concentration; Shortest half-life of immunoglobulins.
 IgG	"Y-shaped" monomer; Four IgG subclasses	Activation of the complement cascade; Neutralisation of toxins and viruses.	Predominant immunoglobulin; Slow development after infection but lifelong; Crosses placenta barrier; Used in serological diagnosis to determine previous exposure.
 IgM	Pentameric	Primary immune response; Opsonising (coating) antigen for destruction; Complement fixation.	Poly-reactive; Does not cross the placenta barrier; Rapid response; Used in serological diagnosis to determine acute infection.

* Table 1 adapted from Marshall (8) and Schroeder(13).

Antigen-specific IgG become detectable shortly after IgM and rises rapidly in both quantity and avidity (12). The IgG response to different antigens from the same organism can rise and fall at different times. As examples, the IgG response to rubella envelope 2 antigen follows the response to envelop 1 antigen (**Figure 1**) (14). In Epstein-Barr virus (EBV) infection, anti-viral capsid antigen (VCA) specific IgM is the first antibody

response to be detectable and decreases relatively quickly over time, whereas the IgG antibodies to VCA is detectable shortly after the onset of symptoms and remains detectable for life in immunocompetent individuals (15). The IgG response to the "early antigen" is also present early in infection and declines to undetectable levels over time. In contrast, the IgG antibody response to the nucleocapsid antigen is usually detectable approximately four weeks post symptoms and remains detectable lifelong. Given the predicable rise and fall of these antibodies, the stage of disease post infection with EBV can be determined serologically with a single bleed. This is also the case with the serology of many other infections, notably hepatitis B virus, where the detection of hepatitis B surface antigen (HBsAg) and envelope antigen (HBeAg), and the levels of antibodies developed against hepatitis B surface antigen (anti-HBs), envelope antigen (anti-HBe) and core antigens (anti-HBc), can be utilised for determining the stage of disease as well as whether the infected individual has developed chronic hepatitis or has resolved infection. Detection of anti-HBs in isolation indicates past vaccination.

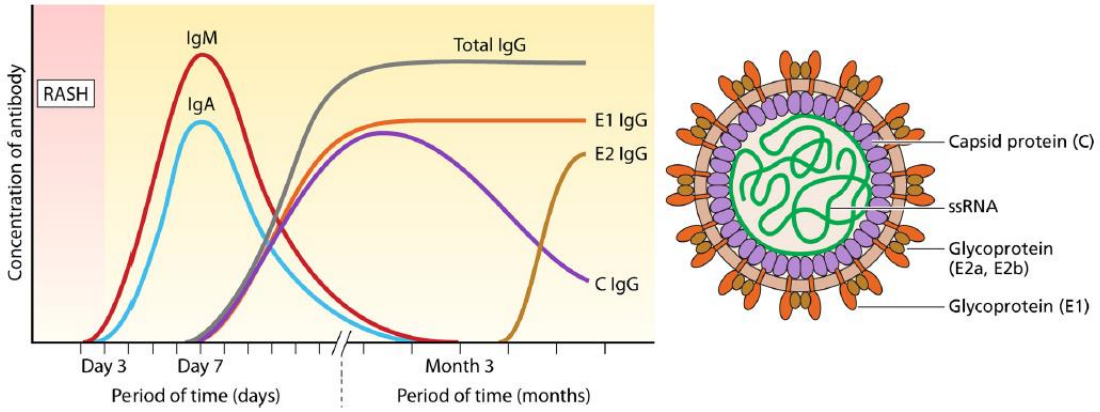


Figure 1. At the right is a schematic diagram detailing the structure of the rubella virus, including the three immunogenic antigens, i.e., two envelope (E1 and E2) antigens and a capsid (C) antigen, and single-stranded RNA (ssRNA). At the left is a plot of a normal immune response to rubella virus infections over time. (Reproduced with permission: Dimech et al (14))

Laboratory testing is employed throughout the disease process, from initial infection to disease resolution, to diagnose infection, monitor the disease progression or the efficacy of medical intervention and therapy, and to determine the prevalence of disease in the population. Direct detection of an infecting organism though culture or through direct detection of proteins or nucleic acid associated with the organism can be used to identify infection through laboratory testing of samples taken from the site of infection (4). A recent example is SARS-CoV-2 infection, where nasal or nasopharyngeal swabs are tested using rapid antigen tests or nucleic acid testing. Very early in the infection, direct detection of the organism is the only laboratory marker of infection, until the antibody response is initiated (**Figure 1**). Testing for antigen-specific IgG and IgM antibodies are most frequently employed in laboratory testing for infectious diseases (**Table 1**), as these antibodies circulate in abundance in human plasma, with IgG accounting for more than 10% of all plasma protein (11). On contact with an antigen, IgM is usually the first detectable immunoglobulin, opsonising (coating) the antigen to facilitating its destruction. IgM can also fix complement (8). The IgM response is usually transient, lasting for period of weeks post infection before becoming undetectable. Therefore, the detection of antigen-specific IgM is considered good evidence of acute infection. However, on re-infection or at times of re-exacerbation of infection, IgM response may become detectable for short period of time. The re-exacerbation of IgM, along with cross-reactivity of IgM antibodies with similar antigens on other organisms, complicate the diagnosis using IgM detection (16). To elucidate in clinically important situations, a positive IgM response can be confirmed using avidity assays, where the avidity of the patient's IgG response to the same organisms is subjected to chaotropic agents such as urea. Low IgG

avidity is an indicator of an acute infection and adds evidence to the diagnosis, whereas high avidity is associated with convalescence (17). The detection of antigen-specific IgG from a single bleed indicates past exposure to the antigen through infection or vaccination.

History of Infectious Disease Testing

There are well-established methods used to support the diagnosis of infection including microscopy, direct antigen detection or the culture and characterisation of organisms using biochemical testing (18) or MALDI-Tof mass spectrometry testing (19, 20). Inferences can be made from haematology full blood analysis of white and red blood cells, or quantification of CD4 makers. The human immune response includes a range of biomarkers such as C-reactive protein, procalcitonin, pancreatic stone protein and interleukins which are non-specific markers of infection (21), and a growing number of miRNAs and other emerging non-specific nucleic acid marker classes have been described (22). However, the diagnosis of most pathogenic infectious diseases depends on the detection of organism-specific immunoglobulin or nucleic acids. As the focus of this thesis is primarily on antibody, antigen and nucleic acid detection, these other markers are out of scope for this thesis.

Prior to the 1980s, the detection of antibodies and antigens associated with infections were predominantly conducted using laboratory-developed tests such as haemagglutination inhibition assays (HAI), complement fixation tests, plaque neutralisation assays (PNT), immunofluorescence, radial haemolysis or Ouchterlony double immunodiffusion. These bioassays detect the functional reactivity between antibodies and antigens and were the first test systems employed for the detection and semi-quantification of infectious disease serological testing (14). These test systems were labour intensive and require specific technical skills. The results varied depending on reagents used and their quality (23-25). The antigen source was usually whole organisms, and the test systems detect, without differentiation, all circulating antibody isotypes, in particular IgG, IgM and IgA. A lack of standardisation of these tests was reported (14, 26) and some professional bodies sought to introduce standard techniques (27) and quality assurance programs that improved commutability of results across laboratories (28, 29). However, these efforts were not universally applied. Over time, these technologies have been replaced with more efficient, automated options as detailed below. The introduction of these new technologies facilitated the introduction of automated processes, increased standardisation and allowed the development of more precise and accurate testing but introduced new issues relating to quantification of biological analytes, the standardisation of quantification of these measurements and the processes related to the control of automated testing. My work, over the past decade, has identified, elucidated, and highlighted these issues, and proposed alternative approaches to address the impact of the issues.

All infectious disease testing systems use the biological functions of the antibody response to the antigen to generate a detectable signal. PNT uses infectious virus inoculated onto a cell culture to create viral plaques. In the presence of a serial dilution of patient sample, the dilution that reduces the number of plaques by 50% is used to determine the end point of the test. HAI uses the ability of some antigens to agglutinate red blood cells of specific animal species. When the antigen is incubated with specific antibodies, all the binding sites become unavailable and the haemagglutination effect of the antigen is inhibited. In a similar manner, complement fixation tests employs a combination of the patient sample, a standardised concentration of the target antigen and complement. The complement is “fixed” when the antigen:antibody complex is formed. If the complement remains unfixed due to a lack of antigen-specific antibodies (either because the patient has no antibodies or the antibodies have been diluted out by serial dilution), the free complement lyses antibody-coated red blood cells added to the test system. For each of these biological test systems, a serial dilution of patient samples can be tested to determine an “end point”; the dilution of patient serum at which the signal is no longer detectable. Traditionally, a two-fold increase

in doubling dilution end points in patient samples obtained 10-14 days apart indicates a recent (acute) infection.

In 1979, Voller *et al* reported the use of an enzyme-linked immunoassay for the detection of specific antibodies (30). This technique has since been applied to the detection of viral antigens such as hepatitis B surface antigen and HIV p24 (31, 32). The introduction of immunoassays (IAs) has revolutionised infectious disease serology. Chemical detection systems used in IAs (substrates) included enzyme-based colour reactions (EIA), radioactive labels, fluorescence and chemiluminescence (33). Usually, the test system has the target antigen bound to a solid, immovable phase (e.g. a microtitre plate, magnetic beads, plastic beads). The patient sample is incubated with the bound antigens and antibodies specific to the target antigen(s) are bound. After washing, a labelled anti-human conjugate is added, and the conjugate binds to the patient antibodies. After a further wash the chemical substrate is added and the resulting signal detected. Generally, there is a dose response, where the signal increases with the amount of bound antibody. However, depending on the assay design, there are plateaus in the dose response prior to the antibody concentration being detectable and another plateau when the test system is saturated. Between these two plateaus, the dose response is usually linear in nature with the rise in signal proportional to the increase in concentration of the antibody being detected (33). This dose response curve is often presented graphically as a sigmoidal curve. Importantly, this dose response is specific to the test system and cannot be assumed to behave with the same dynamics across different test systems, even when detecting the same analyte.

In the past two decades, immunoassays have become increasingly automated, with numerous continuous-access robotic platforms routinely used in clinical and blood screening laboratories. These test systems are well-controlled, highly sensitive, specific and precise, facilitating increased control of test results. During the early 1980s Mullis *et al* described a process called polymerase chain reaction (PCR), allowing the amplification of specific nucleic acid sequences (34). This technology initiated or further developed molecular techniques, since applied to infectious diseases, including the use of chemical probes to detect amplified nucleic acid; reverse transcription for the detection of RNA, and real-time or quantitative PCR (qPCR). These technologies allowed a transition away from slow and labour-intensive viral culture or direct antigen detection and are now applied for the detection of the causative agents for most infectious diseases. As with the developments associated with serological methods, NAT has allowed for the automation, standardisation and control of the test systems.

It is notable that the emergence of IAs and NAT coincided with the onset of the Human Immunodeficiency Virus (HIV) epidemic. This situation elicited large investments into research and development of technologies used for infectious disease testing. The use of HIV viral load to initiate and monitor treatment drove a need for standardisation across NAT, leading to the creation of a WHO international standard for HIV viral load (35). It also saw a strengthening of the regulatory environment (36-38), not only in clinical laboratories but especially in blood screening where several notable failures to apply emerging technologies to HIV, and subsequently Hepatitis C virus (HCV), had severe medicolegal consequences (39, 40). Governments of most developed countries introduced regulations to control the supply of *in-vitro* diagnostic devices (IVDs) used to detect pathogens that have a high risk to the community. In an increasing number of jurisdictions, manufacturers and testing laboratories are now required to have accreditation or certification to relevant International Organization for Standardization (ISO) standards. National regulatory bodies and associated frameworks were strengthened, requiring IVDs to comply with set performance standards, such as the European Union Common Technical Specifications (36). A Global Harmonisation Task Force, now replaced by the International Medical Device Regulators Forum, was established to standardise IVD requirements across jurisdiction and therefore reduce regulatory burden on manufactures (41). To

support countries lacking a national regulatory framework and to guide IVD procurement by funding and implementing bodies such as Global Fund, UNDP, World Bank and Clinton Foundation, the World Health Organization (WHO) established the Prequalification of IVDs, coordinated through the Department of Essential Medicines and Health Products. This activity focuses on testing for priority diseases in resource-limited settings, such as HIV, malaria and TB (42). Many countries with immature regulatory systems use WHO prequalification as a guide for selection of test kits. National regulatory systems such as the Therapeutics Goods Administration in Australia and the Food and Drug Administration in USA are independent and do not refer to WHO prequalification, whereas WHO prequalification process does consider registration of IVDs by national regulatory systems.

Standardisation and Control of Clinical Chemistry Testing

Over time, efforts to standardise and control medical testing were championed by clinical chemists, beginning with external quality assessment (EQA) schemes, introduction of standard methods and metrological traceability of measurands of higher order reference materials through ISO 17511 standard (28). National Measurement Institutes measure pure, high-grade analytical materials such as glucose or potassium using certified reference methods, such as high-performance liquid chromatography (HPLC) or atomic absorption, to produce certified reference standards, often traceable to an SI measurement. Through a chain of commutability, secondary standards are produced and these, in turn, are used to create calibrators and standards for use in medical test systems (**Figure 2**). Using this traceability hierarchy, results in clinical chemistry can assure patient samples tested on different test systems report the same result within a known confidence (43). EQA programs can monitor the success or failure of this process by systematically assessing the results of testing by laboratories using different test systems to detect the same measurand (28). Commutability of multiple clinical chemistry analytes has been a focus of professional bodies (44, 45).

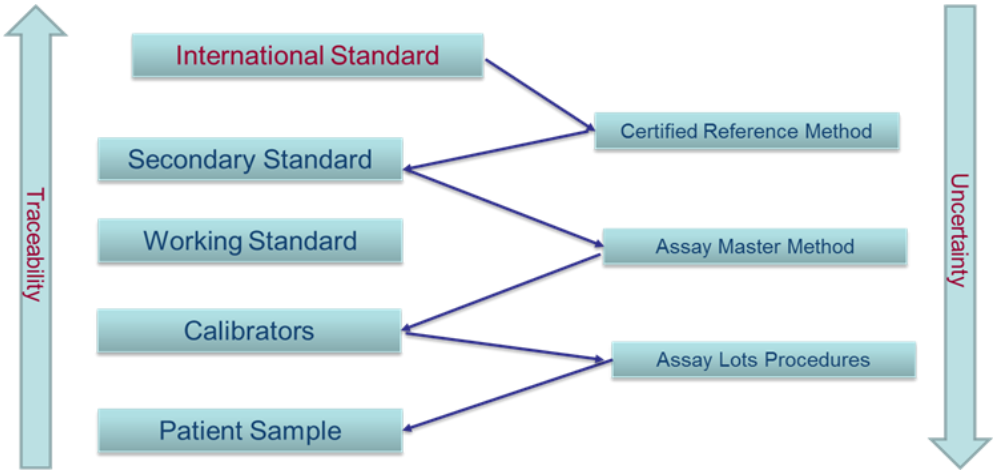


Figure 2. Hierarchy of development of international standards indicating the relationship of international standards (primary), secondary standards and tertiary (working) standards.

To control clinical chemistry test systems, modifications of statistical process controls created by Walter Shewhart and championed by W. Edwards Deming in post-war Japan, were introduced into the laboratory setting (46, 47). Assuming the results of repeated testing of the same sample would have Gaussian (normal) distribution, 95% of all normal results will be within +/- two standard deviations (SD) of the mean of the results and 99.7% within +/-3SD. Traditional quality control (QC) monitoring methods use this principle to

establish QC acceptance limits, noting that 5% of "true" results will be falsely rejected if mean +/-2SD is used as acceptance limits. Results outside of mean +/-2SD are further investigated to determine the root cause of the variation or are rejected and the test repeated. James Westgard published a set of rules, commonly known as Westgard Rules, which created a framework of decision making around the acceptance or rejection of QC results. These rules are almost universally adopted in clinical chemistry testing (48-50). Briefly, the Westgard rules consist of six or more separate rules to identify if a QC result is "out of control" (**Table 2**). The mean and SD of a QC dataset is calculated. If the subsequent QC results are within the mean +/- 2SD, the results are accepted, and testing continues. However, if the results deviates, the Westgard rules are used to accept or reject the QC results or signal a warning that the QC results should be closely observed until results return to expected levels. However, the application of Westgard rules is based on several assumptions, including that the QC results are normally distributed, and the QC results are commutable with patient samples. Standards promoting the use of these rules suggest as few as 20 QC results can be used to establish the mean and SD (51-53). This assumes that these 20 results are indicative of the inherent variation within that test system and are reflective of all subsequent QC results. The papers presented in **Chapter 4** demonstrate that none of these assumptions are true for infectious disease serology.

Table 2. Description of the common Westgard rules and the actions required.

Rule Notation	Rule Description	Action
$\mathbf{1_{3S}}$	One QC result is greater than mean +/- 3SD	Reject run
$\mathbf{1_{2S}}$	One QC result is greater than mean +/- 2SD	Warning
$\mathbf{2_{2S}}$	Two consecutive QC results greater than mean +/- 2SD	Reject run
$\mathbf{R_{4S}}$	One QC result exceeds 2SD and the subsequent QC result exceeds 2SD in the other direction	Reject run
$\mathbf{4_{1S}}$	Four consecutive QC results are greater than mean +/-1SD	Reject run
$\mathbf{10_{X}}$	Ten consecutive QC results fall one side of the calculated mean	Reject run

More recently, a risk-based approach to monitoring the performance of medical test systems has been encouraged (54), with some commentators promoting six sigma principles (55-57). Six sigma, developed in 1986 for the Japanese manufacturer Motorola, is a process that identifies and measures the numbers of defects in a process and then implements design changes to reduce the defect rate. Six sigma assumes that, when implemented successfully, the process controls will produce fewer than 0.0005% defects, or less than 3.4 defects per million times the process is undertaken. There have been several attempts to utilise this concept in medical testing (56, 58), utilising existing metrics, such as external quality assessment results and quality control data to develop tolerance limits (59). These limits can then use used to identify, and then measure, the number of defects encountered when performing laboratory testing. However, at the time of writing this thesis, no systematic review of these rules' applicability to other disciplines such as infectious disease serology has been undertaken.

There are significant differences between the measurement of clinical chemistry analytes and infectious diseases serology. The underlying reason for these differences is that, when testing for an inert chemical such as glucose, the test system is determining the actual quantity (*how much*) glucose is present. Although there is a dose response when testing for antibodies, the test system is also determining the efficacy of binding (*how well*) antibodies to the antigen. A patient sample having low levels of antibodies with high

affinity and avidity to a specific antigen could have a higher level of reactivity compared with a sample with a high concentration of low-avidity antibodies.

As infectious disease testing became more automated, test systems used in clinical chemistry have been adapted for the detection of infectious disease antibodies and antigens. Many larger pathology laboratories have introduced the concept of a "core laboratory", where samples for an increasing range of analytes, including infectious diseases, are tested on the platforms linked by a "track", increasing the efficiency of the laboratory. As these core laboratories are commonly overseen by clinical chemists, it is not surprising that methods used to standardise and control clinical chemistry testing were implemented for infectious disease testing. Unfortunately, it has become evident that these methods, notably the standardisation of quantitative antibody testing and the use of Westgard Rules, are not appropriate for infectious disease testing. Over the past three decades, a significant body of work has provided insight into these deficiencies and have culminated into the abandonment of the use of the WHO International Standard for the calibration of anti-rubella IgG assays and in the development of an alternative and more appropriate method to monitor and interpret the results of infectious disease serology QC. This thesis outlines my contribution to the understanding of this situation and identifies the significant impact my work has contributed to the standardisation and control of infectious disease testing.

Variability of Markers for Infectious Diseases

When discussing measurable quantities, the International Vocabulary of Metrology utilises the term "measurand" to describe the particular quantity subject to measurement. The specification of a measurand may require statements about quantities such as time, temperature and pressure (60). Many clinical chemistry measurands, such as glucose, are small molecules, have minimal heterogeneity and can be described as a chemical formula (e.g. $C_6H_{12}O_6$). These are referred to a "type A" analytes. In contrast, analytes like antibodies that measure functional, biological activity, called "type B" analyte (**Table 3**), are heterogeneous and are not directly traceable to SI units (61, 62). Unlike clinical chemistry analytes, the standardisation of testing for infectious diseases has a chequered history. Although numbers of international standards for infectious diseases have been developed since the 1960s and have subsequently been used to try to standardise serological tests, these efforts have, by and large, been unsuccessful (13, 55). This is due to many issues. Of note, many international standards for serology were developed without due consideration to metrological principles (54, 56-58); serological assays are generally qualitative detecting the presence or absence of antibodies. The analytes being measured are complex, biological and polyclonal in nature (13, 54, 59). Antibodies can be of variable classes/subclasses, fragmented, polyclonal or monoclonal, free or complexed and have variable affinity and avidity (54). Unfortunately, although these factors have been recognised, and warnings published (54, 59, 60), it has been assumed that the method of standardisation used for "type A" clinical chemistry measurands would be suitable for application to antibody testing, resulting in unforeseen consequences that have taken decades to resolve (13, 61). As detailed in the publications presented in Chapter 2 of this thesis (62, 63), the consequences include a lack of comparability of quantitative test results reported by assays measuring what is assumed to be the same analyte; poor clinical interpretation of results; difficulties in the design of assays and a general inaccurate acceptance that "type B" analytes behave in the same manner as "type A" analytes (64).

Pathogenic organisms are complex and variable in structure and often mutate over time. All organisms, whether viruses, bacteria, fungi or parasites, have one or more immunogenic sites. Humans elicit an immune response when exposed to these antigens. For example, rubella virus has three immunogenic proteins; two envelope proteins designated E1 and E2 and one capsid protein, but has only a single serotype (14). Similarly, wild-type measles virus has eight clades containing 24 genotypes based on the nucleotide

sequences of their hemagglutinin and nucleoprotein genes, which are the most variable genes in the viral genome (63).

Table 3. Differences between clinical chemistry and infectious disease serology testing

Clinical Chemistry	Infectious Disease Serology
"Type A" inert analyte • Known molecular structure • Known molecular weight • Invariable composition • No change over time	**"Type B" functional biological analyte** • Variable structures • Different classes and subclasses • Antibody response matures over time • Antibodies may be fragmented, polyclonal or monoclonal, free or complexed • Variable avidity and affinity
Several medical decision points • e.g. 3.9 – 5.6 mmol/L mean fasting glucose to differentiate hypo- and hyper-glycemia; • 0.74 to 1.35 mg/dL creatinine normal range for adult males to assess kidney function and diet; • 3.6 to 5.2 mmol/L potassium, where hypokalemia or hyperkalemia can lead to heart failure	**Single decision point** • Determining the absence of presence of antibodies by use of a single cut-of value • Viral load result being above the limit of quantification of the assay
Quantitative • Determining absolute amount of analyte	**Qualitative** • Determining binding efficiency • Use chemical signal to detect measurand
Single homogeneous molecule • No or minimal heterogeneity • Test developed for specific molecular composition	**Multiple and varying antigens** • Different genotypes/serotypes • Antigenic mutations
Lower level of regulation • Generally low-risk analytes • Assessed as Class 2/B IVDs by regulatory authorities • Self-declared evidence • No batch release testing required • No risk to community • Low to moderate risk to individual	**Highly regulated** • Generally moderate to high risk • Assessed as Class 3 or 4/C or D by regulatory authorities • Mandatory technical file review by authorities • Class 4/D IVDs undergo mandatory batch release testing • Mod/High risk to community • Mod/High risk to individual
Linear dose response curve • Highly sensitive tests detect low levels of analyte • Normal samples have relatively high levels of circulating analyte	**Non-linear dose response curve** • No response if analyte concentration is low • No increase in response if test system is saturated
Adjust for reagent lot variation (Bias) • Can re-calibrate test system to adjust for bias • Calibrators traceable to international standard	**Cannot adjust for reagent lot variation (Bias)** • Tests are highly regulated not allowing modification • No international standards • Lack of commutability to international standards
International standards available • Well defined international standards available • Secondary standards traceable to international standard	**Poor or no international standards** • No international standards for many analytes • Lack of commutability to international standards • Many tests are not calibrated to internationals standard when they exist
Certified reference methods (CRM) • Well established CRM • e.g. Atomic absorption, HPLC	**No certified reference methods** • No CRM available • Variable quantitative results between test systems

The eight clades are designated A to H, with numerals identifying the individual genotypes. Measles virus is immunologically stable with no detectable serological variation (64). In contrast, HIV elicits antibody responses to group-specific antigen protein p24 and its precursor p55; antibodies to the envelope precursor protein gp160 and proteins gp120 and gp41. Antibodies to the polymerase gene products p31, p51 and p66 are also commonly detectable in infected patient samples. HIV-1 has four groups – M (Major), O (Outlier), N (non-M, non-O) and P. Group M has nine subtypes (65). HIV antibody assays usually test for antibodies to HIV-1 and HIV-2 as well detect HIV p24 antigen. The assays are designed to detect each of these analytes but report a single positive or negative result. So, unlike clinical chemistry, the test systems are complex, detecting several different analytes concurrently.

Generally, the number of immunogenic sites increases with the complexity of the organism. Like HIV, many organisms have different genotypes and/or serotypes. When organisms share a large percentage of genome but vary in the immunogenic regions, the difference in the immune response they elicit can be detected and differentiated. As an example, whereas rubella virus has a single serotype, dengue virus has four serotypes DEN-1, DEN-2, DEN-3 and DEN-4, which can be differentiated serologically (66). The variation in the genome of pathogens can also be detected by nucleic acid testing. HCV has six common genotypes designated 1-6; their distribution varying around the world. So differentiation of HCV can be achieved by either serotyping or genotyping (67).

The immunogenicity of organisms also changes over time through mutation. Recent experience with SARS-CoV-2 has highlighted the importance of the ever-changing nature of the virus (68). SARS-CoV2 remained relatively evolutionary stable for the first 11 months, after which there was a rapid emergence of variants globally. The nucleotide mutation rate of the S gene was $8.066 \times 10{-}4$ substitution per site per year, which was at a medium level compared with other RNA viruses (69). Changes in the spike protein, in particular the variable receptor-binding domain which bind to receptors found in the respiratory system, gastrointestinal tract, heart, and kidneys are of greatest significance, reducing the neutralising antibody response and increasing infectivity (5). Whereas single point mutations were commonly detected, multiple mutations were found in emerging variants of concern (70). Mutations of SARS-CoV-2 have been associated with increased transmission and virulence of the virus and have the potential to evade vaccine-induced and/or natural immunity (71) due to escape mutants. Changes to the virus potentially can impact on the efficacy of both serology and molecular diagnostic tests.

The influenza virus is well known for its antigenic “drifts” and “shifts”. Antigenic drifts are small changes in the genes of the virus over time (29). This phenomenon is also seen in HIV (65), which is one of the reasons the development of a vaccine has proven difficult for both organisms. Antigenic shift is a major change in the virus. In Influenza virus, a shift is commonly associated with the combination of the genomes of two influenza viruses derived from different animal species, creating a strain that can evade previously developed antibody responses. Some viruses are prone to mutation. Hepatitis B virus (HBV) has a unique life cycle which includes an error-prone enzyme, reverse transcriptase, and a very high virion replication rate. Antiviral treatment of blood-borne infections such as HIV, HCV and HBV can select out sub-populations creating a change in immunogenicity over time (72). Therefore, infectious disease test systems testing for a specific organism needs to be designed to detect variable antibodies responses and nucleic acids present in ever-changing organisms.

Current IAs are developed to detect antibodies subclasses to a specific antigen e.g. anti-EBV viral capsid antigen IgM or anti-hepatitis B core total (IgM and IgG) antibodies. The manufacturer uses various sources of antigen (whole virus, disrupted virus, purified viral antigens or recombinant antigens) and conjugates, which may be polyclonal, across multiple subclasses (total antibodies) or class specific, or monoclonal antibodies directed to a specific viral epitope. Several monoclonal antibodies could be used in the design

of an assay. Test systems for antibodies must account for this array of variables, emphasising the difference between infectious disease serology and testing analytes such as glucose or potassium. Unlike inert clinical chemistry analytes, the measurand detected in infectious disease serology is extremely variable in individual immune responses, antigen targets, antibodies detected, assay design, and the affinity and avidity of antibodies. Therefore, unless test systems use the same components (antigen source, conjugate and substrate), the measurand is specific to each test system and quantitative results between assays cannot be compared. As an example, it has been well established that the quantitative results of anti-rubella IgG assays are not comparable (14, 73-76). The principles of standardisation (**Chapter 2**) and control (**Chapter 4**) traditionally applied to clinical chemistry analytes cannot be used in infectious disease serology (77).

Quantification of Antibodies to Infectious Diseases

The level of antibody in a patient is rarely useful in a clinical setting. In biological test systems, a rise in antibody titre, from negative to detectable (seroconversion) can be used to confirm an acute infection. In some infectious disease tests, such as syphilis rapid plasma reagin, a doubling dilution titre of greater or equal to 1:8 is indicative of untreated infection. Occasionally, clinicians will use a low-level antibody response to trigger re-vaccination, however, this is not necessarily appropriate (78). Rarely does an IA manufacturer's instructions for use suggest a rise in signal reported by their IA indicate an acute infection.

Although there are many international standards for infectious disease serology, routine reporting of results in international units (IU) and having a defined clinical cut-off associated with this unitage has been limited to three main analytes: anti-rubella IgG, anti-hepatitis B surface antigen (HBsAb) and anti-measles IgG. The cut-off for immunity for anti-rubella IgG and HBsAb is assumed to be 10 IU/mL and 10 mIU/mL respectively. As a result of my published work, the cut-off for rubella has been seriously questioned and is under review (77, 79). Number of studies have identified poor correlation between the quantitative anti-rubella IgG test results reported by different test systems. One study tested a panel of 325 samples that had been reported as having undetectable anti-rubella IgG in seven other quantitative rubella IgG tests and an immunoblot. Only 129 (39.7%) of the samples were found to be anti-rubella IgG negative by all the other tests. However, 59% of the samples were reported as immunoblot positive, indicating initial false negative reactivity. Of note, there was up to a ten-fold difference in the quantitative results reported by the tests (**Table 4**) (79).

Results of different HBsAb assays have also been shown to report different quantitative results when testing the same samples (80), the author concluded "*Thus, our study shows that levels of anti-HBs determined by one assay system cannot be compared with those determined by other systems, although all the assays are calibrated with the same international standard*". The measles PNT titre that corresponded to the protective titre was found to be ≥120 mIU/mL in a study that reviewed protective immunity of university students after a measles outbreak (81). PNT assays have been standardised against the WHO measles antibody international standard (currently the WHO 3rd international standard; NIBSC 97/648). However, the package insert of that International Standard states "*This preparation has not been calibrated for use in ELISA assays and/or a unitage assigned for this use*."

The history of the development of international standards for rubella IgG serology is informative (**Figure 3**). Serological international standards were first produced in the 1960s, primarily for the purpose of assessing the potencies of vaccines for rubella and measles. Studies on the preparation of standards are referred to the WHO Expert Committee on Biological Standards for adoption (82-85).

Table 4. Comparison of qualitative and quantitative results of eight commercial anti-rubella IgG tests when testing patient samples having low levels of rubella IgG confirmed by immunoblot. Quantitative results, expressed as international units per millilitre, and the qualitative interpretation are presented.

Test System (Equivocal range in IU/mL)																
Mikrogen Immuno-blot	Beckmann-Coulter DxI 10-14		Abbott Architect 5-9		VIDAS bioMérieux 10-15		Siemens Enzygnost 5-6		DiaSorin Liaison 5-9		Roche Cobas 6000 No equivocal range		Siemens Centaur 5-10		Serion 10-20	
P[1]	11,1	E[2]	1,8	N[3]	13	E	16	P	21,9	P	4,3	N	42,1	P	28,4	P
P	12,8	E	4,3	N	13	E	6	E	5,4	E	11,6	P	11,1	P	7,36	N
P	12,2	E	4,1	N	11	E	5	E	8,8	E	10,5	P	25,1	P	14,5	E
P	9,4	N	5	E	10	E	6	E	3,5	N	60,4	P	10,7	P	8,11	N
P	9,8	N	7,6	E	13	E	8	P	5,5	E	5	N	11,7	P	10,8	E
P	7,7	N	4,8	N	9	N	5	E	6,3	E	61,1	P	13,3	P	9,35	N
P	6,8	N	4,2	N	7	N	5	E	<3	N	11,8	P	9,3	E	6,1	N
P	8,9	N	5	E	14	E	8	P	5,7	E	41,2	P	17,1	P	10,6	E
P	8,3	N	4,8	N	11	E	8	P	8,8	E	11,4	P	13,6	P	12,1	E
P	12	E	4,1	N	12	E	7	P	8,6	E	7,7	N	23,5	P	12,5	E
P	12,2	E	7	E	10	E	13	P	4,9	N	>500	P	14,1	P	10,8	E
P	9,5	N	6,1	E	12	E	8	P	4,4	N	19,2	P	7,4	E	11,4	E

(Adapted from Bouthry (79))

P Positive
E Equivocal
N Negative

The first international reference preparation of anti-rubella serum was prepared in 1966 using a pool of convalescent-phase human sera (86). It was replaced in 1968, with the second international reference preparation of anti-rubella serum designated BS/96.1833, also known as RUBS (82, 83, 87). Despite its name referring to the standard as being of human serum origin, implying it was the same matrix of patient samples, it was prepared from normal concentrated, human immunoglobulin. The current rubella IgG international standard, designated RUB-1-94, was introduced in 1995, but was originally developed by the Staten Serum Institut (Copenhagen, Denmark) in the 1970s (82, 88). It was also prepared from pooled, concentrated human immunoglobulin in equal parts with saline. It should be highlighted that RUBS and

RUB-1-94 preparations are purified immunoglobulin rather than normal human serum or plasma (84) and therefore a different sample type (matrix) i.e. serum, plasma or purified immunoglobulin, to the patient's samples.

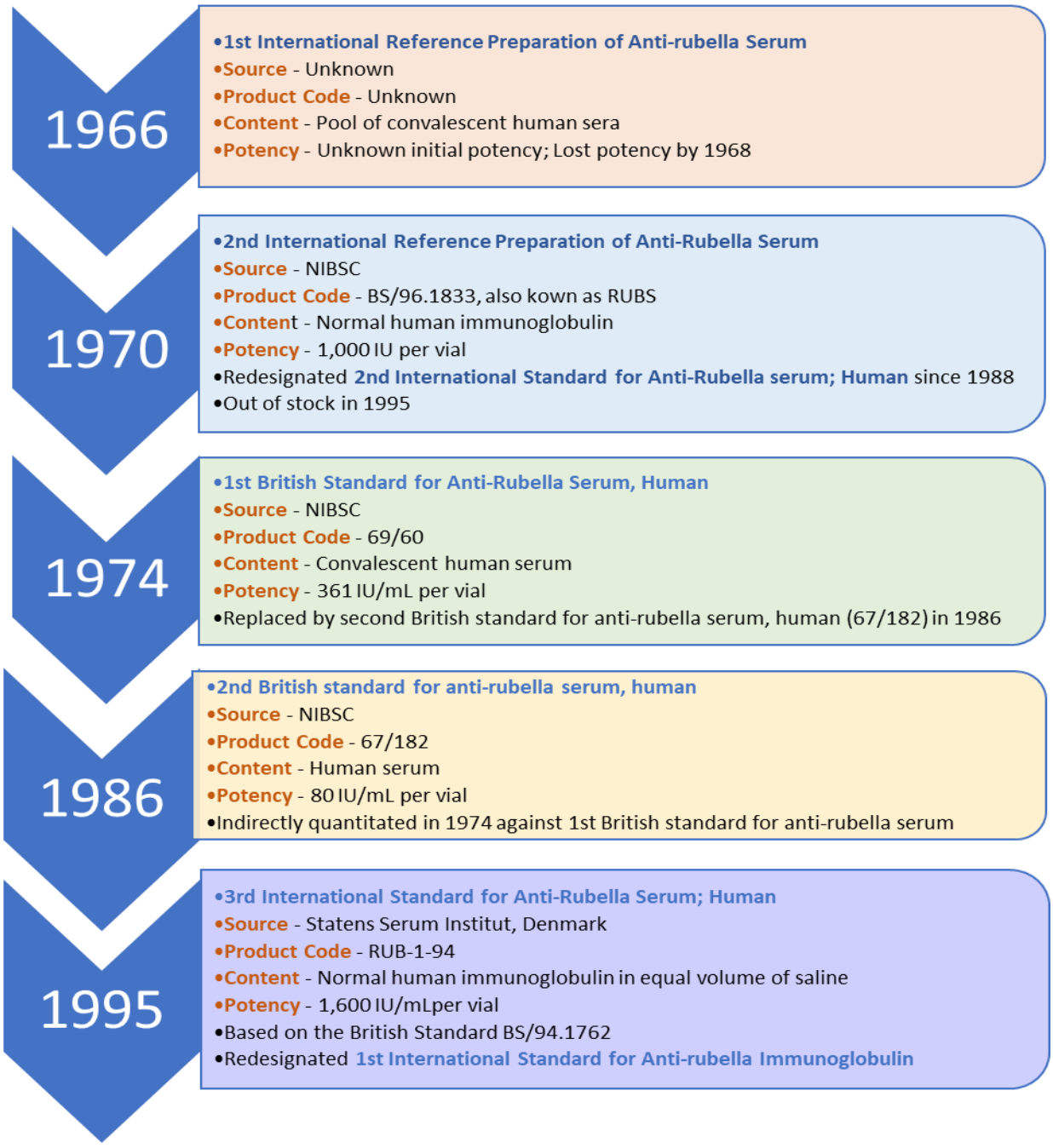


Figure 3. History of the development of anti-rubella standards, where potency if the standard is the designated amount of analyte (in this case anti-rubella IgG) measured in international units per millilitre.

In parallel with the development of international standards for anti-rubella, National Institute of Biological Standards and Controls (NIBSC, Potters Bar, UK) created and released British standards. The 1st British Standard for Anti-Rubella Serum, Human, designated 69/60, was developed in 1974 using convalescent serum (89). At the time of development, several candidate samples were compared in parallel including

sample designated 67/182, which was later selected to be the 2nd British Standard for Anti-Rubella Serum, Human in 1986 when the first standard was depleted (90). Importantly, this standard was calibrated against the 2nd WHO International Reference Preparation of Anti-Rubella Serum, Human (91).

The potency of RUB-1-94 was assessed in a multicentre trial, including 11 laboratories from seven countries. The laboratories used biological HAI or radial haemolysis assays or first generation EIAs. Since this time, most commercial anti-rubella IgG assays have been calibrated against this international standard and report results in IU/mL (14, 88). However there have been many studies demonstrating a lack in correlation of the quantitative results between assays reporting anti-rubella IgG results in IU/mL (14, 73, 75, 76, 92-96). Often, these discrepancies in quantification lead to different clinical interpretations (**Table 4**), with individuals being assumed to be immune when they are not, or visa-versa (75). Unfortunately, this situation can have adverse clinical outcomes or incorrect interpretations, resulting in unnecessary terminations of pregnancy or, conversely delivery of babies with congenital rubella syndrome (14, 75, 76).

More recently, the First WHO International Standard for anti-SARS-CoV-2 immunoglobulin (human) (Code: 20/136) has been released by NIBSC. It is made from human plasma collected from convalescent individuals and has an assigned value of 1000 IU per vial. The instructions for use indicate that the intended use of the standard is "*for the calibration and harmonisation of serological assays detecting anti-SARS-CoV-2 neutralising antibodies*" and that "*for binding antibody assays, an arbitrary unitage of 1000 binding antibody units (BAU)/mL can be used to assist the comparison of assays detecting the same class of immunoglobulins with the same specificity (e.g. anti-RBD IgG, anti-N IgM, etc.*)". Only months after release of the standard, commercial manufacturers had released "quantitative SARS-CoV-2" EIAs. Over the following years, the emergence of numerous SARS-CoV-2 variants will have also added complexity to the standardisation of SARS-CoV-2 antibody testing (68). I have initiated an investigation comparing the quantitative results reported by these assays, with an aim to determining whether standardisation of SARS-CoV-2 serology assays experiences issues similar to that identified with rubella IgG and HBsAb. Results of this testing have been finalised but data analysis and submission of a manuscript for peer review is yet to be performed.

Difficulties in the Standardisation of Antibody Quantification

The theoretical reasons underlying the difficulties in standardisation of "type B" analytes such as antibodies quantification is detailed elsewhere (62, 97, 98). The three elements of a "type B" analyte is the system (e.g., serum), the component (e.g., anti-rubella virus IgG, being the antigen specificity and the antibody isotype(s)), and the kind of quantity (essentially the biological response or biological activity); together making up the measurand (62). Metrological principles for establishing standards, which are difficult to apply to biological standards, require nominated reference laboratories to prepare standards comprised of the same matrix used in the testing system. The measurand must be defined and be in an invariable form in the sample. The amount of measurand should be measured using a certified reference method and ideally expressed in SI units (99, 100). These principles cannot be achieved in the case of biological, "type B" analytes because the measurand is variable, not in pure from and there are no certified reference methods available.

For almost all biological standards, one or more of these elements varies across test systems. As described above, the antibodies that develop in response to infection vary due to antigenic differences across genotype and subtypes, the stage of disease progression, antibody avidity and affinity, the test systems used (different antigen sources or detection systems), as well as functional variability within the test. Therefore, the measurands differ across test systems used to detect and quantify the same analyte e.g. anti-rubella IgG. It is not surprising that it has proved impossible to standardise quantitative antibody

testing. Each test system is quantifying different measurands, and therefore the results cannot be compared. It should be noted that, even though serological assays used to detect the antibodies directed at the same organism are difficult to standardise, they usually have comparable clinical sensitivity and specificity, where the clinical sensitivity measures the ability of the test kit to detect analytes when they are present, and the clinical specificity measure the ability to report a negative result when the analyte is absent. These performance criteria are assessed by the IVD regulator prior to the product being sold into the market (36, 38, 101). Other performance criteria assessed include the analytical sensitivity (or limit of detection) which determines the lowest amount of an analyte that will cause a positive signal in the test kit, analytical specificity that assesses cross-reactions to other analytes that may be present in the sample, the effects of interfering substances such as excessive lipid, haemoglobin or bilirubin and linearity, a measure of the dose response. The positive and negative predictive value determines the percentage of positive and negative results were truly positive and truly negative, respectively. Both predictive values are influenced by the prevalence of the analyte (102). So, although a patient sample may have an anti-rubella IgG result of 20 IU/mL on assay one and a result of 200 IU/mL on assay two, but both assays report a positive result. This difference in quantification of rubella antibodies has been described previously (73, 79, 92, 93, 95). Fortunately, clinical decisions are made on the qualitative result and rarely on the quantitative result.

Although standardisation of serological tests has been fraught, there are some factors compelling manufacturers to attempt calibration using the WHO International standards. Arguably, this began in 1995 with the WHO Expert Committee on Biological Standards 45th report, which stated that there was a need for a rubella standard "*for the calibration of diagnostic kits*" (83). Although not explicit, both European and USA IVD directives suggest that all IVDs be traceable to a higher order standard where one exists. The 1998 European Directive stated "*The traceability of values assigned to calibrators and/or control materials must be assured through available reference measurement procedures and/or available reference materials of a higher order*" (38). Therefore, many IVD manufactures have used the WHO international standards to calibrate a range of assays to report in IU/mL, including tests for antibodies for toxoplasma, cytomegalovirus (CMV) and syphilis, although quantitative reporting has rarely been used for clinical decision-making outside anti-rubella IgG and HBsAb testing. There are currently many International Standards available for antibody testing (**Appendix A)**.

Standardisation of Antigens and Nucleic Acid Testing

The detection of antigens has been clinically useful, with a recent example being rapid testing for SARS-CoV-2 antigens as a public health tool (103). Rapid tests detecting influenza virus, sexually transmitted infections such as chlamydia and gonorrhoea as well as testing for malaria and dengue are available. These tests are qualitative and therefore out of scope of this thesis. Only a few quantitative antigen tests are available, and these are addressed below. There are several international standards for infectious disease antigens, such as Hepatitis B surface antigen (HBsAg) (104) and HIV p24 (105), which have been used to calibrate IAs. However, the quantitative results of these tests are not commonly used in clinical decision making, mainly because they have been superseded by nucleic acid viral load tests. Limited evidence would suggest that standardisation of antigen testing is less problematic than antibody testing (32). Whereas HBsAg quantification may have some clinical use (106), HBV viral load testing is more commonly used to monitor therapy and disease progression (107). Similarly, HIV viral load testing is more commonly used in clinical decisions compared with HIV p24 quantification, and only a few studies had investigated the relationship between HIV p24 levels and disease progression (31) before the viral load assays replaced quantitative HIV p24 assay.

The standardisation of viral load quantification has been more successful than infectious disease serology. When detecting nucleic acid, primers are chosen to target conserved regions of the organism genome. The

signal of the assays is proportional to the amount of target in the sample. Therefore, even if assays target different conserved genome sequences, it is expected that signal of assays calibrated with an international standard, will be comparable. A comprehensive review of standardisation of molecular testing for infectious agents was recently published (108). Standardisation of viral load testing has been championed by Standardisation of Genome Amplification Technology, formed by the WHO and coordinated by the NIBSC and the Paul Ehrlich Institute (Langen, Germany). The Joint Committee for Traceability in Laboratory Medicine was formed in 2002 and facilitates the traceability to higher order standards. Results reported by viral load assays calibrated against international standards have demonstrated commutability, especially for blood-borne viruses HCV and HIV RNA and HBV DNA, which were the first organisms to have international standards created; in 1997 (HCV RNA) (109) and 1999 (HIV RNA and HBV DNA) (35, 110). Since this time, more than 20 international standards for viral and parasitic (malaria and toxoplasma) nucleic acids have been released (108). The potency of these international standards is determined by collaborative studies, whereby candidate materials, usually plasma from infected donors, are tested by numbers of laboratories using different technologies (35, 108-112). The consensus potency is calculated and expressed as in International Units, a measure of biological functionality rather than being expressed as SI units. This approach assumes that variation in extraction efficiency; region, length and conservation of the target; and detection systems are accounted for. However, some institutes have utilised digital droplet PCR to more accurately quantify the numbers of copies of a reference standard and, accounting for extraction efficiency of the test system, compare the copy numbers with the amount, by weight, of RNA or DNA using HPLC (113). In this way the viral load can be expressed in SI units (ng/*u*L), reported in a similar manner to "type A" measurands (114).

Although standardisation of nucleic acid testing has been successful, it is not without issues. Creation of molecular standards need to consider amplification efficiency of qPCR. The reverse transcription step used in the detection and quantification of RNA is known to be variable (115) and an estimation of the efficiency of the reverse transcription step and amplification should be factored into the final quantification (116). Digital droplet PCR has been promoted by some as an alternative to quantify nucleic acids (117). NIBSC manufactures and distributes international standards, providing them on request to IVD manufacturers and testing or research facilities. Although there are some restrictions around the supply of international standards, some standards have been exhausted and have required replacement. For example, the current HCV international standard is the sixth standard. To assess commutability between each new standard, NIBSC compared the new release with the previous version, rather than the initial version. An argument could be made that a better approach would be to reserve a single international standard that is used solely to create secondary standards and have these provided to manufacturers and testing facilities. In this way, commutability from standard to standard could be strengthened. Fragmentation of the target genome found in the patient sample can also cause difficulties in the quantification of nucleic acid, as experienced with herpesviruses such as EBV and CMV DNA (118, 119). Even in these circumstances, it has been demonstrated that results from tests calibrated with an international standard correlate better than uncalibrated tests. Only through standardisation of assays can clinical thresholds be created for monitoring treatment efficacy or intervention (108, 120, 121).

Control of Infectious Disease Testing

Quality control processes are an important mechanism used to monitor the performance of a test system over time (49, 122-124). Quality control samples, usually one having reactivity close to the medical decision point(s), are tested frequently and the results monitored using Levey-Jennings graphs (**Figure 4**).

All test systems experience normal variation due to changes in instrument calibration and maintenance, operator processes, conditions of reagent transport and storage and environmental conditions such as

temperature and humidity. In infectious disease testing, the greatest variable effecting quality control test results is due to changes in reagent lot numbers over time. As these reagents contain complex biological components including antigens and antibodies, maintaining exact reproducibility when components are replaced during manufacture is not possible. Therefore, especially when new master lots are introduced, significant but normal variation in quality control results is observed (122, 125, 126).

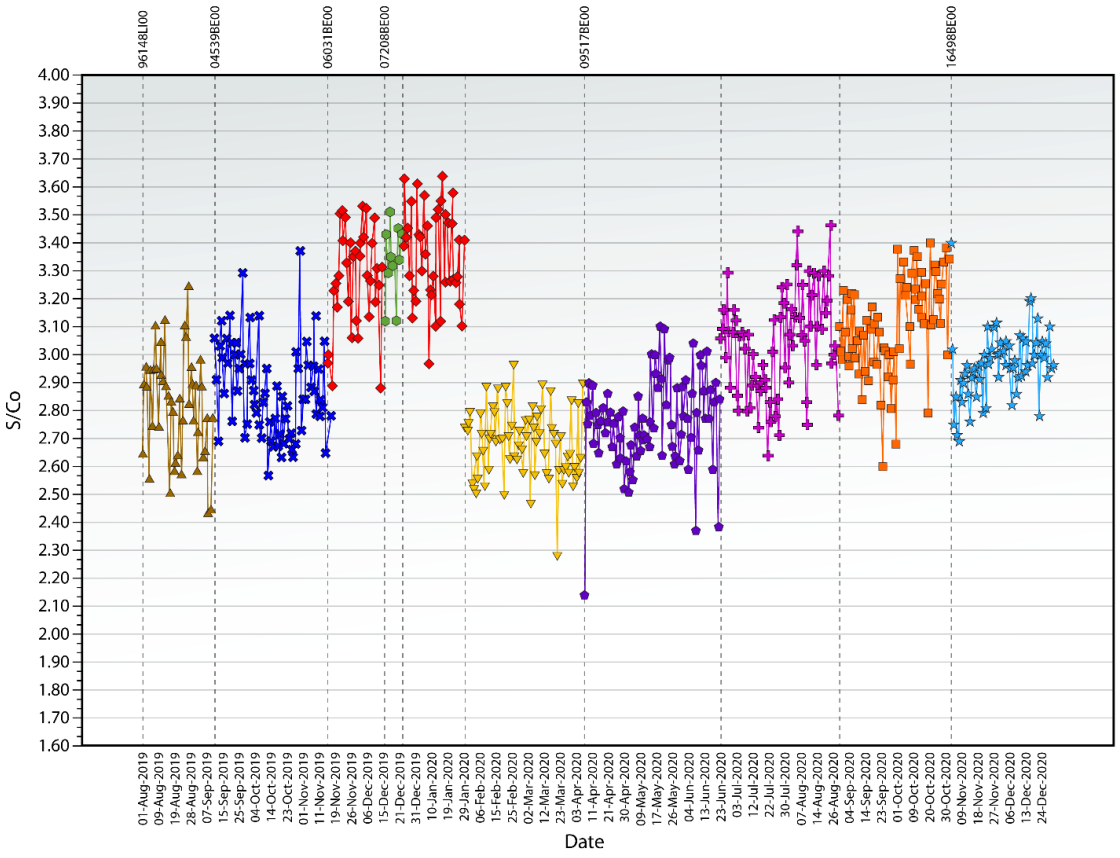


Figure 4. Levey-Jennings chart of a quality control (QC) tested over time. The Y axis represents the signal to cut-off (S/Co) of the QC result. The X axis represents the date of testing. Each colour represents results obtained from different test lot numbers. The changes in the QC test results indicates the variation experienced by the test system. Lower and upper acceptance are not displayed.

The quality control of infectious disease testing suffers from a different, but related, set of issues compared with attempts of standardisation of serology. The application of control processes such as Westgard rules, which are almost universally accepted for use in clinical chemistry, have been applied to infection disease testing without systematic studies to determine their applicability to tests measuring biological functionality. Internationally accepted guidelines for quality control use in medical pathology have been written by clinical chemists, for clinical chemistry (51, 52, 54, 127, 128). However, QC results for infectious disease serology do not have Gaussian (normal) distribution required by these QC methods, because bias is introduced by each new reagent lot. Therefore, these commonly accepted control principles do not apply to infectious disease testing, especially serology (125). Until recently, no systematic review of the applicability of these QC rules to infectious disease serology has been undertaken. In fact, apart from my peer-reviewed papers, jointly published with the colleagues from the National Serology Reference Laboratory, Australia (NRL) and other collaborators, there have been very few peer-reviewed publications on the application of QC principles on infectious disease serology in the past two decades, although a paper reviewing the results of CMV, EBV, and HIV viral load QC results was recently published (57).

When establishing acceptance limits, guidelines suggest using 20 to 30 data points to calculate the mean and SD (49-53, 127, 129). This principle can be applied to clinical chemistry where standardisation protocols facilitate traceability of patient's results to an international standard through a series of calibrations of secondary and working standards, to manufacturer assay calibrators (99). Recalibration of the test system accounts for any bias associated with changes in assay component, resulting in minimal reagent lot-to-lot differences in result. Therefore, patient or QC samples tested on different reagent lots are expected to report comparable quantitative values, within the precision of the test. The 20 to 30 data points are therefore predictive of all future QC test results.

Infectious disease serology test results are qualitative, being derived from a quantitative measurement of a signal. However, the quantitative results, which are usually a ratio of the signal to a manufacturer defined cut-off, are assay-specific and not traceable to an international standard. As described above, international standards have not proven useful in the calibration of IAs across different test platforms. The issue is further complicated because, when testing the same QC sample over time, the introduction of new reagent lot numbers causes a change in the reactivity of the QC result (122, 124, 130). Previous publications have demonstrated that changes in reactivity due to a lot-to-lot change is more significant a source of variation than within-lot imprecision (122-124). The mean of results calculated from previous lots of reagents and used to establish QC acceptance limits no longer apply to the new reagent lots, subsequently causing QC rejection, as represented in **Figure 4**. This situation creates difficulties in interpretation for laboratories using Westgard rules, as guidelines are silent as to what approach should be used when reagent lots cause a change in QC reactivity.

Proponents for the use of traditional QC methods for monitoring infectious disease serology suggested that serological assays are "*just another assay which follows the same laws as any other*" (131). Evidence rejects this statement (132). It has been suggested that SD could be calculated on each new reagent lot. This approach is impractical, as changes in lot number occur frequently and the calculated SD would be valid for only a short period of time prior to the next change of reagent lot. A pooling of SDs from multiple reagent lot has also been suggested (131), but both of these approaches account only for imprecision and not bias, whereas bias is of significant importance due to the reagent lot-to-lot variation. Others have suggested that a national approach evaluating new reagent lots (130), or that patient-based real-time quality control, where the results of patient testing is monitored over time (133) could be useful alternatives to the traditional approaches to QC. However, these have not been validated for use in infectious disease serology.

According to CLSI EP23A guideline, QC monitoring should be based on the risk of occurrence and the severity of harm caused (54). In infectious disease serology, the greatest risk is the reporting of a false negative result for diseases that pose a risk to both the individual and the community, such as HIV or hepatitis. This is especially true in a blood donor screening environment. A false positive test result may cause unnecessary distress to the patient, medicolegal complication and possibly inappropriate treatment. A false negative result may lead to transfusion-transmitted infections. Some test results may lead to unnecessary medical intervention, such as a termination of pregnancy in the case of rubella testing.

The development of a well-designed, risk-based QC process is essential for testing for infectious diseases. Variation in testing is commonplace and the extent and frequency of variation should be monitored over time. Variation is derived from changes in reagents (123), processes, consumables (134) and equipment. By collecting metadata with the QC results and systematically graphing the results, investigations into the cause of unacceptable variation can be facilitated. This approach would be applicable to semi-quantitative serological assays, NAT and PoCT (135).

Outline of Thesis

Testing for infectious diseases through serology and NAT have developed substantially over the past two decades; changing from tests utilising biological functions (such as haemagglutination inhibition or complement fixation) to high throughput, robotic, autoanalyser. As these changes were implemented, and testing moved from the microbiology to clinical chemistry departments, the complexity of testing for infectious diseases was somewhat ignored, especially where standardisation and control of these tests are concerned. This is evident in the lack of professional guidelines or standards specifically referencing infectious disease standardisation and control. Given my unique opportunity to work for an organisation that seeks to improve quality of testing for infectious diseases, access to numerous networks of infectious disease testing laboratories and overseeing several quality assurance programs that systematically collect comprehensive testing data, I have sought to better understand the nature of infectious disease testing and propose novel methods and concepts based on the analysis of these data.

In the subsequent chapters, I present publications that highlight issues and propose solutions to standardisation and control of both serology and NAT infectious disease testing. In **Chapter 2**, I focus on the standardisation of rubella serology, as a representative model of the failures of standardisation of serological testing. Along with HBsAb, anti-rubella IgG testing results are routinely reported in IU/mL. Unlike HBsAb testing, anti-rubella IgG test results can lead to significant medical intervention such as termination of pregnancy. In this chapter, I present two significant first authored papers, supplemented by several other first and co-authored papers that demonstrate the issues and consequences associated with poor standardisation of serology tests and questions whether standardisation of antibody testing is possible. Whereas the feasibility of standardisation of serology assays has been questioned, an additional first-authored publication confirming the appropriateness of applying standardisation principles to nucleic acid, using CMV DNA testing as a model, is presented in **Chapter 3**.

Apart from standardisation, my published works have a strong focus on the monitoring of variation in infectious disease testing using quality control. Since early 2000's, NRL has collected quality control test results and associated metadata from hundreds of clinical and blood screening laboratories globally, employing a large variety of IVDs. As the senior scientist managing these endeavours, several original and significant concepts have been developed and published, creating a novel model of quality control that is more fit for purpose that traditional processes used in clinical chemistry. **Chapter 4** presents two first-authored published papers which are representative of my body of work in this area. The first paper describes a novel method of establishing acceptance criteria for quality control of infectious disease testing (QConnect™) and the second describes a comparison of QConnect™ with other, more traditional methods; demonstrating QConnect™ as being a more suitable quality control process compared with the traditional alternatives. **Chapter 5** presents a first-authored publication that demonstrates the utility of QConnect™, describing a real-life situation where an anti-HCV test kit failure was detected and subsequently the root cause identified and corrected. Using this situation, the consequences of a real-life failure detected by a quality control system were examined.

This thesis demonstrates an original and significant contribution to scientific knowledge and understanding through my prior publications, impacting on WHO guidelines, challenging traditional approaches, and implementing novel QC methodologies that are now used globally. The relevance of this body of work has been highlighted with the emergence of SARS-CoV-2 and the release of First WHO International Standard for anti-SARS-CoV-2 immunoglobulin (human) and on-going difficulties experienced by laboratories that use traditional approaches for establishing control limits and applying Westgard rules. The publications presented in this thesis establishes a better understanding of the unresolved issues faced with the standardisation and control of infectious diseases and offers validated alternative to traditional methods.

Chapter Two: Standardisation of rubella serology

In the early 1990's, my interest in standardisation of serology testing started when the results of an EQA for anti-rubella IgG reported by my laboratory were assessed as aberrant. This situation, in retrospect, was due to several reasons. At the time, the Royal Melbourne Hospital microbiology serology laboratory used the Dade Behring anti-rubella IgG (Marburg, Germany) microtiter plate EIA, whereas the majority of Australian laboratories used the Abbott AxSYM anti-rubella IgG assay (Chicago, Ill, USA), an automated immunofluorescent assay. The EQA provider used a "consensus" approach to assess results that were specifically aimed at around the cut-off of the assay. This method of analysis compared the qualitative outcome for samples at the low end of the assay sensitivity from participating laboratories using various test systems. The qualitative results for a particular sample reported by more than 80% of the participants was deemed as being the "correct" result, irrespective of the test system(s) used. However, when most participants use the same test system, an erroneous result could inappropriately be deemed the "correct" result.

Observing all laboratories' EQAS results, ranging from negative to high positive, it was clear that different test systems clearly reported different quantitative results, indicating a lack of standardisation between assays. The qualitative results reported by any single test system, however, were similar. Some samples provided in the EQA were low positive samples. For these samples, results from different test systems reported a range of qualitative results, with the test system being used in our laboratory reporting a negative result, whereas the test system used by the majority of participants reporting positive results. Until that time, no peer-reviewed papers had demonstrated the difference in the quantification of anti-rubella IgG results reported by various test systems, although several papers had reviewed differences in qualitative reporting of low positive samples, especially as compared with the "gold standard" HAI (96, 136, 137).

To investigate whether test systems calibrated with the same international standard demonstrated standardisation of results and to measure the differences between reported quantitative results, I coordinated a multicentre comparison of results reported by different commercial and in-house antirubella IgG test kits, which was undertaken and published in 1992 (73). The hypothesis was that the use of an international standard failed to standardise the results reported by different test systems. Seven local public and private laboratories, each using different anti-rubella IgG tests (five commercial IAs and one in-house IA) that were calibrated with the WHO international standard RUB-1-94, tested the same set of positive and negative samples and the quantitative results of testing were compared. The samples were also tested on HAI. The test systems were evaluated for linearity, repeatability and reproducibility. The quantitative results of 40 individual samples, reported by each test, were compared. Results of the different commercial assays had a Pearson's correlation coefficient ranging between 0.64 to 0.75 when compared with results obtained from the in-house EIA, which was also calibrated using the same standard. When the results of the commercial tests were compared against each other, the correlation ranged between 0.63 to 0.93. A correlation coefficient of 0.8 was considered to be a strong correlation, therefore the correlations between most test systems was considered moderate. The result of 0.93 was reported when comparing two tests from the same manufacturer. The conclusion of this study was "*the international units reported by the commercial kits are insufficiently consistent to be of practical use in diagnostic clinical microbiology*", therefore setting the scene for future standardisation publications.

As State, and then National Convenor of the Australian Society for Microbiology, Serology Special Interest Group, the topic of rubella IgG standardisation was often explored at national conferences. A first-author publication on the topic, not included in the thesis was also published in 1995 (138). This study evaluated a novel single point calibration method, called the alpha method, developed by Dade-Behring for use in a microtiter-plate EIA for the quantification of anti-rubella IgG. The results of testing in the commercial EIA were compared with results of an in-house EIA calibrated with the WHO international standard. A total of 40 serum samples obtained from routine antenatal testing were tested in both the commercial and in-house assays and the quantitative results compared using Pearson's coefficient of correlation. The correlation between the two assays was estimated as 0.7, indicating a moderate correlation. Over the next decade, HAI testing for anti-rubella antibodies was superseded by commercial, microtiter plate EIAs and automated platforms, almost all being calibrated with RUB-1-94 WHO international standard. By 2008, the test systems included in the 1992 publication were obsolete, mainly due to the adoption of automation and the use of different detection technologies. During the same period, reporting of anti-rubella IgG results in IU/mL was adopted universally by IVD manufacturers, professional bodies, medical laboratories and used in clinical interpretations. Even so, the results of EQA demonstrated that the lack of standardisation remained (139).

In a conference abstract, the providers of the UKNEQAS quality assessment scheme reviewed the results from 48 samples distributed between 2002 and 2006. Of these 48 samples, 10 had low levels of anti-rubella IgG. Over 3,000 test results from multiple test systems for these 10 samples were reported, and up to 6% of participants reported these as being non-immune (139). The author concluded "*Even though there has been an international standard for rubella available for many years there remains a difference in the quantitative results reported for different kits. For specimens containing low levels of antibodies this variation can result in the specimens being classified as coming from a patient who is nonimmune*." To confirm this assumption, I coordinated a new comparison of anti-rubella IgG results reported in IU/mL for eight commercial immunoassays in use at that time. The study was designed to assess if automated anti-rubella IgG assays reported different quantitative results when testing the same samples. The hypothesis being that, although technology has changed, the use of an international standard to calibrate these new assays would not result in standardisation of quantitative results reported by the tests. A first-authored study similar to that published in 1992 was undertaken and published in 2008 (93) and is the first paper of this chapter. This study also identified that the comparator qualitative assays, HAI and a western blot (WB), were suitable confirmatory assays, and were utilised in future studies.

Arising from these studies, a particular clinical interest was how to interpret a fixed, quantitative cut-off given a historical lack of standardisation of quantitative results reported by anti-rubella IgG assays at the limit of detection of test systems. Although previous publications demonstrated similar sensitivity and specificities across test kits based on qualitative results, laboratory scientists and medical microbiologists were concerned that low-level, quantitative anti-rubella IgG results may not be protective (139). To address this concern, I established a study to contribute to the interpretation of low-level quantitative anti-rubella IgG results (92). To determine the true antibody status 100 samples, found to have low-level reactivity in each of five commercial anti-rubella IgG assays reporting in IU/mL (total of 500 samples), were testing in HAI, in-house EIA and WB assays previously developed in our laboratory. The findings of this study were used to advise Australian medical laboratories establishing "grey-zones" for anti-rubella IgG testing. This study is presented in this thesis as the second paper contributed to **Chapter 2**.

JOURNAL OF CLINICAL MICROBIOLOGY, June 2008, p. 1955–1960 Vol. 46, No. 6
0095-1137/08/$08.00+0 doi:10.1128/JCM.00231-08

Evaluation of Eight Anti-Rubella Virus Immunoglobulin G Immunoassays That Report Results in International Units per Milliliter[▽]

Wayne Dimech,[1]* Lena Panagiotopoulos,[1] Barbara Francis,[1] Nicholas Laven,[2] Joan Marler,[2] David Dickeson,[3] Tony Panayotou,[4] Kim Wilson,[1] Robyn Wootten,[2] and Elizabeth M. Dax[1,5]

National Serology Reference Laboratory, Australia, St. Vincent's Institute, 4th Floor, Healy Building, 41 Victoria Parade, Fitzroy, Victoria 3065,[1] Department of Serology, Symbion Health, Dorevitch Pathology, 18 Banksia Street, Heidelberg, Victoria 3084,[2] Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, New South Wales 2145,[3] Microbiology Department, Southern Health, 246 Clayton Road, Clayton, Victoria 3168,[4] and Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria 3052,[5] Australia

Received 4 February 2008/Returned for modification 9 April 2008/Accepted 14 April 2008

An evaluation of anti-rubella virus immunoglobulin G (IgG) immunoassays that report in international units per milliliter (IU/ml) was performed to determine their analytical performance and the degree of correlation of the test results. A total of 321 samples were characterized based on results from a hemagglutination inhibition assay. The 48 negative and 273 positive samples were used to determine the sensitivity and specificity of the assays. When equivocal results were interpreted as reactive, the sensitivity of the immunoassays ranged from 98.9 to 99.9% and the specificity ranged from 77.1 to 95.8%. All assays had positive and negative delta values of less than 2. A significant difference between the mean results of all assays was demonstrated by analysis of variance. However, post hoc analysis showed there was good correlation in the mean results expressed in IU/ml between some of the assays. Our results show the level of standardization between anti-rubella virus IgG immunoassays reporting results expressed as IU/ml has improved since a previous study in 1992, but further improvement is required.

Rubella virus causes a relatively benign childhood rash and fever. However, primary maternal infection during the first trimester is associated with a 80 to 90% risk of congenital rubella syndrome (2, 3, 25). In developed countries, the risk of congenital rubella syndrome has been minimized through vaccination programs (22–24) and by testing pregnant women for evidence of rubella virus immunoglobulin G (IgG) at their first antenatal visit (10, 11). Since the isolation of rubella virus in 1962, rubella testing has developed continuously, with the hemagglutination inhibition (HAI) assay often being considered the reference method (4, 15, 29).

Since the 1980s, rubella virus IgG assays have been calibrated against the same World Health Organization (WHO) international standard rubella virus serum (second standard preparation) and test results have been reported in international units per milliliter (IU/ml). The introduction of quantitative measurement of rubella virus IgG had the potential to increase standardization and facilitate the comparison between the results of different tests.

In 1992, we published a multicenter evaluation comparing commercial immunoassays used to measure rubella virus IgG antibodies (9). The conclusion was that, although there was a moderate degree of correlation, reporting anti-rubella virus IgG levels in IU/ml had insufficient practical use. At that time, we concluded that the results of rubella virus antibody testing be confined to a statement concerning immunity rather than a numerical value. More than 15 years later, the assays compared in the 1992 study are no longer in common usage in Australia and have generally been replaced with random-access analyzers that perform a range of immunoassays of multiple disciplines. A comparison of six random-access and two microtiter plate (MTP) immunoassays that report anti-rubella virus IgG levels in IU/ml was undertaken to review analytical performance and determine whether the standardization of reporting in the newer assays had improved. While the standardization of reporting for rubella virus IgG levels is greater with the introduction of automated immunoassays, further improvement is needed.

MATERIALS AND METHODS

Samples. A total of 321 serum or plasma samples were included in the study. The samples were from 201 plasma packs obtained from Australian blood donors and 83 serum samples from individuals presenting for routine pathology tests that were prescreened by the HAI assay and found to have low levels of rubella virus IgG. Another 28 sera from individuals with serological evidence of acute rubella infection were included in the study. They included 13 individual samples and 15 samples from three seroconversion panels. Nine sera containing anti-toxoplasma IgM antibodies were also tested.

Serum or plasma samples used in the study were collected and stored at −20°C. Samples were thawed and aliquoted into single-use vials that were refrozen and stored at −20°C until they were used. Before testing, thawed aliquots

* Corresponding author. Mailing address: National Serology Reference Laboratory, Australia, 4th Floor Healy Building, 41 Victoria Parade, Fitzroy 3065, Victoria, Australia. Phone: 61 (0)3 9418 1132. Fax: 61 (0)3 9418 1155. E-mail: wayne@nrl.gov.au.

▽ Published ahead of print on 23 April 2008.

were held at 4°C for up to 3 weeks or until use and then were discarded. No sample underwent more than three freeze-thaw cycles.

Tests. All samples were tested by the HAI assay and eight commercially available immunoassays. Selected samples were tested further using an in-house Western blot assay.

(i) HAI assay. All samples were tested in an HAI assay (1, 13, 29). Briefly, samples were treated with kaolin to remove nonspecific agglutinins. A twofold serial dilution of each sample was made in phosphate-buffered saline buffer. Fresh pigeon red blood cells coated with rubella virus antigen obtained from Dade Behring (Marburg, Germany) were used as the indicator. The results were expressed as the reciprocal of the titer. Each sample was tested in duplicate, and the results were read by two independent readers. If a reading that exceeded a difference of 1 doubling dilution between duplicate tests or between readers was obtained, the sample was retested.

(ii) Immunoassays. Six of the assays were automated immunoassays using random-access instruments that could perform a range of infectious disease and biochemical assays, and two were 96-well MTP immunoassays. All tests were performed as instructed by the manufacturer. The manufacturer's cutoff was applied to determine reactivity of the sample. All assays reported the test results in IU/ml.

The random-access immunoassays were Access Rubella IgG (Beckman Coulter, CA), AxSYM Rubella IgG (Abbott Diagnostics, IL), Advia Centaur Rubella G (Bayer HealthCare, NY), Immulite 2000 Rubella Quantitative IgG (Diagnostic Products Corporation, CA), Liaison Rubella IgG (DiaSorin, Saluggia, Italy), and Vidas Rub IgG II (bioMérieux, Marcy l'Etoile, France).

Access Rubella IgG (Beckman Coulter, CA) (Access) had rubella virus membrane antigen bound to paramagnetic particles. Anti-rubella virus IgG bound to the particles was detected using alkaline phosphatase-conjugated monoclonal antibody and a chemiluminescent substrate (Lumi-Phos). The light produced was proportional to the amount of bound patient anti-rubella virus IgG, and the results were calibrated against a multipoint calibration curve standardized against the WHO second international standard preparation for anti-rubella virus serum.

AxSYM Rubella IgG (Abbott Diagnostics, IL) (AxSYM) (8, 21) used a microparticle solid phase coated with rubella virus antigen. Anti-human rubella virus IgG bound to the solid phase was detected with an anti-human IgG conjugated to alkaline phosphatase and a 4-methylumbelliferyl phosphate substrate.

Advia Centaur Rubella G (Bayer HealthCare, NY) (Centaur) (6, 7, 17) employed a sandwich immunoassay using direct chemiluminometric technology. An anti-human IgG monoclonal antibody was coupled to paramagnetic particles acting as the solid phase. Rubella virus was labeled with acridinium ester. The test sample was simultaneously incubated with the solid phase and labeled rubella virus, and the resulting antibody-antigen complex was detected through the addition of acid and base reagents.

Immulite 2000 Rubella Quantitative IgG (Diagnostic Products Corporation, CA) (Immulite) (8, 20, 30) used beads coated with inactivated rubella virus as the solid phase and alkaline phosphatase conjugated to monoclonal murine anti-human IgG as the conjugate. Chemiluminescent substrate was used to detect the antibody-antigen reaction.

Liaison Rubella IgG (DiaSorin Saluggia, Italy) (Liaison) (21) had rubella virus antigen coated onto magnetic particles as the solid phase. The secondary antibody was a mouse monoclonal antibody linked to an isoluminol derivative. To detect bound anti-rubella virus IgG, a starter reagent was added and a flash chemiluminescence reaction was induced. The resulting light signal was measured using a photomultiplier and converted to relative light units that were proportional to the amount of anti-rubella virus IgG present in the sample.

Vidas Rub IgG II (bioMérieux, Marcy l'Etoile, France) (Vidas) (21, 30) combined a two-step sandwich immunoassay method using a fluorescence detection system. The solid-phase receptacle acted as both the solid phase and pipetting device. The conjugate was an alkaline phosphatase-labeled monoclonal anti-human IgG (mouse), and the substrate was 4-methylumbelliferyl phosphate.

The two MTP immunoassays used 96-well plates coated with rubella virus antigen and a series of standards to calibrate the assay. They were ETI-RUBEK-G Plus (DiaSorin, Saluggia, Italy) and Enzygnost Anti-Rubella-Virus/IgG (Dade Behring, Marburg, Germany).

ETI-RUBEK-G Plus (DiaSorin, Saluggia, Italy) (DiaSorin) (9) was a MTP immunoassay using rubella virus coated to the MTP wells. Anti-rubella virus antibodies in the sample were bound to the rubella virus and were detected using a protein A conjugated to horseradish peroxidase tracer and a tetramethylbenzidine-hydrogen peroxide substrate, giving a color change that was proportional to the amount of anti-human IgG bound to the solid phase.

Enzygnost Anti-Rubella-Virus/IgG (Dade Behring, Marburg, Germany) (Enzygnost) was a MTP immunoassay consisting of duplicate test wells, one coated with rubella virus and the other control well coated with noninfected cell culture. The secondary antibody was a rabbit antibody conjugated with peroxidase. The antibody-antigen reaction was detected using a tetramethylbenzidine-hydrogen peroxide substrate. The optical density of the control well antigen was subtracted from that of the antigen-coated well for each sample to reduce the effect of nonspecific reactivity.

Western blot assay. Samples that were negative by the HAI assay but had an equivocal or positive test result in one or more immunoassays were tested by Western blotting if sufficient sample remained. Western blots were performed by running rubella virus lysate on a nonreducing 10% sodium dodecyl sulfate-polyacrylamide gel, transferring the proteins to nitrocellulose, and probing for antibody in plasma samples that were diluted 1 in 100 in buffer (16, 31, 32).

Analysis. (i) Sensitivity and specificity. The results of the HAI assay were used to assign a negative or positive status to the samples. An HAI titer of less than 8 was considered negative for anti-rubella virus immunoglobulin. A titer of 8 or more was considered positive (15). The analytical sensitivity and specificity of each immunoassay were estimated by comparing the immunoassay qualitative results with the sample's assigned positive or negative status. The sensitivity and specificity of the immunoassays were calculated twice, first interpreting equivocal results as negative and then as positive. The Western blot result did not change the status of any samples and did not affect the estimations of sensitivity or specificity.

(ii) Statistical analyses. The results expressed as IU/ml were used in statistical analyses. Where an assay produced a result expressed as greater than the highest limit of the assay, the result was assigned a value of the limit level plus one. For example, if the assay's highest reportable result was 300 IU/ml, results greater than 300 IU/ml were assigned a value of 301 IU/ml prior to statistical analyses. The exception was results from Centaur, because the highest reportable result changed from 500 IU/ml to 175 IU/ml midway through the evaluation. All results greater than 175 IU/ml were converted to 176 IU/ml for statistical analysis for this assay.

For all immunoassays, the mean values of the results expressed in IU/ml were compared by analysis of variance, and Tamhane's T2 post hoc test using SPSS (SPSS Inc., Chicago, IL) version 15.0. A P value of <0.05 was considered statistically significant. Positive and negative delta values were estimated for each immunoassay (5). The delta value is the distance the mean of the positive and negative populations of a data set is removed from the cutoff and is measured in standard deviations. To calculate the delta value, a cutoff of 10 IU/ml was used for each assay, with all results greater than or equal to 10 IU/ml being analyzed as positive. Results less than 10 IU/ml were analyzed as negative.

RESULTS

Of the 321 samples that were tested by the HAI assay, 48 samples had HAI results less than 8 and were considered to have a negative status for anti-rubella virus antibodies. The 273 samples with an HAI titer of 8 or greater were considered positive and consisted of 20 samples with a titer of 8, 27 samples with a titer of 16, 56 samples with a titer of 32, 63 samples with a titer of 64, 52 samples with a titer of 128, 28 samples with a titer of 256, 19 samples with a titer of 512, and 8 samples with a titer of 1,024 or greater. The sensitivity and specificity for each immunoassay were estimated twice, once interpreting equivocal results as positive and again interpreting equivocal results as negative (Table 1).

There were 48 negative samples tested in all eight immunoassays. When the immunoassay equivocal results were assigned a negative status, the specificity of the eight immunoassays ranged from 85.4 to 95.8%. Twenty-eight of the 48 negative samples were negative in all eight immunoassays. The results for the other 20 samples are presented in Table 2. A Western blot assay could not be performed on 6 of the 20 samples. Of the remaining 14 samples, 10 were negative and 4 were positive by Western blotting. All four Western blot-positive samples had a positive or equivocal test result in three or more immunoassays. One sample was positive in all eight immunoassays. Two of the six samples that were not tested by Western blotting

TABLE 1. Sensitivity and specificity of eight immunoassays testing 48 negative and 273 positive samples for anti-rubella virus IgG[a]

Assay	Sensitivity (%) (95% confidence limits [%]) with equivocal results assigned as:		Specificity (%) (95% confidence limits [%]) with equivocal results assigned as:	
	Negative	Positive	Negative	Positive
Access	96.0 (92.7–97.9)	99.3 (97.1–99.9)	95.8 (84.6–99.3)	91.7 (79.1–97.3)
AxSYM	98.2 (95.5–99.3)	99.3 (97.1–99.9)	85.4 (71.6–93.5)	77.1 (62.3–87.5)
Centaur	99.3 (97.1–99.9)	99.6 (97.7–99.9)	93.8 (81.8–98.4)	87.5 (74.1–94.8)
Enzygnost	99.6 (97.7–99.9)	99.6 (97.7–99.9)	91.7 (79.1–97.3)	85.4 (71.6–93.5)
Immulite	99.3 (97.1–99.9)	99.6 (97.7–99.9)	91.7 (79.1–97.3)	81.3 (66.9–90.6)
Liaison	98.2 (95.5–99.3)	98.9 (96.6–99.7)	95.8 (84.6–99.3)	91.7 (79.1–97.3)
DiaSorin	98.9 (96.6–99.7)	99.9 (98.3–99.9)	87.5 (74.1–94.8)	83.3 (69.2–92.0)
Vidas	96.7 (93.6–98.4)	99.6 (97.7–99.9)	95.8 (84.6–99.3)	95.8 (84.6–99.3)

[a] Positive and negative status was defined by HAI test results.

had six or more positive or equivocal immunoassay test results. There were 14 samples that were reactive by either one or two immunoassay test results. Access, Centaur, Liaison, and Vidas gave negative results for all of these 14 samples, whereas Enzygnost, Immulite, DiaSorin, and AxSYM gave a positive or equivocal result for 1, 4, 5, and 7 of these samples, respectively.

Of the 273 samples assigned a positive status, 255 were positive by all immunoassays. The other 18 positive samples were negative or equivocal in one or more immunoassays. When equivocal immunoassay results were considered positive, the assay's sensitivities ranged from 98.9 to 99.9%. Five samples were reactive in one immunoassay only, eight samples were reactive in two immunoassays, three samples were reactive in three immunoassays, and two samples were reactive in four immunoassays (Table 3). Of the 18 samples, six samples had an HAI titer of 8 or 16, four samples had a titer of 32, and one sample each had a titer of 128 and 256. No sample assigned a positive status was negative or equivocal in more than four immunoassays. No assay reported a negative result on more than 3 of the 18 samples.

The mean of the results expressed in IU/ml of all eight immunoassays were significantly different ($F = 8.375$ with 7 df [$P < 0.001$]) (Fig. 1). Post hoc tests showed that the results of the DiaSorin assay were significantly different from the results from the other assays. The Access and AxSYM results demonstrated no significant difference compared to the results of all other assays, except DiaSorin. A significant difference between results of various combinations of the other assays was demonstrated (Table 4).

The negative and positive delta values were calculated for each immunoassay (Table 5). All assays had a delta value below 2, indicating that the mean of the negative and positive sample populations was less than 2 standard deviations removed from the cutoff of 10 IU/ml. DiaSorin had a negative delta value of less than 1, and AxSYM had a positive delta value of less than 1. The smaller the delta value, the greater the potential of false-positive or negative test results. However, the delta values may have been skewed because a number of test results were reported as greater or less than the limits of reporting for each assay.

DISCUSSION

Testing samples for the presence of anti-rubella virus IgG to determine immune status is performed routinely, especially on women presenting for their first antenatal visit. For the past 15 years, most commercial immunoassays have been calibrated against a WHO international standard, and their results were reported in IU/ml, with 10 IU/ml generally considered the cutoff between immune and nonimmune status (14, 15, 26). In 1992, a multicenter evaluation of immunoassays reporting anti-rubella virus IgG results in IU/ml determined that only a moderate correlation between the results reported by each assay existed (9). Since this time, there have been few published investigations into anti-rubella virus IgG testing, even though many new assays have become available. The present study was undertaken to determine the performance characteristics of some new assays and to establish whether correlation between results reported in IU/ml had improved.

The status of all samples was assigned by HAI testing. Although HAI testing is often considered to be the reference method for rubella virus antibody detection, it must be noted that, unlike rubella virus IgG-specific immunoassays, HAI testing detects both IgG and IgM (13, 15). There were 28 samples from individuals known to have acute rubella infection included in the study, and all these samples were reactive by both the HAI test and immunoassays. Therefore, any discrepancy between the HAI test and immunoassays was not due to the detection of rubella virus IgM.

When a cutoff of 10 IU/ml was used, all assays included in the study had comparable sensitivity and specificity, with overlapping 95% confidence intervals. All assays reported some false-negative and false-positive test results. It is more acceptable to report a false-negative anti-rubella virus IgG test result than a false-positive result. Clinically, a false-negative result may give rise to unnecessary vaccination or, at worst, anxiety for a pregnant woman who has had contact with rubella. A false-positive result may lead to a susceptible person not being vaccinated and result in an infection if she is subsequently exposed to the virus. If a woman is in the first trimester of pregnancy, a congenitally acquired rubella infection may ensue.

When the equivocal test results were considered reactive, all assays had a sensitivity of 98.9% or greater, offering confidence in their ability to detect the presence of anti-rubella virus IgG. A relatively small percentage of false-negative results would be reported using any of the assays evaluated. However, the specificity of the assays would result in a higher percentage of false-positive results, with up to 22% false-positive results with AxSYM. The AxSYM specificity reported by Diepersloot et al.

TABLE 2. Qualitative and quantitative anti-rubella virus IgG test results and corresponding Western blot test results of 20 samples[a]

Sample no.	Western blot result[b]	Test result[c] for immunoassay															
		Access		AxSYM		Centaur		Enzygnost		Immulite		Liaison		DiaSorin		Vidas	
		Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)
352494	NT	Negative	7.9	Negative	3.1	Negative	2.6	Equivocal	6	Equivocal	9.1	Negative	<5.0	Negative	8.8	Negative	7
352516	NT	Negative	2.3	Negative	0.2	Negative	0.2	Negative	<4	Equivocal	6.5	Negative	<5.0	Negative	5.9	Negative	0
352514	NT	Negative	1.5	Equivocal	5.3	Negative	0.2	Negative	<4	Negative	<5.0	Negative	<5.0	Negative	2.3	Negative	1
351150	NT	Negative	6.7	Negative	4.9	Negative	0.0	Negative	<4	Positive	10.8	Negative	<5.0	Negative	3.5	Negative	4
351157	NT	Equivocal	11.6	Positive	16.8	Equivocal	8.2	Positive	17	Positive	13.0	Equivocal	9.6	Negative	8.4	Negative	9
352505	NT	Positive	18.6	Positive	14.6	Positive	41.6	Positive	25	Positive	20.3	Negative	5.4	Positive	25.9	Positive	24
352492	Negative	Negative	1.7	Negative	0.0	Negative	0.2	Negative	<4	Equivocal	5.2	Negative	<5.0	Equivocal	10.4	Negative	1
331811	Negative	Negative	0.6	Equivocal	7.1	Negative	0.3	Negative	<4	Negative	<5.0	Negative	<5.0	Equivocal	9.2	Negative	1
331930	Negative	Negative	1.1	Equivocal	8.7	Negative	2.6	Negative	<4	Negative	<5.0	Negative	<5.0	Negative	1.9	Negative	1
331443	Negative	Negative	1.4	Positive	23.5	Negative	0.3	Negative	<4	Negative	<5.0	Negative	<5.0	Negative	0.6	Negative	1
331442	Negative	Negative	0.9	Positive	14.8	Negative	0.9	Negative	<4	Negative	<5.0	Negative	<5.0	Negative	1.4	Negative	1
331479	Negative	Negative	1.1	Positive	15.6	Negative	0.6	Negative	<4	Negative	<5.0	Negative	<5.0	Negative	3.1	Negative	4
329934	Negative	Negative	1.1	Positive	28.1	Negative	0.7	Negative	<4	Negative	<5.0	Negative	<5.0	Negative	3.5	Negative	1
331590	Negative	Negative	1.4	Negative	1.5	Negative	0.3	Negative	<4	Negative	<5.0	Negative	<5.0	Positive	20.8	Negative	0
349450	Negative	Negative	1.3	Negative	0.6	Negative	0.0	Negative	<4	Negative	<5.0	Negative	<5.0	Positive	47.9	Negative	0
331812	Negative	Negative	1.1	Negative	0.7	Negative	0.4	Negative	<4	Negative	<5.0	Negative	<5.0	Positive	108.7	Negative	0
349703	Positive	Negative	7.8	Negative	4.0	Equivocal	7.6	Equivocal	6	Negative	<5.0	Equivocal	9.1	Negative	5.9	Negative	5
349418	Positive	Equivocal	10.9	Negative	2.9	Positive	17.9	Equivocal	4	Equivocal	8.7	Positive	25.8	Positive	16.5	Negative	7
331441	Positive	Negative	7.1	Equivocal	5.3	Equivocal	5.1	Positive	8	Equivocal	7.8	Negative	<5.0	Negative	5.7	Negative	5
348611	Positive	Positive	25.6	Positive	28.0	Positive	81.7	Positive	31	Positive	31.3	Positive	45.9	Positive	35.8	Positive	35

[a] The 20 samples were negative in an HAI assay that had at least one reactive test result when tested in eight anti-rubella virus IgG immunoassays as part of an evaluation of the analytical sensitivity of those assays.
[b] NT, not tested.
[c] Qualitative (Qual.) and quantitative (Quant.) test results are given for each immunoassay.

TABLE 3. Results of 18 HAI positive samples that tested negative or equivocal in one or more of eight rubella IgG immunoassays

Sample no.	HAI test result	Result[a] of immunoassay															
		Access		AxSYM		Centaur		Enzygnost		Immulite		Liaison		DiaSorin		Vidas	
		Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)	Qual.	Quant. (IU/ml)
352519	8	Equivocal	13.8	Equivocal	9.3	Equivocal	8.8	Positive	10	Positive	15.0	Positive	11.1	Positive	11.4	Equivocal	12
352498	16	Equivocal	11.2	Equivocal	6.8	Positive	15.0	Positive	7	Positive	10.9	Equivocal	9.8	Positive	12.5	Equivocal	11
349091	8	Positive	29.9	Negative	3.6	Negative	0.0	Positive	79	Positive	23.1	Positive	61.0	Positive	39.9	Negative	2
351148	16	Positive	15.1	Positive	16.4	Positive	16.3	Positive	11	Positive	10.7	Negative	8.1	Equivocal	10.8	Equivocal	12
331447	32	Equivocal	13.6	Positive	23.7	Positive	16.8	Positive	11	Positive	13.6	Equivocal	10.1	Equivocal	10.6	Positive	16
348632	8	Equivocal	13.6	Positive	16.7	Positive	39.4	Positive	16	Positive	17.5	Positive	21.1	Positive	20.9	Positive	16
351149	32	Positive	19.9	Positive	65.4	Positive	24.5	Positive	27	Positive	20.3	Negative	6.3	Positive	19.2	Positive	17
349624	256	Positive	148	Positive	98.4	Positive	333.8	Negative	<4	Positive	131.0	Positive	178.0	Positive	74.9	Positive	184
351164	32	Equivocal	14.7	Positive	35.8	Positive	37.5	Positive	23	Positive	74.4	Positive	15.2	Positive	19.7	Positive	20
349456	128	Negative	9.1	Positive	202.3	Positive	>175.0	Positive	80	Positive	261.0	Positive	267.0	Positive	114.2	Positive	279
351176	16	Positive	18.7	Positive	30.4	Positive	29.7	Positive	25	Positive	14.1	Negative	<5.0	Positive	20.1	Equivocal	13
351151	16	Equivocal	14.8	Positive	18.4	Positive	13.0	Positive	20	Positive	14.6	Positive	13.0	Positive	11.7	Equivocal	13
351107	32	Equivocal	12.5	Positive	13.1	Positive	16.6	Positive	25	Equivocal	6.9	Positive	41.2	Positive	20.9	Positive	31
351159	16	Positive	15.7	Positive	13.7	Positive	17.4	Positive	15	Positive	13.8	Positive	13.4	Equivocal	10.9	Equivocal	14
351153	16	Equivocal	11.4	Positive	15.4	Positive	143.1	Positive	11	Negative	<5.0	Positive	14.1	Positive	11.9	Positive	19
348839	8	Positive	19.5	Negative	0.3	Positive	18.5	Positive	22	Positive	12.9	Positive	11.6	Positive	13.6	Equivocal	14
348566	8	Negative	9.4	Equivocal	6.9	Positive	11.4	Positive	8	Positive	13.1	Positive	12.6	Positive	14.7	Positive	19
349320	8	Equivocal	14.7	Positive	12.7	Positive	29.1	Positive	12	Positive	16.3	Positive	30.3	Positive	21.2	Equivocal	12

[a] Qualitative (Qual.) and quantitative (Quant.) test results are given for each immunoassay.

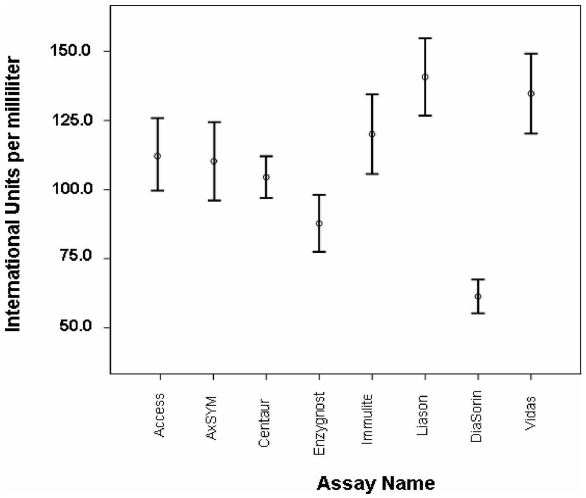


FIG. 1. Mean and 95% confidence intervals of results reported for 321 samples tested with eight immunoassays for anti-rubella virus IgG tests giving results in international units per milliliter.

TABLE 5. Estimated delta values of 48 negative samples and 273 positive samples tested for the quantification of anti-rubella virus IgG in eight immunoassays

Immunoassay	Estimated delta value[a]	
	Negative	Positive
Access	−1.64	1.68
AxSYM	−1.08	0.86
Centaur	−1.83	1.71
Enzygnost	−1.74	1.54
Immulite	−1.47	1.27
Liaison	−1.46	1.30
DiaSorin	−0.62	1.79
Vidas	−1.29	1.75

[a] Estimated delta values expressed as standard deviations.

was 81.5% (8). In Australia, many laboratories use the manufacturer's cutoff to determine nonimmune status but apply a "gray zone" to express doubt over the levels of immunity conferred by low levels of reactivity. The "gray zone" used differs widely between laboratories.

Of the 48 samples with an HAI result of <8, four samples were subsequently found to be Western blot positive. Two samples had insufficient volume for Western blotting but were reactive in at least six of the eight immunoassays. It has been noted previously that using methods other than HAI testing, immune individuals with specific but low levels of rubella virus antibodies have been identified (12, 15, 18, 27, 28). Of the remaining 42 samples, Access, Centaur, Liaison, and Vidas reported no false-positive results; Enzygnost had one equivocal result (6 IU/ml). Immulite gave one positive test result (10.8 IU/ml) and three equivocal test results for the 42 samples. The application of a "gray zone" to these assays may not be necessary. AxSYM gave four positive and three equivocal results, the highest results being 28 IU/ml. This confirms previous findings which indicated about 1% of AxSYM positive results could not be confirmed (19). A "gray zone" of 30 IU/ml and a comment indicating doubt in immune status with results between 10 and 30 IU/ml may be considered for this assay. DiaSorin gave three positive test results (20.8, 47.9, and 108.7 IU/ml) and two equivocal test results. The spread of results overlapped considerably with the results reported for positive samples. An application of a "gray zone" to this assay would be impractical.

Statistical comparisons of the results reported by all eight immunoassays suggested that several assays gave comparable results. Results from the automated immunoassays Access, AxSYM, Centaur, and Immulite assays compared well, as did the Immulite, Liaison, and Vidas assays. Results from the MTP immunoassay DiaSorin (and, to a lesser extent Enzygost) did not compare well with any other assay. These results indicate that standardization of some anti-rubella virus IgG assays that report in IU/ml has occurred, but greater standardization throughout all immunoassays is required.

Positive and negative delta values were calculated for all immunoassays. The delta value describes the distance the mean of the positive and negative populations of a data set is removed from the cutoff and is measured in standard deviations. Therefore, an assay with a positive delta value of 4.0 has a mean of the results of positive samples 4 standard deviations from the cutoff. All immunoassays had a delta value of less than 2, implying poor separation of the negative and positive population, potentially leading to the false-negative and false-positive test results. Assays undergo variation from test event to test event, arising from changes of reagent batches, variation in the volume of reagents pipetted, temperature, length of incubation, and other process changes. A low delta value in-

TABLE 4. Tamhane's post hoc analysis of the means of anti-rubella virus IgG results[a]

Immunoassay	Significance (P value)[b] of the difference of the means of immunoassays							
	Access	AxSYM	Centaur	Enzygnost	Immulite	Liaison	DiaSorin	Vidas
Access								
AxSYM	1.000							
Centaur	1.000	1.000						
Enzygnost	0.129	0.351	0.302					
Immulite	1.000	1.000	0.892	**0.013**				
Liaison	0.127	0.092	**0.000**	**0.000**	0.797			
DiaSorin	**0.000**	**0.000**	**0.000**	**0.001**	**0.000**	**0.000**		
Vidas	0.591	0.469	**0.010**	**0.000**	0.998	1.000	**0.000**	

[a] Anti-rubella virus IgG results were expressed as international units per milliliter from 321 samples tested on eight immunoassays.
[b] Significance is defined as $P < 0.05$, and significant P values are indicated by bold font.

dicates an increased possibility of these variations in the test system affecting the sensitivity or specificity of the assay.

Immunoassays used for the quantification of rubella virus IgG are standardized to the WHO international standard rubella virus serum (second standard preparation) and report results in IU/ml. A report expressed in IU/ml implies traceability from one assay to another, much in the manner of many biochemistry assays. This and previous investigations indicate that the assumption of transferability of IU/ml is incorrect. Therefore, greater standardization of assays reporting rubella virus IgG in IU/ml is required.

ACKNOWLEDGMENTS

We acknowledge the provision of assays at reduced or no charge by manufacturers and distributors. This study was funded by the Australian Government, Department of Health and Ageing, Diagnostics and Technology Branch, as part of "Quality Assurance Systems for Laboratories: Generalizing Their Use for Improving Pathology Services in Australia."

Carlie Di Camilo and Sau-Wan Chan provided technical assistance.

REFERENCES

1. **Anonymous.** 1978. Haemagglutination-inhibition test for the detection of rubella antibody. J. Hyg. (London) **81:**373–382.
2. **Best, J. M.** 2007. Rubella. Semin. Fetal Neonatal Med. **12:**182–192.
3. **Best, J. M., S. O'Shea, G. Tipples, N. Davies, S. M. Al-Khusaiby, A. Krause, L. M. Hesketh, L. Jin, and G. Enders.** 2002. Interpretation of rubella serology in pregnancy—pitfalls and problems. BMJ **325:**147–148.
4. **Cradock-Watson, J. E.** 1991. Laboratory diagnosis of rubella: past, present and future. Epidemiol. Infect. **107:**1–15.
5. **Crofts, N., W. Maskill, and I. D. Gust.** 1988. Evaluation of enzyme-linked immunosorbent assays: a method of data analysis. J. Virol. Methods **22:**51–59.
6. **Dati, F.** 2004. An update on laboratory productivity with infectious disease assays on the Bayer ADVIA Centaur immunoassay system. Clin. Lab. **50:** 53–61.
7. **Dati, F., G. Denoyel, and J. van Helden.** 2004. European performance evaluations of the ADVIA Centaur infectious disease assays: requirements for performance evaluation according to the European directive on in vitro diagnostics. J. Clin. Virol. **30**(Suppl. 1)**:**S6–S10.
8. **Diepersloot, R. J., H. Dunnewold-Hoekstra, J. Kruit-Den Hollander, and F. Vlaspolder.** 2001. Antenatal screening for hepatitis B and antibodies to *Toxoplasma gondii* and rubella virus: evaluation of two commercial immunoassay systems. Clin. Diagn. Lab. Immunol. **8:**785–787.
9. **Dimech, W., A. Bettoli, D. Eckert, B. Francis, J. Hamblin, T. Kerr, C. Ryan, and I. Skurrie.** 1992. Multicenter evaluation of five commercial rubella virus immunoglobulin G kits which report in international units per milliliter. J. Clin. Microbiol. **30:**633–641.
10. **Francis, B. H., L. I. Hatherley, J. E. Walstab, and L. I. Taft.** 1982. Rubella screening and vaccination programme at a Melbourne maternity hospital. A five-year review. Med. J. Aust. **1:**502–504.
11. **Francis, B. H., A. K. Thomas, and C. A. McCarty.** 2003. The impact of rubella immunization on the serological status of women of childbearing age: a retrospective longitudinal study in Melbourne, Australia. Am. J. Public Health **93:**1274–1276.
12. **Kleeman, K. T., D. J. Kiefer, and S. P. Halbert.** 1983. Rubella antibodies detected by several commercial immunoassays in hemagglutination inhibition-negative sera. J. Clin. Microbiol. **18:**1131–1137.
13. **Maroto Vela, M. C., M. C. Bernal Zamora, A. Levya Garcia, and G. Piedrola.** 1986. Detection of specific IgG and IgM antibodies in the haemagglutination inhibition test and the enzyme-linked immunoassay for the diagnosis of rubella infection. Infection **14:**159–162.
14. **Mitchell, L. A., M. K. Ho, J. E. Rogers, A. J. Tingle, R. G. Marusyk, J. M. Weber, P. Duclos, M. L. Tepper, M. Lacroix, and M. Zrein.** 1996. Rubella reimmunization: comparative analysis of the immunoglobulin G response to rubella virus vaccine in previously seronegative and seropositive individuals. J. Clin. Microbiol. **34:**2210–2218.
15. **National Committee for Clinical Laboratory Standards.** 1997. Detection and quantitation of rubella IgG antibody: evaluation and performance criteria for multiple component test products, specimen handling, and use of test products in the clinical laboratory; approved guideline. NCCLS guideline I/LA6-A, vol. 17, p.25. National Committee for Clinical Laboratory Standards, Villanova, PA.
16. **Nedeljkovic, J., T. Jovanovic, S. Mladjenovic, K. Hedman, N. Peitsaro, and C. Oker-Blom.** 1999. Immunoblot analysis of natural and vaccine-induced IgG responses to rubella virus proteins expressed in insect cells. J. Clin. Virol. **14:**119–131.
17. **Okrongly, D.** 2004. The ADVIA Centaur immunoassay system—designed for infectious disease testing. J. Clin. Virol. **30**(Suppl. 1)**:**S19–S22.
18. **O'Shea, S., J. M. Best, and J. E. Banatvala.** 1983. Viremia, virus excretion, and antibody responses after challenge in volunteers with low levels of antibody to rubella virus. J. Infect. Dis. **148:**639–647.
19. **O'Shea, S., H. Dunn, S. Palmer, J. E. Banatvala, and J. M. Best.** 1999. Automated rubella antibody screening: a cautionary tale. J. Med. Microbiol. **48:**1047.
20. **Owen, W. E., T. B. Martins, C. M. Litwin, and W. L. Roberts.** 2006. Performance characteristics of six IMMULITE 2000 TORCH assays. Am. J. Clin. Pathol. **126:**900–905.
21. **Petersen, E., M. V. Borobio, E. Guy, O. Liesenfeld, V. Meroni, A. Naessens, E. Spranzi, and P. Thulliez.** 2005. European multicenter study of the LIAISON automated diagnostic system for determination of *Toxoplasma gondii*-specific immunoglobulin G (IgG) and IgM and the IgG avidity index. J. Clin. Microbiol. **43:**1570–1574.
22. **Plotkin, S. A.** 2001. Rubella eradication. Vaccine **19:**3311–3319.
23. **Robertson, S. E., S. L. Cochi, G. A. Bunn, D. L. Morse, and S. R. Preblud.** 1987. Preventing rubella: assessing missed opportunities for immunization. Am. J. Public Health **77:**1347–1349.
24. **Robertson, S. E., F. T. Cutts, R. Samuel, and J. L. Diaz-Ortega.** 1997. Control of rubella and congenital rubella syndrome (CRS) in developing countries. Part 2. Vaccination against rubella. Bull. W. H. O. **75:**69–80.
25. **Robertson, S. E., D. A. Featherstone, M. Gacic-Dobo, and B. S. Hersh.** 2003. Rubella and congenital rubella syndrome: global update. Rev. Panam Salud Publica **14:**306–315.
26. **Skendzel, L. P.** 1996. Rubella immunity. Defining the level of protective antibody. Am. J. Clin. Pathol. **106:**170–174.
27. **Steece, R. S., M. S. Talley, M. R. Skeels, and G. A. Lanier.** 1985. Comparison of enzyme-linked immunosorbent assay, hemagglutination inhibition, and passive latex agglutination for determination of rubella immune status. J. Clin. Microbiol. **21:**140–142.
28. **Steece, R. S., M. S. Talley, M. R. Skeels, and G. A. Lanier.** 1984. Problems in determining immune status in borderline specimens in an enzyme immunoassay for rubella immunoglobulin G antibody. J. Clin. Microbiol. **19:**923–925.
29. **van Weemen, B., and J. Kacaki.** 1976. A modified haemagglutination inhibition test for rubella antibodies, using standardized, freeze-dried reagents. Report of a comparative multi-centre trial. J. Hyg. (London) **77:**31–42.
30. **Vlaspolder, F., P. Singer, A. Smit, and R. J. Diepersloot.** 2001. Comparison of Immulite with Vidas for detection of infection in a low-prevalence population of pregnant women in The Netherlands. Clin. Diagn. Lab. Immunol. **8:**552–555.
31. **Wilson, K. M., C. Di Camillo, L. Doughty, and E. M. Dax.** 2006. Humoral immune response to primary rubella virus infection. Clin. Vaccine Immunol. **13:**380–386.
32. **Zhang, T., C. A. Mauracher, L. A. Mitchell, and A. J. Tingle.** 1992. Detection of rubella virus-specific immunoglobulin G (IgG), IgM, and IgA antibodies by immunoblot assays. J. Clin. Microbiol. **30:**824–830.

CVI
Journals.ASM.org

Investigation into Low-Level Anti-Rubella Virus IgG Results Reported by Commercial Immunoassays

Wayne Dimech, Nilukshi Arachchi, Jingjing Cai, Terri Sahin, Kim Wilson

NRL, Melbourne Australia, Victoria, Australia

Since the 1980s, commercial anti-rubella virus IgG assays have been calibrated against a WHO International Standard and results have been reported in international units per milliliter (IU/ml). Laboratories testing routine patients' samples collected 100 samples that gave anti-rubella virus IgG results of 40 IU/ml or less from each of five different commercial immunoassays (CIA). The total of 500 quantitative results obtained from 100 samples from each CIA were compared with results obtained from an in-house enzyme immunoassay (IH-EIA) calibrated using the WHO standard. All 500 samples were screened using a hemagglutination inhibition assay (HAI). Any sample having an HAI titer of 1:8 or less was assigned a negative anti-rubella virus antibody status. If the HAI titer was greater than 1:8, the sample was tested in an immunoblot (IB) assay. If the IB result was negative, the sample was assigned a negative anti-rubella virus IgG status; otherwise, the sample was assigned a positive status. Concordance between the CIA qualitative results and the assigned negative status ranged from 50.0 to 93.8% and 74.5 to 97.8% for the assigned positive status. Using a receiver operating characteristic analysis with the cutoff set at 10 IU/ml, the estimated sensitivity and specificity ranged from 70.2 to 91.2% and 65.9 to 100%, respectively. There was poor correlation between the quantitative CIA results and those obtained by the IH-EIA, with the coefficient of determination (R^2) ranging from 0.002 to 0.413. Although CIAs have been calibrated with the same international standard for more than 2 decades, the level of standardization continues to be poor. It may be time for the scientific community to reevaluate the relevance of quantification of anti-rubella virus IgG.

Infection with the rubella virus usually results in a mild childhood illness. However, infection during the first trimester of pregnancy can result in the neonate developing congenital rubella syndrome (1). For this reason, rubella vaccination programs have been established (2–5). In Australia, most diagnostic testing for rubella immunity is performed as part of an antenatal screen to ensure that the mother has protective levels of antibody. A hemagglutination inhibition titer greater than or equal to 1:16 and/or an antibody concentration greater than 10 or 15 IU/ml, depending upon the assay, is considered protective (6, 7). Some laboratories choose to report a "gray zone" to indicate uncertainty in the degree of protection conferred by low anti-rubella virus IgG levels. In Australia, the most frequently used gray-zone range is 10 to 30 IU/ml.

Since the 1980s, commercial assays used for the quantification of anti-rubella virus IgG have been calibrated against the World Health Organization (WHO) international standard rubella virus serum and test results have been reported in international units per milliliter (IU/ml) (8, 9). In theory, the calibration of assays should lead to standardization of quantitative results (10). However, several reports have indicated that quantitative anti-rubella virus IgG results reported by different assays are not always comparable (8, 9, 11). Consequently, individuals presenting to laboratories using different assays may be given different clinical interpretations, especially if their anti-rubella virus IgG levels are low.

Recently, several new commercial anti-rubella virus IgG assays have become available. New assays require validation prior to introduction into routine use (12). A common approach to validation used by laboratories is the comparison of results obtained from the new assay with those obtained from the assay routinely used by the laboratory (11, 13, 14). However, if a difference in qualitative or quantitative results obtained with the two assays occurs, it is difficult to elucidate which assay's result is correct (11).

In this study, the results of a hemagglutination inhibition assay (HAI) (15, 16) and a rubella immunoblot (IB) assay (17–20) were used to assign an anti-rubella virus IgG status. Samples with test results less than or equal to 40 IU/ml were collected from collaborating laboratories using each of five different commercial anti-rubella virus IgG immunoassays (CIAs). The 500 samples were tested by HAI and IB assays, and the qualitative results from the CIAs were compared with the assigned anti-rubella virus IgG status. To investigate the accuracy of quantification of anti-rubella virus IgG levels by the CIAs, all samples were tested in an in-house enzyme immunoassay (IH-EIA) developed to detect and quantify low levels of anti-rubella virus IgG.

MATERIALS AND METHODS

Samples. For each of five CIAs used routinely in collaborating laboratories, 100 routine clinical samples giving an anti-rubella virus IgG test result of 40 IU/ml or less, totaling 500 individual samples, were collected. After initial CIA testing in the collaborating laboratories, the samples were transported at ambient temperature to a central laboratory where they were stored at −20°C until further testing was performed. Prior to testing, the samples were thawed and tested in a testing strategy as described below. All testing, including repeat testing, was performed within 1 week of thawing. After testing, the remaining volume was divided into aliquots

Received 21 October 2012 Returned for modification 13 November 2012
Accepted 2 December 2012
Published ahead of print 19 December 2012
Address correspondence to Wayne Dimech, wayne@nrl.gov.au.

doi:10.1128/CVI.00603-12

TABLE 1 Summary of the assay characteristics of five commercial immunoassays

Assay characteristic	Description or value for indicated commercial immunoassay				
	Abbott Architect	Abbott AxSYM	Roche Elecsys	bioMérieux Vidas	Ortho Vitros
Solid phase	Microparticles	Microparticles	Magnetic beads	Solid-phase receptacles	Plastic wells
Antigen	Partially purified rubella virus	Partially purified rubella virus (strain HPV77)	Rubella-like particles and recombinant E1 antigen	Rubella antigen (strain MR 383)	UV-treated rubella antigen from cell culture
Detection system	Chemiluminescence	MUB[a]	Chemiluminescence	MUB	Luminescence
No. of calibrators	6	6	2	1	4-parameter logistic curve
Calibration range (IU/ml[b])	0–500	0–500	0.17–500	0–250	0–350
Standard	RUB-1-94[c]	WHO standard (not specified)	RUB-1-94	RUB-1-94	RUB-1-94
Negative range (IU/ml)	<4.9	<5.0	<10.0	<5.0	<9.99
Equivocal range (IU/ml)	5.0–9.9	5.0–9.9	NA[d]	5.0–10.0	NA
Low positive (IU/ml)	NA	NA	NA	NA	10.0–14.9
Positive range (IU/ml)	≥10.0	>10.0	≥10.0	≥10.0	≥15.0

[a] MUB, methylumbelliferyl.
[b] IU/ml, international units per milliliter.
[c] RUB-1-94, WHO standard (1st International Standard).
[d] NA, not applicable.

in 1-ml vials for future use. No sample underwent more than three freeze-thaw cycles.

Tests. Samples were tested using the manufacturer's instructions in one of five CIAs: Architect rubella IgG (Abbott Diagnostics, Abbott Park, IL), AxSYM rubella IgG assay (Abbott Diagnostics, Abbott Park, IL), Vidas Rub IgG II (bioMérieux, Durham, NC), Vitros rubella IgG (Ortho Clinical Diagnostics, Buckinghamshire, United Kingdom), and Elecsys rubella IgG (Roche Diagnostics, Mannheim, Germany). A description of the characteristics of each CIA is presented in Table 1. To determine the anti-rubella virus IgG status, all 500 samples were tested in a commercial HAI and IB assay. To assess the accuracy of quantification, the samples were tested in the IH-EIA.

(i) HAI assay. All 500 samples were tested in the Siemens RubeHIT (Siemens Health Care, Marburg, Germany). Briefly, nonspecific agglutinins were removed by incubating samples and controls in a kaolin suspension (250 g per liter) for 20 min at room temperature. The samples were centrifuged at 3,000 × *g* for 10 min, and the supernatant was retained. Doubling dilutions of the supernatant were made in a $CaCl_2$-bovine albumin-NaCl-$MgSO_4$ (CANM) saline solution, provided with the kit. A standard concentration of rubella antigen, at 4 to 8 hemagglutinating units per 25 μl, was added to each dilution, and the reaction mixture was incubated for 1 h at room temperature. A 4% (vol/vol) suspension of human erythrocytes was added to each well, and the reaction mixture was incubated overnight at room temperature. The highest dilution in which the hemagglutination was inhibited was considered the endpoint titer. All test results were scored by each of two different individuals without reference to the results determined by the other. Any discrepant results were reviewed by both readers, and a consensus result was obtained. Samples with HAI results of 1:16 or greater were deemed to be HAI positive. To ensure that any lack of reactivity was not due to technical error, tests of samples with a HAI titer of 1:8 or less were repeated in duplicate. Samples with HAI results repeatedly less than or equal to 1:8 were considered HAI negative.

With each HAI test run, negative, low-positive, and high-positive controls, provided by the manufacturer, were tested. Negative, low-positive, and high-positive external controls with known reactivity were also tested in each test run. Each control was required to produce an HAI titer within 1 doubling dilution of its target result for the run to be considered valid. No runs were invalid.

(ii) Immunoblot analysis. All HAI-positive samples were tested in the recombBlot rubella IgG (Mikrogen Diagnostik, Neuried, Germany) according to the manufacturer's instructions. The nitrocellulose strips provided by the manufacturer contained recombinant rubella antigens which were separated by SDS-PAGE and transferred to the nitrocellulose membrane. Briefly, 20 μl of sample or control was diluted in 2 ml of Tris buffer and incubated with the nitrocellulose strip containing rubella antigen overnight at room temperature with gentle shaking. After incubation, the strips were washed and a rabbit anti-human IgG-horseradish peroxidase conjugate was added. The strips were incubated for 1 h at room temperature and washed. Tetramethylbenzidine substrate was then added, forming a color reaction where anti-rubella virus IgG present in the sample was bound to the rubella antigens (E1, E2, c, and an E1/E2 complex) on the strip. An E2 weak-positive control, provided by the manufacturer, was tested with each set of 20 samples. The intensity of the color reaction of the E2 control acted as an assay cutoff, with any band being considered positive if its intensity was greater than that of the intensity of E2 control band. All results were scored independently by two different individuals, and any discrepant readings were resolved by consensus. Any sample having one or more reactive bands was deemed IB positive. To ensure that any lack of reactivity was not due to technical error, all samples with a negative IB result were retested in a single assay.

(iii) In-house EIA. An IH-EIA was developed by NRL. Microtiter flat-bottom plates (Nunc, Roskilde, Denmark) were coated with 50 μl of rubella virus antigen (HPV-77 strain) (MyBioSource, San Diego, CA) at a concentration of 0.5 μg/well in a carbonate buffer (pH 9.6) and incubated overnight at 37°C. After washing in PBS-T (phosphate-buffered saline [pH 7.4] containing 0.01% [vol/vol] Tween 20), 150 μl of blotto (50 mM Tris-HCl [pH 8.0], containing 5% skim milk powder, 2 mM $CaCl_2$, 150 mM NaCl, and 0.2% Nonidet P-40) was added, and the reaction mixture was incubated for 1.5 h at 37°C to block nonspecific binding. Plates were washed three times with PBS-T, and then 10 μl of control or sample was diluted in 90 μl of blotto and added to each well. A plate shaker was used to ensure adequate mixing before incubation at 37°C for 1 h. After being washed with PBS-T, 100 μl of mouse anti-human IgG conjugated to horseradish peroxidase (Southern Biotechnology Associates, Birmingham, AL) diluted 1:1,000 in blotto was added to each well, and the reaction mixture was incubated for 1 h at 37°C. After washing, 100 μl of substrate, 2 mM ABTS (2′-azinobis [3-ethylbenzthiazoline-6-sulfonic acid]), in 25 mM sodium citrate buffer (pH 4.5) containing 0.3% hydrogen peroxide was added to each well, and the reaction mixture was incubated in the dark at room temperature for 20 min. The reaction was stopped with the ad-

TABLE 2 Qualitative test results obtained from five commercial immunoassays compared with an anti-rubella virus IgG status[a]

Assay	Samples with negative rubella IgG status						Samples with positive rubella IgG status					
	Total no. of samples	No. of samples with indicated assay result			Concordance with status (%) (95% CI)		Total no. of samples	No. of samples with indicated assay result			Concordance with status (%) (95% CI)	
		Negative	Equivocal	Positive	Equivocal result considered negative	Equivocal result considered positive		Negative	Equivocal	Positive	Equivocal result considered negative	Equivocal result considered positive
Abbott Architect	23	13	5	5	69.2 (48.1–84.9)	50.0 (30.3–69.6)	77	1	7	69	89.6 (80.0–95.1)	90.9 (81.6–96.0)
Abbott AxSYM	9	6	2	1	NA[b]	NA	91	2	5	84	92.3 (84.3–96.6)	97.8 (81.5–99.6)
bioMérieux Vidas	6	6			NA	NA	94	6	18	70	74.5 (64.2–82.7)	93.6 (86.1–97.4)
Ortho Vitros	32	29	1	2	93.8 (77.8–98.9)	90.6 (73.8–97.5)	68	6	7[c]	55	80.9 (69.2–89.0)	91.1 (81.1–96.4)
Roche Elecsys	44	29		15	65.9 (50.0–79.1)	NA	56	5		51	91.1 (79.6–96.7)	NA

[a] Concordance of commercial immunoassay results with status, including 95% confidence limits (95% CI), were estimated by considering equivocal results to be both negative and positive.
[b] NA, not applicable.
[c] Seven Ortho Vitros results between 10 and 15 IU/ml were considered low positive rather than equivocal.

dition of 50 μl of 5% oxalic acid. The optical density of each control and sample was read at 405 nm.

Each microtiter plate contained an eight-point standard curve. The standard curve was constructed using duplicate, doubling dilutions of the WHO international anti-rubella immunoglobulin standard (RUBI-1-94) starting at a concentration of 200 IU/ml. An anti-rubella virus IgG external quality control (QC) sample (AcroMetrix, Benicia, CA), calibrated at 20 IU/ml, was tested at least once in each plate. The results of the QC sample were used to determine the assay's repeatability and reproducibility. To determine the linearity of the IH-EIA, a secondary standard, independently calibrated against RUBI-1-94 (2°STD; AcroMetrix, Alkmaar, The Netherlands), and consisting of seven doubling dilutions from 68 to 1.0625 IU/ml was tested and quantified using the standard curve. The IH-EIA was validated using a panel of known anti-rubella virus antibody-positive ($n = 95$) and -negative ($n = 25$) samples.

Testing strategy. All 500 samples were screened using the HAI. All samples having an HAI titer of 1:8 or less were assigned a negative anti-rubella virus antibody status. As HAI detects both rubella virus IgG and IgM, any sample that had an HAI titer of greater than 1:8 was tested in the IB assay. If the IB result was negative, the sample was assigned a negative anti-rubella virus IgG status. If the IB result was positive, that sample was assigned a positive anti-rubella virus IgG status. All 500 samples were then tested on the IH-EIA.

Analysis. The qualitative results of the CIAs were compared with the assigned anti-rubella virus IgG status derived from the HAI and IB testing; where more than 20 results were available for analysis, the percentage of concordance with the assigned status and 95% confidence intervals (95% CI) were determined for each CIA. The quantitative results reported by the CIAs were analyzed using a receiver operating characteristic (ROC) analysis (Analyze-it for Excel; Analyze-it Software, Leeds, United Kingdom) in order to predict the CIAs ' sensitivities and specificities (12, 21). The quantitative results obtained from the CIAs were also compared with those obtained from the IH-EIA using the coefficient of determination (R^2) and Bland-Altman analyses (Analyze-it for Excel; Analyze-it Software, Leeds, United Kingdom) (22–24).

RESULTS

All samples that were initially negative in the HAI and IB were negative on repeat testing. The qualitative results reported by each of the five CIAs were compared with the assigned anti-rubella virus IgG status. The percentage and 95% CI of CIA qualitative results that were concordant with the assigned status were estimated by assuming equivocal results to be either negative or positive (Table 2). As each CIA had a different set of 100 samples analyzed, different proportions of the 100 samples were assigned a positive or negative status. The range of quantitative test results reported by each CIA for samples assigned a negative or positive status is represented graphically in Fig. 1. Only the bioMérieux Vidas reported all samples with a negative status as negative ($n = 6$) but reported 6 and 18 samples assigned a positive status ($n = 94$) as negative and equivocal, respectively. The Roche Elecsys reported 5 of 56 samples with a positive status as negative and 15 of 44 samples with a negative status as positive.

A total of 23 samples assigned a negative status, ranging from 0 to 15 samples per assay, were reported as positive by a CIA (Table 3). Although assigned a negative status by the testing strategy, 19 of the 23 samples had a HAI titer of 16 or greater. All 19 had a negative IB results. HAI detects antibody reactivity against E1 antigens but not E2 or core. Three of the remaining four samples had an HAI titer of 8 and a positive IB result, with evidence of antibody reactivity to E2 antigen. Only 1 of the 23 samples had negative HAI and IB results. The highest positive CIA test result obtained on a sample with a negative status was 36 IU/ml, reported by the Abbott AxSYM. The 15 Roche Elecsys-positive results obtained from samples with a negative status ranged from 10 to 35 IU/ml. A total of 20 samples assigned a positive status, ranging from 1 to 6 samples per assay, were reported as negative by a CIA. All 20 samples were positive for both the HAI and IB tests.

The CIA results were used to perform ROC analysis. Using a cutoff of 10 IU/ml, the ROC analysis was used to determine the predicted sensitivity and specificity, including the 95% CI, of the CIAs for this population of samples (Table 4). The predicted sensitivity for the CIAs ranged from 70.2% for the bioMérieux Vidas to 91.2% for both the Ortho Vitros and the Abbott AxSYM. The predicted specificity ranged from 65.9% for the Roche Elecsys to 100% for the bioMérieux Vidas. However, it is noted that the confidence limits, especially for the specificity calculations, were large due to the relatively low number of samples assigned a negative status. Further, the sensitivity and specificity are not reflective of the assays' performances when testing a normal population, as the samples in this study were selected as having low positive reactivity.

The 23 QC test results obtained from 10 test runs gave a mean of 25.1 IU/ml and a coefficient of variation (CV), expressed as a percentage, of 19.6%. A further 39 QC test results, obtained from a single run, gave a mean of 26.5 IU/ml and a CV of 9.4%.

The 2°STD was tested in two test runs, once in duplicate and once in a single assay, for a total of three test results for each of the seven panel members. When the results of the 2°STD test were

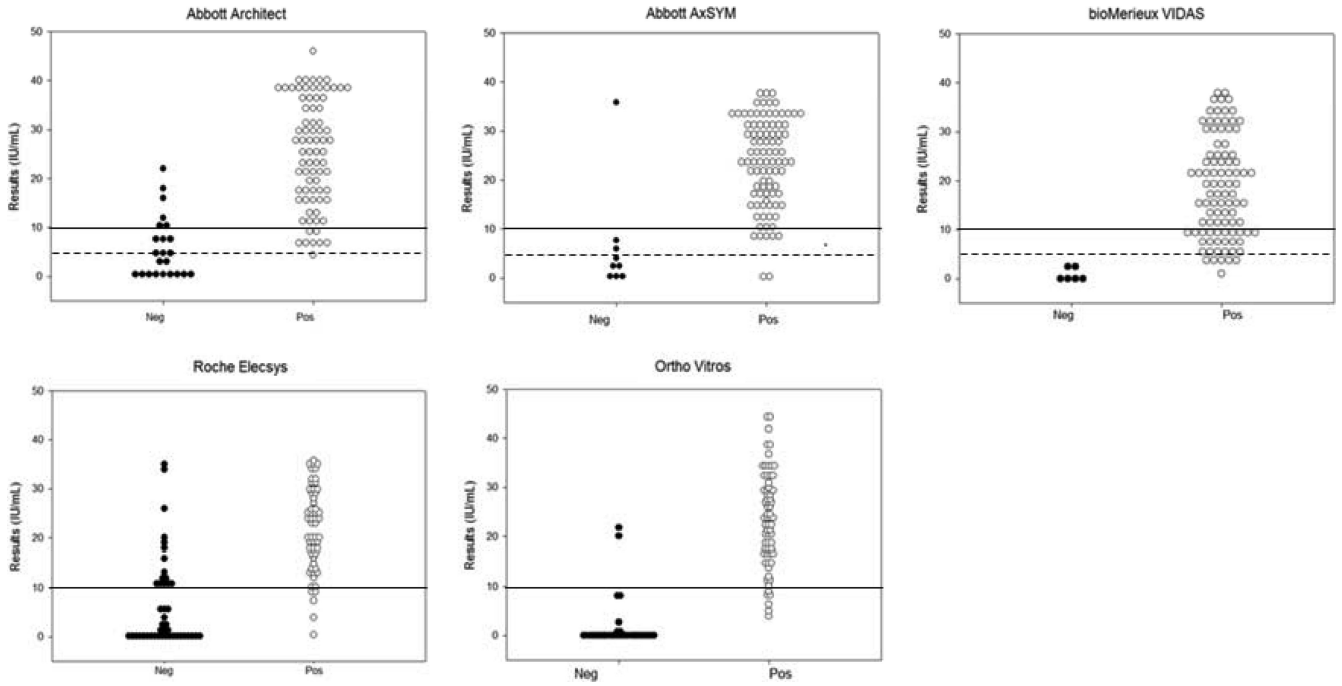


FIG 1 Dot histograms of the quantitative test result, expressed in international units per milliliter (IU/ml), of 100 low-positive (Pos) and negative (Neg) results obtained from each of five commercial immunoassays, for a total of 500 results, plotted against an assigned negative or positive status. The assay's cutoff is represented with a horizontal line, and equivocal ranges are represented with a hashed line.

plotted against the expected value, the R^2 was 0.99 and the equation describing the correlation was $y = 0.90x + 6.26$.

Of the 500 samples, 497 were tested in the IH-EIA. Three samples had insufficient volume to complete the testing. Of the 497 samples tested, 115 were assigned a negative status and 382 a positive status when tested in the HAI and the IB assay. The quantitative results reported by the CIAs were compared with those reported by the IH-EIA. Using Bland-Altman analysis, the IH-EIA quantitative results were within the 95% confidence limits of agreement for all CIA quantitative results up to approximately 60 IU/ml. At concentrations higher than this, the IH-EIA consistently reported levels of anti-rubella virus IgG that were higher than those reported by the CIAs. The R^2, slope, and bias of the comparison of IH-EIA quantitative results with the corresponding CIA result were estimated with and without the IH-EIA results greater than 60 IU/ml (Table 5). When the samples having an IH-EIA result of greater than 60 IU/ml were removed, the R^2 ranged from 0.210 to 0.421, indicating a lack of correlation between the results of the IH-EIA and each of the CIAs. When all test results were used to calculate the R^2, the R^2 values were even lower. Using Bland-Altman analysis, the IH-EIA had a positive bias compared with each of the CIAs, ranging from 3.71 to 9.11 IU/ml.

TABLE 3 The number and range of quantitative test results reported by five commercial anti-rubella virus IgG immunoassays that were discordant compared with an assigned anti-rubella virus IgG status[a]

	CIA-positive results for samples assigned a negative status		CIA-negative results for samples assigned a positive status	
Assay	No.	Range (IU/ml)	No.	Range (IU/ml)
Abbott Architect	5	11–22	1	4.3
Abbott AxSYM	1	36	2	0.0–0.5
bioMérieux Vidas	0		6	1.0–4.0
Ortho Vitros	2	20–22	6	3.9–9.1
Roche Elecsys	15	10–35	5	0.4–9.5

[a] CIA, commercial anti-rubella virus IgG immunoassay.

DISCUSSION

The instructions for use (IFU) of the international rubella standard state that "RUB-1-94 consists of freeze-dried residuals of 2.0 ml of a mixture of normal human immunoglobulin and an equal volume of sterile distilled water." It was calibrated against the second International Standard for Rubella Serum and was estimated to contain 1,600 IU per ampoule. This standard has been used to calibrate commercial anti-rubella virus IgG assays since 1995 (8–10). The European *in vitro* diagnostics directive states that "the traceability of values assigned to calibrators and/or control materials must be assured through available reference measurement procedures and/or available reference materials of a higher order" (25). However, others have questioned the practicality of standardizing biologicals that are traceable not to Système International d'Unités (SI) units but to arbitrary units such as WHO international standards, stating that "immunogenic proteins such as viral proteins are generally highly complex and heterogeneous mixtures in biological fluids" (26). In these cases, the reference materials and the biological fluids are therefore "nonidentical," which consequently invalidates the basic principle of traceability: to compare like with like (27). Indeed, the WHO standard is polyclonal in nature and it is generally not possible to create a secondary standard or certified reference material that would behave in the same manner when tested in the same immunoassay. The shape of the curve generated by the secondary standards would always differ from that generated by the WHO stan-

TABLE 4 Estimation of the sensitivity and specificity, using receiver operating characteristic analysis with a cutoff set at 10 international units per milliliter, of five commercial immunoassays using samples having low-level and negative results to anti-rubella virus IgG[a]

ROC analysis parameter	Abbott Architect	Abbott AxSYM	bioMérieux Vidas	Roche Elecsys	Ortho Vitros
Curve area	0.93	0.88	1.00	0.88	0.97
Predicted specificity (%)	78.3	88.9	100	65.9	93.8
95% confidence interval	56.3–92.5	51.8–99.7	54.1–100.0	50.1–79.5	79.2–99.2
Predicted sensitivity (%)	89.6	91.2	70.2	89.3	91.2
95% confidence interval	80.6–95.4	83.4–96.1	59.9–79.2	80.4–97.0	81.8–96.7

[a] ROC, receiver operating characteristic.

dard, and this would be most apparent at the lower region of the curve (28). Considering the differences in the formulations of the CIAs (Table 1), in particular, in the antigen(s) bound to the solid phase, a lack of standardization between CIAs is not surprising.

Past studies have demonstrated a lack of standardization between some assays used to quantify anti-rubella virus IgG in serum (8, 9). In conjunction with a lack of standardization, vaccination programs have resulted in lower levels of anti-rubella virus IgG being detected in the population (2, 3). This lack of standardization and the number of individuals having vaccine-induced low-level anti-rubella virus IgG can cause difficulties in the interpretation of the results, especially when the result is close to the cutoff of the assay (7, 29). Studies indicate that the sensitivities and specificities of many commercial EIAs are similar (9, 14, 30, 31). However, as most CIAs use 10 IU/ml as a cutoff for immunity, different qualitative results for the same sample are reported by different assays. Therefore, results generated from an individual's sample that is tested in one assay cannot be compared with results obtained in other assays. So when results are obtained from acute and convalescent samples and each sample is tested in different assays, the results may resemble a seroconversion to anti-rubella virus IgG, with the early sample testing negative in one assay and later samples testing positive in another. Potentially, these results may be interpreted as evidence of a recent rubella virus infection. This situation may lead to anxiety for the patient or even a recommendation for termination of pregnancy. Therefore, consecutive samples from the same individual should be tested together with the same assay (11).

As new commercial immunoassays are introduced to the market, scientists comparing the results obtained from the new assay with those obtained from their routine assay experience difficulty in elucidating discordant test results. Testing samples with discordant test results on a third assay is not recommended (32, 33). Comparison of qualitative results with those obtained from a gold standard reference test is preferred (12, 28). For anti-rubella virus IgG testing, HAI, viral neutralization, and Western blot analyses have been considered appropriate reference tests, although very few laboratories worldwide retain the expertise for neutralization testing (6, 15, 16, 19). Further, these tests are manual and complex and also subject to variation (15, 16, 34). Of the 23 samples assigned a negative status but having a positive CIA result, 22 had a positive result in either HAI or IB testing but a negative result in the other test. The negative result may have been due to a lack of sensitivity of the assay. Also, HAI detects only antibodies to E1 antigen. This phenomenon caused three samples to be assigned a negative status in the testing strategy, whereas the CIA reported a positive result and the IB assay had evidence of E2 antibody reactivity. The HAI can detect anti-rubella IgM, whereas the IB test and the CIA detect only IgG-specific antibodies. The presence of anti-rubella virus IgM may explain why some samples were HAI positive but IB negative. These results underline the difficulty of selecting a reference testing strategy to confirm qualitative anti-rubella virus IgG results.

The qualitative test results of the CIAs were compared with a status determined by HAI and IB testing. If the equivocal results in the CIAs were assumed to be positive for anti-rubella virus IgG, the qualitative results of all CIAs gave greater than 90% concordance with the assigned positive status. If the equivocal results in the CIAs were assumed to be negative, the percentages of concordance of the qualitative test results on the Abbott Architect and Roche Elecsys with the assigned negative status were poor, at 69.2 and 65.9%, respectively. However, the concordance of the qualitative results on the Ortho Vitros with assigned negative status was higher at 83.8%. These findings are not indicative of the sensitiv-

TABLE 5 Summary of the analysis of correlation between quantitative results, reported as international units per milliliter obtained from five commercial anti-rubella virus IgG immunoassays and an in-house immunoassay calibrated with the World Health Organization international standard rubella virus serum RUB-1-94[a]

	Commercial EIA				
Analysis category and parameter	Abbott Architect	Abbott AxSYM	bioMérieux Vidas	Roche Elecsys	Ortho Vitros
Samples with IH-EIA results less than 60 IU/ml					
Coefficient of determination (R^2)	0.421	0.210	0.276	0.227	0.420
Slope	0.58	0.52	0.42	0.41	0.49
Intercept (IU/ml)	2.48	14.04	5.78	6.41	4.42
All samples					
Coefficient of determination (R^2)	0.354	0.106	0.169	0.063	0.110
Slope	1.35	0.90	1.12	0.60	0.94
Intercept (IU/ml)	10.50	15.12	15.12	20.57	23.41

[a] IH-EIA, in-house immunoassay; IU/ml, international units per milliliter.

ities and specificities of the assays for testing an unselected population, as the samples in the study were preselected for low positive reactivity. Generally, CIAs have sensitivity greater than 98% and specificity of greater than 85% (9, 31).

The IFU for the two Abbott assays and bioMérieux Vidas specify equivocal ranges from 5.0 to 9.9 IU/ml, whereas the Ortho Vitros IFU specify an equivocal range of 10.0 to 14.9 IU/ml. The Roche Elecsys IFU do not state an equivocal range. Only the Abbott AxSYM and the Roche Elecsys reported test results greater than 30 IU/ml. No CIA reported a positive result greater than 36 IU/ml for a sample that was assigned a negative status. The use of a strict cutoff of 10 IU/ml invariably results in the reporting of false-positive and -negative test results. Many Australian laboratories testing for anti-rubella virus IgG choose to report results between 10 and 30 IU/ml as positive but add a clinical interpretation stating that the clinical significance of the result is in doubt. The sources for this uncertainty are the imprecision of the assay and biological false reactivity of some samples in the assay. Establishing and reporting a gray zone allows laboratories to follow the manufacturer's instructions while acknowledging the uncertainty of the clinical interpretation at low levels of antibodies.

The qualitative nature of the reference test results does not aid in determining the accuracy of quantification of antibodies. The present study evaluated quantitative CIA results by testing sets of samples having negative or low-level rubella virus IgG in an IH-EIA, which was directly calibrated using RUBI-1-94 and designed to be linear at low levels of anti-rubella virus IgG. Results obtained from the IH-EIA were shown to be precise, with repeatability (within-run precision) of less than 10% and reproducibility (between-run precision) of 20%. Excellent correlation with results obtained from the independent secondary standard, 2°STD, indicated that the IH-EIA was accurate to a level of about 60 IU/ml. When the quantitative results of CIAs were compared with those obtained from the IH-EIA, the level of correlation was very poor for all CIAs and the IH-EIA had a positive bias ranging from 3.7 to 9.1 IU/ml. This would support the theory that CIAs lack standardization, particularly at low levels of rubella virus IgG, even though the calibrators of all but one CIA were traceable to RUBI-1-94.

All assays experience imprecision (35). In our experience conducting QC programs for infectious disease testing for more than 10 years, results of testing the same QC sample on manual microtiter plate assays generally show variation of less than 20% whereas those from instrument-based serology assays show variation of less than 15%. The results of a peer-comparison QC program using a QC sample calibrated at 20 IU/ml against RUBI-1-94 demonstrated that the imprecision of CIAs ranged between 5 and 20%, with a measurement uncertainty (MU) of about 2 to 5 IU/ml (unpublished data). That is, when the MU is 5 IU/ml, a result of 10 IU/ml has a 95% confidence of being between 5 and 15 IU/ml. Therefore, the interpretation of low-positive test results is difficult when both the imprecision and lack of standardization (bias) of the assays are considered.

This study has demonstrated the difficulty in both standardizing assays designed to detect and quantify antibodies and, consequently, using quantitative results to set immune/nonimmune cutoffs. Both imprecision and bias contribute to these difficulties (35). However, the main contributing factor to the lack of standardization is the poor implementation of traceability protocols. When the WHO standard was developed, there was insufficient definition of the analyte as required in traceability. The approach taken did not consider factors such as biological variation and the complexity of the proteins being detected, the reactivity of these antigens with antibodies of differing levels of avidity and affinity, the characteristics of the assay being used, and the stage of disease of the patient being tested (26, 27). Anti-rubella virus IgG testing has no quantitative reference method with known and defined uncertainty. Indeed, the results of testing of RUBI-1-94 were never published, no details of the methods for calibration are available, and the manufacturer's IFU indicates that "This study has almost been completed."

Although an international reference standard has been available since the 1980s and has been used to calibrate assays for the detection and quantification of anti-rubella virus IgG, it has not led to greater levels of standardization between commercial assays and issues reported in 1992 remain. It may be time for the scientific community to question the relevance of quantification of anti-rubella virus IgG. It may be possible for manufacturers of commercial assays to assign a cutoff for their assays by maximizing sensitivity and specificity using well-validated panels of samples with a known status, disregarding the WHO standard, and no longer reporting the anti-rubella virus IgG results in IU/ml but as a qualitative result.

ACKNOWLEDGMENTS

We acknowledge all laboratories providing samples for this study: Douglass Hanly Moir Pathology, San Pathology, Eastern Health Pathology, Pathology One, and Adelaide Pathology in Australia and Hackensack University Medical Center in the United States.

Equal financial sponsorship was provided by each manufacturer.

REFERENCES

1. **Gregg NM.** 1991. Congenital cataract following German measles in the mother. 1941. Aust. N. Z. J. Ophthalmol. **19**:267–276.
2. **Francis BH, Hatherley LI, Walstab JE, Taft LI.** 1982. Rubella screening and vaccination programme at a Melbourne maternity hospital. A five-year review. Med. J. Aust. **1**:502–504.
3. **Francis BH, Thomas AK, McCarty CA.** 2003. The impact of rubella immunization on the serological status of women of childbearing age: a retrospective longitudinal study in Melbourne, Australia. Am. J. Public Health **93**:1274–1276.
4. **Reef SE, Cochi SL.** 2006. The evidence for the elimination of rubella and congenital rubella syndrome in the United States: a public health achievement. Clin. Infect. Dis. **43**(Suppl 3):S123–S125.
5. **Robertson SE, Cutts FT, Samuel R, Diaz-Ortega JL.** 1997. Control of rubella and congenital rubella syndrome (CRS) in developing countries, part 2: vaccination against rubella. Bull. World Health Organ. **75**:69–80.
6. **NCCLS.** 1997. Detection and quantification of rubella IgG antibody: evaluation and performance criteria for multiple component products, specimen handling, and use of test products in the clinical laboratory; approved guideline. NCCLS, Wayne, PA.
7. **Skendzel LP.** 1996. Rubella immunity. Defining the level of protective antibody. Am. J. Clin. Pathol. **106**:170–174.
8. **Dimech W, Bettoli A, Eckert D, Francis B, Hamblin J, Kerr T, Ryan C, Skurrie I.** 1992. Multicenter evaluation of five commercial rubella virus immunoglobulin G kits which report in international units per milliliter. J. Clin. Microbiol. **30**:633–641.
9. **Dimech W, Panagiotopoulos L, Francis B, Laven N, Marler J, Dickeson D, Panayotou T, Wilson K, Wootten R, Dax EM.** 2008. Evaluation of eight anti-rubella virus immunoglobulin G immunoassays that report results in international units per milliliter. J. Clin. Microbiol. **46**:1955–1960.
10. **Forsgren M.** 1985. Standardization of techniques and reagents for the study of rubella antibody. Rev. Infect. Dis. **7**(Suppl 1):S129–S132.
11. **Guillet G, Grangeot-Keros K, Vauloup-Fellous C.** 2012. Discrepancies between rubella IgG results: which assay is right, which assay is wrong? Abstr. 22nd European Congress of Clinical Microbiology and Infectious Diseases (ECCMID), poster P963. John Wiley and Sons, Inc., Hoboken, NJ.

12. **CLSI.** 2008. Clinical evaluation of immunoassays; approved guidelines—2nd ed. CLSI document I/LA21-A2 28. CLSI, Wayne, PA.
13. **Diepersloot RJ, Dunnewold-Hoekstra H, Kruit-Den Hollander J, Vlaspolder F.** 2001. Antenatal screening for hepatitis B and antibodies to *Toxoplasma gondii* and rubella virus: evaluation of two commercial immunoassay systems. Clin. Diagn. Lab. Immunol. **8**:785–787.
14. **Vlaspolder F, Singer P, Smit A, Diepersloot RJ.** 2001. Comparison of Iimmulite with Vidas for detection of infection in a low-prevalence population of pregnant women in The Netherlands. Clin. Diagn. Lab. Immunol. **8**:552–555.
15. **Public Health Laboratory Services Standing Advisory Committee on Viral Reagents.** 1978. Haemagglutination-inhibition test for the detection of rubella antibody. J. Hyg. (Lond) **81**:373–382.
16. **Cradock-Watson JE.** 1991. Laboratory diagnosis of rubella: past, present and future. Epidemiol. Infect. **107**:1–15.
17. **Cusi MG, Metelli R, Valensin PE.** 1989. Immune responses to wild and vaccine rubella viruses after rubella vaccination. Arch. Virol. **106**:63–72.
18. **Cusi MG, Rossolini GM, Cellesi C, Valensin PE.** 1988. Antibody response to wild rubella virus structural proteins following immunization with RA 27/3 live attenuated vaccine. Arch. Virol. **101**:25–33.
19. **Nedeljkovic J, Jovanovic T, Mladjenovic S, Hedman K, Peitsaro N, Oker-Blom C.** 1999. Immunoblot analysis of natural and vaccine-induced IgG responses to rubella virus proteins expressed in insect cells. J. Clin. Virol. **14**:119–131.
20. **Wilson KM, Di Camillo C, Doughty L, Dax EM.** 2006. Humoral immune response to primary rubella virus infection. Clin. Vaccine Immunol. **13**:380–386.
21. **Buttner J.** 1993. Evaluation of diagnostic and prognostic measures, p 25–30. *In* Haeckel R (ed), Evaluation methods in laboratory medicine. HCH, Weinhiem, Germany.
22. **Bland JM, Altman DG.** 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet **i**:307–310.
23. **Bland JM, Altman DG.** 1995. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet **346**: 1085–1087.
24. **Bland JM, Altman DG.** 1999. Measuring agreement in method comparison studies. Stat. Methods Med. Res. **8**:135–160.
25. **European Commission.** 1998. Directive 98/79/EC of the European Parliament and the Council on in vitro diagnostics (IVD) medical devices 1998. Official Journal of the European Communities. European Commission, Enterprise and Industry DG, Brussels, Belgium.
26. **Lequin RM.** 2002. Standardization: comparability and traceability of laboratory results. Clin. Chem. **48**:391–393.
27. **American Association for Laboratory Accreditation.** 2012. P102-A2LA policy on measurement traceability. The American Association for Laboratory Accreditation, Frederick, MD.
28. **Davies C.** 1994. Principles, p 23–46. *In* Wild D (ed), The immunoassay handbook. Stockton Press, New York, NY.
29. **O'Shea S, Dunn H, Palmer S, Banatvala JE, Best JM.** 1999. Automated rubella antibody screening: a cautionary tale. J. Med. Microbiol. **48**:1047.
30. **Greenwood NP, Ovsyannikova IG, Vierkant RA, O'Byrne MM, Poland GA.** 2010. A qualitative and quantitative comparison of two rubella virus-specific IgG antibody immunoassays. Viral Immunol. **23**:353–357.
31. **Medici MC, Martinelli M, Albonetti V, Chezzi C, Dettori G.** 2008. Evaluation of rubella virus immunoglobulin G (IgG) and IgM assays with the new Vidia instrument. J. Clin. Microbiol. **46**:1847–1849.
32. **Hadgu A.** 1996. The discrepancy in discrepant analysis. Lancet **348**:592–593.
33. **Miller WC.** 1998. Bias in discrepant analysis: when two wrongs don't make a right. J. Clin. Epidemiol. **51**:219–231.
34. **Schmidt NJ, Lennette EH.** 1970. Variables of the rubella hemagglutination-inhibition test system and their effect on antigen and antibody titers. Appl. Microbiol. **19**:491–504.
35. **Dimech W, Francis B, Kox J, Roberts G.** 2006. Calculating uncertainty of measurement for serology assays by use of precision and bias. Clin. Chem. **52**:526–529.

Chapter Two - Conclusion

The results of the first paper (93) found some of the assays under investigation reported comparable quantitative results whereas others did not. The study concluded "*Our results show the level of standardization between anti-rubella virus IgG immunoassays reporting results expressed as IU/ml has improved since a previous study in 1992, but further improvement is required*." The findings of this study were subsequently confirmed by other European colleagues (75, 76, 94). These studies also highlighted issues relating to the use of a quantitative cut-off and further raised the issue regarding the potential of false test results leading to misinterpretation and poor clinical outcomes. Indeed, this seems to be the case in a paper "evaluating a possible re-emergence of rubella" where, of 298 women, it was concluded that 19 (6.4%) "*seroconverted from no detectable anti-rubella IgG to detectable antibodies during pregnancy"* (140). An assessment of the paper's methodology indicated that the change in detectable antibody levels was most likely due to testing patient samples on different test systems, where the initial sample contained low levels of anti-rubella IgG but that was identified as being negative and non-immune in some test systems, whereas subsequent bleeds from the same individual post-partum reported a positive result. The differences were due to the use of different assays when testing the pre- and post-partum samples, rather than a true seroconversion. In serology, acute and convalescent samples must always be tested at the same time (in parallel). Although the paper identified some neonatal abnormalities, these were not due to congenital rubella and no definitive evidence of the mothers having rubella infection was presented. There are medical consequences to a misleading anti-rubella serology result. A false positive result can give a pregnant woman a false sense of immunity, especially if they are exposed to the infection. A false negative result, or a pseudo seroconversion, can lead to an unnecessary termination of pregnancy.

The results of the second, subsequent study (92) provided a better understanding of the clinical relevance of low-level anti-rubella IgG reactivity. The study demonstrated that no assay reported a positive result greater than 36 IU/ml for any of 500 sample that were assigned a negative status by the HAI and WB reference testing. This finding, along with other studies presented at National conferences, established a "grey-zone" of 10 to 30 IU/mL for interpreting anti-rubella IgG results as being "doubtful immunity". This range was adopted by many Australian laboratories when reporting rubella serology results. Following the second paper included in **Chapter 2** (92), Picone *et al* conducted a significant study investigating the humoral and T cell response to rubella infection in antenatal women (95). Women having low-level anti-rubella IgG immune responses of 1 to 10 IU/mL on antenatal testing, were vaccinated post-partum and subsequently tested with an immunoblot, a T cell stimulation assay, anti-rubella IgG and IgM assays, as well as anti-rubella IgG avidity testing, to determine if the immune response was primary or secondary. The study found that 52.2% of the women with low levels of anti-rubella IgG elicited a secondary response on vaccination, indicating that the original result of between 1 and 10 IU/mL was reflective of true anti-rubella IgG prior to vaccination. Indeed, the suggested grey-zone of 10-30 IU/mL was found to be conservative, and individuals having detectable levels of anti-rubella IgG of less than 10 IU/mL in most assays elicited a secondary response when challenged by vaccination.

The significant impact of original work presented in **Chapter 2** was multi-fold. An incidental but useful finding from these collective studies demonstrated that anti-rubella IgG immunoblot results are highly specific and sufficiently sensitive to be used by others as a confirmatory test for anti-rubella IgG (76, 95). Moreover, the detection of anti-rubella E1 antibodies on immunoblot was indicative of a secondary immune response following re-immunisation, irrespective of the level of antibodies (95).

More significantly, the papers include in this thesis and others, highlighted the lack of commutability of the WHO International standard RUB-1-94. Its use as a calibrator over a period of two decades has failed to standardise anti-rubella IgG IAs. Multiple publications by myself and colleagues highlighted this lack of

standardisation of anti-rubella IgG test results and the concern raised by these publications resulted in the convening of three extra-ordinary WHO-sponsored meeting in Barcelona (2014), Copenhagen (2015) and Geneva (2017) to address the issue (**Appendix B**). These meetings identified the need for a comprehensive review of rubella standardisation, which I first authored (14), and the development of a panel of anti-rubella IgG negative samples that manufacturers could use to maximise assay specificity (141). Ultimately, the impact of my work, and that of others, resulted in a submission to the WHO Expert Committee on Biological Standardisation, who authorised an amendment to the RUB-1-94 International standard IFU (**Appendix C**) to include the statement "*IVD manufacturers, regulators and assay users should be made aware of RUBI-1-94 potential lack of commutability when used as a calibrant*" (85).

My studies have focused attention on the use of a quantitative cut-off for anti-rubella IgG testing. A paper co-authored by me and representatives from WHO, USA CDC, USA FDA, Paul Ehrlich Institute and NIBSC reported the recommendations of the WHO Expert Committee on Biological Standards. It concluded that stakeholders "*should be encouraged to reconsider the appropriateness of quantitative anti-rubella measurement for the determination of immune status and use of 10 IU/mL as a cut-off point for assessing immune protection*" (77, 85). Manufacturers of commercial anti-rubella IgG assays are encouraged to adopt qualitative reporting of results. However, there is little evidence of this happening, although anecdotally there are some considerations of this approach by certain manufacturers. Developing tests with qualitative results will allow easier differentiation of positive and negative populations, thereby reducing the need for a "indeterminate" or "doubtful significant" result. There is still significant work required to elucidate the understanding of rubella testing. Until then, clinicians will need to interpret "grey-zone" results with caution. Indeed, an alternative outcome is the abandonment of routine screening for immunity to rubella in very low prevalence settings, as has occurred in the UK in 2016 (142).

However, the issue of lack of standardisation of serology tests has not been resolved, as evident by the recent release of a WHO international standard for SARS-CoV-2 antibodies (143, 144). A recent study questions the effectiveness of this standard (145). As the lead investigator of a comprehensive evaluation of anti-SARS-CoV-2 serology test kits, in collaboration with the WHO, studies investigating the efficacy of the calibration of commercial immunoassays using the new international standard are currently in progress.

Chapter Three: Standardisation of nucleic acid testing for infectious diseases

Processes used to standardise viral load NAT have proven more successful than those applied to infectious disease serology (108). Numerous international standards have been developed and utilised to calibrate commercial assays. Results reported by different test systems calibrated against these standards have been shown to correlate (108). This is because, generally, molecular assays are designed to detect segments of the viral genome that are highly conserved, whereas in serology, the measurand varies considerably from one assay to another even when detecting antibodies against the same organism. So even though different molecular assay may detect different regions of the genome, a similar number of copies of that genome are present prior to amplification and therefore similar quantitative values are reported, meaning the measurand responds more like a "type A" analyte.

It has been observed that differences in quantification associated with different genotypes may occur when an assay has been optimised for a predominant genotype (146). Another potential cause of poor standardisation of NAT is due to the presence of circulating genomic fragments in the patient's sample. This is particularly problematic in the quantification of CMV and EBV, most commonly tested post solid tissue transplantation. (118, 119). The cause of this phenomenon is unknown (118). The consequence is that these non-infectious fragments of DNA circulate and, depending on the target region and primers used, these fragments may or may not be detectable and therefore measured. This situation potentially impacts the ability to standardise test systems, as each test system is potentially detecting different measurands (i.e. including or excluding different fragments depending on the target). In 2017, CMV DNA quantification was performed by laboratories globally using a mixture of commercial and in-house tests, as evident by results reports to multiple proficiency test providers described in the paper presented in **Chapter 3**. To investigate the effectiveness of the international standard in calibrating CMV viral load results of commercial assays, I lead a study of three international EQA providers, UKNEQAS (Colindale, UK), Instand (Duesseldorf, Germany) and NRL to test the hypothesis that all CMV DNA test systems, testing the same samples, would report comparable quantitated results. Although the 1st WHO International Standard for Human Cytomegalovirus for Nucleic Acid Amplification Techniques (Code 09/162), was released by NIBSC in 2010, only some commercial assays were calibrated against the standard, and other assays allowed the user to report in either IU/mL or copies/mL. Test systems that utilised calibrators standardised against the WHO international standard could report results in IU/mL, whereas all other tests systems not calibrated against the WHO international standard reported in copies/mL. The process that manufacturers of commercial assays and laboratories using in-house tests used for establishing the copies/mL is not well defined. Possible methods could include the use of commercial standards with results expressed in copies/mL or using droplet digital PCR or next generation sequencing. Other non-standardised assays only reported in copies/mL. In partnership with the two other EQA providers, the same set of samples were sent to participants globally and the results reported in IU/mL and copies/mL from different assays were compared (147). The panel was comprised of five samples – three dilutions of CMV culture supernatant (AD169 strain); a dilution of a clinical sample obtained from infected amniotic fluid and a sample negative for CMV virus.

The paper presented in **Chapter 3** adds to the understanding of the effectiveness of the WHO international standard in standardising CMV viral load assays and determines the relationship between CMV viral load assays that express results in IU/mL and copies/mL, indicating the equivalence of the two units of measure. However, the study also found that there was no significant difference determined between the results from assays reporting in IU/mL, whereas there were significant differences between assays reporting in c/mL, highlighting the utility of the international standard in standardising CMV DNA viral load results.

Journal of Virological Methods 252 (2018) 15–23

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Results of cytomegalovirus DNA viral loads expressed in copies per millilitre and international units per millilitre are equivalent

Wayne Dimech[a], Liza M. Cabuang[a,*], Hans-Peter Grunert[b], Vanessa Lindig[d], Vivienne James[c,1], Brigitte Senechal[c,1], Giuseppe A. Vincini[a], Heinz Zeichhardt[d,e,f,1]

[a] National Serology Reference Laboratory (NRL), Melbourne, Australia
[b] GBD Gesellschaft fuer Biotechnologische Diagnostik mbH, Berlin, Germany
[c] UK National External Quality Assessment Service (UKNEQAS) for Microbiology, Public Health England, London, United Kingdom
[d] Institut fuer Qualitaetssicherung in der Virusdiagnostik – IQVD der GBD mbH, Berlin, Germany
[e] INSTAND Gesellschaft zur Foerderung der Qualitaetssicherung in medizinischen Laboratorien e.V., Duesseldorf, Germany
[f] Charité-Universitaetsmedizin, Institut fuer Virologie, Berlin, Germany

ARTICLE INFO

Keywords:
Cytomegalovirus
Viral load
International standard
Copies per millilitre
Standardization

ABSTRACT

Quantification of Cytomegalovirus (CMV) DNA is required for the initiation and monitoring of anti-viral treatment and the detection of viral resistance. However, due to the lack of standardisation of CMV DNA nucleic acid tests, it is difficult to set universal thresholds. In 2010, the 1st WHO International Standard for Human Cytomegalovirus for Nucleic Acid Amplification Techniques was released. Since then CMV DNA viral load assays have been calibrated using this standard. Three external quality assessment (EQA) providers sent the same five samples to their participants and analysed the results to determine the equivalence of reporting CMV DNA results in international units per millilitre (IU/mL), and compared the difference in results reported in IU/mL with those reported in copies per millilitre (c/mL), and to determine the rate of adoption of IU/mL. About 78% of participants continue to report results in c/mL even though six of the 12 commercial assays are calibrated against the standard. The range of the results reported in IU/mL was less than those reported in c/mL indicating that the adoption of the WHO standard successfully improved the reporting of the CMV viral load. The variation in individual sample results reported by different assays, irrespective of whether in IU/mL or c/mL, is still great and therefore more standardisation of the assays is needed to allow the setting of treatment and monitoring thresholds. This study can act as a bench mark to determine rate of future adoption if reporting CMV DNA viral load results in IU/mL.

1. Background

Human cytomegalovirus (CMV) is a double-stranded DNA virus of the Herpesviridae family (Ramanan and Razonable, 2013; Ross et al., 2011) causing mild or asymptomatic infection in healthy humans and becomes latent systemically in the host (Ross et al., 2011). Transmission occurs through various routes, including person-to-person, vertically from mother to child or through blood or solid organ donation (Ramanan and Razonable, 2013; Razonable and Hayden, 2013). CMV is the most common cause of congenital infection (Ross et al., 2011).

CMV can cause serious infection in the immunocompromised (Ramanan and Razonable, 2013; Razonable and Hayden, 2013; Babady et al., 2015). To reduce the risk of transmission to the immunocompromised, recipient are treated with antiviral drugs. The success of treatment is determined by testing for CMV DNA viral load (VL) using nucleic acid tests (NAT) (Ramanan and Razonable, 2013; Razonable and Hayden, 2013). Accurate CMV DNA VL quantification is critical in management and the prevention of CMV infection (Kraft et al., 2012). In 2010 the 1st WHO International Standard for Human Cytomegalovirus for Nucleic Acid Amplification Techniques (Code 09/162) (NIBSC, Potters Bar, UK) (Freyer et al., 2010) was released. Whilst many commercially manufactured assays that measure CMV DNA VL now report test results in international units per millilitre (IU/mL), the transition from reporting results in copies per millilitre (c/mL) to IU/mL has taken a relatively long period of time.

Assessment of the performance of CMV DNA VL testing can be determined through participation in External Quality Assessment (EQA) schemes (Hayden et al., 2012). One drawback with these schemes is

* Correspondence to: 4th Floor Healy Building, 41 Victoria Parade, FITZROY, 3065, Victoria, Australia.
E-mail address: wayne@nrl.gov.ai (W. Dimech).
[1] Retired.

https://doi.org/10.1016/j.jviromet.2017.11.001
Received 31 July 2017; Received in revised form 29 October 2017; Accepted 1 November 2017
Available online 07 November 2017

Table 1
Summary of results, expressed as log_{10} copies per millilitre (log_{10} c/mL) and International Units per millilitre (log_{10} IU/mL), reported by participants testing four samples positive for human cytomegalovirus DNA, distributed by three external quality assessment (EQA) scheme providers.

Sample	EQAS provider	Units	Count	Mean	Median	SD*	Minimum	Maximum	Range
1	Instand	log_{10} c/mL	54	4.98	4.99	0.40	3.56	6.08	2.52
		log_{10} IU/mL	27	5.07	5.09	0.24	4.34	5.34	1.00
	NRL	log_{10} c/mL	9	4.94	4.94	0.42	4.24	5.93	1.68
		log_{10} IU/mL	3	5.07	4.95	0.34	4.73	5.53	0.80
	UKNEQAS	log_{10} c/mL	84	5.03	5.00	0.46	2.83	6.53	3.70
		log_{10} IU/mL	11	4.94	4.95	0.20	4.62	5.28	0.67
	All Data	log_{10} c/mL	147	5.00	4.99	0.44	2.83	6.53	3.70
		log_{10} IU/mL	41	5.03	5.03	0.24	4.34	5.53	1.19
2	Instand	log_{10} c/mL	53	3.97	3.95	0.42	0.00	5.11	5.11
		log_{10} IU/mL	27	4.14	4.13	0.25	3.44	4.68	1.24
	NRL	log_{10} c/mL	9	4.00	4.00	0.39	3.36	4.85	1.48
		log_{10} IU/mL	3	3.95	3.95	0.11	3.82	4.09	0.27
	UKNEQAS	log_{10} c/mL	84	4.03	4.07	0.41	2.94	5.28	2.34
		log_{10} IU/mL	12	3.84	3.97	0.51	2.34	4.48	2.13
	All Data	log_{10} c/mL	146	4.03	4.01	0.41	2.30	5.28	2.98
		log_{10} IU/mL	42	4.04	4.09	0.37	2.34	4.68	2.34
4	Instand	log_{10} c/mL	54	3.97	3.95	0.40	2.30	4.99	2.69
		log_{10} IU/mL	27	4.10	4.15	0.23	3.55	4.52	0.97
	NRL	log_{10} c/mL	9	4.00	3.99	0.44	3.20	4.95	1.76
		log_{10} IU/mL	3	4.00	4.07	0.13	3.82	4.12	0.30
	UKNEQAS	log_{10} c/mL	84	4.10	4.05	0.42	2.49	5.41	2.92
		log_{10} IU/mL	12	3.95	3.99	0.36	3.08	4.41	1.34
	All Data	log_{10} c/mL	147	4.05	4.03	0.42	2.30	5.41	3.12
		log_{10} IU/mL	42	4.05	4.09	0.28	3.08	4.52	1.44
5	Instand	log_{10} c/mL	54	5.04	5.07	0.49	3.41	6.18	2.76
		log_{10} IU/mL	27	5.15	5.25	0.30	4.00	5.72	1.72
	NRL	log_{10} c/mL	9	5.08	5.06	0.40	4.45	6.00	1.55
		log_{10} IU/mL	3	5.46	5.26	0.36	5.16	5.97	0.81
	UKNEQAS	log_{10} c/mL	84	5.12	5.12	0.46	3.38	6.46	3.08
		log_{10} IU/mL	11	4.90	4.93	0.37	4.15	5.45	1.30
	All Data	log_{10} c/mL	147	5.09	5.11	0.47	3.38	6.46	3.08
		log_{10} IU/mL	41	5.11	5.16	0.36	4.00	5.97	1.97

*SD Standard deviation.

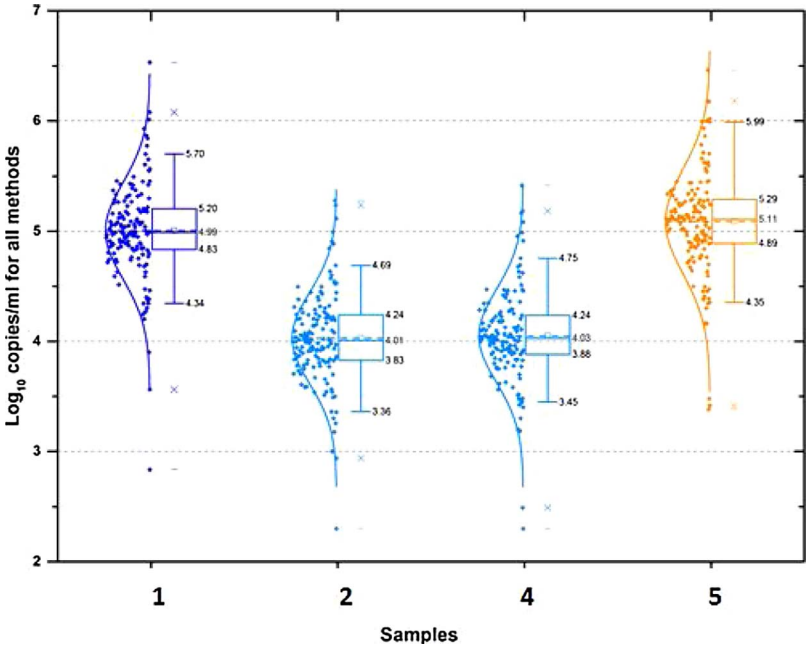


Fig. 1. Accumulated results for quantitative CMV DNA testing for all four CMV DNA positive samples (n = 147 for each sample) from the participants of three EQA providers, INSTAND, NRL and UKNEQAS: results of all applied methods expressed in Log_{10} copies/ml.

that different sample sets are distributed by different scheme providers. Three of scheme providers, INSTAND Gesellschaft zur Foerderung der Qualitaetssicherung in medizinischen Laboratorien e.V. (INSTAND e.V., Duesseldorf, Germany), UK National External Quality Assurance Scheme (UK NEQAS) for Microbiology (London, UK) and the National Serology Reference Laboratory, Australia (NRL, Melbourne, Australia) collaborated by providing the same set of samples to all participants in their respective CMV DNA VL schemes, represents the first and largest global assessment of variation in CMV DNA VL test results reported internationally.

2. Objectives

To compare results of CMV DNA VL assays reporting in IU/mL and

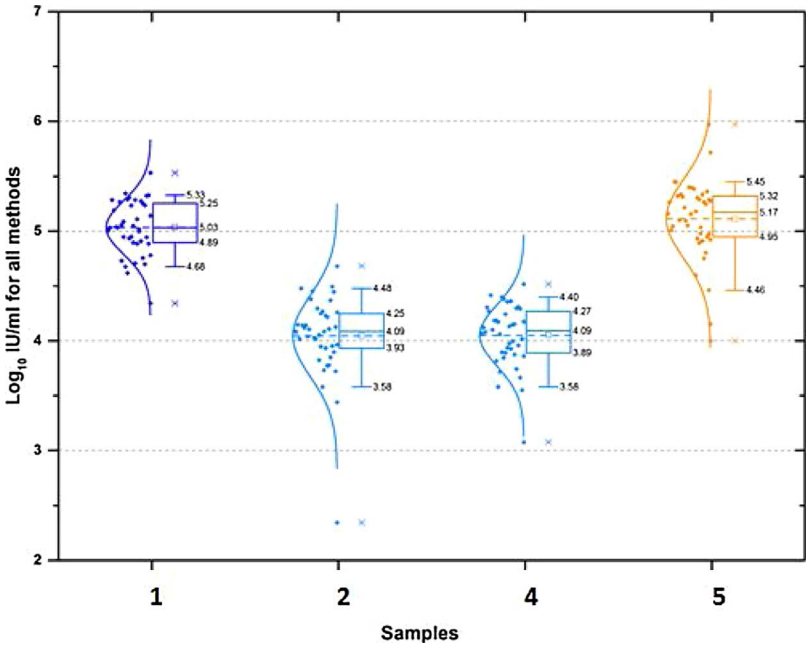


Fig. 2. Accumulated results for quantitative CMV DNA testing for all four CMV DNA positive samples (n = 42 for each sample) from the participants of three EQA providers, INSTAND, NRL and UKNEQAS: results of all applied methods expressed in Log_{10} IU/ml.

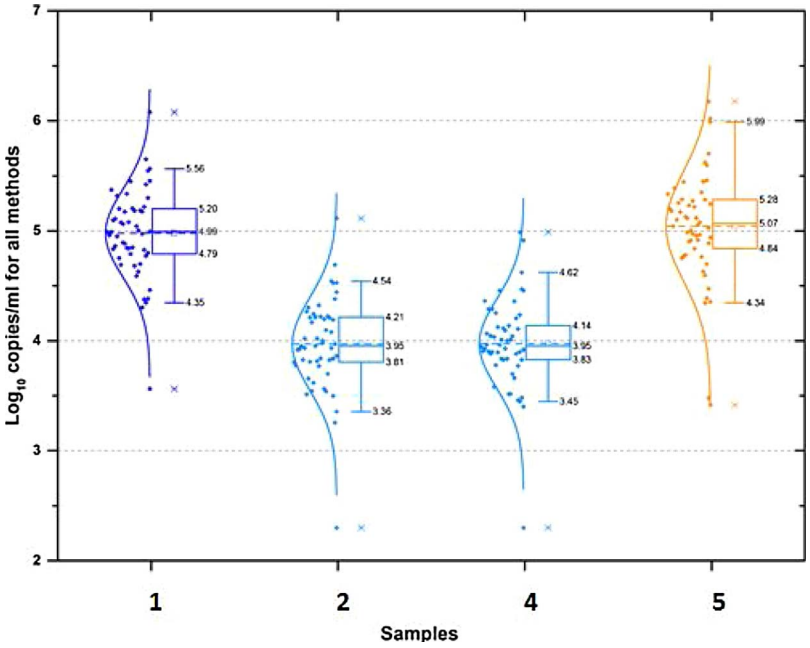


Fig. 3. Selected results for quantitative CMV DNA for all four CMV positive samples (n = 54 for samples 1, 4 and 5; 53 for sample 2) from the participants of EQA provider INSTAND: results of all applied methods in Log_{10} copies/ml.

c/mL to determine equivalence and to assess the rate of adoption of reporting CMV DNA VL results in IU/mL.

3. Study design

3.1. EQA providers

Three EQA providers participated in the project; INSTAND, NRL and UKNEQAS. All three organisations are not-for-profit, scientific organisations providing a range of interdisciplinary EQA programmes to hundreds of laboratories each, with the aim of monitoring the performance of the participating laboratories and the assays used. The three organisations routinely provide CMV VL EQA to participants globally, with participants ranging across Europe, Asia Pacific and elsewhere.

3.2. Samples

Five samples were provided to the participants of each of the collaborating EQA providers. Sample one was a dilution of CMV culture supernatant (AD169 strain) at a viral load of 5 log_{10} copies/mL, diluted in normal human plasma (NHP) negative for CMV DNA and CMV-specific antibodies. Samples two and four were duplicate ten-fold dilutions of sample one (4 log_{10} copies/mL) also diluted in NHP. Sample three was NHP tested negative for CMV DNA. Sample five was a dilution of a clinical sample, obtained from infected amniotic fluid having an extremely high VL, diluted in negative NHP to a concentration of 5 log_{10} copies/mL. Post production testing was conducted by UKNEQAS to determine the accuracy of the sample production prior to distribution. The collaboration did not use the WHO international standard as the sample source for several reasons. Primarily, the viral load of the standard is 5 × 6 log_{10} IU/mL, therefore many vials would have been

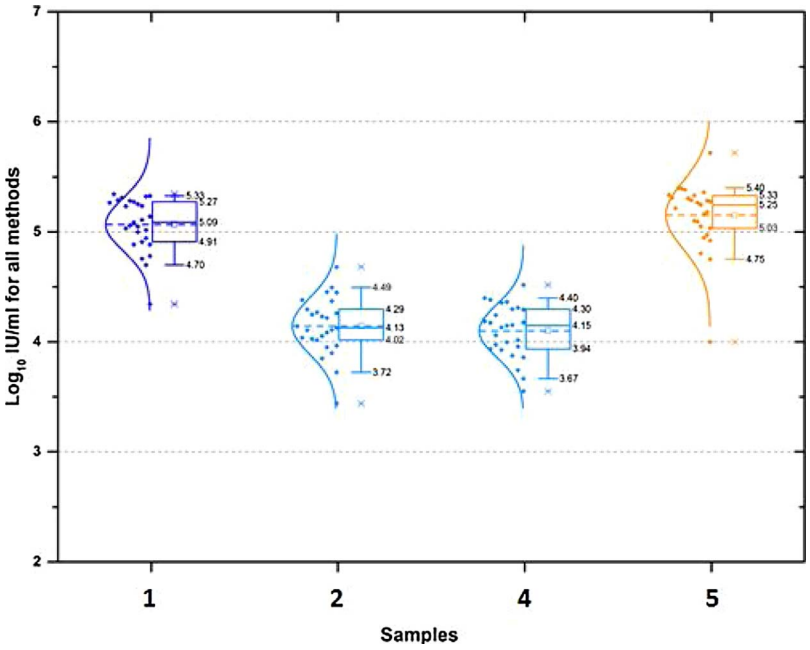


Fig. 4. Selected results for quantitative CMV DNA for all four CMV positive samples (n = 27 for all samples) from the participants of EQA provider INSTAND: results of all applied methods in Log_{10} IU/ml.

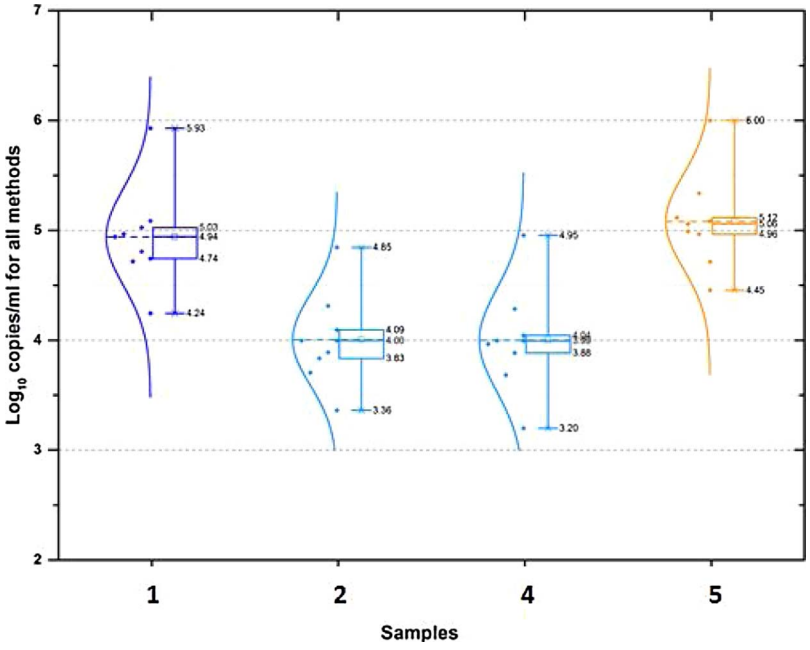


Fig. 5. Selected results for quantitative CMV DNA for all four CMV positive samples (n = 9 for all samples) from the participants of NRL: results of all applied methods in Log_{10} copies/ml.

required to prepare all the samples. Also, it is not encouraged to use WHO standards in this way. The standards are a primary reference material and researchers are expected to use the standard to calibrate a secondary standard for use in research.

Sample vials were prepared and lyophilised by UKNQAS prior to shipping appropriate numbers of panels to the other providers. In addition all samples were pre-tested by four INSTAND expert laboratories before distribution of the panels to the participants. Each EQA provider distributed panels to their participants at ambient temperature. The participants were given written instructions on how to test the samples and report the test results and associated information regarding testing protocols to their EQA provider. The associated information included the testing system used, detailing extraction, amplification and detection test processes, the reagent lot numbers and the unit of measure of the result.

All information was sent to INSTAND for compilation and analysis. All data were stratified by EQA provider and by unit of measure (e.g c/mL or IU/mL). The median, mean, standard deviation (SD) and other summary statistics were estimated for each category. The results obtained from the participants, expressed as c/mL and as IU/mL, were analysed by plotting distribution and box and whisker graphs for each positive sample. The test results obtained by each EQA provider were analysed individually and as combined data sets. For replicate samples two and four, the results submitted by the participants, expressed as c/mL or IU/mL, were plotted against each other and identified by test method (assay). The mean of data stratified by EQA provider and by assay were compared using the Tukey-Kramer HSD comparison test as were results reported for duplicate samples 2 and 4 and samples 1 and 5, which had the same target VL but had different sources of CMV.

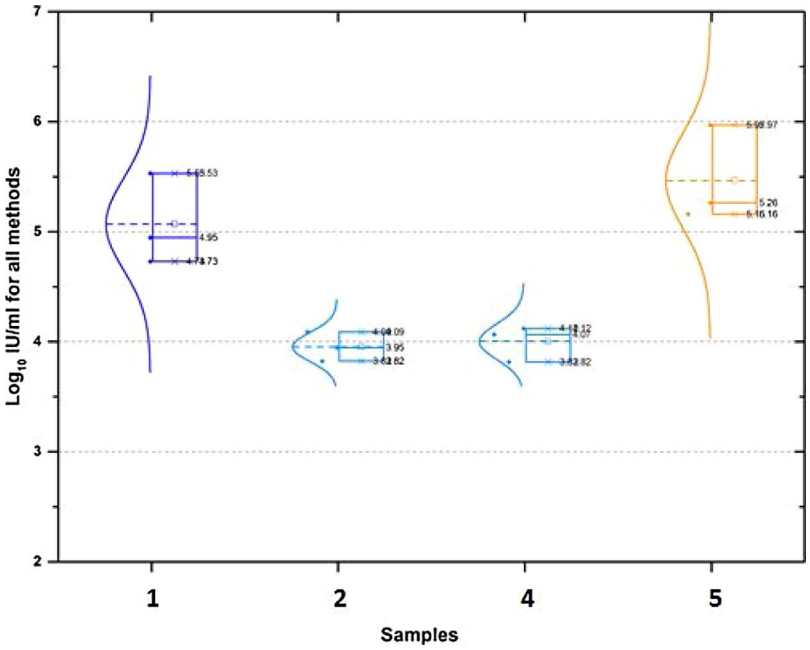


Fig. 6. Selected results for quantitative CMV DNA for all four CMV positive samples (n = 3 for all samples) from the participants of NRL: results of all applied methods in Log10 IU/ml.

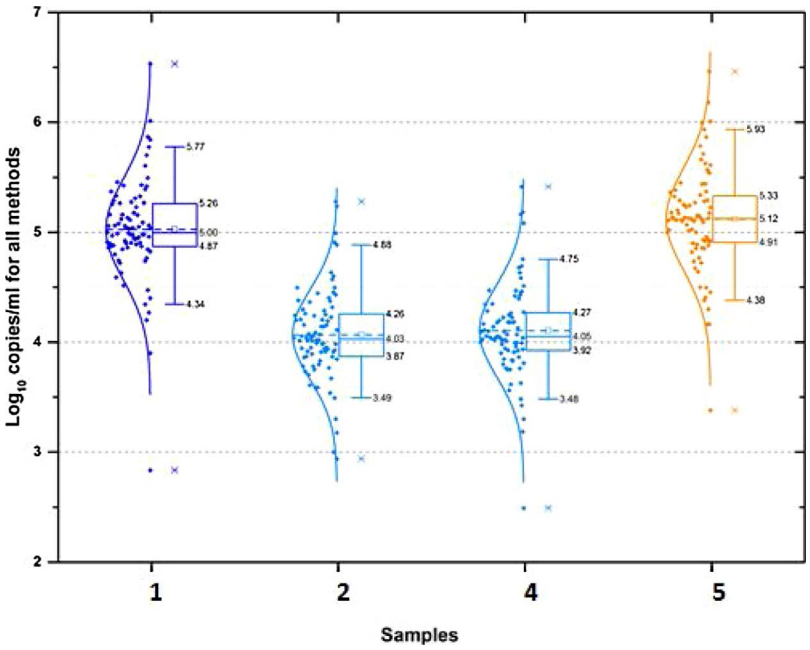


Fig. 7. Selected results for quantitative CMV DNA for all four CMV positive samples (n = 84 for all samples) from the participants of UKNEQAS: results of all applied methods in Log10 copies/ml.

4. Results

There were 96 UKNEQAS participants reporting 96 sets of quantitative results; 12 reported results in IU/mL and 84 reported results as c/mL. INSTAND received results from 91 participants, reporting 81 datasets of quantitative results; 27 as IU/mL and 54 as c/mL; in addition 10 INSTAND participants reported qualitative data only. NRL had 12 participants reporting 12 quantitative data sets; 3 IU/mL and 9 c/mL. Overall only 22% of the participants reported CMV VL expressed in IU/mL.

Only two participants reported detecting CMV DNA in the CMV negative sample 3. One participant, testing with the Abbott Real-time CMV DNA assay, reported sample 3 having CMV DNA at a concentration of 1.46 $\log_{10}$ IU/mL and another participant, using the Argene CMV DNA assay reported detected CMV DNA at a concentration of 2.68 $\log_{10}$ IU/mL.

The results for each of the four positive samples were combined and further analysed (Table 1). The ratio of the mean of total results reported in different units (i.e. mean IU/mL divided by mean c/mL) for the four positive samples, irrespective of the assay used or EQA provider, ranged from 0.9999 to 1.0059; effective demonstrating equivalence. The results reported for each positive sample, as c/mL and IU/mL, were graphed using a distribution and box and whisker graphs (Figs. 1–8).

Sample 1 had a target concentration of 5 $\log_{10}$ c/mL. The mean and median of 147 participants test results reported for sample 1 was 5.00 and 4.99 $\log_{10}$ c/mL respectively. The range of results reported in c/mL was 3.7 $\log_{10}$ c/mL. This was compared with the results for sample 1 obtained from the 41 participants reporting in IU/mL, having both a mean and median of 5.03 $\log_{10}$ IU/mL with a smaller range of 1.19 $\log_{10}$ IU/mL.

Samples 2 and 4 were duplicate samples, made using a 10-fold

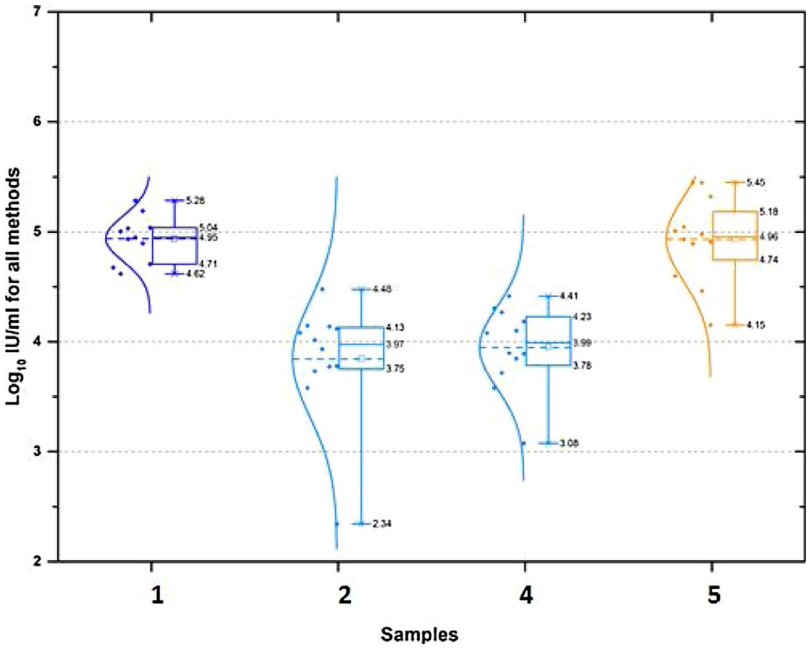


Fig. 8. Selected results for quantitative CMV DNA for all four CMV positive samples (n = 11 for samples 1 and 4, 12 for samples 2 and 5) from the participants of UKNEQAS: results of all applied methods in Log10 IU/ml.

Table 2
Summary of the quantitative results of testing four samples positive for human cytomegalovirus DNA on different commercial and in-house assays.

Assay	Number of results	Average [SD*] of result (log_{10} copies/mL)			
		Sample 1	Sample 2	Sample 4	Sample 5
In house assays	42	5.1 [0.6]	4.1 [0.5]	4.1 [0.5]	5.1 [0.6]
artus CMV RG PCR Kit	36	4.9 [0.3]	4.0 [0.4]	4.0 [0.4]	5.0 [0.4]
COBAS AmpliPrep/COBAS TaqMan CMV Test	16	4.8 [0.3]	3.8 [0.2]	3.9 [0.2]	4.8 [0.3]
Nanogen AD: R-T Alert	13	5.1 [0.3]	4.2 [0.3	4.2 [0.3]	5.2 [0.3]
ARGENE real-time CMV DNA PCR	11	5.5 [0.2]	4.5 [0.4]	4.5 [0.3]	5.6 [0.3]
Other manufacturer	7	4.9	3.8	3.8	5.0
Abbott RealTi*m*e CMV	6	4.9	3.9	4.0	5.1
ELITech CMV ELITe MGB® Kit	5	5.1	4.1	3.9	5.1
Priv. Inst. Immunol. Molekulargen. Amplignost CMV	5	4.7	3.9	3.8	4.9
GeneProof Cytomegalovirus (CMV) PCR Kit	2	4.6	3.6	4.2	4.2
RealStar® CMV PCR	1	5.1	3.9	4.0	5.3
Focus Diagnostics Simplexa CMV Kit	1	4.4	3.5	3.6	4.5
Mikrogen CMV DNA Real Time PCR	1	5.1	4.3	4.1	5.2
Roche LightCycler CMV Quant kit	1	5.6	4.6	4.7	5.8
Combined results	147	5.0	4.0	4.0	5.1

Assay	Number of results	Average [SD*] of result (log_{10} IU/mL)			
		Sample 1	Sample 2	Sample 4	Sample 5
RealStar CMV PCR	17	5.1 [0.3]	4.2 [0.3]	4.2 [0.2]	5.2[0.3]
In house assays	11	5.0 [0.2]	4.1 [0.2]	4.1 [0.2]	5.1 [0.3]
COBAS AmpliPrep/COBAS TaqMan CMV Test	4	4.8	3.8	3.8	5.0
Abbott RealTi*m*e CMV	3	5.1	4.1	4.1	5.3
ARGENE® real-time CMV DNA PCR	2	4.8	3.2	3.5	4.6
Other manufacturer	2	5.0	4.0	4.1	5.2
artus CMV RG PCR Kit	2	5.1	4.3	4.0	5.1
Focus Diagnostics Simplexa CMV Kit	1	5.5	4.1	4.1	5.2
Combined results	42	5.0	4.0	4.0	5.1

*SD Standard deviation (datasets greater than 10).

dilution of sample 1, having a target VL of 4 log_{10} c/mL. The mean of results reported in c/mL for sample 2 (n = 146) and sample 4 (n = 147) was 4.03 and 4.05 log_{10} c/mL respectively. The mean of results reported in IU/mL for the same samples (n = 42 for both samples) was 4.04 and 4.05 respectively. Using Tukey-Kramer HSD comparison test, there was no significant difference is test results reported in c/mL or IU/mL for sample 2 compared with sample 4 (p = 0.98, p = 0.99 respectively) indicating excellent reproducibility of assays.

Sample 5 was included as a clinical sample known to contain high levels of CMV DNA, in part to determine if the quantification of a clinical strain differs from that of derived laboratory-adapted strain. The mean of all results for sample 5 reported in c/mL (n = 147) was 5.09 log_{10} c/mL with the results ranging from 3.38 to 6.46 log_{10} c/mL. The mean of results reported as IU/mL (n = 42) was 5.11 log_{10} IU/mL with a range from 4.00 to 5.97 log_{10} IU/mL. There was no significant difference in test results reported in c/mL or IU/mL for samples 1 or 5

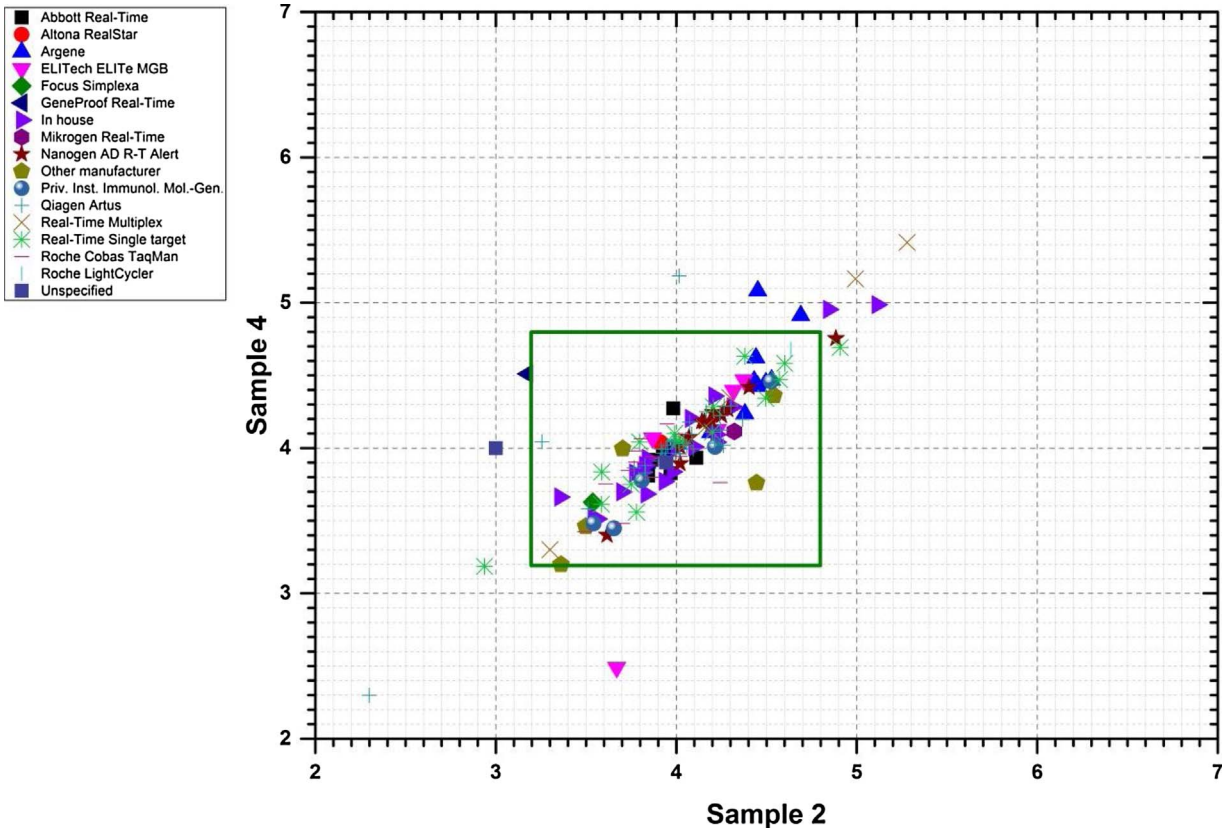


Fig. 9. Results for the duplicate CMV DNA positive samples 2 and 4 (n = 146) expressed in Log_{10} copies/ml summarized using a Youden plot. The results of sample 2 are in the X axis and the results of sample 4 on the Y axis. The methods used to produce the result. s are coded using colors and symbols.

using Tukey-Kramer HSD comparison test (p = 0.34, p = 0.67 respectively).

A total of 12 different commercial assays were used by the participants; the results for the four positive samples summarised in Table 2. Some participants reported using in-house assays or did not specify the manufacturer of the assay used. All commercial assays reported the results in c/mL and six of the 12 assays also reported in IU/mL. Of note, the greatest number of results expressed in c/mL was obtained from participants testing on the Qiagen ARTUS CMV DNA assay (n = 36; 24.5%), whereas only two of the 42 (4.8%) participants reporting in IU/mL used the Qiagen assay. A total of 42 participants reporting in c/mL (28.6%) and 11 of 42 participants reporting in IU/mL (26.2%) used an in-house assay. The Altona Realstar CMV DNA assay was the most frequently used assay where participants reported in IU/mL. Using the Tukey-Kramer HSD comparison test, a significant difference in test results reported in c/mL for the ARGENE® real-time CMV DNA PCR compared with those reported by the GeneProof Cytomegalovirus (CMV) PCR Kit (p = 0.027), AmpliGnost CMV PCR Kit (Priv. Inst. fuer Immunologie und Molekulargenetik, Karlsruhe, Germany) (p = 0.011), QIAGEN artus® CMV RG PCR Kit (p = 0.0002), COBAS® AmpliPrep/COBAS TaqMan® CMV Test (p ≦ 0.0001), in-house (p = 0.01) and non-specified assays (p = 0.003). Notably, no difference was found in the test results by different assays when results were reported in IU/mL.

Youden plots were graphed comparing the mean result of duplicate samples 2 and 4 when results were reported in c/mL (Fig. 9) and IU/mL (Fig. 10). A square, representing 0.80 log_{10} c/mL of the median of all results reported in c/mL for these two samples was overlayed on the Youden plot. Of the 146 participants, reporting results in c/mL for both samples, 16 (11,0%), obtained from seven different assays, reported results outside the square. In contrast, only one of the 42 participants reporting results in IU/mL reported a result outside the 0.8 log_{10} IU/mL square.

No significant difference between results reported in c/mL or IU/mL by each EQA provider participant was found using Tukey-Kramer HSD comparison test.

5. Conclusion

CMV DNA quantification is to guide the initiation and monitoring of anti-viral therapy and to indicate the risk of clinical relapse or the emergence of viral drug resistance (Babady et al., 2015; Kraft et al., 2012; Hayden et al., 2012). To use CMV VL testing for these purposes, universal VL thresh-holds need to be established and accurately defined and the assays used need to be standardised (Razonable and Hayden, 2013; Razonable et al., 2013). Several authors reported a lack of standardisation of CMV DNA VL test results prior to 2010 (Razonable and Hayden, 2013; Hayden et al., 2012; Pang et al., 2009). Therefore clinicians had resorted to laboratory-specific treatment and monitoring thresholds (Pang et al., 2009). The introduction of a WHO international standard for CMV DNA was expected to aid in the standardisation of assays (Kraft et al., 2012). Five years after the release of the WHO standard, three EQA providers distributed the same five samples to their participants and analysed the results to determine the rate of adoption of IU/mL and compare the results reported in c/mL and IU/mL.

The introduction of an international standard has resulted in the standardisation of test results reported by different manufacturers of HCV RNA, HBV DNA and other viral load assays (Hayden et al., 2012; Fryer and Minor, 2009; Hayden et al., 2008). The 1 st WHO International Standard for Human Cytomegalovirus for Nucleic Acid Amplification Techniques (Code 09/162), released in 2010 was developed using the Human CMV Merlin strain and has concentration of 5,000,000 IU/vial (Freyer et al., 2010). The mean potency estimate for the candidate standard in the original study was 5 × 106 (6.7 log10) c/

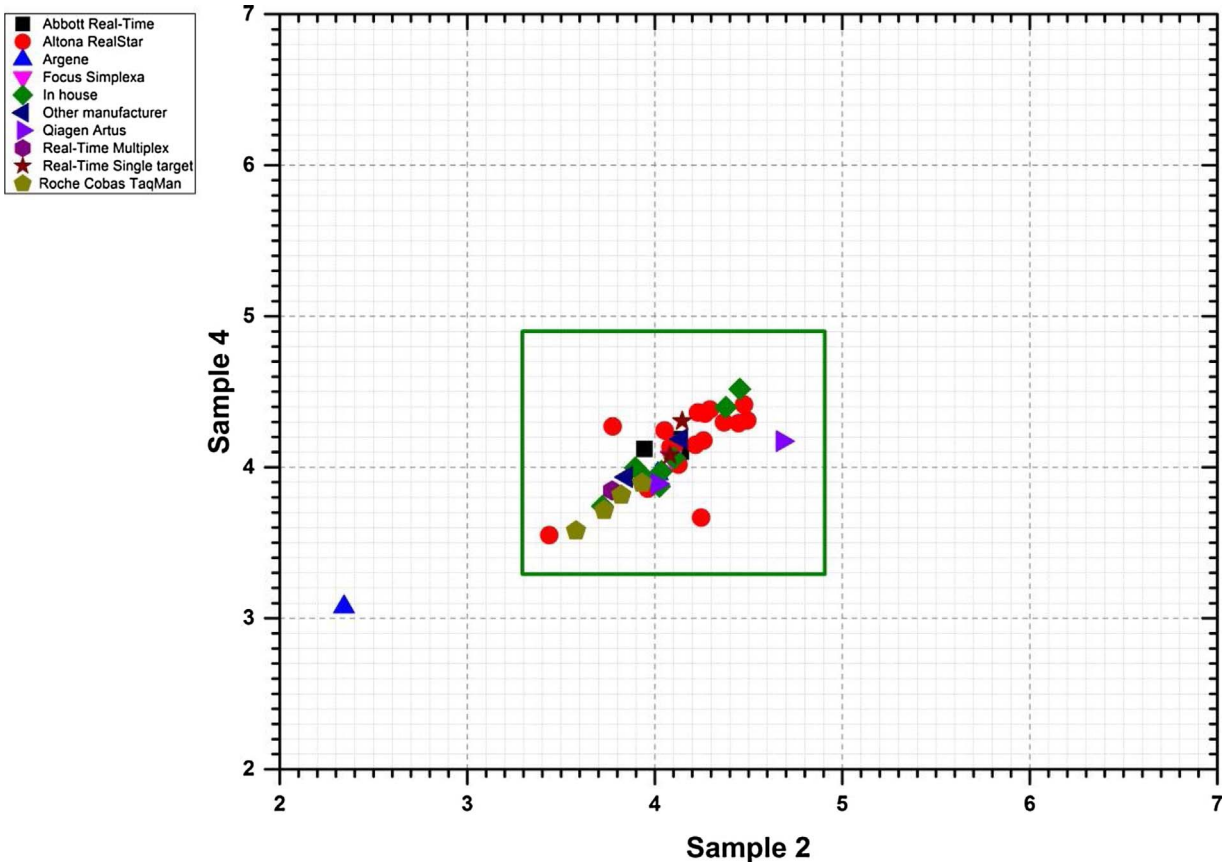


Fig. 10. Results for the duplicate CMV DNA positive samples 2 and 4 (n = 42) expressed in Log_{10} IU/ml summarized using a Youden plot. The results of sample 2 are in the X axis and the results of sample 4 on the Y axis. The methods used to produce the results are coded using colors and symbols.

mL. Therefore, it is expected that the results reported in IU/mL and c/mL should be equivalent.

Many factors potentially contribute to a lack of standardisation of NATs. Each assay may differ in position, length and number of targets sites, the probes and chemistries used, the extraction methods and its efficiencies, and sample types and volumes (Kraft et al., 2012; Hayden et al., 2012). CMV nucleic acid is usually freely circulating and highly fragmented (Preiksaitis et al., 2016). Therefore, assays that target large amplicons may not efficiently detect smaller fragments. Also the commutability of the WHO standard compared with patient samples differs from assay to assay (Hayden et al., 2015). This may be due to the fact that the WHO standard is cell culture derived whole virus, rather than fragmented circulating free DNA (Preiksaitis et al., 2016). To determine if there was a detectable difference in the results reported for cell-culture derived and clinical samples, CMV from amniotic fluid was used in sample 5. However, no assessment of the amount of viral fragmentation was assessed. It is possible that little or no fragmentation is found in this sample type. Providers of EQA should consider viral fragmentation in the selection of samples for CMV viral load programmes.

Prior to the introduction of the WHO standard, a multi-centre study compare the results of different CMV DNA VL assays. The mean of test results from different assays testing the same sample were similar but the range of results was large, sometimes up to 4.3 log_{10} c/mL (Pang et al., 2009). A study comparing the results of dilutions of the WHO international standard reported a range of variance was 0.69–1.35 IU/mL; mean difference being 0.94 IU/mL (Preiksaitis et al., 2016). The greatest variation was found in samples with lower VL (Pang et al., 2009; Hirsch et al., 2013). In the current study, the variation in results reported in c/mL was greater than those reported in IU/mL. For duplicate samples 2 and 4 (4 log_{10} c/mL), the range of test results reported in c/mL and IU/mL were 2.98 and 3.12 c/mL and the 2.34 and 1.44 IU/mL respectively. For samples 1 and 5 (5 log_{10} c/mL), the ranges were 3.70 and 3.08 c/mL and 1.19 and 1.97 IU/mL respectively. The Youden plots had fewer results fell outside the 0.8 log10 IU/mL square, indicating greater standardisation when reporting in IU/mL. Comparing the results for each individual sample tested on each assay, no significant difference was determined between the results reported in IU/mL, whereas there were differences between assays reporting in c/mL. This may be in part due to a greater number of participants reporting in c/mL, and therefore a greater possibility of laboratory to laboratory variation. Although the variation in results reported in IU/mL was less than those reported in c/mL, the range was still unacceptably large.

Notably, the mean of the results reported by each individual assay, and the mean of the combined results, were very close to the target VL (Table 2). The mean of the combined results for all four samples were within one decimal place of the target result, whether reported in c/mL or IU/mL. Generally, the ARGENE® real-time CMV DNA PCR and the Roche LightCycler CMV Quant kits overestimated and the GeneProof Cytomegalovirus (CMV) PCR Kit and Focus Diagnostics Simplexa™ CMV Kits underestimated the viral loads when reporting in c/mL.

There was little difference in the results expressed in c/mL and IU/mL, the units being effectively equivalent in the current study. This differs somewhat from a previous report that found the conversion factor was 1.16 when the WHO standard was diluted in PBS and tested on the Abbott RealTi*me* CMV DNA assay. The manufacture of that assay recommends a conversion factor of 1.56 copies/IU. The Roche COBAS AmpliPrep/COBAS TaqMan CMV test package insert recommends conversion factors of 1.1 copies/IU for that assay. However, for the samples included in the current study, the mean of the results, from all assays, reported in c/mL were equivalent to the mean of results reported in IU/mL.

The introduction of the WHO international standard for CMV DNA

is expected to reduce the variation in test results obtained from different assays. Our data demonstrates that not all participating laboratories using commercial assays report results in IU/mL even when the assay is calibrated against the WHO standard. The majority of participants reported results in c/mL, even though the variation in results reported a c/mL is greater. The results reported in this study demonstrates that variation in tests quantifying CMV DNA is still a limitation in setting thresholds for implementation and monitoring of anti-viral therapy.

Acknowledgements

We would like to thank the following INSTAND expert laboratories for pre-testing the CMV panels before distribution of the panels to the EQA participants: Dr Oliver T Keppler Dr Holger F Rabenau, Dr Annemarie Berger, Dr Martin Stürmer, Institute of Medical Virology, University Hospital of Frankfurt, Frankfurt, Germany; Dr Martin Enders, Dr Gunnar Schalasta, Laboratory Enders & Partners, Stuttgart, Germany; Dr Thomas Mertens, Dr Detlef Michel, Institute of Virology, University Hospital Ulm, Ulm, Germany; Dr Gerhard Jahn, Dr Klaus Hamprecht, Institute for Medical Virology and Epidemiology of Viral Diseases, University Hospital Tübingen, Tübingen, Germany.

References

Babady, N.E., Cheng, C., Cumberbatch, E., Stiles, J., Papanicolaou, G., Tang, Y.W., 2015. Monitoring of cytomegalovirus viral loads by two molecular assays in whole-blood and plasma samples from hematopoietic stem cell transplant recipients. J. Clin. Microbiol. 53 (4), 1252–1257.

Freyer, J.F., Heath, A.B., Anderson, R., Minnor, P.D., 2010. Collaborative Study to Evaluate the Proposed 1st [first] WHO International Standard for Human Cytomegalovirus (HCMV) for Nucleic Acid Amplification (NAT)-based Assays, vol. WHO/BS/10.2138 World Heath Organization, Geneva, Switzerland.

Fryer, J.F., Minor, P.D., 2009. Standardisation of nucleic acid amplification assays used in clinical diagnostics: a report of the first meeting of the SoGAT Clinical Diagnostics Working Group. J. Clin. Virol. 44 (2), 103–105.

Hayden, R.T., Hokanson, K.M., Pounds, S.B., et al., 2008. Multicenter comparison of different real-time PCR assays for quantitative detection of Epstein-Barr virus. J. Clin. Microbiol. 46 (1), 157–163.

Hayden, R.T., Yan, X., Wick, M.T., et al., 2012. Factors contributing to variability of quantitative viral PCR results in proficiency testing samples: a multivariate analysis. J. Clin. Microbiol. 50 (2), 337–345.

Hayden, R.T., Preiksaitis, J., Tong, Y., et al., 2015. Commutability of the first World Health Organization international standard for human cytomegalovirus. J. Clin. Microbiol. 53 (10), 3325–3333.

Hirsch, H.H., Lautenschlager, I., Pinsky, B.A., et al., 2013. An international multicenter performance analysis of cytomegalovirus load tests. Clin. Infect. Dis. 56 (3), 367–373.

Kraft, C.S., Armstrong, W.S., Caliendo, A.M., 2012. Interpreting quantitative cytomegalovirus DNA testing: understanding the laboratory perspective. Clin. Infect. Dis. 54 (12), 1793–1797.

Pang, X.L., Fox, J.D., Fenton, J.M., et al., 2009. Interlaboratory comparison of cytomegalovirus viral load assays. Am. J. Transplant. 9 (2), 258–268.

Preiksaitis, J.K., Hayden, R.T., Tong, Y., et al., 2016. Are we there yet? Impact of the first international standard for cytomegalovirus DNA on the harmonization of results reported on plasma samples. Clin. Infect. Dis. 63 (5), 583–589.

Ramanan, P., Razonable, R.R., 2013. Cytomegalovirus infections in solid organ transplantation: a review. Infect. Chemother. 45 (3), 260–271.

Razonable, R.R., Hayden, R.T., 2013. Clinical utility of viral load in management of cytomegalovirus infection after solid organ transplantation. Clin. Microbiol. Rev. 26 (4), 703–727.

Razonable, R.R., Asberg, A., Rollag, H., et al., 2013. Virologic suppression measured by a cytomegalovirus (CMV) DNA test calibrated to the World Health Organization international standard is predictive of CMV disease resolution in transplant recipients. Clin. Infect. Dis. 56 (11), 1546–1553.

Ross, S.A., Novak, Z., Pati, S., Boppana, S.B., 2011. Overview of the diagnosis of cytomegalovirus infection. Infect. Disord. Drug Targets 11 (5), 466–474.

Chapter Three - Conclusion

The mean of the results obtained from all assays that reported in copies/mL were equivalent to the mean of results reported in IU/mL, indicating equivalency of the units. The study concluded "*Our data demonstrates that not all participating laboratories using commercial assays report results in IU/mL even when the assay is calibrated against the WHO standard. The majority of participants reported results in c/mL, even though the variation in results reported a c/mL is greate*r". Although CMV viral load testing is affected by circulating DNA fragments, this study demonstrated that quantitative results reported between test system were comparable, irrespective whether they reported in IU/mL or copies/mL.

Quantitative molecular assays experience many different sources of variation both within and between test systems. They are designed to detect different targets, have different chemistries, method of nucleic acid extraction and amplification efficiencies. Both the extraction and amplification efficiencies of the methods used can influence the amount of target nucleic acid as well as "*differences in primer/probe sequences, alternate DNA/RNA extraction methodologies, PCR instrumentation, the source of calibrators that facilitate quantification, and even the fragmentation of circulating nucleic acid for a particular pathogen and specimen type*" (148). For CMV and other viruses, circulation of fragmented DNA is another separate source of variation (149). Without a point of reference, the results from each assay are less likely to be comparable, as indicated by this study. The use of an international standard provides that point of reference for the manufacturers to optimise their assays.

This study highlighted a reluctance of some manufacturers to calibrate their assays using the international standard, and of some laboratories to report in IU/mL even when the assay being used was calibrated. The study found that results reported in IU/mL from difference assays did not have statistically significant differences whereas those reporting in copies/mL were statistically significant, indicating that the use of the standard was effective harmonising viral load results across assays. However, the results between assays reporting in IU/mL were still too wide to allow assay-to-assay result comparisons for clinical use (150, 151). Several studies highlighted a difference in the reporting of results of CMV DNA viral load leading to different quantitative results or clinical outcomes (152, 153). However, others suggest that the progress towards standardisation of CMV viral load has enables clinicians to utilise the quantitative results to establish thresholds for patient management if the results are expressed in IU/mL and the commutability of results are established (149).

Errata: Results of Cytomegalovirus DNA Viral Loads Expressed in Copies per Millilitre and International Units per Millilitre are Equivalent.

- Legend for Figure 9 should read: "The methods used to produce the results are coded using colors and symbols."
- Last sentence of page 53 should read "The mean potency estimate for the candidate standard in the original study was 5×10^6 (6.7 log10) c/mL"

Chapter Four: Quality control of infectious disease testing

NRL has provided a QC program for infectious disease testing for decades. In 2001, the first internet-based QC monitoring software globally (EDCNet™ - www.nrlquality.org.au/qconnect) was developed and deployed (122). It was re-written in 2014 by Ashvins Group (Miami, Fl.). As senior scientist and project manager, I contributed significantly to the design specifications of both the original and subsequent release of the software, project managed the software development and coordinated the validation of the system. Briefly, this peer-comparison software allows for the data entry of QC results (both from manufacturer-provided kit controls and third-party external controls) and associated metadata from participating laboratories into a single database. QC results from laboratories testing the same QC sample on the same assay (peer group) can be compared using a range of graphical and tabular reports. Metadata, such as instrument identification, operator details, reagent and QC lot numbers, are associated with each QC result as line-listed data, allowing the results to be analysed, reviewed and/or filtered, based on the user's specific selections (122, 124). Unlike other internet-based QC monitoring systems that were subsequently released by other providers, EDCNet™ is designed predominantly for infectious disease serology and NAT (122, 154). It is now used globally and has been instrumental in the detection of several, significant adverse IVD performance issues (123, 134). Data have also been used to determine the uncertainty of measurement of serology assays, a requirement for ISO 15189 accredited laboratories (155).

From reviewing QC data submitted into EDCNet™, it became obvious the results of QC testing for infectious diseases did not follow the rules universally applied to clinical chemistry, including the Westgard rules (48-50) described in detail in **Chapter 1**. Indeed, this situation was found to cause difficulty for infectious disease testing laboratories trying to comply with traditional QC guidelines (51-53, 127, 156). Therefore, with the support of colleagues and a biostatistician, I developed and published an alternative to Westgard rules (125). Using data accumulated by the NRL QC program, I led the development of a novel method to establish QC acceptance criteria, called QConnect™. QConnect™ uses historical QC data to determine acceptance limits for QC testing for infectious diseases and is based on several principles. QC samples using the QConnect™ principles are optimised for each assay. As highlighted in the **Chapter 1**, an infectious disease serology measurand is assay specific. Therefore, a patient or QC sample will elicit differing quantitative test results when tested on different assays. To develop a QConnect™ QC sample, a series of dilutions of a stock material containing the analyte to be monitored is tested on each assay to obtain low-level reactivity on the linear part of each assay's dose response curve, where the variation inherent in the test system will be greatest. QCs that are too high or too low will be on the plateau of the dose response curve and not be sensitive to change. Once the dilution is selected, each new lot of the assay-specific QC sample is manufactured using the same stock material at the same concentration. In this way, the same low-level reactivity is obtained from each lot of QC sample, allowing for continuous monitoring of the assay's performance over time. The QC results are entered into EDCNet™ by participating laboratories (122, 124). Using these historical data, a method for establishing acceptance limits, calculated for each QC/assay combination, based on many thousands of data points (rather than 20 to 30 data point used by traditional QC methods) was developed (125) and the publication presented in **Chapter 4** of this thesis. Any results outside these QConnect™ limits are unexpected and should trigger a root cause investigation. The introduction of QConnect™ represents the first alternative to traditional approaches to establishing acceptance criteria for QC test results from both serology and NAT infectious disease testing.

The traditional QC approaches are supported by professional body guidelines (51-54, 127) and, as such, laboratory technical auditors expect them to be used. To assess the respective utility of different approaches to establishing QC acceptance limits, I led a study which reviewed the number of QC rejection when 103 QC data sets were subjected to different QC methods, including QConnect™, Westgard rules and the German RiliBÄK rules (157). The data used were collected from clinical or blood screening laboratories that participated in the NRL QConnect™ program and entered data into EDCNet™ for the calendar year of 2015. In total, 21,510 QC test results, tested in five different countries, from 14 different assays testing for six different analytes commonly used in infectious disease serology testing laboratories were analysed. All results were indicated as being "valid" by the contributing laboratory. A "valid" result was a result of a QConnect™ sample tested in an assay meeting the manufacturer's release criteria. Therefore, the numbers of failures were expected to be low. Where there were more than 10 laboratories using the same test and QC combination, ten datasets used were randomly selected for analysis. Analysis of results for all analytes, submitted by all laboratories, indicated that the Westgard and RiliBÄK rules reported numerous "rejections", with many of the 103 data sets having more than 20% of QC results failing the respective rules. It was concluded that these QC rules were not fit for purpose for the assays being monitored. The consequence of laboratories using these rules would be unnecessary investigations into "valid" QC results. In contrast, QConnect™ limits reported only two of the 103 datasets as having more than 20% rejections. Of these two datasets, both demonstrated obvious unacceptable variation on review. These examples are described in detail on the paper presented in **Chapter 4**.

This study is important as it utilises real QC data obtained from multiple laboratories testing multiple analytes. Therefore, it compares the QC methodologies using "real-world" data, rather than theoretical assumptions commonly employed in QC discussions. QConnect™, was found to be the most appropriate method of establishing QC acceptance limits for serology testing. The use of historical QC data to set limits accounted for the lot-to-lot variation that is inherent in serology testing. This variation is ignored by Westgard and RiliBÄK rules, accounting for the unacceptable level of "false rejections". The study validated the approach designed and implemented at NRL, under my guidance. QConnect™ is the first, and currently the only, published, validated QC method fit-for-purpose for serology testing.

Wayne Dimech*, Giuseppe Vincini and Marina Karakaltsas

Determination of quality control limits for serological infectious disease testing using historical data

DOI 10.1515/cclm-2014-0546
Received May 23, 2014; accepted July 22, 2014; previously published online August 12, 2014

Abstract

Background: An effective quality control (QC) program requires the establishment of control limits within which the results of the QC sample is expected to fall. Traditionally, the mean plus/minus two standard deviations calculated for a set of QC sample results is used to establish control limits. Allowable total error (TEa) and Westgard rules aid in interpreting QC sample results. Westgard rules assume QC sample results are normally distributed and TEa assumes commutability between the QC sample and patient results. None of these paradigms apply to infectious disease testing.

Methods: Results from the NRL's QC program were extracted and sorted into assay/QC lot number-specific data. Control limits for selected QC samples used to monitor 64 commonly used serological assays were calculated and validated using the within- and between-QC lot variance of data from each of the assay/QC combinations.

Results: No assay/QC combination had more than 10% of results less than the lower control limit or greater than the upper control limit. Of the 423 assay/QC lot combinations, 14 (3.3%) had more than 5% of results less than the lower limit and 48 (11.3%) had more than 5% of results greater than the upper limit calculated for that assay/QC combination.

Conclusions: The control limits, established by this novel method, are based on more than a decade of QC test results from >300 laboratories from 30 countries and provides users of the NRL QC program evidence-based control limits that can be applied in isolation or in conjunction with more traditional methods for establishing control limits.

***Corresponding author: Wayne Dimech,** NRL, 4th Floor Healy Building, 41 Victoria Parade, Fitzroy, Victoria, Australia,
Phone: +61 3 94181132, Fax: +61 3 94181155,
E-mail: wayne@nrl.gov.au
Giuseppe Vincini and Marina Karakaltsas: NRL, Fitzroy, Victoria, Australia

Keywords: control limits; infectious diseases; QConnect; quality control; serology.

Introduction

In infectious disease serology, the use of a quality control (QC) sample, independent of the assay manufacturer's kit controls is a requirement in some jurisdictions [1–3] and highly recommended by others [4–6]. Standards and guidance documents direct laboratory staff to develop and maintain a system of monitoring the results of the QC samples [1, 4–10]. In this way, unexpected shifts or trends can be detected and corrective action taken. A critical aspect of control charts is the establishment of lower and upper control limits. QC results outside these limits may indicate unexpected variation in the testing system and trigger a review of test processes.

Variation in QC test results arise from several common sources: changes in assay production, especially the introduction of new assay master lot components, differences in transport and storage of reagents and consumables, instrumentation, divergent operator processes and environmental conditions. The monitoring of results of QC testing over time is sensitive to these changes. Using a Levey-Jennings chart will monitor variation within an individual laboratory and give a graphical representation of the imprecision of the test.

The establishment of control limits in laboratory medicine has traditionally relied on calculating the mean of a set of QC test results and applying a range determined as the mean $\pm x$ standard deviations (SD) of the same QC test results, where x is usually equal to two or three [4, 5, 7, 8, 11, 12]. A consensus agreement of experts suggested quality specifications of a laboratory test system should be determined using a hierarchy of criteria starting with an evaluation of the effect of the analytical performance of the test system on clinical outcomes in specific clinical settings, professional recommendations or goals set by regulatory bodies or external quality assessment schemes [13–15].

If the effect of the change in QC test results is assumed to measure change in the analytical performance of a test system, there must similarly be an assumption that the patient test results are affected proportionally to those of the QC sample due to that change, i.e., commutability. That is, if the QC test results decrease unexpectedly, the results of the patient population tested in the same conditions as the QC sample would be expected to decrease in a proportional manner [14, 16–19]. Although this may occur in clinical chemistry, there is evidence that it is not universal [20] and, to our knowledge, there are no systematic evaluations of this phenomenon for infectious disease serology except that reported by Dimech et al. [21].

Our institution, NRL (Melbourne Australia) has conducted an international QC program (www.nrl.gov.au/qconnect) for laboratories performing infectious disease serology for over a decade [22]. QConnect QC samples (previously known as PeliSpy) are selected to give low-level reactivity on specific assays [21, 22]. Each lot of QC is manufactured using the same stock material at the same concentration, thereby minimising the lot-to-lot variation. Laboratories testing the QC sample enter results and associated information into an internet-based program (EDCNet; NRL, Melbourne, Australia). Results are monitored through a range of tabular and graphical reports and QC test results for laboratories testing the same assay and QC combination (peer group) can be compared [23]. In this way, NRL has collected thousands of QC sample results from different lots of QC for each peer group.

Establishing QC control limits for serological assays is problematic. When a traditional mean $\pm x$ SD is used to establish the control limits for the QC samples used to monitor serological assays, large shifts in reactivity, often seen with the introduction of a new assay lot number, can cause all subsequent QC test results to fail QC rules, even though the assay kit control results are valid as per the manufacturer's instructions for use [12, 24]. Unlike clinical chemistry, it is not possible to correct the bias introduced by the change in reagent lot numbers using a calibrator referenced to an international standard. The QC officer is required to recalculate the control limits, or use the manufacturer's instructions alone to validate the test run.

An alternative method of setting control limits for specific assay/QC combinations has been developed. This manuscript describes an improved approach to establishing control limits that uses all QC test results for each peer group, and incorporates the variance in QC test results arising from within and between each QC lot. The method uses weighting to account for different numbers of test results obtained for different QC lots. As the method incorporates 10 years of QC test result data, it included all expected sources of variation in the test system including variation derived from multiple changes in assay lot numbers, and from many instruments, operators and environmental conditions.

The aim of the study was to develop lower and upper control limits for assay/QC combinations based on data collected over a period of time >10 years.

Materials and methods

QC samples

QC samples for a range of analytes were developed and manufactured by AcroMetrix, a part of Thermo Fisher (Benicia, CA, USA) under ISO 13485 conditions [25]. Generally, the QC samples were manufactured by diluting a stock material, having a high concentration of the required analyte, in a base matrix of normal human, defibrinated, delipidated, plasma. Large numbers of aliquots of the stock material were stored in liquid nitrogen prior to use so that the same stock material could be used over many QC production lots. Different concentrations of the stock material were used to achieve low-level reactivity, usually 2–3 times the assay manufacturer's cut-off value in different assays. QC samples used to monitor serological assays testing for blood-borne viruses (HBsAg and anti-HIV, -HCV, -HBc and -HTLV) were manufactured as multimarker samples, i.e., a single QC sample with optimised reactivity for multiple analytes. Other QC samples, such as those for anti-rubella IgG, syphilis and HIV-1 p24 Ag, were manufactured as single analyte QC samples. Where relevant, the QC sample was calibrated against the WHO International Standard (NIBSC, Potters Bar, UK) using an ISO 17511 traceability claim [26]. The QC sample could therefore be used as a secondary standard [27].

Participating laboratories

More than 300 laboratories from 30 countries have participated in the NRL QC program since January 2001.

Data collection

Participants in the NRL QC program tested the QC samples periodically, usually daily, and entered the results into EDCNet. All QC test results were extracted into MS Excel in January 2013. The QC sample results were sorted into assay/QC sample lot number-specific data (assay/QC lot) sets which were combined to form assay/QC sample-specific data sets (assay/QC). For example, multiple lots of the QC sample QConnect Blue were tested on the Abbott Architect anti-HCV assay. When these data are analysed for individual QC lots (not combined) they are referred to as 'assay/QC lot' data. When combined these assay/QC lot data are referred to as assay/QC data.

Data removal

Data from some assay/QC lot combinations were removed from analysis because the manufacture of these lots did not comply with standard procedure. In addition, assay/QC lot data sets that had fewer than 10 results and assay/QC combinations with a single QC lot were excluded from the calculations.

Data analysis

For each assay/QC lot data set the number of replicates, mean and SD were calculated. An analysis of variance (ANOVA) was used to estimate the within- and between-lot variance components. The total variance was calculated as the sum of the within- and between-lot variances and the total SD (SD_{qc}) was estimated as the square root of the total variance. The overall mean ($\bar{x}_{qc}$) for the assay/QC data set was calculated as the weighted average of the within-lot averages (with weights being dependent upon the within-lot replicate counts). The lower assay/QC limit was $\bar{x}_{qc} - 2SD_{qc}$ and the upper assay/QC-specific limit was $\bar{x}_{qc} + 2SD_{qc}$.

The overall mean was a weighted average of the prior mean (the mean of all previously tested lots) and the mean of the data with the new lot. The weight for the prior mean is 1 over the lot-to-lot variance, so if the lot-to-lot variance is large, the prior mean will get relatively little weight whereas if prior lots were very consistent (small variation), it will take more new data to overpower the historical results. The weight given to the new data increases as n increases (n=number of observations with new lot) and depended on the within-lot variability of that lot of QC.

Control limit validation

The calculation of each assay/QC control limit was based on results obtained from multiple lots of QC. It was possible that the calculated control limits would not be appropriate for some lots of QC samples, especially if lot-to-lot variation of QC samples was too great. Therefore the limits were validated by applying them to the results obtained for each individual assay/QC lot dataset to determine the number and percentage of individual QC sample results that would have fallen outside these limits. Assay/QC lot data sets that had more than 10% of results outside the control limits were noted. The total number of QC sample results falling outside the control limits for any assay/QC combination, irrespective of the QC lot was expected to be <5%.

Results

There were a total of 255 assay/multimarker QC lot combinations included in the study. A total of 11 lots of multimarker QC (4.3%) were excluded from analysis for one or more analytes. One lot of QC was excluded for three analytes; anti-HIV, -HCV and HBsAg. A second lot was excluded for two analytes; anti-HIV and HBsAg. Seven lots were excluded for HBsAg only and two lots excluded for anti-HIV only. The exclusion of QC lots for HBsAg testing was due to the QC samples having unacceptable levels of anti-HBs in the base matrix resulting in low reactivity in most HBsAg assays. Other QC lots were excluded due to unexplained high reactivity in some assays for anti-HIV and anti-HCV, probably due to deviations in manufacturing processes, a change in the components of the assays, or due to results being entered into EDCNet under the incorrect assay. No lot of a single analyte QC sample was excluded.

Control limits were established for four single analyte and four multi marker QC samples, each used for multiple assays (Table 1). Control limits for a total of 64 commonly used serological assays were established. The number of QC lots used to establish the control limits ranged from 2 to 22. The number of QC test results for any assay/QC combination ranged from 83 for the QConnect Syphilis QC tested on the Siemens ADVIA Centaur Syphilis ChLIA to 112,241 for the QConnect Yellow QC tested on the Abbott PRISM HCV ChLIA.

The control limits for each assay/QC combination were determined (Table 1). In general, the difference between the lower and upper control limits was <5 units for each assay/QC combination and the lower control limit was close to the assay's cut-off in most cases. Some exceptions were noted. Results for the QC sample tested on the Roche Elecsys Anti-HCV ECLIA were high, with QC limits determined as 235–400 CIO. QConnect Green tested in the Bio-Rad Genscreen HIV-1/2 Version 2 EIA had QC limits determined as 0.7–10.4 signal to cut-off ratio (S/Co). QConnect anti-HBs and QConnect RubG were calibrated against the relevant WHO international standard at a concentration of 10 mIU/mL and IU/mL, respectively, the cut-off value for most of these assays. Different assays produced a range of results for these two QC samples. The QC limits determined for QConnect Syphilis tested in the Sekisui Mediace TPLA were 27.9–47.7 S/Co. Of the 64 assay/QC combinations, more than 5% of results were less than the lower QC limit in two instances and more than 5% of results greater than the upper QC limit in nine instances. None of the 64 assay/QC combinations had more than 10% of results less than the lower limit or greater than the upper limit.

To validate the control limits that were calculated using the results for all QC lot numbers combined, the control limits were applied to the results reported for each QC lot number individually. For example, the control limits of S/Co 2.0–3.0 for the Abbott Architect anti-HCV CMIA/QConnect Blue assay/QC combination were determined using approximately 86,000 QC test results from 11 QC lots. When these QC limits were applied to the results obtained

Table 1 The number of QC lots and QC sample test results used to determine the lower and upper control limits for each assay/quality control combination.

QC sample	Assay	No. QC lots	Calculated control limits	Unit	No. QC results
QConnect Red	Abbott AxSYM HBsAg Version 2 MEIA	14	1.7–3.0	S/Co	36,187
	Abbott AxSYM HCV Version 3.0 MEIA	16	2.1–3.8	S/Co	46,871
	Abbott AxSYM HIV Ag/Ab Combo MEIA	14	2.4–5.1	S/Co	32,308
	bioMerieux VIDAS HBsAg ELFA	6	0.3–1.0	TV	628
	bioMerieux VIDAS HBsAg Ultra ELFA	2	0.2–0.6	TV	423
	Bio-Rad Access HIV 1/2 New ChLIA	5	3.5–8.2	S/Co	295
	Ortho Vitros Anti-HCV Assay	14	1.5–4.1	S/Co	39,706
	Ortho Vitros Anti-HIV 1+2 Assay	14	0.9–2.9	S/Co	36,472
	Ortho Vitros HBsAg Assay	13	1.8–5.1	S/Co	27,922
	Roche Elecsys HBsAg ECLIA	8	3.0–7.4	COI	5682
	Roche Elecsys HBsAg II ECLIA	5	2.8–8.0	COI	1967
QConnect Yellow	Abbott PRISM HBsAg ChLIA	16	1.7–2.9	S/Co	80,159
	Abbott PRISM HBcore ChLIA	9	0.1–0.3	S/Co	11,864
	Abbott PRISM HCV ChLIA	22	2.0–4.5	S/Co	112,241
	Abbott PRISM HIV O Plus ChLIA	22	1.3–4.2	S/Co	108,757
	Abbott PRISM HTLV-I/HTLV-II ChLIA	10	1.3–3.8	S/Co	44,064
	Abbott ARCHITECT Anti-HBc II CMIA	3	6.0–8.7	S/Co	698
	Abbott ARCHITECT rHTLV-I/II CMIA	6	1.8–4.2	S/Co	8637
	bioMerieux VIDAS Anti-HBc Total II ELFA	3	0.2–0.3	TV	246
	DiaSorin Murex HTLV I+II EIA	6	1.9–5.0	S/Co	3307
QConnect Blue	Abbott ARCHITECT Anti-HCV CMIA	11	2.0–3.0	S/Co	85,960
	Abbott ARCHITECT HBsAg CMIA	8	0.1–0.3	S/Co	39,719
	Abbott ARCHITECT HBsAg Qualitative CMIA	4	1.7–4.7	S/Co	31,333
	Abbott ARCHITECT HIV Ag/Ab Combo CMIA	13	2.0–5.1	S/Co	108,931
	Roche Elecsys Anti-HCV ECLIA	5	235–400	COI	6539
	Roche Elecsys HIV Combi ECLIA	4	2.3–4.6	COI	6122
	Siemens ADVIA Centaur HBsAg ChLIA	7	3.0–6.2	IV	11,205
	Siemens ADVIA Centaur HBsAgII ChLIA	2	2.2–5.6	IV	1366
	Siemens ADVIA Centaur HCV Assay	8	2.2–4.7	IV	13,080
	Siemens ADVIA Centaur HIV1/O/2 Enhanced (EHIV) Assay	6	1.0–2.3	IV	6874
	Siemens ADVIA Centaur HIV Ag/Ab Combo (CHIV) ChLIA	3	1.8–2.7	IV	1317
	DiaSorin Murex HBsAg Version 3 EIA	5	1.2–3.9	S/Co	848
	DiaSorin Murex HIV-1.2.O EIA	9	1.5–4.3	S/Co	2365
	DiaSorin Murex anti-HCV (version 4.0) EIA	12	0.9–3.3	S/Co	15,418
QConnect Green	Bio-Rad Genscreen HIV-1/2 Version 2 EIA	7	0.7–10.4	S/Co	1752
	Bio-Rad Genscreen Plus HIV Ag-Ab EIA	3	1.8–5.1	S/Co	551
	Bio-Rad Genscreen ULTRA HIV Ag-Ab EIA	8	0.8–5.7	S/Co	9622
	Bio-Rad MONOLISA anti-HCV Plus Version 2 EIA	3	1.2–3.0	S/Co	605
	Bio-Rad MONOLISA HBsAg ULTRA EIA	3	0.2–4.1	S/Co	5858
	Bio-Rad MONOLISA HCV Ag-Ab ULTRA EIA	5	1.2–4.4	S/Co	9846
	Bio-Rad MONOLISA anti-HBc Plus EIA	4	4.4–6.9	S/Co	557
QConnect AntiHBs	Abbott ARCHITECT Anti-HBs CMIA	5	3.0–16.0	mIU/mL	16,552
	Abbott AxSYM AUSAB MEIA	5	7.0–18.0	mIU/mL	8146
	bioMerieux VIDAS Anti-HBs Total Quick ELFA	2	9.4–14.4	mIU/mL	225
	Ortho Vitros Anti-HBs Assay	5	4.0–18.0	mIU/mL	996
	Siemens ADVIA Centaur Anti-HBs ChLIA	5	5.0–22.0	mIU/mL	2535
QConnect Syphilis	Abbott ARCHITECT Syphilis TP CMIA	5	1.2–3.4	S/Co	22,489
	Sekisui Mediace TPLA	2	27.9–47.7	TU	107
	Siemens ADVIA Centaur Syphilis (SYPH) ChLIA	2	5.9–9.3	S/Co	83
	DiaSorin Murex ICE* Syphilis EIA	5	1.3–3.2	S/Co	139
QConnect RubG	Abbott ARCHITECT Rubella IgG CMIA	6	6.0–15.0	IU/mL	18,604
	bioMerieux VIDAS RUB IgG II ELFA	4	1 0.0–21.0	IU/mL	1665
	Ortho Vitros Rubella IgG Assay	3	20.0–35.0	IU/mL	1008
	Roche Elecsys Rubella IgG ECLIA	2	15.0–34.0	IU/mL	329

(**Table 1** Continued)

QC sample	Assay	No. QC lots	Calculated control limits	Unit	No. QC results
	Siemens ADVIA Centaur Rubella G Assay	4	14.0–39.0	IU/mL	4310
	Abbott AxSYM Rubella IgG MEIA	5	6.5–17.7	IU/mL	7816
	Siemens IMMULITE 2000 Rubella Quantitative IgG CLEIA	3	11.0–19.0	IU/mL	955
QConnect HIVp24	Abbott ARCHITECT HIV Ag/Ab Combo CMIA	6	1.1–5.9	S/Co	31,933
	Abbott AxSYM HIV Ag/Ab Combo MEIA	6	1.2–2.8	S/Co	5204
	Abbott PRISM HIV Ag/Ab Combo ChLIA	2	1.0–3.5	S/Co	426
	Roche Elecsys HIV Combi ECLIA	2	0.7–1.7	COI	2571
	Roche Elecsys HIV combi PT ECLIA	4	1.4–4.4	COI	3488
	Bio-Rad Genscreen ULTRA HIV Ag-Ab EIA	6	1.1–4.9	S/Co	1239
	Bio-Rad Genscreen HIV-1 Ag EIA	3	1.8–6.4	S/Co	240

from each of the 11 QC lots individually, >5% of results were less than S/Co=2.0 in two lots and >5% of results greater than S/Co=3.0 in one lot. Overall, of the almost 90,000 QC test results analysed to validate the range, only 3375 (3.75%) were less than S/Co=2.0 and 1553 (1.73%) results were greater than S/Co=3.0. Overall, there were 423 assay/QC lot combinations. Of these, 14 QC lots (3.3%) had more than 5% of results less than the lower limit determined for the assay/QC combination and 48 QC lots (11.3%) had more than 5% of results greater than the upper limit determined for the relevant assay/QC combination (Table 2).

Discussion

QC programs have been used in clinical chemistry for decades to quantify the repeatability and precision of measurement procedures [28–32]. Traditionally, a defined minimum set of QC values, often cited as 20 values [4, 5, 7], are used to calculate control limits; often using the mean plus and minus two or three SDs of that sample's results [4, 5, 7]. This approach assumes that the results obtained were normally distributed and that there will be approximately 5% (mean±2SD) or 1% (mean±3SD) of results falsely rejected [4, 7]. The approach also assumes that these approximately 20 values used to establish the control limits will be representative of future QC test results. Westgard rules are often applied to aid the interpretation of QC sample test results [8, 9, 32]. Allowable total error (TEa) is another means of establishing control limits [4, 7, 9, 10, 33]. Some authors have called for the simplification of QC processes to increase compliance [34, 35].

Table 2 The number of assay/QC lots analysed for each QConnect sample, the number of each assay/QC lot having more than 5% of results less than the lower control limit or greater than the upper control limit and percentage of the total.

QC sample	Number of assay/QC lots	Number assay/QC lots with more than 5% of results	
		Less than lower control limit	Greater than upper control limit
QConnect Red	119	5	9
QConnect Blue	96	3	10
QConnect Yellow	84	4	10
QConnect Green	33	2	7
QConnect AntiHBs	22	Nil	4
QConnect RubG	26	Nil	5
QConnect HIVp24	29	Nil	2
QConnect Syphilis	14	Nil	1
Total	423	14 (3.3%)	48 (11.3%)

The accurate measurement of analytes, such as glucose, relies on several concepts. Certified reference materials (CRM) are prepared from high-grade analytical materials and are quantified using sophisticated methodologies such as isotope dilution mass spectrometry and strict metrological traceability principles [36, 37]. IVD manufacturers use CRM to produce calibrators for assays, thereby allowing the results to be reported in SI units, such as mmol/L. Only about 30 routine biochemical measurands are traceable to SI in this way [38]. Generally, the levels of these analytes in a normal population have a Gaussian distribution (Figure 1). Using population studies, 'reference intervals' are established, with medical decision points assigned, often adjusted for gender and age. If the results of testing exceed these values, medical intervention may be initiated [39].

When testing a QC sample targeted at the medical decision level, a change in QC sample results may predict a change in patient results at that same level, if commutability is assumed. In this case, TEa is a useful metric when establishing control limits.

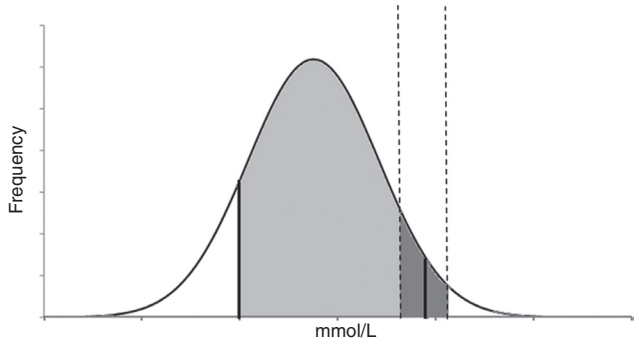


Figure 1 Schematic diagram of the distribution of fasting blood glucose in healthy adults.
The area shaded in light grey highlights the normal range for blood glucose. The dotted lines indicate a possible change in patient test results associated with a proportional change in QC sample results. The dark grey indicates the proportion of individuals that could be mis-classified due to the change in QC sample results.

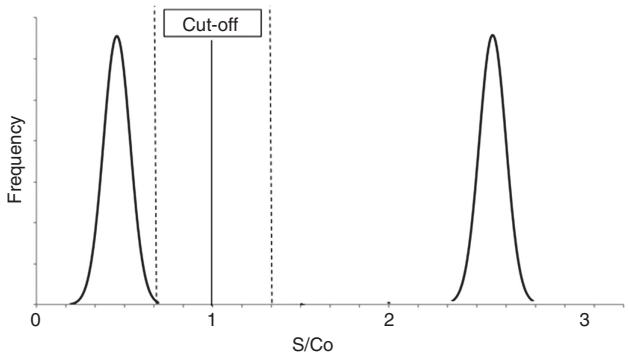


Figure 2 Schematic diagram of the bi-modal distribution of anti-HIV antibodies in an adult population, demonstrating a population of negative results (S/Co<1) and positive results (S/Co>1).
The dotted lines indicate a possible change in patient test results associated with a proportional change in QC sample results. Theoretically very few true negative or positive results will be misclassified if a proportional change in patient and QC sample results is assumed, because of the distance of the populations from the assay cut-off.

In infectious disease serology, such as the detection of anti-HIV antibodies, the paradigm above does not apply [21, 23]. There are generally no CRM to facilitate quantification; the exception being analytes such as anti-hepatitis B surface antigen and anti-rubella IgG [12]. The cut-off of the assay is determined by applying a manufacturer-defined algorithm. The cut-off divides the population into a dichotomous grouping; those with the analyte under investigation and those without. Therefore, unlike the inert analytes discussed above, the detection of antibodies (or antigens) is bimodal, not normally distributed. Assay manufacturers design serological assays to separate the positive and negative populations as widely as possible (Figure 2). This separation can be quantified as the Δ Value (DV), which is the number of SDs the mean of the positive (DV+) or negative (DV–) population's results are removed from the cut-off [40]. Third and fourth generation serological assays have DVs >10 so that very few true negative or positive test results fall close to the cut-off. Indeed, in a previous study, over a 4-year period, of a total of 5 million blood donations for hepatitis B surface antigen, only five of 281 confirmed positive donations had initial test results with S/Co results of 2.00 or less [21].

Serological assays used to test for infectious diseases often experience considerable variation associated with the introduction of new assay lot numbers [12, 23, 24, 41]. Therefore, after setting control limits using the traditional method of determining the mean ±2 or 3 times the SD from a previous approximately 20 QC sample results, subsequent results frequently fall outside these control limits when a new assay lot number is introduced. In these cases, should the patients' results be rejected even though the assay is valid as per the manufacturer's instructions, or should the laboratory re-establish the QC limits? In our experience conducting QC programs for infectious disease testing for over a decade, this situation frequently arises.

The principles used in establishing control limits when testing for inert analytes does not apply to infectious disease serology. Even assuming commutability of QC and patient samples, if there was a proportional change in patient and QC test results, the number of falsely reported patient results will be minimal because there are very few true patient results around the cut-off. The concept of TEa does not apply because there is no error allowed in the detection of infectious diseases. That is, a positive patient sample should always be reactive and negative patient samples reported non-reactive, especially in a blood donor screening setting. Although traditional QC principles have been applied to infectious disease serology, there have been very few publications on their application [12, 41, 42].

NRL's QC program, called QConnect (www.nrl.gov.au/qconnect) has allowed the collection of hundreds of thousands of QC test results for many assay/QC combinations. Uniquely, each lot of QC was manufactured using the same stock materials and, where possible, calibrated against a reference standard using ISO 17511 traceability claims [26]. Therefore the QC lot-to-lot variability was minimised. Using these data, NRL sought to establish assay/QC limits using the within- and between-QC lot variation in place of traditional methods of establishing control limits.

In the QConnect program all laboratories using the same assay also used the same QC, allowing comparison of QC sample test results across an assay/QC peer group.

Analysis of the results of assay/QC combinations confirmed that the lot-to-lot variation in QC reactivity was low, given that the results reflect the variation in manufacturing of the assay, along with multiple instruments, operators and environmental conditions over a long period of time. For the majority of the 64 assay/QC combinations <5% of QC results were outside the determined control limits. None had more than 10% results outside the range. A greater number of individual QC lots had more than 5% of the results outside the established limits, with 30 of the 432 assay/QC lot number combinations having more than 10% of results outside the established limits.

There were 11 QC lots that had were removed from the data set establishing QConnect limits for one or more assays. This represented approximately 4% of QC lots. Each QC lot removed was manufactured under processes that did not conform with the standard manufacturing process. Nine of the 11 QC lost were removed because the base matrix used to dilute the stock material had low levels of anti-HBs, resulting in unacceptable low level reactivity of the HBsAg. The other lots were removed only for specific assays, indicating either a change in the assays' performance characteristics or that the participants entered data under the incorrect assay.

Now that control limits have been established, they can be used to determine the QC lot release criteria. When new lots of QC are produced, results of pre-release testing would be expected to fall within this pre-determined range, thereby eliminating the risk of QC test results being outside the limits due to variation in QC manufacturing. In the future, QC production lots having unexpected results, such as those that were excluded from this analysis, will fail lot release and not be provided to participants in the QC programme. In order to set control limits for new assays, results of at least two lots of QC are required.

Although widely used, traditional methods of setting control limits have yet to be proven to apply to infectious disease serology, as there are few publications critically analysing whether these practices are applicable. Widely fluctuating levels of QC sample reactivity with changes in assay lot numbers and the fact that results of patient's samples are bimodal rather than Gaussian have not been considered when applying mean ±2 or 3 SD.

The approach of setting control limits based on historical data takes into account total variation expected in a test system and offers a more robust and meaningful mechanism for setting control limits, which can be used in isolation or in conjunction with other methods. The QC method described will improve the outcomes of QC monitoring because it: 1) provides evidence-based limits developed from tens of thousands of test results over a long period of time; 2) gives clear guidance to the method users; and 3) simplifies and demystifies QC, making it more accessible to laboratory workers.

Acknowledgments: The stock materials used for the manufacture of the PeliSpy/QConnect multimarker products were originally developed by Dr. Nico Lelie and Dr. Harry van Drimmelen. Carl Schaper provided support for the development of the statistical approach.
Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Financial support: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organisation(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. AS 4633 (ISO 15189) Field application document – medical testing; supplemental requirements for accreditation, 2013.
2. TGA. Australian Code of Good Manufacturing Practice – human blood and tissue. In: TGA, editor. Canberra: Commonwealth of Australia, Australian Government Department of Health and Ageing, 2000.
3. ISO 15189: 2012 Medical laboratories – Requirements for quality and competence, 2012.
4. Statistical quality control for quantitative measurement procedures: principles and definitions; approved guidelines, 2006.
5. Public Health England. Quality assurance in the diagnostic virology and serology laboratory. UK Standards for Microbiology Investigation. London: Standards Unit, 2013.
6. Laboratory quality control based on risk management; approved guidelines, 2011.
7. Brooks Z. Performance-driven quality control. Washington: AACC Press, 2001.
8. Westgard JO. Internal quality control: planning and implementation strategies. Ann Clin Biochem 2003;40:593–611.
9. Westgard JO. Statistical quality control procedures. Clin Lab Med 2013;33:111–24.
10. Yundt-Pacheco J, Parvin CA. Validating the performance of QC procedures. Clin Lab Med 2013;33:75–88.
11. Harel O, Schisterman EF, Vexler A, Ruopp MD. Monitoring quality control: can we get better data? Epidemiology 2008;19:621–7.
12. Kruk T, Ratnam S, Preiksaitis J, Lau A, Hatchette T, Horsman G, et al. Results of continuous monitoring of the performance of rubella virus IgG and hepatitis B virus surface antibody assays using trueness controls in a multicenter trial. Clin Vaccine Immunol 2012;19:1624–32.
13. Fraser CG, Kallner A, Kenny D, Petersen PH. Introduction: strategies to set global quality specifications in laboratory medicine. Scand J Clin Lab Invest 1999;59:477–8.

14. Kaplan LA. Determination and application of desirable analytical performance goals: the ISO/TC 212 approach. Scand J Clin Lab Invest 1999;59:479–82.
15. Petersen PH. Quality specifications based on analysis of effects of performance on clinical decision-making. Scand J Clin Lab Invest 1999;59:517–21.
16. Fraser CG, Petersen PH. Desirable standards for laboratory tests if they are to fulfill medical needs. Clin Chem 1993;39:1447–53; discussion 53–5.
17. Fraser CG, Petersen PH. Desirable performance standards for imprecision and bias in alternate sites. The views of laboratory professionals. Arch Pathol Lab Med 1995;119:909–13.
18. Fraser CG, Petersen PH. Analytical performance characteristics should be judged against objective quality specifications. Clin Chem 1999;45:321–3.
19. Schoen I, Graham GA, Custer E, Bandi Z, Georgewill D. Clinically useful limits (CUL) criteria as a basis for quality control including minimal and optimal goals for quality control. Am J Clin Pathol 1989;91:172–81.
20. Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. Clin Chem 2011;57:76–83.
21. Dimech W, Freame R, Smeh K, Wand H. A review of the relationship between quality control and donor sample results obtained from serological assays used for screening blood donations for anti-HIV and hepatitis B surface antigen. Accred Qual Assur 2012;18:11–8.
22. Walker S, Dimech W, Kiely P, Smeh K, Francis B, Karakaltsas M, et al. An international quality control programme for PRISM chemiluminescent immunoassays in blood service and blood product laboratories. Vox Sang 2009;97:309–16.
23. Dimech W, Walker S, Jardine D, Read R, Smeh K, Karakaltsas M, et al. Comprehensive quality control programme for serology and nucleic acid testing using an internet-based application. Accred Qual Assur 2004;9:148–51.
24. Pham TD, Karakaltas M, Dax EM, editors. Variation in the performance of reagent batches of the Abbott Architect HIV Ag/Ab Combo CMIA detected in an international quality control programme. XXXth International Congress of the ISBT; 2008; Macau: Vox Sang.
25. International Organization for Standardization. ISO 13485:2003; Medical devices – quality management systems – requirements for regulatory purposes. Geneva: International Organization for Standardization, 2003.
26. International Organization for Standardization. ISO 17511:2003 In vitro diagnostic medical devices – measurement of quantities in biological samples – metrological traceability of values assigned to calibrators and control materials. Geneva: International Organization for Standardization, 2003.
27. Vesper HW, Miller WG, Myers GL. Reference materials and commutability. Clin Biochem Rev 2007;28:139–47.
28. Grannis GF, Miller WG. On the design of clinical chemistry quality-control sera. Clin Chem 1976;22:500–12.
29. Barnett RN. Error limits and quality control. Arch Pathol Lab Med 1989;113:829–30.
30. Parvin CA. Quality-control (QC) performance measures and the QC planning process. Clin Chem 1997;43:602–7.
31. Parvin CA, Gronowski AM. Effect of analytical run length on quality-control (QC) performance and the QC planning process. Clin Chem 1997;43:2149–54.
32. Westgard JO, Stein B. Automated selection of statistical quality-control procedures to assure meeting clinical or analytical quality requirements. Clin Chem 1997;43:400–3.
33. Burnett D, Ceriotti F, Cooper G, Parvin C, Plebani M, Westgard J. Collective opinion paper on findings of the 2009 convocation of experts on quality control. Clin Chem Lab Med 2010;48:41–52.
34. Howanitz PJ, Tetrault GA, Steindel SJ. Clinical laboratory quality control: a costly process now out of control. Clin Chim Acta 1997;260:163–74.
35. Parvin CA, Jones JB. QC design: it's easier than you think. MLO Med Lab Obs 2013;45:18, 20, 22.
36. Muller MM. Implementation of reference systems in laboratory medicine. Clin Chem 2000;46:1907–9.
37. Miller WG, Tate JR, Barth JH, Jones GR. Harmonization: the sample, the measurement, and the report. Ann Lab Med 2014;34:187–97.
38. Royal College of Pathologist Australasia. Australian Pathology Units and Terminology Sydney: Royal College of Pathologists, Australasia; 2014 [4 May 2014]. Available from: http://www.rcpa.edu.au/Library/Practising-Pathology/PTIS/APUTS-Downloads. Accessed on August 5, 2014.
39. Zhou J, Li H, Ran X, Yang W, Li Q, Peng Y, et al. Establishment of normal reference ranges for glycemic variability in Chinese subjects using continuous glucose monitoring. Med Sci Monit 2011;17:CR9–13.
40. Crofts N, Maskill W, Gust ID. Evaluation of enzyme-linked immunosorbent assays: a method of data analysis. J Virol Methods 1988;22:51–9.
41. Kim J, Swantee C, Lee B, Gunning H, Chow A, Sidaway F, et al. Identification of performance problems in a commercial human immunodeficiency virus type 1 enzyme immunoassay by multiuser external quality control monitoring and real-time data analysis. J Clin Microbiol 2009;47:3114–20.
42. Green GA, Carey RN, Westgard JO, Carten T, Shablesky L, Achord D, et al. Quality control for qualitative assays: quantitative QC procedure designed to assure analytical quality required for an ELISA of hepatitis B surface antigen. Clin Chem 1997;43:1618–21.

The final publication is available at DOI 10.1515/cclm-2014-0546

Wayne Dimech*, Marina Karakaltsas and Giuseppe A. Vincini

Comparison of four methods of establishing control limits for monitoring quality controls in infectious disease serology testing

https://doi.org/10.1515/cclm-2018-0351
Received February 12, 2018; accepted April 23, 2018; previously published online May 25, 2018

Abstract

Background: A general trend towards conducting infectious disease serology testing in centralized laboratories means that quality control (QC) principles used for clinical chemistry testing are applied to infectious disease testing. However, no systematic assessment of methods used to establish QC limits has been applied to infectious disease serology testing.

Methods: A total of 103 QC data sets, obtained from six different infectious disease serology analytes, were parsed through standard methods for establishing statistical control limits, including guidelines from Public Health England, USA Clinical and Laboratory Standards Institute (CLSI), German Richtlinien der Bundesärztekammer (RiliBÄK) and Australian QConnect. The percentage of QC results failing each method was compared.

Results: The percentage of data sets having more than 20% of QC results failing Westgard rules when the first 20 results were used to calculate the mean ± 2 standard deviation (SD) ranged from 3 (2.9%) for R_{4S} to 66 (64.1%) for 10_X rule, whereas the percentage ranged from 0 (0%) for R_{4S} to 32 (40.5%) for 10_X when the first 100 results were used to calculate the mean ± 2 SD. By contrast, the percentage of data sets with >20% failing the RiliBÄK control limits was 25 (24.3%). Only two data sets (1.9%) had more than 20% of results outside the QConnect Limits.

Conclusions: The rate of failure of QCs using QConnect Limits was more applicable for monitoring infectious disease serology testing compared with UK Public Health, CLSI and RiliBÄK, as the alternatives to QConnect Limits reported an unacceptably high percentage of failures across the 103 data sets.

Keywords: infectious disease; QConnect Limits; run control; serology; Westgard rules.

Introduction

The use of third-party controls to monitor the performance of testing systems in medical pathology is generally considered best practice [1–4]. Most clinical chemistry laboratories use third-party controls (also called internal quality controls, run controls or quality controls [QC]) routinely. This is not the case for infectious disease serology testing, where the use of QC is sporadic. An effective QC system requires several components: QC samples reactive at an appropriate level, access to a single QC lot number over a long period of time, a process for plotting, visualizing and analyzing the QC test results and a system for setting control limits within which the QC test results are expected to fall [3–7]. Control limits set too tightly will lead to unnecessary investigations and delays in patients' reports whereas control limits set too widely potentially allow reporting incorrect patients' results [3, 7].

Traditionally, setting control limits has relied on principles established by Walter A. Shewhart and S. Levey and E.R. Jennings and further developed by James Westgard, who created a set of statistics-based rules, commonly known in medical testing laboratories as "Westgard rules" [8, 9]. With the advent of "core" laboratories, an increasing number of organizations are incorporating testing for infectious disease serology onto automated platforms that perform cross-discipline testing. As a result, the traditional approach to monitoring QC is being applied to infectious disease tests (in this document referred to as "assays").

A growing number of alternative methods for monitoring test systems are being adopted, and Westgard rules are no longer applied in a blanket fashion [2, 7]. However, laboratory scientists and managers rely on standards and guidelines written by experts and published by

***Corresponding author: Wayne Dimech,** National Serology Reference Laboratory, 4th Floor Healy Building, 41 Victoria Parade, Fitzroy, Victoria 3065, Australia, Phone: +61 3 9418 1132, Fax: +61 3 9418 1155, E-mail: wayne@nrlquality.org.au
Marina Karakaltsas and Giuseppe A. Vincini: National Serology Reference Laboratory, Fitzroy, Victoria, Australia

professional bodies when adopting processes such as setting acceptable limits for monitoring QC results. Some guidelines suggest that control limits be established by testing each new lot of QC sample 15–20 times on separate days and calculating the mean and standard deviation (SD) of the results [3, 4, 6, 7]. Some recent guidelines state that 20 results are appropriate without any caveat [4]; others, such as German Richtlinien der Bundesärztekammer (RiliBÄK) standards [10], suggest using the first 15 results. Other standards suggest control limits, initially calculated using 20 data points, be reviewed periodically and where necessary adjusted over time [6, 7], although no specific instructions are given on how long a time period or how many data points are required. Anecdotally, many laboratories persist with using 20 data points to set their control limits and apply Westgard rules to infectious disease serology testing assuming a normal (Gaussian) distribution over time [9]. However, there has been no report of a systematic review of the suitability of this approach to monitoring infectious disease testing.

More recently, the National Serology Reference Laboratory (NRL), Australia, developed a comprehensive QC program for serology laboratories testing for infectious diseases [11, 12] and devised and published a method of determining assay-specific control limits, called QConnect Limits, within which QC sample results were expected to fall [13]. Briefly, QConnect Limits combine the variance of QC results reported for each QC lot number with the variance between the mean of results reported for each QC lot number tested on the same assay by multiple laboratories [11, 12].

The purpose of this study was to perform a systematic review of the percentage of QC test results, obtained from routine infectious disease serology testing, that would be flagged as out of control when subjected to the unmodified processes described in each of four different standards for establishing control limits and monitoring QC test results. The Clinical and Laboratory Standards Institute (CLSI) guideline C24 states "Empirical evaluation of QC rules performance can also be done by obtaining a large series of QC results from the measurand procedure that has been operating in a stable condition over a sufficiently long period of time interval to include all major sources of variability in the data" [7]. As only QC test results obtained from test runs that passed the manufacturer's validation criteria were used, it was expected that the percentage of flagged results would be small. The information obtained will provide those involved with selecting infectious disease testing QC monitoring processes evidence to support an informed choice between published standards.

Materials and methods

Single or multimarker QConnect™ QC samples were tested by participating clinical or blood screening laboratories testing for infectious diseases. The results of the QC sample testing were entered into internet-based QC monitoring software system, EDCNet (NRL, Australia; www.nrlquality.org.au), along with associated data such as date of testing, QC and reagent lot numbers, instrument serial number and operator identification. A detailed description of the QC program can be found elsewhere [12]. Only QC results from test runs fulfilling the manufacturer's criteria for acceptance were analyzed.

Data submitted for the calendar year 2015 were extracted. Results from the QC lot that was predominantly used for a particular assay during that time period were selected for the study. Assuming laboratories would establish control limits when each new lot of QC was introduced, any QC that was being used in the first week of January had further data from 2014 extracted to include data from the date of introduction of that QC lot. All assay/QC combinations with 50 or more test results for the selected period of time were analyzed. If there were more than 10 laboratories with >50 results for an assay/QC combination, results from 10 laboratories using that assay/QC combination were randomly selected to be included in the analysis. Where a laboratory used more than one of the same instrument, data from each instrument were analyzed separately. In this paper, QC results from an assay/QC/laboratory/instrument are referred to as data sets, i.e. QC results reported by a laboratory testing a QC sample on a particular assay and a particular instrument.

The extracted data sets were subjected to four different methods for setting QC limits. The standard published by Public Health England [4] states that 20 QC results be used to establish mean ± 1, 2 and 3 SDs and Westgard rule be applied. The CLSI guide C24, 4th ed. [7] states "initial estimates of SD are obtained by measuring at least 20 data points on separate days" and that Westgard rules be applied. The guide acknowledges that these initial estimates have "limited reliability" and that "cumulative SD over the first several months of operation gives a better estimate of SD". To this end, the mean and SD used to establish limits for Westgard rules was estimated using both the first 20 ($Westgard_{20}$), as per the Public Health England guideline and 100 ($Westgard_{100}$) QC test result in line with the CLSI guidance document.

To establish QC limits in accordance with the RiliBÄK standard, the equation $\Delta_{MAX} = \sqrt{K^2 * S_{ep}^2 + \delta_{ep}^2}$ was used, where K is the coverage factor for calculating the internal laboratory deviation limits, S_{ep} is the empirical SD of the control sample measurements used in the calculations during the preevaluation period and δ_{ep} is the systematic deviation of measurement of the QC measurements used in the calculations during the evaluation period, i.e. the bias. Given that infectious disease serology QC samples do not have a "true value", for the purposes of this study, the equation $\Delta_{MAX} = \sqrt{K^2 * S_{ep}^2}$ was used, where the RiliBÄK control limits were mean of the first 15 test results $\pm \Delta_{MAX}$. Each data set was also subjected to the relevant QConnect Limits, which have been previously established using historical data.

Using a Microsoft® Excel spreadsheet, each data set was parsed through the four methods $Westgard_{20}$, $Westgard_{100}$, RiliBÄK and QConnect. Where applicable, the QC results were subjected to the Westgard rules 1_{2S}, 1_{3S}, 2_{2S}, 4_{1S}, 10_{X} and R_{4S}. The total number of results, the mean, the SD, and the number of QC test results failing each rule were calculated for each data set. The percentage of QC results from each data set that was outside the respective limits was categorized into <10%, 10%–20% and >20%.

The total number of reagent lots used in each data set was calculated and the mean number of QC test results reported for each reagent lot number was estimated.

Results

A total of 21,510 QC test results from 14 different assays that are commonly used in infectious disease serology testing laboratories were analyzed (Table 1). The participants represented hospital-based, private, blood screening and reference laboratories from five countries: Australia, Canada, Israel, Poland and New Zealand. In total, 103 assay/QC/laboratory/instrument data sets were analyzed across six of the most commonly tested infectious disease serology analytes, including assays that detect antibodies and antigens. The percentage of QC results failing each rule was divided into three categories: <10%, between 10% and 20%, and >20%. A representative data set of results of testing QConnect™ Blue, lot number 424521, on the Siemens ADVIA Centaur hepatitis C virus assay is given in Table 2. In this example, seven data sets from four laboratories are shown. The number of QC results for each data set ranged from 58 to 400. Only one data set has more than 20% of QC results failing a single rule (Participant 877; Westgard$_{20}$; 1_{2S} rule). Similar tables for the remaining 102 data sets were developed.

For each assay, the number of data sets within each category of percentage failures was tabulated for each rule (Supplementary Data, Table S1), and the percentage of 103 data sets with more than 20% of results failing each rule was calculated (Table 3). Of the 103 data sets, the number of data sets that had more than 20% of QC results fail Westgard$_{20}$ rules ranged from 3 (2.9%) for R_{4S} to 66 (64.1%) for 10_X rule. Of the 103 data sets, 79 had more than 100 results, allowing the application of Westgard$_{100}$. The number of data sets that had more than 20% of QC results failing each Westgard$_{100}$ rule ranged from 0 (0%) for R_{4S} to 32 (40.5%) for 10_X rule. By contrast, the number of data sets with >20% failing the RiliBÄK control limits was 25 (24.3%). The study showed that 95 (92.2%) of the 103 datasets had <10% of results failing the QConnect Limits. There were only two data sets with more than 20% of results outside the QConnect Limits: one participant using the Abbott ARCHITECT HBsAg Qualitative II CMIA and another using the Roche Elecsys HBsAg II ECLIA. These two instances were investigated further.

A participant testing in the ARCHITECT HBsAg assay reported 180 QConnect Blue test results between July and December 2015. Of these, 81 (45.0%) were greater than the upper QConnect Limit of signal to cutoff ratio (S/Co) of

Table 1: Summary of the QConnect Limits for each of the quality control sample lots tested on each assay.

Assay	Analyte	QConnect QC name	QConnect QC lot number	QConnect Limits, units	Number of laboratories	Number of instruments
Abbott ARCHITECT anti-HBs CMIA	Anti-HBs	QConnect Anti-HBs	422701	3.0–16.0 mIU/mL	10	10
Abbott ARCHITECT anti-HCV CMIA	Anti-HCV	QConnect Blue	424521	1.7–2.9 S/Co	10	13
Abbott ARCHITECT HBsAg Qualitative II CMIA	HBsAg	QConnect Blue	424521	3.1–6.0 S/Co	10	13
Abbott ARCHITECT Rubella IgG CMIA	Anti-rubella IgG	QConnect Rubella G	330101	10.5–15.0 IU/mL	10	11
DiaSorin Murex Anti-HCV (version 4.0) EIA	Anti-HCV	QConnect Blue	424521	0.8–3.5 S/Co	5	NA
DiaSorin Murex HTLV I+II EIA	Anti-hTLV	QConnect Yellow	433507	1.9–5.0 S/Co	4	NA
Ortho VITROS anti-HCV assay	Anti-HCV	QConnect Red	434212	3.1–4.9 S/Co	7	7
Ortho VITROS anti-HIV 1 + 2 assay	Anti-hiv	QConnect Red	434212	2.0–3.3 S/Co	6	6
Siemens ADVIA Centaur HBsAg II ChLIA	HBsAg	QConnect Blue	424521	1.5–5.1 S/Co	4	7
Siemens ADVIA Centaur HCV assay	Anti-HCV	QConnect Blue	424521	2.2–4.7 S/Co	4	7
Siemens ADVIA Centaur Rubella G assay	Anti-rubella IgG	QConnect Rubella G	330101	30–70 IU/mL	3	4
Roche Elecsys Anti-HCV II ECLIA	Anti-HCV	QConnect Blue	424521	13.4–138.2 COI	10	10
Roche Elecsys HBsAg II ECLIA	HBsAg	QConnect Red	434212	2.8–8.0 COI	4	4
Roche Elecsys Rubella IgG ECLIA	Anti-rubella IgG	QConnect Rubella G	330101	15.0–34.0 IU/mL	2	2

Table 2: Results of testing QConnect Blue, lot number 424521 on the Siemens ADVIA Centaur HCV assay.

Laboratory identification (instrument identification)	Number of results	Failure above or below acceptance range	Number of results failing Westgard rules												QConnect Limits	RiliBÄK control limit
			First 20 results						First 100 results							
			1_{2S}	1_{3S}	2_{2S}	4_{1S}	10_{X}	R_{4S}	1_{2S}	1_{3S}	2_{2S}	4_{1S}	10_{X}	R_{4S}		
15	153	Low	8	5	0	3	2		0	0	0	0	0		4	10
		High	20	15	15	16	11		1	0	0	0	0		0	1
		Total	28	20	15	19	13	12	1	0	0	0	0	1	4	11
		Percentage	18.30%	13.10%	9.80%	12.40%	8.50%	7.80%	0.70%	0.00%	0.00%	0.00%	0.00%	0.70%	2.90%	7.20%
47 (SN 3)	58	Low	0	0	0	0	1		0	0	0	0	0		0	0
		High	0	0	0	0	6		0	0	0	0	0		0	0
		Total	0	0	0	0	7	0	0	0	0	0	0	0	0	0
		Percentage	0.00%	0.00%	0.00%	0.00%	12.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
47 (C1)	390	Low	10	3	0	3	12		24	5	0	6	55		4	2
		High	8	0	3	4	24		5	0	5	2	2		0	0
		Total	18	3	3	7	36	15	29	5	5	8	57	24	4	2
		Percentage	4.60%	0.80%	0.80%	1.80%	9.20%	3.80%	7.40%	1.30%	1.30%	2.10%	14.60%	6.20%	1.00%	0.50%
47 (C4)	400	Low	8	0	0	1	9		20	3	0	10	9		0	0
		High	1	0	0	0	1		9	1	3	1	1		0	0
		Total	9	0	0	1	10	9	29	4	3	11	10	26	0	0
		Percentage	2.30%	0.00%	0.00%	0.30%	2.50%	2.30%	7.30%	1.00%	0.80%	2.80%	2.50%	6.50%	0.00%	0.00%
47 (XPT)	63	Low	0	0	0	0	2		0	0	0	0	0		0	0
		High	0	0	0	0	0		0	0	0	0	0		0	0
		Total	0	0	0	0	2	0	0	0	0	0	0	0	0	0
		Percentage	0.00%	0.00%	0.00%	0.00%	3.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
115	203	Low	3	0	0	1	3		3	0	0	0	0		0	0
		High	6	1	2	5	10		1	0	1	0	6		0	1
		Total	9	1	2	6	13	6	4	0	1	0	6	2	0	1
		Percentage	4.40%	0.50%	1.00%	3.00%	6.40%	3.00%	2.00%	0.00%	0.50%	0.00%	3.00%	1.00%	0.00%	0.50%
877	144	Low	21	8	0	10	16		4	0	0	0	6		0	8
		High	22	8	15	7	2		0	0	1	0	0		0	6
		Total	43	16	15	17	18	28	4	0	1	0	6	3	0	14
		Percentage	29.90%	11.10%	10.40%	11.80%	12.50%	19.40%	2.80%	0.00%	0.70%	0.00%	4.20%	2.10%	0.00%	9.70%

Table 3: Summary of the percentage of data sets with <10%, 10%–20% and >20% of results failing each control limit for each assay investigated.

Number of data sets investigated (>100 results)	Percentage outside range	Percentage of results failing Westgard rules										QConnect Limits	RiliBÄK control limit
		First 10 results					First 100 results						
		1_{2s}	1_{3s}	2_{2s}	4_{1s}	10_{x}	R_{4s}	1_{2s}	1_{3s}	2_{2s}	4_{1s}	10_{x}	R_{4s}		
103 (79)	<10%	48.5%	60.2%	51.5%	31.1%	21.4%	61.2%	70.9%	92.4%	87.3%	64.6%	46.8%	96.2%	92.2%	58.3%
	10%–20%	25.2%	18.4%	20.4%	21.4%	14.6%	35.9%	16.5%	5.1%	6.3%	15.2%	12.7%	3.8%	4.9%	17.5%
	>20%	26.2%	21.4%	28.2%	47.6%	64.1%	2.9%	12.7%	2.5%	6.3%	20.3%	40.5%	0.0%	2.9%	24.3%

6.0 (Supplementary Data, Figure S1). Over the same time period, additional 22 participants reported QConnect Blue results using the same reagent lot numbers. Of the 22, only 6 (27.3%), including the participant under investigation, reported any results greater than the QConnect upper limit. The other five participants reported few results above S/Co 6.0, the most being 9 (15.0%) of 60 results (Supplementary Data, Figure S2).

The participant testing in the Roche Elecsys HBsAg assay reported 105 QConnect Red test results, of which 72 (68.6%) were greater than the upper QConnect Limit of cutoff index (COI) 8.0 for that assay/QC combination (Figure 1). Over the same time period, five other participants tested the same QC sample in the same assay reagent lots, and none reported results outside the QConnect Limits (Figure 2). For both examples, the investigation indicated the source of variation causing results to be outside the QConnect Limits was confined to the participant's laboratories, not the reagents.

The failures observed when applying Westgard$_{20}$ rules to each data set were mainly due to the introduction of different reagent lots. A possible procedure to accommodate these changes would be to reestablish the mean and SD when each new reagent lot was introduced. To assess the feasibility of this approach, the numbers of QC test results submitted for each reagent lot were assessed for each of the 103 assay/QC/laboratory/instrument data sets (Supplementary Data, Table S2). The mean number of QC test results reported for each reagent lot varied from assay to assay but ranged from 19 to 83. However, of the 638 reagents lots analyzed, 373 (58.5%) had ≤30 QC test results and 504 (79.0%) had ≤50 QC test results, meaning that setting new limits based on the first 20 results obtained for each new reagent lot was impractical, as the reagent lot would be exhausted before, or soon after, the 20 data points were collected.

Discussion

Monitoring the variation of testing systems using third-party controls is considered best practice [1, 3, 4, 7, 14]. Laboratory scientists and managers rely on standards and associated commentaries [5, 6, 8, 15–18] when establishing procedures. These standards, as well as auditors, direct laboratory staff to utilize the processes outlined without modification. In the present study, four different methods were applied to 103 data sets obtained from routine infectious disease serology testing: Westgard$_{20}$, as specified by Public Health England; (Westgard$_{100}$) in line with CLSI,

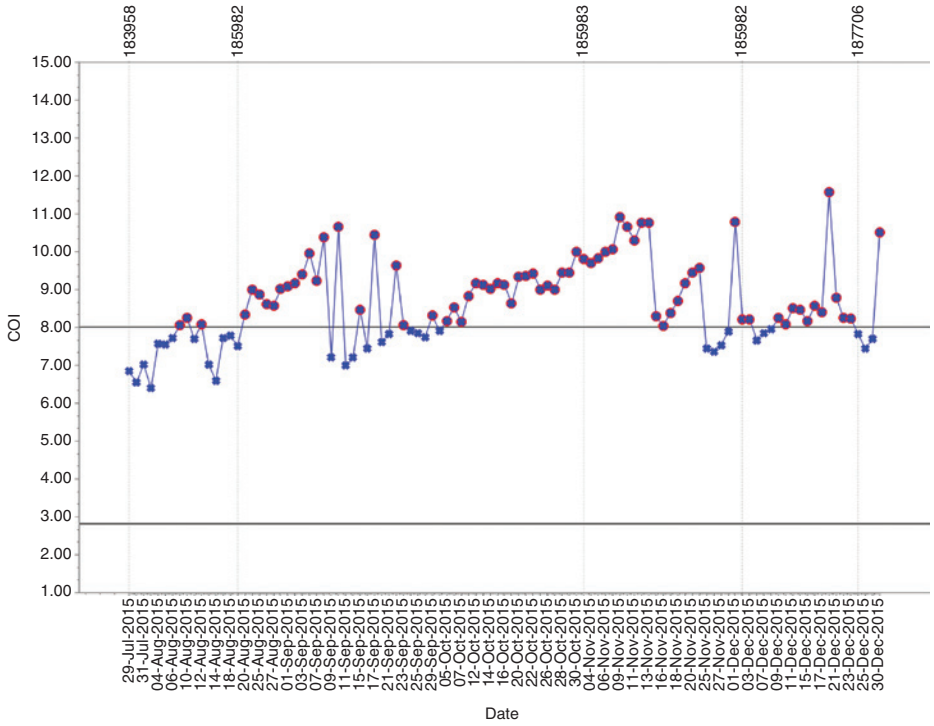


Figure 1: Levey-Jennings chart of results from Participant 174 testing QConnect Red lot number 434212 in the Roche Elecsys HBsAg II ECLIA over the period 29 July 2015 to 31 December 2015.
The top of the graph shows the change of reagent lot numbers. The horizontal bar at 2.8 and 8.0 cutoff index (COI) represent the lower and upper QConnect Limits for that assay/QC combination.

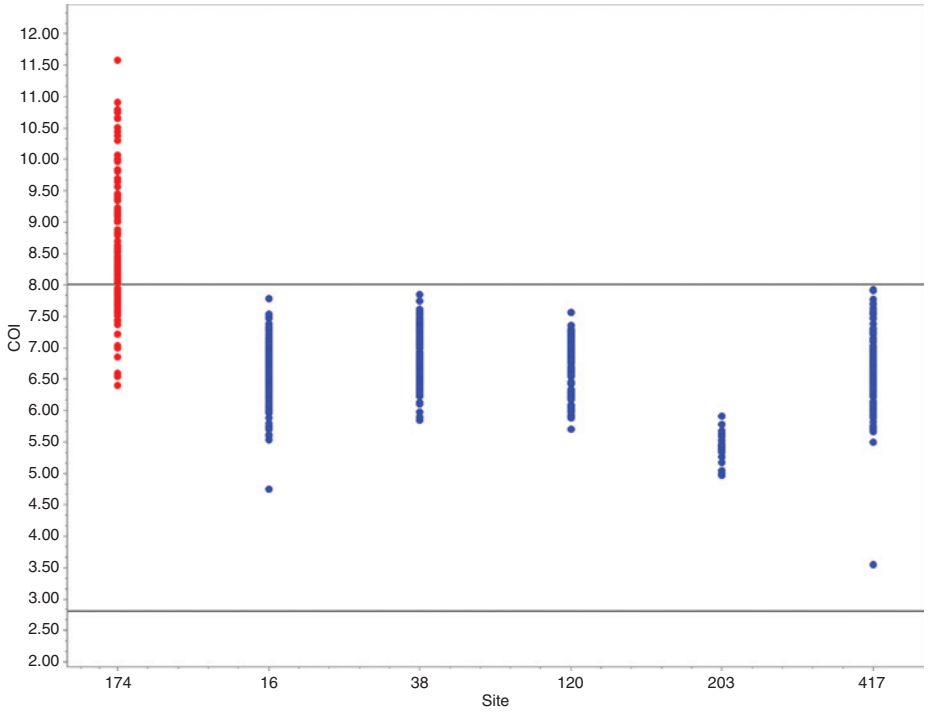


Figure 2: Scatter plot of the test results, expressed as signal to cutoff index (COI), from all participants testing QConnect Red lot number 434212 in the Roche Elecsys HBsAg II ECLIA on the same reagent lots used by Participant 174 over the period 1 January 2015 to 31 December 2015.
The red dots indicate results reported by the participant under investigation.

USA; and the German RiliBÄK standard and Australian QConnect Limits. All QC results in the data sets were from tests deemed valid by the manufacturers' instruction for use (IFU).

All test systems experience "normal" variation. There are many sources of variation including, but not limited to, changes in reagent lots, instrument and equipment calibration and maintenance, operator processes, storage and transport conditions of reagents and consumables and/or environmental conditions [12]. A comprehensive study by Algeciras-Schimnich demonstrated significant reagent lot variation in insulin-like growth factor 1 assays across several years, leading to changes in the percentage of patient results reported as being unexpectedly high [19]. Reagent lot change is the major source of variation in infectious disease serology testing [12, 20], and unlike clinical chemistry, variation between reagent lots cannot be corrected by recalibration. However, it is unlikely that reagent lot change contributes significantly to changes in clinical sensitivity or specificity of infectious disease assays.

The guideline published by Public Health England, which is specific for infectious disease testing, is proscriptive and directs laboratories to "test the control material in 20 separate assay runs" to establish mean ± 2 SD and to use the Westgard rules to monitor QC results [4]. The CLSI guidelines state that limits should initially be established using the mean ± 2 SD of the first 20 QC test results [3, 7]. Once control limits are established, tools including Westgard rules can be applied to data to determine if future results are acceptable [3, 6, 7, 17]. The more recent CLSI guideline suggests that QC limits should be recalculated once several months' of data are collected [7]. The RiliBÄK standard [10] states that 15 QC results should be used to establish QC limits using a specific calculation.

These standards have been developed for and used in clinical chemistry testing for decades and have been adopted for infectious disease serology testing for various reasons, such as the introduction of automation allowing such testing to be performed in a central or "core" laboratory and the lack of an alternative method to establish control limits for infectious disease serology. This study is the first to systematically assess the suitability of published guidelines for establishing QC limits. Indeed, only four previous studies focusing on QC for serological assays, other than those originating from NRL, were found in a literature search of English language publications [20–23] since the mid-1990s.

When a new infectious disease serology assay reagent lot is introduced, frequently the results of the QC consistently fall either above or below the mean and often fall outside the 2 SD range established using the first 20 results. The present study confirms this theory with more than 64% of data sets failing the 10_X rule more than 20% of the time, and 26.2%, 21.4%, 28.2% and 47.6% failing 1_{2S}, 1_{3S}, 2_{2S} and 4_{1S} rules more than 20% of the time, respectively. Therefore, it is concluded that using the first 20 data points to establish control limits based on mean ± SD is not appropriate, as the limits established using this method include insufficient "normal" variation to predict future test results.

Using a larger number of data points (in our case the first 100 results) increases the amount of "normal" variation when establishing the control limits and therefore increases the range of the control limits. Although this approach reduced the percentage of data sets with more than 20% failures, 40.5% of all data sets failed the 10_X rule and 20.3% failed the 4_{1S} rule rendering the universal application of Westgard 100 rules inappropriate. The selective use of the 1_{2S} (12.7% failing more than 20% of the time), 1_{3S} (2.5%), 2_{2S} (6.3%) and 4_{4S} (0.0%) rules could be deemed appropriate. However, only 79 of the 103 data sets investigated had more than 100 QC results. Also, if testing on a daily basis, collection of results to establish the control limits would take more than 3 months. Most standards are silent on what method a laboratory should use to monitor their QC results while they are collecting sufficient data to establish control limits.

The RiliBÄK standard is essentially a variation of mean ± 3SD, if the correction for bias is removed as is necessary for infectious disease serology. Using mean ± 3SD to establish control limits increases the range compared with ±2 SD, thereby reducing the number of failures. About 24% of data sets had a failure rate of >20%, and 17.5% of data sets had a failure rate between 10% and 20%. We believe this is still unacceptably high for QC results originating from valid test runs. The advantage of the RiliBÄK standard is that it uses only 15 results to establish and is easy to apply, although 15 results is still not representative of future results.

By contrast, the QConnect Limits, established by NRL for each assay/QC combination, had fewer failures and can be applied to all data from the time of introducing a new lot of QC. All but eight of the 103 data sets (7.7%) analyzed had QConnect Limit failure rates of less than 10%. Investigations into the two data sets (1.9%) that reported >20% failure rate indicated that the variation experienced was laboratory specific. Due to the frequent changes in infectious disease assay reagent lot numbers and the consequential change in QC reactivity, the approach to setting control limits based on small data sets is not fit for purpose. When the mean ± SD values are established using

approximately 20 QC test results, the resulting control limits will only be suitable until the introduction of a reagent lot that performs differently. The person monitoring the QC test results must then decide on their response. Anecdotally, actions range from approving the QC test results and ignoring the failure; resetting the limits [20]; retesting the QC sample or recalibrating the instrument with the hope it will change the QC result; or reporting the situation to the manufacturer, initiating an expensive and often futile investigation [6, 15, 24]. In extreme circumstances, laboratories have requested certain reagent lots to be exchanged, even though the assay is performing within manufacturers' specifications.

It is theoretically possible to recalculate the mean ± 2 SD at the change of every reagent batch to avoid the influence of reagent batch-to-batch variation. This approach was suggested previously [20] where a small set of data were investigated. In this study, Westgard$_{20}$ rules were applied to nine assays, with only 60 failures from 1808 test runs (3.3%). However, only one assay experienced a reagent lot change and many of the failures were associated with this change. The authors recalculated the mean and SD for the new lot.

It is not possible to determine if the failures flagged in this study by any of the methods described were "true" or "false". Further experiments would be required. What is known is that the QC results examined were derived from test runs that were valid by the manufacturers' IFU. The high percentage of failures using the Westard$_{20}$, Westgard$_{100}$ and RiliBÄK approaches was therefore unexpected. It is possible that the QConnect Limit approach failed to detect "true failures". However, all the results analyzed were obtained from test runs validated by the manufacturer's instructions, and the investigations into the two data sets with >20% failures using QConnect Limits did highlight laboratory specific issues.

The use of QC to monitor serological assays is highly recommended [1–4, 6]. The traditional approach to QC monitoring using mean ± 2 SD and Westgard rules does not fulfill these criteria when applied to infectious disease serology testing. The Public Health, England guideline [4], which is specific for virology and serology laboratories testing for infectious diseases, suggests using mean ± 2 SD without caveat. This guideline should be retracted and replaced. Other guidelines [7] that suggest using mean ± 2 SD and then periodically recalculating using additional QC results should be modified to specify the limitations of this approach. Those laboratories applying RiliBÄK standards [10] to monitor QC results should be aware of its limitations. Our investigation indicated that using QConnect Limits was easily applied without delay of collecting results and is an appropriate method for monitoring QC results from infectious disease serology testing.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Australian Standards. Medical laboratories – requirements for quality and competence. Vol. AS ISO 15189-2013. Sydney, Australia: Standards Australia, 2013.
2. CLSI. Laboratory quality control based on risk management; approved guideline. Vol. EP23-A. Wayne, PA: CLSI, 2011.
3. CLSI. Statistical quality control for quantitative measurement procedures: principles and definitions; approved guidelines, 3rd ed. Vol. C24-A3. Wayne, PA: CSLI, 2006.
4. Public Health England. Quality assurance in the diagnostic virology and serology laboratory. In: Standards Unit MS, editor, Vol. Colindale, UK: Public Health England, 2015.
5. Parvin CA. Quality-control (QC) performance measures and the QC planning process. Clin Chem 1997;43:602–7.
6. Kinns H, Pitkin S, Housley D, Freedman DB. Internal quality control: best practice. J Clin Pathol 2013;66:1027–32.
7. CLSI. Statistical quality control for quantitative measurement procedures: principles and definitions. Vol. CLSI Guideline C24. 4th ed. Wayne, PA: Clinical and Laboratory Standards Institute, 2016.
8. Westgard JO. Selecting appropriate quality-control rules. Clin Chem 1994;40:499–501.
9. Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. Clin Chem 1981;27:493–501.
10. German Medical Association. Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK" (unauthorized translation). J Lab Med 2015;39:26–69.
11. Dimech W, Walker S, Read S, Smeh K, Karakaltsas M, Dent B, et al. Comprehensive quality control programme for serology and nucleic acid testing using an internet-based application. Accred Qual Assur 2004;9:148–51.
12. Walker S, Dimech W, Kiely P, Smeh K, Francis B, Karakaltsas M, et al. An international quality control programme for prism chemiluminescent immunoassays in blood service and blood product laboratories. Vox Sang 2009;97:309–16.
13. Dimech W, Vincini G, Karakaltsas M. Determination of quality control limits for serological infectious disease testing using historical data. Clin Chem Lab Med 2015;53:329–36.

14. Westgard JO. Internal quality control: planning and implementation strategies. Ann Clin Biochem 2003;40:593–611.
15. Badrick T. The quality control system. Clin Biochem Rev 2008;29(Suppl 1):S67–70.
16. Parvin CA, Jones JB. QC design: it's easier than you think. MLO Med Lab Obs 2013;45:18, 20, 22.
17. Westgard JO. Statistical quality control procedures. Clin Lab Med 2013;33:111–24.
18. Westgard JO, Westgard SA. Quality control review: implementing a scientifically based quality control system. Ann Clin Biochem 2016;53:32–50.
19. Algeciras-Schimnich A, Bruns DE, Boyd JC, Bryant SC, La Fortune KA, Grebe SK. Failure of current laboratory protocols to detect lot-to-lot reagent differences: findings and possible solutions. Clin Chem 2013;59:1187–94.
20. Gray JJ, Wreghitt TG, McKee TA, McIntyre P, Roth CE, Smith DJ, et al. Internal quality assurance in a clinical virology laboratory. II. Internal quality control. J Clin Pathol 1995;48:198–202.
21. Green GA, Carey RN, Westgard JO, Carten T, Shablesky L, Achord D, et al. Quality control for qualitative assays: quantitative QC procedure designed to assure analytical quality required for an ELISA of hepatitis B surface antigen. Clin Chem 1997;43:1618–21.
22. Kim J, Swantee C, Lee B, Gunning H, Chow A, Sidaway F, et al. Identification of performance problems in a commercial human immunodeficiency virus type 1 enzyme immunoassay by multiuser external quality control monitoring and real-time data analysis. J Clin Microbiol 2009;47:3114–20.
23. Kruk T, Ratnam S, Preiksaitis J, Lau A, Hatchette T, Horsman G, et al. Results of continuous monitoring of the performance of rubella virus IgG and hepatitis B virus surface antibody assays using trueness controls in a multicenter trial. Clin Vaccine Immunol 2012;19:1624–32.
24. Howanitz PJ, Tetrault GA, Steindel SJ. Clinical laboratory quality control: a costly process now out of control. Clin Chim Acta 1997;260:163–74.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2018-0351).

The final publication is available at https://doi.org/10.1515/cclm-2018-0351

Chapter Four - Conclusion

The two publications comprising **Chapter 4** build on each other. The first describes QConnect™ principles and offers it as an alternative to traditional QC processes, including Westgard rules. Our approach uses large data sets and includes all normal sources of variation, including reagent lot changes, over time. This publication demonstrates that "*QC method described will improve the outcomes of QC monitoring because it: 1) provides evidence-based limits developed from tens of thousands of test results over a long period of time; 2) gives clear guidance to the method users; and 3) simplifies and demystifies QC, making it more accessible to laboratory workers*". When traditional methods for QC monitoring are applied to serology testing, laboratory staff are commonly faced with an unacceptable level of rejections and are obligated by accreditation and good laboratory practice to investigate the source of these issues. This results in a waste of staff time and a frustration by scientific staff. While professional guidance documents, designed for clinical chemistry, requires the laboratories to implement traditional QC methods, the outcome results in confusion as to why they cause false rejections, or lead to inappropriate and/or unvalidated manipulation of the QC methods to limit the negative impact.

The studies presented in **Chapter 4** provides evidence that will support the acceptance of QConnect™ as an alternative approach to traditional methods. The second paper provides further evidence that QConnect™ is a more appropriate QC approach for infectious diseases, and that Westgard rules and other traditional methods were not-fit for-purpose. This study supported the theory that Westgard and RiliBÄK rules generated an unacceptably high levels of QC rejections, which would cause infectious disease laboratories employing these methods considerable waste in resources investigating inappropriate QC rejections. Of the 103 data sets analysed, the number of data sets that had more than 20% of QC results fail Westgard rules when the first 20 data points were used to establish the acceptance limits ranged from 3 (2.9%) for R_{4S} to 66 (64.1%) for 10_X rule. The number of data sets that had more than 20% of QC results failing each Westgard rule when the first 100 datapoints were used to establish the acceptance limits ranged from 0 (0%) for R_{4S} to 32 (40.5%) for 10_X rule. By contrast, the number of data sets with more than 20% failing the RiliBÄK control limits was 25 (24.3%). QConnect™ identified only two of the 103 data sets as having more than 20% of the QC results rejected. On further investigation, the rejection of these two data sets were found to identify testing issues within the reporting laboratory, with the majority of results submitted by those laboratories being outside the QConnect™ limits. Other laboratories, testing the same assay/QC combination reported all results within the QConnect limits. Therefore, QConnect™ accurately identified unacceptable variation in the relevant test system.

The impact of these two studies is significant. The QConnect™ concept is the first scientifically validated method for monitoring QC test results for infectious diseases. The data collected by EDCNet™ is continually reviewed by NRL staff and the data periodically added to the historical dataset used to calculate the QConnect limits. In this way, unexpected QC results can be investigated proactively, reducing the risk of a catastrophic failure of testing. The second study provides evidence that the QConnect™ concept is fit for purpose and therefore offers testing laboratories a genuine alternative to traditional QC methods. The success of this approach is demonstrated in **Chapter 5**. The concept is now used by several hundred laboratories globally and has been adopted by two major QC manufacturers through contractual licensing agreements, Technopath Clinical Diagnostics (Ballina, Ireland) and DiaMex (Heidelberg, Germany).

However, acceptance of QConnect™ as an alternative to Westgard rules must overcome entrenched opinions of professional bodies. NRL staff will advocate for change through publications, presentations at international conferences, communications directly to key opinion leaders and professional bodies and continue to publish evidence-based studies. Further studies reviewing the cost-benefit of the QConnect™ approach will be undertaken with key opinion leaders in diagnostic and blood screening laboratories. An

opinion piece, highlighting the need for an alternative to traditional methods for QC, authored by myself and colleagues in Europe and the USA is in draft. Some acceptance is evident with the Public Health England, who referenced QConnect™ in their standard for quality assurance in the diagnostic infection sciences laboratory (128) as an alternative approach to establishing QC limits stating "*However, other methods available include, but are not limited to, use of national guidelines such as the Richtlinien der Bundesärztekammer (RiliBÄEK) guidelines used in Germany or the use of larger datasets such as the entire historical dataset for a given assay/QC combination*" referencing the papers presented in **Chapter 4**.

Chapter Five: Understanding the relationship between QC and patient samples

Monitoring variation of a test system using QC measures the amount of imprecision in the test system over time. If the mean of QC test results from a laboratory is compared with the mean of QC results of the peer group (laboratories using the same QC and assay combination) then a sense of the inherent bias can be determined. However, there was an unresolved question about the commutability of the QC samples and patient results; that is, if QC results were unexpectedly high or low due to a source of variation, does this increase the possibility of false positive or negative patient results? Put another way, can QC results predict changes in sensitivity and specificity of a test system. Several co-authored studies were undertaken to test this assumption. These studies used QC data and blood donor results from serological testing for blood-borne infections, the majority of which were negative. In one set of studies, changes in the QC sample reactivity were mapped against the changes in reactivity of a population of blood donor samples screened negative for blood-borne viruses (158-160). No significant commutability between the QC and negative donor samples was observed. This is not surprising because, in serological assays screening for blood-borne infections, by design, the mean negative donor result is far removed from the assay cut-off. Additionally, any donor at risk of infection is deferred prior to donation by a questionnaire of risk factors, and any donor with an initial biological false reaction is also deferred from future donations. It is noted that this study was performed on only tests for HBsAg and anti-HIV antibodies from the same assay manufacturer (158). Assays with different characteristics may have a different outcome, so the outcomes cannot be generalised. A method for the visualisation of the relationship between QC and donor sample results, as well as other correlations, have previously been described (159, 160).

More important was the question of the possibility of a false negative patient result if the QC reactivity decreased significantly. True low positive IgG serology test results are found during early infection when the immune response is developing, and late in infection when antibodies may wane. On exposure to an antigen, the immune response develops a humoral response, creating circulating IgG antibodies rising to a detectable level (window period). Once detectable, the low-level IgG reactivity within that test system lasts less than 72 hours, as the levels of IgG rise rapidly during early infection. It is during this period that a false negative result may be possible if a reagent lot produces lower than expected reactivity. The QC samples provided by NRL are tested daily by the participating laboratories, and the results are expected to fall within the QConnect™ limits. The QC results that fall outside the QConnect™ limits indicates unacceptable variation. In this paper **(Chapter 5)**, I used an unexpected decrease in QC results associated with a change of reagent lots of a commercial anti-HCV assay to assess the likelihood of false negative patient results (126).

To investigate if the change associated with the affected reagent lot numbers changed the reactivity of true, low positive patient samples, we tested selected seroconversion panel samples. Seroconversion panels are samples collected from individuals known to be undergoing a seroconversion event. Blood samples are drawn from the individuals from the time of exposure to the time of a detectable antibodies. Of these panels, samples with very low, true positive reactivity were tested in both the affected and unaffected reagent lots and the results compared. Six reagent lots were found to produce lower than expected QC reactivity. A total of 44 low-positive seroconversion samples, obtained from 35 different individuals, were tested in affected and unaffected assay lots. Only three of the 44 samples reported results below the assay cut-off when tested on two of the six affected assay lot. A further sample had results below the cut-off for only one affected lot. The risk of false negative results when screening clinical or donor populations due to changes in reagent lot reactivity was concluded to be low, as the possibility of a true seroconversion sample

at very early stages of disease is rare and less than 10% of these would be falsely negative. The study provides evidence that QConnect™ acceptance criteria are effective in identifying true analytical errors, even further reducing the risk of changes in reagent reactivity.

Wayne J. Dimech*, Giuseppe A. Vincini, Liza M. Cabuang and Megan Wieringa

Does a change in quality control results influence the sensitivity of an anti-HCV test?

https://doi.org/10.1515/cclm-2020-0031
Received January 12, 2020; accepted February 7, 2020; previously published online March 7, 2020

Abstract

Background: Laboratories use quality control (QC) testing to monitor the extent of normal variation. Assay lot number changes contribute the greatest amount of variation in infectious disease serology testing. An unexpected change in six lots of an anti-HCV assay allowed the determination of the effect these lot changes made to the assay's clinical sensitivity.

Methods: Two sets of seroconversion samples comprising of 44 individual samples and 9 external quality assessment scheme (EQAS) samples, all positive to anti-HCV, were tested in affected and unaffected assay lots, and the difference in the quantitative and qualitative results of the samples was analyzed.

Results: Of 44 low-positive seroconversion samples tested in affected and unaffected assay lots, only three samples had results reported below the assay cutoff when tested on two of the six affected assay lot. A further sample had results below the cutoff for only one affected lot. None of the EQAS samples reported false-negative results. Samples having a signal to cutoff value of less than 6.0 generally had lower results in the affected lots compared with the unaffected lots.

Conclusions: Unexpected changes in QC reactivity related to variation, in particular assay lot changes, may affect patient results. This study demonstrated that QConnect Limits facilitated the detection of an unexpectedly large variation in QC test results, allowed for the identification of the root cause of the change, and showed that the risk associated with the change was low but credible. The use of evidence-based QC program is essential to detect changes in test systems.

***Corresponding author: Wayne J. Dimech,** NRL, Australia, 4th Floor Healy Building, 41 Victoria Parade, Fitzroy, 3065 Melbourne, VIC, Australia, Phone: +61 3 9418 1132, Fax: +61 3 9418 1155, E-mail: wayne@nrlquality.org.au
Giuseppe A. Vincini and Liza M. Cabuang: NRL, Australia, Melbourne, VIC, Australia
Megan Wieringa: Monash Health, Monash Medical Centre, Melbourne, VIC, Australia

Keywords: anti-hepatitis C assay; clinical sensitivity; quality control; seroconversion.

Introduction

When monitoring infectious diseases serological testing quality control (QC) results, a change in assay lot number is often associated with a change in reactivity of the QC sample [1–4]. Although this change should be considered normal in most cases, occasionally a QC reactivity change associated with the new assay lot is greater than previously experienced. This QC reactivity change raises the question, what amount of QC reactivity change is allowable before there is an increased probability of reporting an incorrect patient/donor result? The question assumes that the QC results change proportionally with those of the patient/donor specimen. This assumption is yet to be tested in a systematic manner.

QC samples are usually manufactured by diluting a sample with a known, high concentration of the analyte into a negative matrix, often pooled human plasma screened negative to the analyte in question. The QCs typically have low-level reactivity on the test platforms for which they are designed and are often considered to mimic the level of reactivity of a seroconversion sample. However, during a seroconversion event, the antibodies are immature, have low avidity and affinity, and are often produced against a limited number of specific immunogenic proteins. In contrast to those early antibodies, sera from chronically infected individuals used to manufacture the QC samples have a mature antibody response, with high avidity and reactivity to a broader range of antigens and in some cases representing different antibody isotopes to that seen with seroconversion response.

NRL, Australia (www.nrlquality.org.au), provides an international QC monitoring program for laboratories testing for infectious diseases (QConnect; NRL, Australia) [3]. QConnect uses QC samples that are optimized for each test platform; provides access to an online, peer comparison, QC monitoring program (EDCNet, NRL, Australia; https://edcnet.nrlquality.org.au); and incorporates a novel process for establishing control limits, which accounts for normal assay lot-to-lot variation as well as

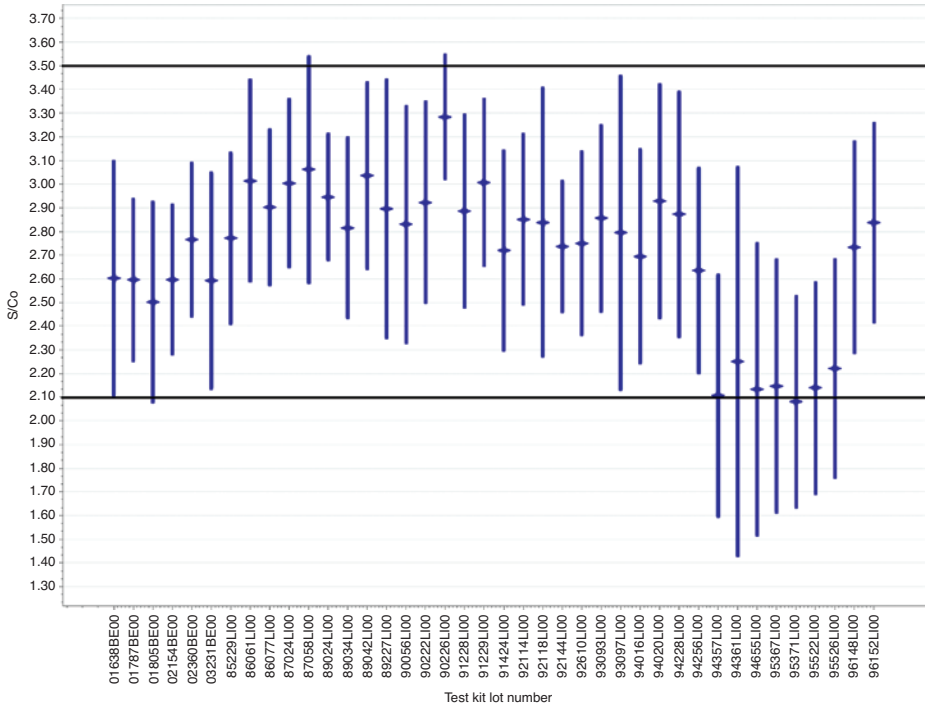


Figure 1: Representative results of QConnect Blue, tested in different lots of ARCHITECT HCV assay, where the central bar is the mean of the QC test results and the vertical bars represent the mean ± 2 standard deviations.
The S/Co value is presented in the y axis and the assay lot numbers on the x axis. Horizontal lines indicate the NRL acceptance range "Connect Limits" for QConnect Blue QC sample testing in the Abbott Architect anti-HCV CMIA.

other sources of variation (QConnect Limits) [5]. However, various international guidelines for monitoring QC results are primarily written for clinical chemistry and do not provide guidance for managing the effects of assay lot changes on serological QC results [6–9].

In February 2019, several assay lots of the Abbott ARCHITECT anti-HCV CMIA (Abbott Laboratories, Chicago, IL, USA), henceforth referred to as ARCHITECT HCV, were associated with a decrease in QConnect Blue (DiaMex GmbH, Heidelberg, Germany), a multimarker QC sample designed to monitor the performance of ARCHITECT blood screening assays. The decrease in reactivity of QConnect Blue was much greater than previously experienced with other ARCHITECT HCV assay lot numbers. From February 2018 to the release of the affected assay lots in February 2019, QConnect Blue was tested on 45 separate ARCHITECT HCV assay lots, each having at least 30 QC tests, for a total of 19,158 results; the overall mean being signal to cutoff (S/Co) 2.81. The mean QC result for each individual assay lot ranged from S/Co 2.35 to 3.08. The QConnect Limits of this QC/Assay combination were S/Co 2.1–3.5 (Figure 1).

There were seven affected assay lots. The mean of 2086 QConnect Blue results on these assay lots ranged from S/Co 2.06 to 2.22 (Figure 1), with the majority of results below the lower QConnect Limit. Using this event, a study was designed to determine whether a change in QC reactivity had a corresponding effect on the quantitative results of seroconversion samples and if there was an increased likelihood of reporting a false-negative patient/donor result when using these assay lot numbers.

Materials and methods

The study was designed in two parts: the comparison of the reactivity of early seroconversion samples tested on affected and unaffected assay lots and the change in reactivity of external quality assessment scheme (EQAS) samples on the same assay lots.

Seroconversion samples

Two sets of seroconversion samples were tested on both affected and unaffected assay lots. In the first set, a total of 25 seroconversion samples obtained from 16 individuals, each having an S/Co of less than 10 reported by the sample supplier or in previous testing at NRL, were tested in three ARCHITECT HCV assay lots: 93093LI00, 95367LI00, and 94655LI00. Assay lot 93093LI00 was not affected by the change in reactivity, whereas assay lots 95367LI00 and 94655LI00 demonstrated a decrease in reactivity of the NRL QConnect Blue QC sample. Each seroconversion sample was tested in singlicate in parallel with the QConnect Blue QC.

The second set of seroconversion samples consisted of 19 individual samples obtained from commercial suppliers, Zeptometrix

(Buffalo, NY, USA) and SeraCare (Milford, MA, USA). Each sample was tested in singlicate on unaffected assay lot 94020LI00 and on four of the six affected lots – 94655LI00, 95526LI00, 94361LI00, and 95367LI00. Of the 19 samples, 6 were also tested on the other two affected lots – 95367LI00 and 95371LI00 only because of the low volume of specimen. Results of each sample tested in each assay lot were compared.

EQAS samples

During the period when laboratories were using the affected assay lots, NRL conducted an EQAS test event. The results of two EQAS (hepatitis serology [HEPM] and multimarker blood screening [MMBS] schemes) were extracted and reviewed. The panel samples for each of the two programs were pooled, undiluted, citrated human plasma obtained from blood donors confirmed as anti-HCV positive. These samples are considered representative of the reactivity normally encountered in a clinical or donor screening setting. The results reported by the participants for each of the four positive panel samples from HEPM and five positive samples from MMBS using different assay lots of the ARCHITECT HCV were compared. Only assay lots with five or more test results and from two or more different laboratories were included in the analysis.

Time of seroconversion

Where available, information provided in the certificates of analyses of the seroconversion panels used in the study was reviewed to determine the period when low-level reactivity lasts. A total of eight seroconversion panels tested in the ARCHITECT HCV and further two panels tested on the Abbott AxSYM had one or more bleeds having an S/Co value between 1.0 and 2.0. The number of days between that bleed and the next subsequent bleed that had an S/Co value >3.0 was determined.

Results

Seroconversion samples

The results of testing the first set of 25 seroconversion samples and the QConnect Blue QC sample are presented in Figure 2. All but one sample had an S/Co value of less than 12, and 19 of 25 samples had an S/Co value of less than 10 when tested on the unaffected assay lot 93093LI00 (Table 1). The reactivity of QConnect Blue was S/Co 3.15 when tested on assay lot 93093LI00, but S/Co 1.98 and 1.74 when tested on affected assay lots 95367LI00 and 94655LI00, respectively. All but three samples had a lower reactivity on assay lot 95367LI00, and all 25 samples had lower reactivity on assay lot 94655LI00 when compared with results obtained when testing on assay lot 93093LI00. No sample results were less than the cutoff (S/Co 1.0) when tested on assay lot 93093LI00; however, four sample results were below the cutoff when tested on assay lot 94655LI00, and three of these four samples were also below the cutoff when tested on assay lot 95367LI00. These results would be interpreted as nonreactive as per the manufacturer's instructions for use (IFU). However, three of the four samples had reactivity of S/Co of 0.8 or above and, thus, may have undergone scrutiny depending on individual laboratory validation of test results. One sample had a high negative reactivity of S/Co 0.62 in one assay lot.

None of the results of the second set of seroconversion samples had reactivity below the cutoff when tested in any of the four affected assay lots (Table 2). The

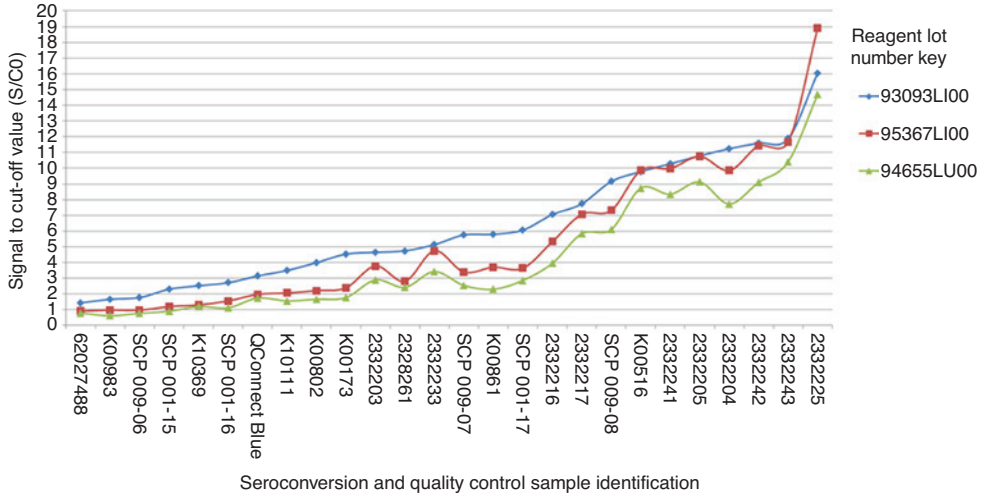


Figure 2: Graph of test results of 25 seroconversion samples (set 1) and QConnect Blue tested in three different ARCHITECT HCV assay lot numbers, sorted in ascending order of the sample reactivity on unaffected assay lot number 93093LI00.

Table 1: Results of seroconversion samples (set 1) and QConnect Blue on three different assay lots of ARCHITECT HCV assay.

Patient number	Seroconversion panel number	Sample identification	Abbott ARCHITECT anti-HCV assay lot number (S/Co)		
			93093LI00	95367LI00	94655LI00
1	N/A	62027488	1.44	0.92	0.80
2	N/A	K00173	4.54	2.40	1.78
3	N/A	K00516	9.78	9.86	8.73
4	N/A	K00802	4.00	2.22	1.67
5	N/A	K00861	5.80	3.71	2.31
6	N/A	K00983	1.67	0.98	0.62
7	N/A	K10111	3.50	2.07	1.57
8	N/A	K10369	2.53	1.31	1.21
9	SCP-HCV-009	SCP 009-06	1.78	0.98	0.78
9	SCP-HCV-009	SCP 009-07	5.76	3.40	2.55
9	SCP-HCV-009	SCP 009-08	9.16	7.32	6.11
10	SCP-HCV-001	SCP 001-15	2.31	1.21	0.91
10	SCP-HCV-001	SCP 001-16	2.72	1.56	1.13
10	SCP-HCV-001	SCP 001-17	6.07	3.64	2.87
11	549	Z328261	4.75	2.82	2.43
12	541	Z332203	4.65	3.79	2.9
12	541	Z332204	11.24	9.85	7.73
12	541	Z332205	10.79	10.75	9.12
13	542	Z332216	7.05	5.33	3.96
13	542	Z332217	7.74	7.07	5.84
14	543	Z332225	16.05	18.9	14.68
15	9045	Z332233	5.14	4.75	3.43
16	9047	Z332241	10.28	9.98	8.35
16	9047	Z332242	11.6	11.43	9.11
16	9047	Z332243	11.88	11.65	10.41
QConnect BLUE (DM18210)			3.15	1.98	1.74

Table 2: Results of seroconversion samples (set 2) and QConnect Blue on three different assay lots of ARCHITECT HCV assay.

Seroconversion panel number	Panel supplier	Abbott ARCHITECT anti-HCV assay lot number (S/Co)				
		94655LI00	95526LI00	94361LI00	95367LI00	94020LI00
HCV-6229	Zeptometrix	1.39	1.62	1.14	1.57	1.67
HCV-6225	Zeptometrix	1.51	1.18	N/A	N/A	1.78
HCV-6226	Zeptometrix	1.55	1.57	1.53	1.92	2.10
HCV-9058	Zeptometrix	1.70	1.63	1.45	1.90	2.30
HCV-6224	Zeptometrix	1.87	1.89	1.99	2.00	2.39
HCV-10041	Zeptometrix	2.14	2.02	2.21	1.95	2.53
HCV-6215	Zeptometrix	2.48	2.23	2.35	2.66	2.78
PHV913	BBI Diagnostics	2.13	2.23	2.08	2.61	3.26
HCV-6216	Zeptometrix	3.04	2.62	2.66	2.47	3.39
HCV-9045	Zeptometrix	2.38	2.32	2.93	2.56	3.47
HCV-9047	Zeptometrix	3.95	3.97	4.23	4.64	4.69
PHV923	BBI Diagnostics	3.62	3.59	3.3	3.45	4.96
HCV-10165	Zeptometrix	3.52	3.43	4.04	4.02	5.27
HCV-6222	Zeptometrix	5.11	5.40	5.36	5.15	6.26
PHV919	BBI Diagnostics	6.65	6.20	6.71	5.73	6.72
HCV-6227	Zeptometrix	6.70	6.34	5.96	6.35	6.92
PHV906	BBI Diagnostics	7.60	6.89	7.36	7.90	8.13
HCV-10185	Zeptometrix	11.58	11.21	11.57	11.45	11.25
HCV-9046	Zeptometrix	11.20	12.60	12.91	13.36	12.79

result of the 19 samples had reactivity ranging from S/Co 1.67 to 12.79 when tested in the unaffected assay lot 94020LI00 (Figure 3). One sample (HCV-6225) had insufficient volume to test on affected assay lots 94361LI00 and 95367LI00. There were seven samples with a result less than S/Co 3.0 when tested on the unaffected assay lot 94020LI00. The lowest result obtained from these seven samples was S/Co 1.14 when sample HCV-6229 was tested in assay lot 94361LI00. Of the 19 samples, 6 were tested on the other two lot numbers of affected reagents 94357LI00 and 95371LI00. When tested on the unaffected lot 94020LI00, the range of reactivity was S/Co 1.78–4.69. The sample with the lowest result on the unaffected lot reported S/Co results of 1.47 and 1.28 when tested on assay lots 94357LI00 and 95371LI00, respectively. None of the six samples had a result less than S/Co 1.0 on any assay lot number.

EQAS samples

The results of the positive EQAS samples tested on different assay lot numbers were analyzed by box and whisker outlier graphs (Figures 4 and 5). EQAS participants tested the four positive HEPM samples in unaffected assay lot 94256LI00 (13 participants) and three affected assay lots 94357LI00 (5 participants), 94655LI00 (13 participants), and 95367LI00 (8 participants). Five positive MMBS samples were tested on unaffected assay lots 92114LI00 and 94016LI00 (eight participants each) and affected lot 94357LI00 (three participants). The number of results reported for any individual sample for each assay lot ranged from 8 (MMBS for assay lot 94357LI00) to 30 (HEPM for assay lot 94256LI00).

Sample I had the lowest mean S/Co values for the HEPM EQAS program. HEPM sample I had a mean value of S/Co 10.26 for unaffected assay lot number 94256LI00 and higher S/Co values of 11.34, 11.94, and 12.51 for affected assay lots 94357LI00, 94655LI00, and 95367LI00, respectively. The mean value of each of the other three positive HEPM samples B, G, and C was higher when tested on three affected lots compared with the mean value when tested on the unaffected lots.

MMBS sample D had a mean S/Co value of 7.20 and 7.22 when tested on unaffected assay lot numbers 92114LI00 and 94016LI00 and a slightly lower mean value of 6.13 when tested affected assay lot 94357LI00. The other four other MMBS-positive samples H, L, O, and S each had mean S/Co greater than 10.0, and the mean results of each of these samples tested on the affected lot 94357LI00 were greater than the mean of each sample tested on unaffected lots 94016LI00 and 92114LI00. It is noted that the results of all HEPM and MMBS EQAS samples, except for MMBS sample D, reported a higher mean when tested in the affected assay lots compared with unaffected reagent.

Time of seroconversion

Only samples with S/Co value of less than 2.31 tested on the unaffected assay lot 93093LI00 had a nonreactive result when tested on affected lot 95367LI00 and/or 94655LI00. Of eight commercial seroconversion panels reviewed, all had at least one bleed having a result between S/Co 1.0 and 2.0. The average time before the subsequent bleed having a result of greater than S/Co 3.0 was 3.5 days (range 2–5 days). The average S/Co value of the subsequent bleed was 5.3 (range 3.73–7.0).

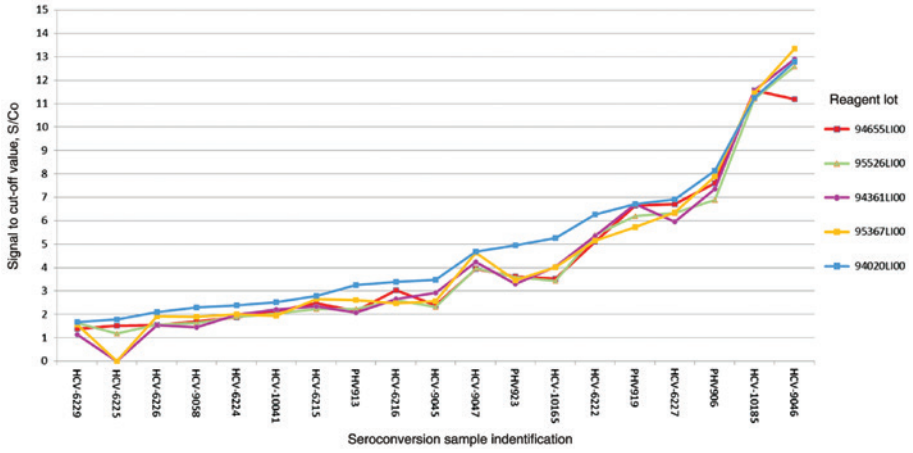


Figure 3: Graph of test results of 19 seroconversion samples (set 2) tested in five different ARCHITECT HCV assay lot numbers, sorted in ascending order of the sample reactivity on unaffected assay lot number 94020LI00.

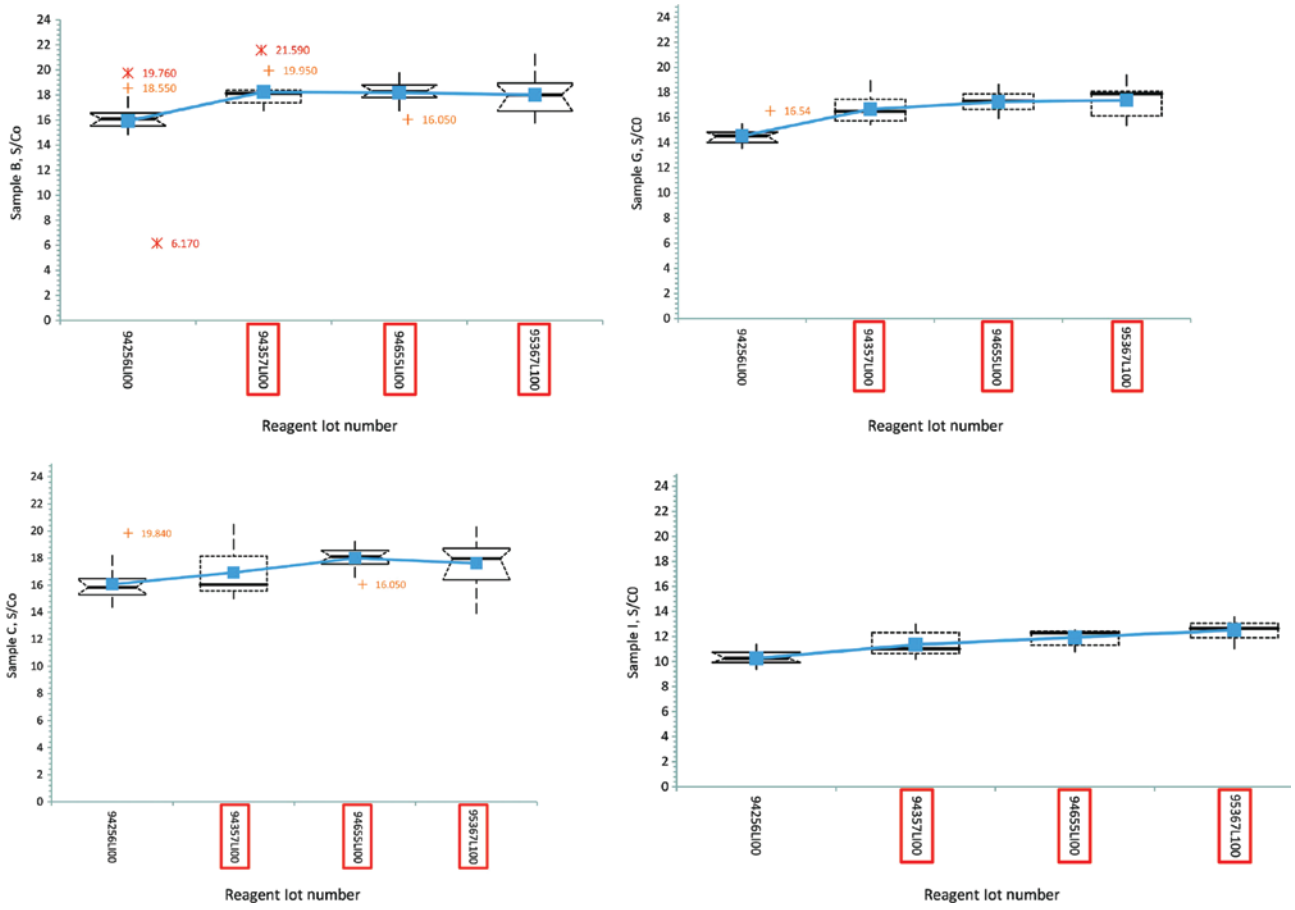


Figure 4: Box and whisker graphs of results reported by participants of the NRL HEPM EQAS when testing positive samples in four different lot numbers of the ARCHITECT HCV assay.

Discussion

All test systems experience normal, expected levels of variation, and laboratories should monitor this variation using an external third party QC, in addition to the manufacturer's kit controls. For infectious disease serology testing, changes in assay lot numbers are the main sources of normal variation. However, the introduction of new lots of reagent is often associated with a change in reactivity of the QC sample [5]. Traditional methods of establishing QC limits based on mean ± *x* standard deviations calculated on small numbers of data points, e.g. 20 results as recommended by international guidelines [6–9], has been demonstrated to be inappropriate for infectious disease serology [1]. This traditional approach to QC does not account for assay lot changes as a source of variation and is heavily reliant on assumptions of normal distribution of QC sample data and the commutability of QC results to patient testing results. NRL developed an alternative approach to setting control limits, called QConnect Limits, based on historical QC peer-group data [5]. The calculations for QConnect Limits often include up to tens of thousands of QC results obtained from many laboratories, instruments, and assay lot and, therefore, include all sources of variation experienced over time for an assay-QC peer-group.

An investigation into the root cause of QConnect Blue reactivity change in six assay lots of ARCHITECT HCV was initiated by the manufacturer, and a further investigation to determine the likelihood of false-negative test results associated with the affected assay lots was undertaken by NRL. Two sets of seroconversion samples were tested. Four of 25 seroconversion samples in set 1 reported nonreactive (S/Co <1.0) test results when tested on affected assay lot 94655LI00, and three of four samples had nonreactive results when tested on assay lot 95367LI00. None of the 19 seroconversion samples in set 2 had a result less than S/Co 1.0. No sample having S/Co of greater than 2.31 would have been reported as nonreactive on the affected lots.

As the S/Co value of the sample tested on unaffected assay lots increased, the difference between sample results tested in unaffected lots and from those tested on

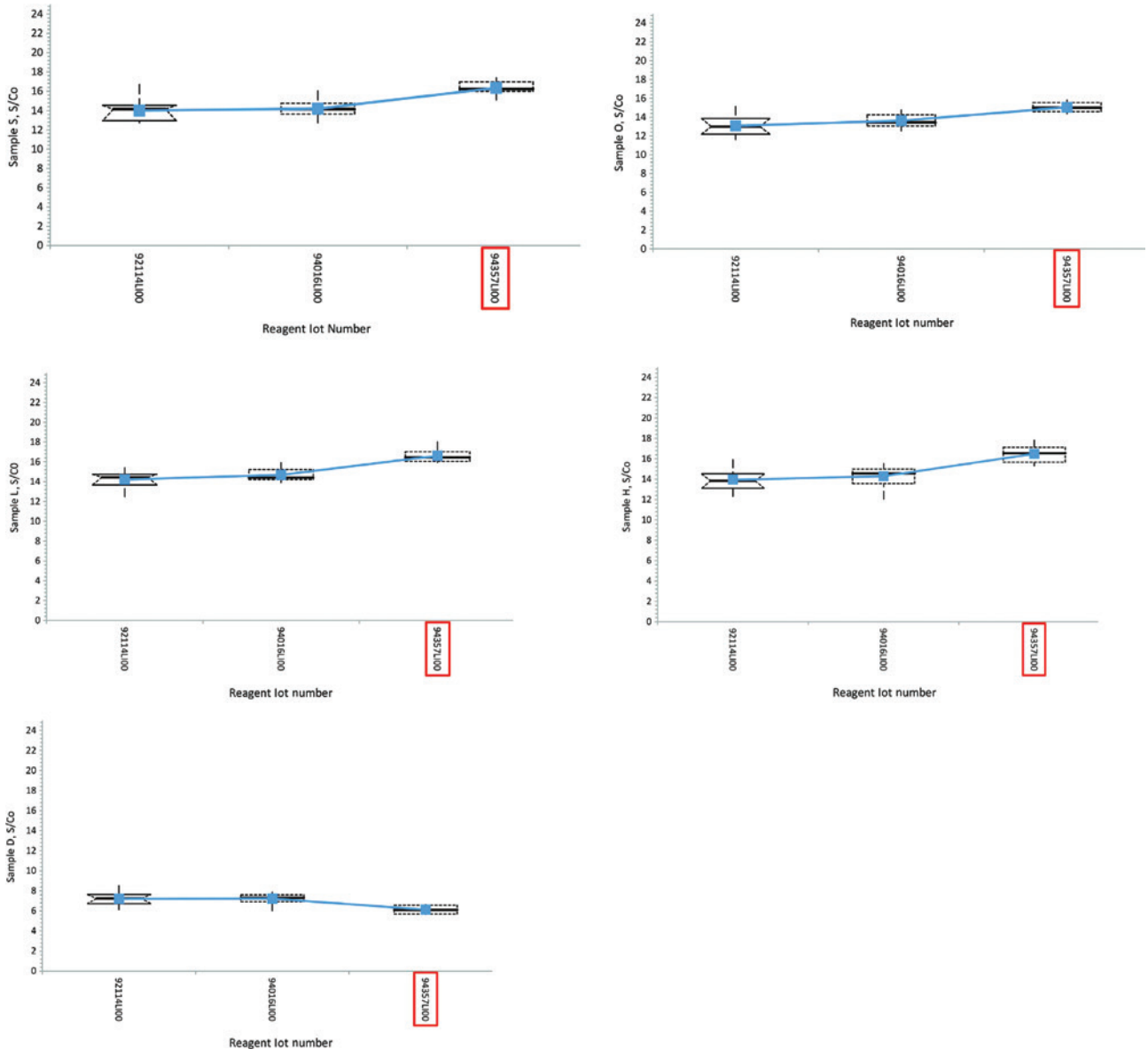


Figure 5: Box and whisker graphs of results reported by participants of the NRL MMBS EQAS when testing positive samples in three different lot numbers of the ARCHITECT HCV assay.

the affected decreased. Once the S/Co value was over 6.0 on the unaffected lot, the results from the affected lots were similar and in some cases higher. This finding was supported by the review of the EQAS results. Each of the positive samples included in the NRL EQAS (HEPM and MMBS) represented samples usually seen in chronically infected individuals. All four positive samples included in the HEPM scheme and four of the five positive samples included in the MMBS program had mean S/Co values that were lower when tested in the unaffected assay lots compared with testing on affected lots. These findings were not consistent with the decrease in reactivity observed when testing seroconversion samples.

This phenomenon may be due to the maturation of the antibody response and subsequent increased avidity, with immature antibodies being more prone to changes than a more mature antibody response, yet this theory was not examined further. However, the reason why low-level samples are more sensitive to assay lot change may be due to the dose-response curve of the assay. As the concentration of the analyte increases, the ability of that sample to demonstrate a significant S/Co change in the test system decreases.

Testing and monitoring the results of an independent third party QC sample is highly recommended [7–13] and is an effective way of detecting unexpected variation in a

test system. Results of QC samples are not a measure of clinical sensitivity and should not be used for this purpose. Although the anti-HCV antibody window period is approximately 6–7 weeks [14], true extremely low-level reactivity samples are only seen during a seroconversion event, and the low levels are present for approximately 48–72 h. Seroconversion samples have different antibody profiles compared with diluted chronically infected samples.

The current study is the only publication that systematically maps the changes in clinical sensitivity associated with an unexpectedly large change in a QC sample result, although specificity changes have been previously mapped without conclusion [15]. The incidence of naïve HCV infection in very high-risk populations varies considerably in the literature. Reported incidence of HCV infection from high-risk populations includes 7.6 and 31 per 100 person years [16, 17]. The incidence in routine diagnostic laboratories, including the chance of obtaining a bleed during the 48–72 h when the S/Co value is between 1.0 and 2.0, would be considerably lower. In a blood screening setting, processes for limiting the possibility of a false-negative serological result include predonation deferral processes, such as questionnaires detecting high-risk activities, and the use of nucleic acid testing. It is therefore extremely unlikely that a true seroconversion sample would be falsely reported as negative, and in these extremely rare cases, other clinical and diagnostic tools would be used to mitigate this situation, including clinical history, antigen-antibody combination assays, repeat testing of high-risk individuals, and, especially in blood screening, nucleic acid testing.

All in vitro diagnostic devices (IVDs) for the detection of anti-HCV and other high-risk devices undergo batch release by the manufacturer and stringent assessment, including performance testing, by regulatory bodies before being placed on the market. In Europe, all high-risk IVDs (IVD class D) such as for the detection of anti-HCV requires a premarket conformity assessment and individual lot release by a notified body according to the EU Regulations [18, 19]. In the United States, all high-risk devices (class III) must also be assessed by the FDA to determine the safety and effectiveness of the IVD, whereas those used for screening blood are assessed by the Center for Biologics Evaluation and Research using the 510(k) process [20]. Similar regulatory requirements are found in most countries with a developed IVD regulatory framework. Manufacturers must provide scientific evidence that the test system meets the intended use and that manufacturing processes comply with ISO 13485 or equivalent. It is important to note that the definition of the test system includes the manufacturer's method of validating the performance of the test, in particular the use of the manufacturer's positive and negative control validation criteria. The sensitivity and specificity provided by the manufacturer to the regulator, and stated in the IFU, relies on the manufacturer controls to be valid. It is therefore assumed that if the manufacturer's controls are within expected limits, the sensitivity and specificity of the assay is as reported in the IFU.

The six affected assay lots passed lot release testing by the manufacturer and the independent notified body and were therefore considered fit for use under the existing regulations. However, although stringent pre- and postmarketing controls on high-risk IVDs are in place, it is important that laboratories continually monitor the performance of their test systems. Variation can be introduced postmanufacture throughout the transport and storage of the reagents, changes in consumables, biological and nonbiological components of the test system, equipment and its maintenance (or lack of) and calibration, as well as human error. NRL QC program frequently detected these sources of variation, which can contribute to the potential of a false result as much as assay lot changes can [21]. This is even more the case for less regulated IVDs used to detect antibodies and antigens to infectious disease such as rubella and toxoplasma. Laboratories that do not use an evidence-based QC monitoring program independent of the manufacturer's kit controls are naive to these variations and, therefore, accept a greater risk of reporting compromised results. It is the responsibility of each laboratory manager to assess and manage the level of risk presented in their laboratory setting [6].

In this case, detection of variation allowed Abbott, Germany, to undertake a root cause analysis, identify a new batch of nonbiological component of the assay as being the source of variation, and act quickly to release new, unaffected lots. Using QConnect Blue results, NRL could confirm the resolution of the situation, with results falling within the QConnect Limits when tested on each subsequent assay lot.

There were several limitations with this study. Obtaining seroconversion samples of low reactivity is difficult and expensive. Each sample was performed in singlicate and not on all six assay lots. The study only presents data on a single manufacturing root cause of variation in QC results, and similar variation cannot be assumed when different changes in manufacturing occurs. NRL has evidence to confirm that all serology assays, irrespective of the manufacturers, experienced detectable, sometimes significant, assay lot-to-lot variations in manufacturing. It cannot be assumed that QC samples provided by other QC manufacturers have the same level of sensitivity to variation because the

QConnect series of QC samples are optimised to be reactive on the linear part of the assay's dose-response curve and are comprised entirely of human plasma, not diluted plasma or plasma-derived components (such as human/bovine serum albumin) in buffer.

However, the findings do represent the first detailed investigation into the relationship between reactivity of QC and patient samples when tested in a qualitative serology assay and the first time the assumption that a decrease in QC reactivity may be associated with the increased possibility of a false-negative patient result has been tested. The findings add further evidence that the QConnect Limits, based on historical QC peer-group data, are appropriate for monitoring infectious disease serology testing.

Acknowledgments: NRL would like to acknowledge Abbott Diagnostics for the collaboration in this study by testing seroconversion panel two and for sharing information that helped better understand the sources of variation. Also, NRL acknowledges DiaMex for performing the testing of some of the samples and contributing seroconversion samples.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Dimech W, Karakaltsas M, Vincini GA. Comparison of four methods of establishing control limits for monitoring quality controls in infectious disease serology testing. Clin Chem Lab Med 2018;56:1970–8.
2. Kruk T, Ratnam S, Preiksaitis J, Lau A, Hatchette T, Horsman G, et al. Results of continuous monitoring of the performance of rubella virus IgG and hepatitis B virus surface antibody assays using trueness controls in a multicenter trial. Clin Vaccine Immunol 2012;19:1624–32.
3. Dimech W, Walker S, Read S, Smeh K, Karakaltsas M, Dent B, et al. Comprehensive quality control programme for serology and nucleic acid testing using an internet-based application. Acc Qual Ass 2004;9:148–51.
4. Kim J, Swantee C, Lee B, Gunning H, Chow A, Sidaway F, et al. Identification of performance problems in a commercial human immunodeficiency virus type 1 enzyme immunoassay by multiuser external quality control monitoring and real-time data analysis. J Clin Microbiol 2009;47:3114–20.
5. Dimech W, Vincini G, Karakaltsas M. Determination of quality control limits for serological infectious disease testing using historical data. Clin Chem Lab Med 2015;53:329–36.
6. CLSI. Laboratory Quality Control Based on Risk Management; Approved Guideline. Vol. EP23-A. Wayne, PA: Clinical and Laboratory Standards Institute, 2011.
7. CLSI. Statistical quality control for quantitative measurement procedures: principles and definitions. Vol. CLSI Guideline C24. 4th ed. Wayne, PA: Clinical and Laboratory Standards Institute, 2016.
8. German Medical Association. Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-Baek" (unauthorized translation). J Lab Med 2015;39:26–69.
9. Public Health England. Quality assurance in the diagnostic virology and serology laboratory. In: Standards Unit Microbiology Services. Colindale, UK: Public Health England, 2015.
10. Australian Standards. Medical laboratories – requirements for quality and competence. Vol. AS ISO 15189-2013. Sydney: Standards Australia, 2013.
11. Badrick T. The quality control system. Clin Biochem Rev 2008;29(Suppl 1):S67–70.
12. Kallner A. International standards in laboratory medicine. Clin Chim Acta 2001;307:181–6.
13. Kinns H, Pitkin S, Housley D, Freedman DB. Internal quality control: best practice. J Clin Pathol 2013;66:1027–32.
14. Warkad SD, Song KS, Pal D, Nimse SB. Developments in the HCV screening technologies based on the detection of antigens and antibodies. Sensors (Basel) 2019;19.
15. Dimech W, Freame R, Smeh K, Wand H. A review of the relationship between quality control and donor sample results obtained from serological assays used for screening blood donations for anti-HIV and hepatitis B surface antigen. Acc Qual Ass 2013;18:11–8.
16. Aitken CK, Agius PA, Higgs PG, Stoove MA, Bowden DS, Dietze PM. The effects of needle-sharing and opioid substitution therapy on incidence of hepatitis C virus infection and reinfection in people who inject drugs. Epidemiol Infect 2017;145:796–801.
17. Micallef JM, Macdonald V, Jauncey M, Amin J, Rawlinson W, van Beek I, et al. High incidence of hepatitis C virus reinfection within a cohort of injecting drug users. J Viral Hepat 2007;14:413–8.
18. European Parliament and the European Council of the European Union. Directive 98/79/EC of the European Parliament and the council on in vitro diagnostic medical devices. Official Journal of the European Communities 1998; Directive 98/79/EC.
19. European Parliament and the European Council of the European Union. Regulation EU 2017/746 of the European Parliament and the council. Regulation (EU) 2017/746 Official Journal of the European Communities, 2017.
20. U.S. Food and Drugs Administration. Code of federal regulations. In: Department of Health and Human Services, Title 21, Volume 7. U.S. Food and Drugs Administration, 2019.
21. NRL Australia. QC investigations. https://www.nrlquality.org.au/qc-investigations. Accessed 24 November 2019.

The final publication is available at https://doi.org/10.1515/cclm-2020-0031

Chapter Five - Conclusion

Laboratory staff often question whether a reduction in QC test reactivity is predictive of a greater risk of reporting false negative patient results. They are acutely aware that, if a low positive sample is reported as negative there could be dire consequences. In the blood screening setting, a false negative result could lead to a transfusion transmitted infection (161). Indeed, donations of whole blood, and particularly plasma, can be transfused into multiple individuals. Donated plasma is also fractionated and used in therapeutic goods. An infected donation could result in hundreds of vials of therapeutic goods being infected. Fortunately, there are many mitigating processes to limit this eventuality. The use of pre-donation donor deferral questionnaires identifies and removes at-risk individuals. Donations are screened by NAT to detect the presence of virus, and fractionated plasma are subject to viral inactivation processes. However, laboratory staff retain a real concern over medicolegal consequences, patient safety, and further spread of infection by transfusion of blood and blood products from donors with false negative results. It is noted that screening donations for blood-borne viruses by NAT is not universal, with many low and middle-income countries, where the burden of disease is high, being unable to afford this testing, thereby raising the risk of reporting of false results (161-163).

Infectious disease serology assays are designed so that positive and negative patient populations are well removed from the cut-off (164). Therefore, if there is a systematic shift in reactivity due to a change in the test system, it is unlikely that a positive patient result will be reported as negative, except when the reactivity of the patient sample is extremely low, as in the case of a seroconversion event, a conclusion supported by others (165). Other possible clinical situations that can lead to low-level antibodies are during treatment (e.g. anti-retroviral therapy or pre-exposure prophylaxis for HIV), immunosuppression therapy or immunodeficiency syndromes or late in infection as antibodies decline. The likelihood of testing a patient at the time of seroconversion for HCV is low, even in high prevalence populations. Our study concluded "*However, the findings do represent the first detailed investigation into the relationship between reactivity of QC and patient samples when tested in a qualitative serology assay and the first time the assumption that a decrease in QC reactivity may be associated with the increased possibility of a false-negative patient result has been tested. The findings add further evidence that the QConnect Limits, based on historical QC peer-group data, are appropriate for monitoring infectious disease serology testing*", thereby supporting the use of QConnect™ concept and providing some comfort to laboratory staff that the risk of a false negative result is minimal.

The impact of this study is significant as it is the only such study published, and therefore is the reference paper for understanding the relation between QC test results and patient results for infectious disease serology. No other organisation has systematically compared the commutability of QC and patient results for serology, although others have for clinical chemistry analytes, also concluding that the level of commutability between the QC samples and patient samples was poor (166). In the case studied, only NRL's QConnect™ detected the change and highlighted the issue to the manufacturer, who was unaware of the situation, thereby validating NRL's approach to QC monitoring. The study also highlighted the importance of having a robust, scientifically validated QC monitoring process like QConnect™ and prompted other to call for a review of QC methods (165).

Conclusion

It is important that testing for infectious diseases by serology and NAT is accurate, has minimal bias and imprecision, and provides meaningful clinical information. Ideally, qualitative results for the presence or absence of an analyte, as is the case for antibody testing, should be comparable from one test system to another. That is, the clinical sensitivity of assays testing for the same analyte should be comparable. Quantitative infectious disease test results, such as viral load testing, are used to monitor the efficacy of therapy, the progress of disease or as a trigger for treatment. It is important that the quantitative results from one test system are comparable to the results reported by the same test system in different locations, and to different tests systems measuring the same analyte. This outcome can be achieved by standardising and controlling testing. However, principles of standardisation commonly applied to clinical chemistry have been shown not to be applicable for infectious disease serology because the detection and quantification of antibodies across test systems measure "type B" functional, biological analytes. Testing for biological functionality must account for a range of factors such as genetic variation and different serovars of the microorganisms, changes in the immune response over time, differences in assay design and components, reagent lot changes, as well as day-to-day variations due to the testing environment. The measurand of serological assays is assay-specific and therefore commutability of quantitative results between assays is yet to be achieved in a clinical meaningful way.

So where to for standardisation of infectious disease serology testing? There is strong evidence that international standards for antibody quantification cannot be used to calibrate serological IAs and regulatory bodies, WHO and other interested parties should state categorically that infectious disease serology assays should only report qualitative results. The package inserts of several serological international standards already have caveats stating they should not be used for the calibration of test kits. However, prior to my work, international standards for anti-rubella IgG and HBsAb have routinely been used for assay calibration, as has the recently released anti-SARS-CoV-2 standard. The publications presented in **Chapters 2 to 4** represents a comprehensive body of work published on the subject of standardisation of infectious disease serology, highlighting the difficulties faced and presenting an understanding of why standardisation of serology is problematic.

The lack of standardisation of rubella IgG assays has now been well documented (1, 14, 73, 74, 76, 92-95, 138, 139). Although less studied, there is evidence that the use of an international standard to calibrate HBsAb testing has also been unsuccessful (80). As a result of the publications presented in this thesis, along with those of other colleagues, in 2017 the WHO Expert Committee on Biological Standardization determined that the rubella IgG international standard should remain available, but IVD manufacturers should be made aware of the lack of commutability of the standard and therefore consider replacing rubella IgG quantitative tests, reporting in IU/mL, with qualitative assays (85). Recent publications, which I co-authored with representatives of WHO, NIBSC, Paul Ehrlich Institut, USA CDC and USA FDA, have noted that results of rubella IgG assays should no longer be expressed in IU/mL (77). After more than a decade of activity by concerned scientists, the rubella RUB-1-94 standard package insert (88) was recently amended to include the statement "*IVD manufacturers, regulators and assay users should be made aware of RUBI-1-94 potential lack of commutability when used as a calibrant. This was highlighted by the WHO Expert Committee on Biological Standardization in TRS 68th Report (Section 3.3.4: 2018)*." This statement is in line with a similar statement on the measles standard. However, no similar statement has yet been included in the HBsAb package insert even though there is evidence that standardisation of HBsAb assays is ineffective (80).

My published works, supported by other colleagues with similar interests, presents a compelling argument for the discontinuation of using WHO international standards for the calibration and attempted standardisation of serology assays used to detect infectious disease antibodies. In summary,

- There are no certified reference methods for infectious disease serology;
- serological tests do not measure the numbers of antibodies, it measures the ability of antibodies in a patient sample to bind to the antigenic target of the test system;
- the quantitative amount of antibodies detected has no clinical use, only the presence or absence of antibodies are clinically important;
- assays testing for the same analyte have different measurands because they target different antibody classes and subclasses, antigenic targets and have different detection systems;
- decades of attempts to use the international standard for standardising rubella IgG have failed, and
- WHO expert committees have conceded this and have subsequently modified the anti-rubella IgG international standard's IFU accordingly.

In spite of the compelling arguments against using serological international standards for assay calibration, this issue recurred with the release by NIBSC of the First WHO International Standard for anti-SARS-CoV-2 immunoglobulin (human) (NIBSC code: 20/136) (143, 167, 168) and associated WHO guidance document (144). Several IVD manufacturers have released quantitative anti-SARS-CoV-2 assays calibrated against this standard. The utility of this standard is yet to be determined. One study reported that, by reporting results relative to the international standard, interlaboratory variation was reduced 50 fold for neutralising antibody assays and 2000 fold for EIAs (168). It was also stated that the standard could be used " *to assist the comparison of assays detecting the same class of immunoglobulins with the same specificity (eg, anti-receptor binding domain IgG, anti-N IgM, etc)*" (168). Others have also lauded the development of the SARS-CoV-2 international standard for serology and advocated its use for standardisation of vaccine potency testing, but have identified "*even between binding assays, antibody titres against two viral proteins (eg, nucleoprotein and S protein) might not necessarily correlate. This factor could be particularly relevant in situations where diagnostic assays (eg, for confirmation of natural infection) are conflate with quantitative assays that measure only vaccine antibody titres" (169).* It should be noted that these statements recognise the fact that binding assays vary in types of immunoglobulins and antigenic targets, my stated causes of lack of standardisation between serology assays. Recently a study reviewed 13 different standards available for anti-SARS-CoV-2. They found varying levels of concordance between diagnostic methods at specific antigen–antibody combinations but concluded "*Our findings indicate that the arbitrary units of the WHO IS are not an accurate means to compare SARS-CoV-2 serology results between different laboratories or methods*" (145).

Although over 2400 vials of the 1st International SARS-CoV-2 standard was manufactured, a replacement batch was required by August 2021. As mentioned previously, SARS-CoV-2 developed numerous variants over the past several years (68). At the beginning of the study to assign unitage to the second standard, there were five variants of concern listed by the WHO (170) and by July 2022, Omicron and its lineage strains were the predominant strains circulating. Several candidate samples were tested in both neutralisation assays and binding assays. Differences were determined when the candidate samples were tested against variants of concern using neutralising assays, having a reduction of potency against beta and delta strains, as compared with ancestral strain. When tested alongside the candidate samples, 1st International standard, WHO IS 20/136, could not detect neutralising antibodies against Omicron (170). The 2nd WHO IS for anti SARS CoV 2 immunoglobulin (21/340) was manufactured from convalescent plasma from individuals infected early 2020 and released for use in October 2021.

This situation highlights the impact on the ability to standardise antibody testing due to emergence of new strains over time. The 2nd WHO IS for anti SARS CoV 2 immunoglobulin (21/340) IFU (171) contains a table, "intended for guidance only", with the recommended potencies against the de-classified variants of concern. It also states that "*The intended use of the International Standard is for the calibration and harmonisation of neutralisation assays using SARS-CoV-2 early isolates (2020). For the calibration of secondary reagents used in neutralisation assay against variants of concerns the 1st WHO International Standard for anti-SARS-CoV-2 variants of concern immunoglobulin, NIBSC code 21/338, should be used*"(171).

New strains may impact the ability of manufacturers to standardise the quantification of SARS-CoV-2 antibodies. Only a rigorous comparison of quantitative results, reported in international units, obtained from different anti-SARS-CoV-2 test kits will determine whether the standard has successfully facilitated standardised reporting.

Whereas assays that quantify the amount of detectable antibody arguably can serve to monitor the levels of antibodies over time, it remains to be demonstrated that the levels reported by one assay correlate with those reported by another, or indeed that a quantitative measurement of binding antibodies has any correlation with protective immunity (168). If the rubella IgG scenario is reproduced with SARS-CoV-2, similar difficulties resulting from lack of standardisation are to be expected. I have coordinated a large-scale, head-to-head evaluation of anti-SARS-CoV-2 serology assays on behalf of the WHO, the Australian TGA and directly with test kit manufacturers. To date, 12 rapid lateral flow test devices (RDT) have evaluated for TGA; 29 RDTs and nine EIAs evaluated for WHO and more than 10 high-throughput serology platforms have been evaluated on behalf of the manufacturers. Test kits under evaluation reported results for various SARS-CoV-2 antibodies including IgG, IgM, IgA, total antibodies, and neutralising antibodies; and to different antigens including spike and nucleocapsid. The summary results published on the NRL website (172) with a final reports currently in draft. Of these assays, several are calibrated against the WHO international standard and report quantitative results. A dilution series of the standard has been tested with these assays to understand the accuracy of quantification compared with the standard. The more important study is to compare the quantitative results, reported in IU/mL, of 199 patients' samples included in the study. This investigation will provide insight into the comparison of quantification across several quantitative assays. Analysis of the data obtained from this series of evaluations will assess the correlation of quantitative results. If hypothesis that serology assays cannot be standardised is supported, it is hoped that future IVD directives, professional bodies and key opinion leaders embrace the concept that the standardisation of serological assays for infectious diseases is not possible and promote the reporting of qualitative results.

My position, which has been relayed to NIBSC and WHO, is that the SARS-CoV-2 antibody standard should have the intended use for the standardisation of viral neutralisation assays used to evaluate the potency of vaccines, and the international standard IFU categorically state that it should not be used as a calibrator of IAs. The use of virus neutralisation tests and plaque reduction neutralisation tests using pathogenic viruses are the gold standard method for measuring vaccine-induced neutralising antibodies, and the need for standardisation of these assays is required (167, 173). In 2020, the Coalition for Epidemic Preparedness Innovations established a global network to establish standardisation of COVID-19 vaccine development (174), where the availability of an international standard would be invaluable.

We have briefly reviewed the potential of using monoclonal antibodies as standard or controls (unpublished data). The argument for monoclonal antibodies serving these purposes is that they can be made in quantity, are stable when stored under appropriate conditions and are epitope specific. Therefore, theoretically at

least, they do not have the variability associated with polyclonal, human derived samples. However, this attribute is also a deficiency, as assays rarely detect a single epitope. Our personal experience is that, when a toxoplasma-specific monoclonal IgM antibody was used as a QC sample, reactivity was at acceptable levels for most assays except one, which did not detect that epitope, underlying potential issues with this approach.

This is not the case for NAT or antigen testing, where the traditional approach used to standardise "type A" analytes has resulted in the successful standardisation of test systems for RNA, DNA and, to a lesser extent, antigen quantification. There is good evidence that the development and use of international standards has harmonised reporting of results across test systems (108, 121). This fact has allowed the establishment of viral load thresholds for implementation or cessation of treatment (120, 150, 175), rejection of donor plasma prior to fractionation when NAT positive for viral pathogens (161) and the ability to compare test results across testing systems and therefore from laboratory to laboratory (108). There are some remaining issues identified relating to the standardisation of NAT, such as fragmentation of nucleic acids and the fact that international standards are produced sequentially rather than using a single standard to calibrate subsequent standards. My work has demonstrated that these are minor and do not materially impact on standardisation efforts for NAT.

The control of test systems used by medical laboratories should be risk-based (54). As serological assays have become more automated, the traditional control mechanisms used by clinical chemists, especially the use of Westgard rules, have been adopted by testing laboratories. However, until recently, no systematic review of the applicability of these rules had been undertaken. My work has demonstrated an unacceptable number of false rejections are encountered if international guidelines for clinical chemistry QC are followed in the context of infectious disease testing (157). In collaboration with others, I lead the development and publication of an alternative approach, called QConnect™, which is more applicable to infectious disease serology (122, 125). QConnect™ method includes all normal variation when calculating acceptance criteria, especially the variation caused by changes of reagent lots over time. A recent study has demonstrated that, when using QConnect™, even when a test system experiences significant changes, as evidenced by a drop in QC reactivity, the risk of a true false negative result during a seroconversion event is minimal (126). The QConnect™ approach reduces unnecessary investigations of "false rejections" of QC results, allows laboratory staff to focus on true errors, and ultimately increases the confidence in test results.

Through my work, I continually highlight the need for infectious disease testing laboratories to change away from traditional QC processes and adopt new approaches for monitoring testing. Recently, the QConnect™ method was referenced as a suitable alternative to traditional QC methods in the UK Standards for Microbiology Investigations, Quality Assurance in the Diagnostic Infection Sciences Laboratory guidance document (128), the first such international recognition of the concept. Although acceptance by professional bodies is muted, the QConnect™ approach has been more successful commercially, with major QC sample manufacturers embracing QConnect™. A working party of international experts, coordinated by and including me, has been assembled and a communique, seeking to highlight the need for change, is in draft, and is expected to be submitted as a publication as an opinion piece in a peer-reviewed journal in 2022. Meanwhile, NRL is conducting further studies aimed at better understanding the imprecision of infectious disease assays. As the QConnect™ concept incorporates all sources of variation, including both bias and imprecision combined, a possible criticism of the approach is that it does not differentiate between these two sources of variation, and therefore the acceptance limits may be perceived to be too wide. Monitoring imprecision in isolation over time and establishing assay-specific acceptable levels of imprecision may address this objection. Line listed data from all test systems and all analytes submitted into EDCNet™, from the beginning of 2017 to end of 2021 have been extracted and are being analysed. This

represents over a million data points from hundreds of analytes and testing platforms, making it the most comprehensive investigation of infectious disease serology QC ever. It is expected that the data obtained from this analysis will inform additional algorithms that will eventually be published in peer-reviewed journals and subsequently built into the QConnect™ concept, strengthening its ability to identify unexpected variation.

Starting three decades ago, and remaining relevant and important today, my publications have addressed the topics of standardisation and control of infectious disease testing, contributing significant and original knowledge to the understanding of medical laboratory testing. The work has highlighted deficiencies in approaches, especially related to infectious disease serology. The work has systematically identified inappropriate application of traditional methods for the standardisation and control of infectious disease serology, undertook to determine the root cause of why traditional approaches do not work and designed, implemented, and evaluated novel alternative approaches. In the case of standardisation of serology testing, my work led to extraordinary WHO meetings, including representatives from USA FDA, USA CDC, Paul Ehrlich Institut and NIBSC, resulting in a WHO expert committee publication (85) and a change to the rubella IgG International Standard IFU (88). For QC, my work has resulted in the development of trademarked internet-based quality control monitoring software EDCNet™, and a novel QC monitoring concept QConnect™, used by hundreds of accredited clinical and national blood screening laboratories globally and licenced to two major QC manufacturers. These developments have been systematically published in peer-reviewed journals, leading to a significant body of work (1, 2, 14, 73, 74, 77, 92, 93, 121-126, 132, 134, 135, 138, 147, 154, 155, 157-160, 176-180).

Infectious disease testing using biological function is different to testing for inert molecules (1). The rules applied to clinical chemistry for standardisation and control cannot and should not be applied without evidence. Indeed, evidence is mounting to indicate that specific approaches to standardisation and control of infectious disease testing is required. More detailed investigations in this area are encouraged. WHO, professional bodies, organisations publishing guidance documents and the IVD industry must heed this evidence and commit resources to developing a set of guidance documents for the standardisation and control of infectious disease testing, rather than applying inappropriate processes without evidence.

Having highlighted the deficiencies of traditional QC methods, I will continue to advocate for the application of scientifically validated approach to monitoring the QC of infectious disease testing through publications, presentations, and communications to professional bodies and peers. I will continue to identify problems and educate interested laboratory staff of the inherent issues in medical laboratory testing due to the inappropriate implementation of approaches to standardisation and control of infectious disease testing, through conferences and company-sponsored webinars, social media posts, and through professional organisations. This is especially true where issues may potentially cause adverse impact on patient safety and clinical outcomes and will be achieved. This PhD thesis serves to summarise concisely the concepts involved in the standardisation and control of infectious disease testing, identify the areas of need and propose the future work required to address these gaps.

References

1. Dimech W. 2021. The standardization and control of serology and nucleic acid testing for infectious diseases. Clin Microbiol Rev 34.
2. Dimech W, Mupunga I. October 2021 2021. Testing Times - A documentary., *on* National Serology Reference Laboratory, Australia. https://www.nrlquality.org.au/about/achievements/. Accessed 8 March 2023.
3. Hengel B, Causer L, Matthews S, Smith K, Andrewartha K, Badman S, Spaeth B, Tangey A, Cunningham P, Saha A, Phillips E, Ward J, Watts C, King J, Applegate T, Shephard M, Guy R. 2021. A decentralised point-of-care testing model to address inequities in the COVID-19 response. Lancet Infect Dis 21:e183-e190.
4. Seventer JM, Hochberg, N.S. 2017. Principles of Infectious Diseases: Transmission, Diagnosis, Prevention, and Control, p *In* (ed), International Encyclopedia of Public Health, ed 2nd, vol 6. Elsevier Inc.,
5. Hosseini A, Hashemi V, Shomali N, Asghari F, Gharibi T, Akbari M, Gholizadeh S, Jafari A. 2020. Innate and adaptive immune responses against coronavirus. Biomed Pharmacother 132:110859.
6. Various. 2019. Manual of Clinical Microbiology, 12th ed. ASM Press, Washington DC.
7. Turvey SE, Broide DH. 2010. Innate immunity. J Allergy Clin Immunol 125:S24-32.
8. Marshall JS, Warrington R, Watson W, Kim HL. 2018. An introduction to immunology and immunopathology. Allergy Asthma Clin Immunol 14:49.
9. Bonilla FA, Oettgen HC. 2010. Adaptive immunity. J Allergy Clin Immunol 125:S33-40.
10. Gutzeit C, Chen K, Cerutti A. 2018. The enigmatic function of IgD: some answers at last. Eur J Immunol 48:1101-1113.
11. Vidarsson G, Dekkers G, Rispens T. 2014. IgG subclasses and allotypes: from structure to effector functions. Front Immunol 5:520.
12. Sette A, Crotty S. 2021. Adaptive immunity to SARS-CoV-2 and COVID-19. Cell 184:861-880.
13. Schroeder HW, Jr., Cavacini L. 2010. Structure and function of immunoglobulins. J Allergy Clin Immunol 125:S41-52.
14. Dimech W, Grangeot-Keros L, Vauloup-Fellous C. 2016. Standardization of assays that detect anti-rubella virus IgG antibodies. Clin Microbiol Rev 29:163-74.
15. De Paschale M, Clerici P. 2012. Serological diagnosis of Epstein-Barr virus infection: Problems and solutions. World J Virol 1:31-43.
16. Sohn MJ, Cho JM, Moon JS, Ko JS, Yang HR. 2018. EBV VCA IgM and cytomegalovirus IgM dual positivity is a false positive finding related to age and hepatic involvement of primary Epstein-Barr virus infection in children. Medicine (Baltimore) 97:e12380.
17. Nurmi V, Hedman L, Perdomo MF, Weseslindtner L, Hedman K. 2021. Comparison of approaches for IgG avidity calculation and a new highly sensitive and specific method with broad dynamic range. Int J Infect Dis 110:479-487.
18. Anonymous. 2019. Manual of Clinical Microbiology, 12th ed. American Society for Micorbiology, Washington, USA.
19. Alizadeh M, Yousefi L, Pakdel F, Ghotaslou R, Rezaee MA, Khodadadi E, Oskouei MA, Soroush Barhaghi MH, Kafil HS. 2021. MALDI-TOF Mass Spectroscopy Applications in Clinical Microbiology. Adv Pharmacol Pharm Sci 2021:9928238.
20. Patel R. 2015. MALDI-TOF MS for the diagnosis of infectious diseases. Clin Chem 61:100-11.
21. Hernández R, Rueda F. 2023. Biomarkers as Prognostic Predictors and Therapeutic Guide in Critically Ill Patients: Clinical Evidence. Journal of Personalized Medicine 13:27.
22. Witwer KW. 2015. Circulating microRNA biomarker studies: pitfalls and potential solutions. Clin Chem 61:56-63.
23. Bradstreet CM, Kirkwood B, Pattison JR, Tobin JO. 1978. The derivation of a minimum immune titre of rubella haemagglutination-inhibition (HI) antibody. A Public Health Laboratory Service collaborative survey. J Hyg (Lond) 81:383-8.
24. Stephenson I, Das RG, Wood JM, Katz JM. 2007. Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: an international collaborative study. Vaccine 25:4056-63.

25. Gust ID, Mathison A, Winsor H. 1973. A survey of rubella haemagglutination-inhibition testing in the southern states of Australia. Bull World Health Organ 49:139-42.
26. Schmidt NJ, Lennette EH. 1970. Variables of the rubella hemagglutination-inhibition test system and their effect on antigen and antibody titers. Appl Microbiol 19:491-504.
27. Hierholzer JC, Suggs MT. 1969. Standardized viral hemagglutination and hemagglutination-inhibition tests. I. Standardization of erythrocyte suspensions. Appl Microbiol 18:816-23.
28. Miller WG. 2018. The standardization journey and the path ahead. J Lab Precis Med 3.
29. Zacour M, Ward BJ, Brewer A, Tang P, Boivin G, Li Y, Warhuus M, McNeil SA, LeBlanc JJ, Hatchette TF, Public Health Agency of Canada, Canadian Institutes of Health Influenza Research Network. 2016. Standardization of Hemagglutination Inhibition Assay for Influenza Serology Allows for High Reproducibility between Laboratories. Clin Vaccine Immunol 23:236-42.
30. Voller A, Bartlett A, Bidwell DE. 1978. Enzyme immunoassays with special reference to ELISA techniques. J Clin Pathol 31:507-20.
31. Gray ER, Bain R, Varsaneux O, Peeling RW, Stevens MM, McKendry RA. 2018. p24 revisited: a landscape review of antigen detection for early HIV diagnosis. AIDS 32:2089-2102.
32. Sonneveld MJ, Rijckborst V, Boucher CA, Zwang L, Beersma MF, Hansen BE, Janssen HL. 2011. A comparison of two assays for quantification of Hepatitis B surface Antigen in patients with chronic hepatitis B. J Clin Virol 51:175-8.
33. Cox KL, Devanarayan V, Kriauciunas A, Manetta J, Montrose C, Sittampalam S. 2004. Immunoassay Methods. *In* Markossian S, Sittampalam GS, Grossman A, Brimacombe K, Arkin M, Auld D, Austin CP, Baell J, Caaveiro JMM, Chung TDY, Coussens NP, Dahlin JL, Devanaryan V, Foley TL, Glicksman M, Hall MD, Haas JV, Hoare SRJ, Inglese J, Iversen PW, Kahl SD, Kales SC, Kirshner S, Lal-Nag M, Li Z, McGee J, McManus O, Riss T, Saradjian P, Trask OJ, Jr., Weidner JR, Wildey MJ, Xia M, Xu X (ed), Assay Guidance Manual. Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD).
34. Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155:335-50.
35. Holmes H, Davis C, Heath A, Hewlett I, Lelie N. 2001. An international collaborative study to establish the 1st international standard for HIV-1 RNA for use in nucleic acid-based techniques. J Virol Methods 92:141-50.
36. The Commission of European Communities. 2009. Common technical specifications (CTS) for In-vitro diagnostic medical devices. Official Journal of the European Union, Brussels.
37. Australian Government. 2015. Classification of IVD Medical Devices. Department of Health Therapeutic Goods Administration, Canberra. https://www.tga.gov.au/sites/default/files/classification-ivd-medical-devices.pdf.
38. The European Parliment and the Council of the European Union. 1998. Directive 98/79/EC of the European Parliment and the Council of 27 October 1998 on in vitro diagnostic medical devices. Official Journal of the European Communities, Brussels.
39. Anonymous. 1995. *In* Leveton LB, Sox HC, Jr., Stoto MA (ed), HIV And The Blood Supply: An Analysis Of Crisis Decisionmaking doi:10.17226/4989, Washington (DC).
40. Bayer R, Feldman E, (ed). 1999. Blood Feuds: AIDS, Blood, and the Politics of Medical Disaster. Oxford University Press, New York.
41. World Health Organization. Global harmonization task force (GHTF), *on* WHO. https://www.who.int/medical_devices/collaborations/force/en/. Accessed 11 October 2020.
42. World Health Organization. 2018. Overview of the WHO prequalifcation of in-vitro diagnostic assessments. Essential Medicines and Health Products, WHO, Geneva.
43. American Association for Laboratory Accreditation. 2012. P102 - A2LA Policy on Measurement Traceability. American Association for Laboratory Accreditation.,
44. Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, Budd JR, Weykamp C, Delatour V, Nilsson G, MacKenzie F, Panteghini M, Keller T, Camara JE, Zegers I, Vesper HW, IFCC Working Group on Commutability. 2018. IFCC Working Group Recommendations for Assessing Commutability Part 1: General Experimental Design. Clin Chem 64:447-454.
45. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, Ceriotti F, Schimmel H, Weykamp C, Keller T, Camara JE, Burns C, Vesper HW, MacKenzie F, Miller WG, Commutability

IWGo. 2018. IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias between a Reference Material and Clinical Samples. Clin Chem 64:455-464.
46. Shewhard WA. 1931. Economic control of manufactured products. Van Nostand, New York.
47. Levey JJ, E.R. 1950. The use of control charts in the clinical laboratory. Am J Clin Pathol 20:1059-1066.
48. Westgard JO. 1994. Selecting appropriate quality-control rules. Clin Chem 40:499-501.
49. Westgard JO. 2003. Internal quality control: planning and implementation strategies. Ann Clin Biochem 40:593-611.
50. Westgard JO. 2013. Statistical quality control procedures. Clin Lab Med 33:111-24.
51. Clincal and Laboratory Standards Institute. 2006. C24-A3 Statistical quality control for quantitative measurement procedures: Principles adn definitions; Approved guidelines - Third edition. CLSI, Wayne PA.
52. CLSI. 2016. Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions. Clinical and Laboratory Standards Institute, Wayne PA.
53. Public Health England. 2021. Quality assurance in the diagnostic virology and serology laboratory. Standards Unit Microbiology Services, London, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005438/Q_2i8.pdf.
54. CLSI. 2011. A Laboratory quality control based on risk management: Approved guidelines. Clinical and Laboratory Standards Institute, Wayne PA.
55. Westgard S, Bayat H, Westgard JO. 2018. Special issue on Six Sigma metrics - experiences and recommendations. Biochem Med (Zagreb) 28:020301.
56. Wauthier L, Di Chiaro L, Favresse J. 2022. Sigma metrics in laboratory medicine: A call for harmonization. Clin Chim Acta 532:13-20.
57. Newton DW, Vandepoele N, Yundt-Pacheco JC, Gauthier D, Bachman MA. 2022. Management of Cytomegalovirus, Epstein-Barr Virus, and HIV Viral Load Quality Control Data Using Unity Real Time. J Clin Microbiol 60:e0167521.
58. Xia Y, Wang X, Yan C, Wu J, Xue H, Li M, Lin Y, Li J, Ji L. 2020. Risk assessment of the total testing process based on quality indicators with the Sigma metrics. Clin Chem Lab Med 58:1223-1231.
59. Westgard S, Bayat H, Westgard JO. 2018. Analytical Sigma metrics: A review of Six Sigma implementation tools for medical laboratories. Biochem Med (Zagreb) 28:020502.
60. International Organization of Legal Metrology Vocabulary. 2007. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM). Paris, France. https://www.oiml.org/en/files/pdf_v/v002-200-e07.pdf.
61. Panteghini M, Forest JC. 2005. Standardization in laboratory medicine: new challenges. Clin Chim Acta 355:1-12.
62. Dybkaer R. 1996. Quantities and units for biological reference materials used with in vitro diagnostic measuring systems for antibodies. Scand J Clin Lab Invest 56:385-91.
63. Bellini WJ, Rota JS, Rota PA. 1994. Virology of measles virus. J Infect Dis 170 Suppl 1:S15-23.
64. Fulton BO, Sachs D, Beaty SM, Won ST, Lee B, Palese P, Heaton NS. 2015. Mutational Analysis of Measles Virus Suggests Constraints on Antigenic Variation of the Glycoproteins. Cell Rep 11:1331-8.
65. Bbosa N, Kaleebu P, Ssemwanga D. 2019. HIV subtype diversity worldwide. Curr Opin HIV AIDS 14:153-160.
66. Ludolfs D, Schilling S, Altenschmidt J, Schmitz H. 2002. Serological differentiation of infections with dengue virus serotypes 1 to 4 by using recombinant antigens. J Clin Microbiol 40:4317-20.
67. van Doorn LJ, Kleter B, Pike I, Quint W. 1996. Analysis of hepatitis C virus isolates by serotyping and genotyping. J Clin Microbiol 34:1784-7.
68. Fergie J, Srivastava A. 2021. Immunity to SARS-CoV-2: Lessons Learned. Front Immunol 12:654165.
69. Wang S, Xu X, Wei C, Li S, Zhao J, Zheng Y, Liu X, Zeng X, Yuan W, Peng S. 2022. Molecular evolutionary characteristics of SARS-CoV-2 emerging in the United States. J Med Virol 94:310-317.

70. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Consortium C-GU, Peacock SJ, Robertson DL. 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol 19:409-424.
71. Thakur S, Sasi S, Pillai SG, Nag A, Shukla D, Singhal R, Phalke S, Velu GSK. 2022. SARS-CoV-2 Mutations and Their Impact on Diagnostics, Therapeutics and Vaccines. Front Med (Lausanne) 9:815389.
72. Caligiuri P, Cerruti R, Icardi G, Bruzzone B. 2016. Overview of hepatitis B virus mutations and their implications in the management of infection. World J Gastroenterol 22:145-54.
73. Dimech W, Bettoli A, Eckert D, Francis B, Hamblin J, Kerr T, Ryan C, Skurrie I. 1992. Multicenter evaluation of five commercial rubella virus immunoglobulin G kits which report in international units per milliliter. J Clin Microbiol 30:633-41.
74. Dimech W. 2016. Where to now for standardization of anti-Rubella virus IgG testing. J Clin Microbiol 54:1682-1683.
75. Bouthry E, Furione M, Huzly D, Ogee-Nwankwo A, Hao L, Adebayo A, Icenogle J, Sarasini A, Revello MG, Grangeot-Keros L, Vauloup-Fellous C. 2016. Assessing Immunity to Rubella Virus: a Plea for Standardization of IgG (Immuno)assays. J Clin Microbiol 54:1720-1725.
76. Huzly D, Hanselmann I, Neumann-Haefelin D, Panning M. 2016. Performance of 14 rubella IgG immunoassays on samples with low positive or negative haemagglutination inhibition results. J Clin Virol 74:13-8.
77. Kempster SL, Almond N, Dimech W, Grangeot-Keros L, Huzly D, Icenogle J, El Mubarak HS, Mulders MN, Nubling CM. 2020. WHO international standard for anti-rubella: learning from its application. Lancet Infect Dis 20:e17-e19.
78. Leuridan E, Van Damme P. 2011. Hepatitis B and the need for a booster dose. Clin Infect Dis 53:68-75.
79. Bouthry EH, D. Revello, M.G. Furione, M. Grangeot-Keros, L. Vauloup-Fellous, C. 2014. An evaluation of nine rubella IgG assays highlights many discrepancies in the interpretation of the results, abstr European Congress of Clinical Microbiology & Infectious Diseases, Barcelona, Spain,
80. Huzly D, Schenk T, Jilg W, Neumann-Haefelin D. 2008. Comparison of nine commercially available assays for quantification of antibody response to hepatitis B virus surface antigen. J Clin Microbiol 46:1298-306.
81. Chen RT, Markowitz LE, Albrecht P, Stewart JA, Mofenson LM, Preblud SR, Orenstein WA. 1990. Measles antibody: reevaluation of protective titers. J Infect Dis 162:1036-42.
82. WHO Expert Committee on Biological Standardization. 1994. WHO Technical Report Series: Report 44. WHO, Geneva.
83. WHO Expert Committee on Biological Standardization. 1995. WHO Technical Report Series: Report 45. WHO, Geneva.
84. WHO Expert Committee on Biological Standardization. 1998. WHO Technical Report Series: Report 47. WHO Geneva.
85. WHO Expert Committee on Biological Standardization. 2018. WHO Technical Report Series: Report 68. WHO, Geneva. https://apps.who.int/iris/bitstream/handle/10665/272807/9789241210201-eng.pdf?ua=1.
86. WHO Expert Committee on Biological Standardization. 1967. WHO Technical Report Series: Report 19. WHO, Geneva.
87. Hansen GA. 1996. Collaborative Study: calibration of a replacement standard for the second international standard for anti-rubella serum, human. International Laboratory for Biological Standards, Staten Serum Institut, Denmark.
88. National Institute of Biological Standards and Controls. 2018. WHO International Standard anti-Rubella immunoglobulin instructions for use, Version 8.0. NIBSC, Potters Bar, UK.
89. Clarke M, Dudgeon JA, Ferris R, Colinet G, Tate H, Perkins FT. 1975. The British Standard for anti-rubella serum. J Biol Stand 3:151-61.
90. National Institute of Biological Standards and Controls. 2009. 2nd British Standard for Anti-Rubella Serum, Human Instructions for use. National Institute of Biological Standards and Controls, United Kingdom.

91. National Institute of Biological Standards and Controls. 2009. Package insert, anti-rubella serum, human (67/182). National Institute of Biological Standards and Controls, London, U.K.
92. Dimech W, Arachchi N, Cai J, Sahin T, Wilson K. 2013. Investigation into low-level anti-rubella virus IgG results reported by commercial immunoassays. Clin Vaccine Immunol 20:255-61.
93. Dimech W, Panagiotopoulos L, Francis B, Laven N, Marler J, Dickeson D, Panayotou T, Wilson K, Wootten R, Dax EM. 2008. Evaluation of eight anti-rubella virus immunoglobulin G immunoassays that report results in international units per milliliter. J Clin Microbiol 46:1955-60.
94. Enders M, Bartelt U, Knotek F, Bunn K, Strobel S, Dietz K, Enders G. 2013. Performance of the Elecsys Rubella IgG assay in the diagnostic laboratory setting for assessment of immune status. Clin Vaccine Immunol 20:420-6.
95. Picone O, Bouthry E, Bejaoui-Olhmann Y, Cordier AG, Nedellec S, Letourneau A, Carbonel M, Brollo M, Grangeot-Keros L, Ayoubi JM, Benachi A, Rouge E, Vauloup-Fellous C. 2019. Determination of rubella virus-specific humoral and cell-mediated immunity in pregnant women with negative or equivocal rubella-specific IgG in routine screening. J Clin Virol 112:27-33.
96. Sautter RL, Crist AE, Johnson LM, Lebar WD. 1994. Comparison of five methods for the determination of rubella immunity. Infect Dis Obstet Gynecol 1:188-92.
97. Ekins R. 1991. Immunoassay standardization. Scand J Clin Lab Invest Suppl 205:33-46.
98. Muller MM. 2000. Implementation of reference systems in laboratory medicine. Clin Chem 46:1907-9.
99. Kallner A. 2001. International standards in laboratory medicine. Clin Chim Acta 307:181-6.
100. De Bièvre P, Dybkaer R, Fajgelj A, Hibbert B. 2011. Metrological traceability of measurement results in chemistry: Concepts and implementation (IUPAC Technical Report). Pure Appl Chem 83:1873–1935.
101. International Medical Device Regulators Forum. 2021. Principles of In Vitro Diagnostic (IVD) Medical Devices Classification. IMDRF, https://www.google.com.au/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwj2pI2Rm5f6AhUHS0EAHRCGAboQFnoECAgQAQ&url=https%3A%2F%2Fwww.imdrf.org%2Fsites%2Fdefault%2Ffiles%2Fdocs%2Fimdrf%2Ffinal%2Ftechnical%2Fimdrf-tech-wng64.pdf&usg=AOvVaw0e4bCHWBsCWErHX2aQ7CyG.
102. Shreffler J, Huecker MR. 2022. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios, StatPearls, Treasure Island (FL).
103. Andryukov BG. 2020. Six decades of lateral flow immunoassay: from determining metabolic markers to diagnosing COVID-19. AIMS Microbiol 6:280-304.
104. Wilkinson DE, Seiz PL, Schuttler CG, Gerlich WH, Glebe D, Scheiblauer H, Nick S, Chudy M, Dougall T, Stone L, Heath AB, Collaborative Study G. 2016. International collaborative study on the 3rd WHO International Standard for hepatitis B surface antigen. J Clin Virol 82:173-180.
105. Prescott GH, J. Atkinson, E. Morris, C. 2018. International Collaborative Study to Establish the 1st WHO International Reference Panel for HIV-1 p24 antigen and HIV-2 p26 antigen. Expert committee on biological standardization, World Health Organization, Geneva.
106. Cornberg M, Wong VW, Locarnini S, Brunetto M, Janssen HLA, Chan HL. 2017. The role of quantitative hepatitis B surface antigen revisited. J Hepatol 66:398-411.
107. Australian Government. 2015. National Hepatitis B Testing Policy v1.2. Department of Health, Australasian Society for HIV Medicine, Viral Hepatitis and Sexual Health Medicine,,
108. Baylis SA, Wallace P, McCulloch E, Niesters HGM, Nubling CM. 2019. Standardization of nucleic acid tests: the approach of the World Health Organization. J Clin Microbiol 57.
109. Saldanha J, Lelie N, Heath A. 1999. Establishment of the first international standard for nucleic acid amplification technology (NAT) assays for HCV RNA. WHO Collaborative Study Group. Vox Sang 76:149-58.
110. Fryer JF, Heath AB, Wilkinson DE, Minor PD, Collaborative Study Group. 2017. A collaborative study to establish the 3rd WHO International Standard for hepatitis B virus for nucleic acid amplification techniques. Biologicals 46:57-63.
111. Saldanha J, Lelie N, Yu MW, Heath A, Group BCS. 2002. Establishment of the first World Health Organization International Standard for human parvovirus B19 DNA nucleic acid amplification techniques. Vox Sang 82:24-31.

112. Fryer JF, Heath AB, Wilkinson DE, Minor PD, Collaborative Study Group. 2016. A collaborative study to establish the 1st WHO International Standard for Epstein-Barr virus for nucleic acid amplification techniques. Biologicals 44:423-33.
113. Bateman AC, Greninger AL, Atienza EE, Limaye AP, Jerome KR, Cook L. 2017. Quantification of BK Virus Standards by Quantitative Real-Time PCR and Droplet Digital PCR Is Confounded by Multiple Virus Populations in the WHO BKV International Standard. Clin Chem 63:761-769.
114. National Measurement Institute Australia (NMIA). 2020. Certified Reference Material Certificate of Analysis: NMIA NA050 to NA055: SARS-CoV-2 Standard., *on* NMIA. https://www.industry.gov.au/sites/default/files/2020-12/sars-cov-2_c_of_a_b200921_feb_2021.pdf. Accessed 18 June 2022.
115. Bustin S, Dhillon HS, Kirvell S, Greenwood C, Parker M, Shipley GL, Nolan T. 2015. Variability of the reverse transcription step: practical implications. Clin Chem 61:202-12.
116. Huggett JF, Cowen S, Foy CA. 2015. Considerations for digital PCR as an accurate molecular diagnostic tool. Clin Chem 61:79-88.
117. Gullett JC, Nolte FS. 2015. Quantitative nucleic acid amplification methods for viral infections. Clin Chem 61:72-8.
118. Boom R, Sol CJ, Schuurman T, Van Breda A, Weel JF, Beld M, Ten Berge IJ, Wertheim-Van Dillen PM, De Jong MD. 2002. Human cytomegalovirus DNA in plasma and serum specimens of renal transplant recipients is highly fragmented. J Clin Microbiol 40:4105-13.
119. Hayden RT, Tang L, Su Y, Cook L, Gu Z, Jerome KR, Boonyaratanakornkit J, Sam S, Pounds S, Caliendo AM. 2019. Impact of Fragmentation on Commutability of Epstein-Barr Virus and Cytomegalovirus Quantitative Standards. J Clin Microbiol 58.
120. Dioverti MV, Lahr BD, Germer JJ, Yao JD, Gartner ML, Razonable RR. 2017. Comparison of standardized cytomegalovirus (CMV) viral load thresholds in whole blood and plasma of solid organ and hematopoietic stem cell transplant recipients with CMV infection and disease. Open Forum Infect Dis 4:ofx143.
121. Baylis S, Nübling CM, Dimech W. 2021. Standardization of Diagnostic Assays, p 52-63, Encyclopedia of Virology, 4th ed, vol 5. Elsevier.
122. Dimech W, Walker S, Jardine D, Read S, Smeh K, Karakaltsas K, Dent B, Dax E. 2004. Comprehensive quality control programme for serology and nucleic acid testing using an Internet-based application. Accred Qual Assur 8:148–151.
123. Pham TD, W. Karakaltsis, M. Dax, E.M. 2008. Variation in the Performance of Reagent Batches of the Abbott Architect HIV Ag/Ab Combo CMIA Detected in an International Quality Control Programme, abstr International Society for Blood Transfusion, Vox Sanguinis,
124. Walker S, Dimech W, Kiely P, Smeh K, Francis B, Karakaltsas M, Dax EM. 2009. An international quality control programme for PRISM chemiluminescent immunoassays in blood service and blood product laboratories. Vox Sang 97:309-16.
125. Dimech W, Vincini G, Karakaltsas M. 2015. Determination of quality control limits for serological infectious disease testing using historical data. Clin Chem Lab Med 53:329-36.
126. Dimech WJ, Vincini GA, Cabuang LM, Wieringa M. 2020. Does a change in quality control results influence the sensitivity of an anti-HCV test? Clin Chem Lab Med 58:1372-1380.
127. Anonymous. 2015. Revision of the “Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK” (unauthorized translation). J Lab Med 39:26-69.
128. Public Health England (PHE). 2021. UK Standards for Microbiology Investigations: UK Standards for Microbiology Investigations. Standards Unit National Infection Service, Public Health England, London, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005438/Q_2i8.pdf.
129. Badrick T. 2008. The quality control system. Clin Biochem Rev 29 Suppl 1:S67-70.
130. Solsvik AE, Kristoffersen AH, Sandberg S, Gidske G, Stavelin AV, Eikeland J, Amundsen E. 2022. A national surveillance program for evaluating new reagent lots in medical laboratories. Clin Chem Lab Med 60:351-360.

131. Badrick T, Parvin C. 2019. Letter to the Editor on article Dimech W, Karakaltsas M, Vincini G. Comparison of four methods of establishing control limits for monitoring quality controls in infectious disease serology testing. Clin Chem Lab Med 2018;56:1970-8. Clin Chem Lab Med 57:e71-e72.
132. Dimech W, Vincini G, Karakaltas M. 2019. Counterpoint to the Letter to the Editor by Badrick and Parvin in regard to Comparison of four methods of establishing control limits for monitoring quality controls in infectious disease serology testing. Clin Chem Lab Med 57:e73-e74.
133. Badrick T, Bietenbeck A, Katayev A, van Rossum HH, Loh TP, Cervinski MA, International Federation of Clinical Chemistry, Laboratory Medicine Committee on Analytical Quality. 2020. Implementation of patient-based real-time quality control. Crit Rev Clin Lab Sci 57:532-547.
134. Kim J, Swantee C, Lee B, Gunning H, Chow A, Sidaway F, Sherlock C, Garceau R, Dimech W, Malloch L, Laboratories C. 2009. Identification of performance problems in a commercial human immunodeficiency virus type 1 enzyme immunoassay by multiuser external quality control monitoring and real-time data analysis. J Clin Microbiol 47:3114-20.
135. Dimech W, Vincini G, Davies K, Karakaltsas M, van Cauwalaert ND, Guichet E, Koppelman M, Cabuang L. 2020. Validation of dried tube sample format quality controls for the monitoring of viral load and blood screening assays. J Virol Methods 285:113957.
136. Field PR, Ho DW, Cunningham AL. 1988. Evaluation of rubella immune status by three commercial enzyme-linked immunosorbent assays. J Clin Microbiol 26:990-4.
137. Kleeman KT, Kiefer DJ, Halbert SP. 1983. Rubella antibodies detected by several commercial immunoassays in hemagglutination inhibition-negative sera. J Clin Microbiol 18:1131-7.
138. Dimech W, Francis B. 1995. Evaluation of a Single-Point Calibration Method to International Units per Millilitre Used in a Rubella Virus Immunoglobulin G Kit. Australian Journal of Medical Laboratory Science 16:139-142.
139. James V, Fagan EJ. 2006. Rubella dilemma: are patients being immunised unnecessarily? J Clin Virol 36.
140. Hutton J, Rowan P, Greisinger A, Mouzoon M. 2014. Rubella monitoring in pregnancy as a means for evaluating a possible reemergence of rubella. Am J Obstet Gynecol 211:534 e1-4.
141. BioMex GmbH. 2016. Rubella IgG Negative Panel; Product code NP-RUB-001. https://www.biomex.de/fileadmin/media/biomex.de/Panel_BX_pdf/BX_Rubella-NP-001_Insert_eng.pdf. Accessed 19 December 2021.
142. UK National Screening Committee. 2012. Rubella Susceptibility Screening in Pregnancy Policy Position Statement. Public Health England, https://view-health-screening-recommendations.service.gov.uk/review/rubella-susceptibility-2011/download-documents/cover_sheet/.
143. National Institute of Biological Standards and Controls. 2020. WHO International Standard First WHO International Standard for anti-SARS-CoV-2 immunoglobulin (human) NIBSC code: 20/136, *on* NIBSC. https://www.nibsc.org/documents/ifu/20-136.pdf. Accessed 20 December 2021.
144. World Health Organization. 2021. Establishment of the WHO International Standard and Reference Panel for anti-SARS-CoV-2 antibody. Expert Committee on Biological Standardization, WHO, Switzerland.
145. Windsor WJ, Roell Y, Tucker H, Cheng C-A, Suliman S, et al. 2022. Harmonization of Multiple SARS-CoV-2 Reference Materials Using the WHO IS (NIBSC 20/136): Results and Implications. Frontiers of Microbiology 13.
146. Colson P, Motte A, Tamalet C. 2006. Broad differences between the COBAS ampliprep total nucleic acid isolation-COBAS TaqMan 48 hepatitis C virus (HCV) and COBAS HCV monitor v2.0 assays for quantification of serum HCV RNA of non-1 genotypes. J Clin Microbiol 44:1602-3.
147. Dimech W, Cabuang LM, Grunert HP, Lindig V, James V, Senechal B, Vincini GA, Zeichhardt H. 2018. Results of cytomegalovirus DNA viral loads expressed in copies per millilitre and international units per millilitre are equivalent. J Virol Methods 252:15-23.
148. Schmitz JE, Stratton CW, Persing DH, Tang YW. 2022. Forty Years of Molecular Diagnostics for Infectious Diseases. J Clin Microbiol 60:e0244621.
149. Kotton CN, Kumar D, Caliendo AM, Huprikar S, Chou S, Danziger-Isakov L, Humar A, The Transplantation Society International CMVCG. 2018. The Third International Consensus Guidelines

on the Management of Cytomegalovirus in Solid-organ Transplantation. Transplantation 102:900-931.
150. Razonable RR, Asberg A, Rollag H, Duncan J, Boisvert D, Yao JD, Caliendo AM, Humar A, Do TD. 2013. Virologic suppression measured by a cytomegalovirus (CMV) DNA test calibrated to the World Health Organization international standard is predictive of CMV disease resolution in transplant recipients. Clin Infect Dis 56:1546-53.
151. Razonable RR, Hayden RT. 2013. Clinical utility of viral load in management of cytomegalovirus infection after solid organ transplantation. Clin Microbiol Rev 26:703-27.
152. Mendez JC, Espy MJ, Smith TF, Wilson JA, Paya CV. 1998. Evaluation of PCR primers for early diagnosis of cytomegalovirus infection following liver transplantation. J Clin Microbiol 36:526-30.
153. Bravo D, Clari MA, Costa E, Munoz-Cobo B, Solano C, Jose Remigia M, Navarro D. 2011. Comparative evaluation of three automated systems for DNA extraction in conjunction with three commercially available real-time PCR assays for quantitation of plasma Cytomegalovirus DNAemia in allogeneic stem cell transplant recipients. J Clin Microbiol 49:2899-904.
154. Dimech W, Vincini G. 2017. Evaluation of a Multimarker Quality Control Sample for Monitoring the Performance of Multiplex Blood Screening Nucleic Acid Tests. Vox Sang 112.
155. Dimech W, Francis B, Kox J, Roberts G, Serology Uncertainty of Measurement Working Party. 2006. Calculating uncertainty of measurement for serology assays by use of precision and bias. Clin Chem 52:526-9.
156. Kruk T, Ratnam S, Preiksaitis J, Lau A, Hatchette T, Horsman G, Van Caeseele P, Timmons B, Tipples G. 2012. Results of continuous monitoring of the performance of rubella virus IgG and hepatitis B virus surface antibody assays using trueness controls in a multicenter trial. Clin Vaccine Immunol 19:1624-32.
157. Dimech W, Karakaltsas M, Vincini GA. 2018. Comparison of four methods of establishing control limits for monitoring quality controls in infectious disease serology testing. Clin Chem Lab Med 56:1970-1978.
158. Dimech W, Freame R, Smeh K, Wand H. 2013. A review of the relationship between quality control and donor sample results obtained from serological assays used for screening blood donations for anti-HIV and hepatitis B surface antigen. Accred Qual Assur 18:11-18.
159. Wand HD, W. Freame, R. Smeh, K. 2015. Visual and statistical assessment of quality control results for the detection of hepatitis B surface antigen among Australian blood donors. Annals of Clinical and Labortaory Research 3.
160. Wand HD, W. Freame, R. Smeh, K. 2017. Identifying the critical cut-points of a quality control process for serological assays: results from parametric and semiparametric regression models. Accred Qual Assur 22:191–198.
161. Busch MP, Bloch EM, Kleinman S. 2019. Prevention of transfusion-transmitted infections. Blood 133:1854-1864.
162. Naidu NK, Bharucha ZS, Sonawane V, Ahmed I. 2016. Nucleic acid testing: Is it the only answer for safe Blood in India? Asian J Transfus Sci 10:79-83.
163. El Ekiaby M, Lelie N, Allain JP. 2010. Nucleic acid testing (NAT) in high prevalence-low resource settings. Biologicals 38:59-64.
164. Crofts N, Maskill W, Gust ID. 1988. Evaluation of enzyme-linked immunosorbent assays: a method of data analysis. J Virol Methods 22:51-9.
165. Galli C, Plebani M. 2020. Quality controls for serology: an unfinished agenda. Clin Chem Lab Med 58:1169–1170.
166. Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. 2011. Commutability limitations influence quality control results with different reagent lots. Clin Chem 57:76-83.
167. Wang Y. 2021. Standardised neutralising antibody assays are needed for evaluating COVID-19 vaccines. EBioMedicine 73:103677.
168. Kristiansen PA, Page M, Bernasconi V, Mattiuzzo G, Dull P, Makar K, Plotkin S, Knezevic I. 2021. WHO International Standard for anti-SARS-CoV-2 immunoglobulin. Lancet 397:1347-1348.
169. Knezevic I, Mattiuzzo G, Page M, Minor P, Griffiths E, Nuebling M, Moorthy V. 2022. WHO International Standard for evaluation of the antibody response to COVID-19 vaccines: call for urgent action by the scientific community. Lancet Microbe 3:e235-e240.

170. Expert Committee on Biological Standardization. 2022. Establishment of the 2nd WHO International Standard for anti-SARS-CoV-2 immunoglobulin and Reference Panel for antibodies to SARS-CoV-2 variants of concern. World Health Organization, Geneva, Switzerland.
171. National Institute of Biological Standards and Controls. 2022. WHO International Standard 2nd International Standard for anti-SARS-CoV-2 immunoglobulin NIBSC code: 21/340 Instructions for use (Version 1.0, Dated 06/12/2022). Medicines and Health Care Products Regulatory Agency, Potters Bar, UK.
172. National Serology Reference Laboratory Australia. 2021. Independent Assessment of In-vitro Diagnostic Devices (IVDs) and Provision of Customised Validation and Verification Panels, Data analysis and Reporting., *on* NRL. https://www.nrlquality.org.au/evaluations. Accessed 28 December 2021.
173. Verch T, Trausch JJ, Shank-Retzlaff M. 2018. Principles of vaccine potency assays. Bioanalysis 10:163-180.
174. Coalition for Epidemic Preparedness Innovations. 2020. CEPI establishes global network of laboratories to centralise assessment of COVID-19 vaccine candidates. https://cepi.net/news_cepi/cepi-establishesglobal-network-of-laboratories-to-centraliseassessment-of-covid-19-vaccine-candidates. Accessed 2 April 2022.
175. Razonable RR. 2013. Management strategies for cytomegalovirus infection and disease in solid organ transplant recipients. Infect Dis Clin North Am 27:317-42.
176. Dimech W, Bowden DS, Brestovac B, Byron K, James G, Jardine D, Sloots T, Dax EM. 2004. Validation of assembled nucleic acid-based tests in diagnostic microbiology laboratories. Pathology 36:45-50.
177. Dimech W, Mulders MN. 2016. A 16-year review of seroprevalence studies on measles and rubella. Vaccine 34:4110-4118.
178. Dimech W, Mulders MN. 2016. A review of testing used in seroprevalence studies on measles and rubella. Vaccine 34:4119-4122.
179. Dimech W, Panagiotopoulos L, Marler J, Laven N, Leeson S, Dax EM. 2005. Evaluation of three immunoassays used for detection of anti-rubella virus immunoglobulin M antibodies. Clin Diagn Lab Immunol 12:1104-8.
180. Pham MD, Wise A, Garcia ML, Van H, Zheng S, Mohamed Y, Han Y, Wei WH, Yin YP, Chen XS, Dimech W, Braniff S, Technau KG, Luchters S, Anderson DA. 2020. Improving the coverage and accuracy of syphilis testing: The development of a novel rapid, point-of-care test for confirmatory testing of active syphilis infection and its early evaluation in China and South Africa. EClinicalMedicine 24:100440.

Appendix A. List of WHO International Standards for Infectious Diseases and Associated Information

Human Anti-Pneumococcal capsule Reference Serum (1st International Standard)

NIBSC code: 007sp

1st International Standard for Human Anti-pneumococcal capsule Reference Serum (nibsc.org)

007sp is for use in the enzyme-linked immunosorbent assay protocol for quantification of human IgG antibodies specific for Streptococcus pneumoniae capsular polysaccharides (Pn PS ELISA). 007sp is a pooled serum from 287 healthy volunteers following vaccination with 23 valent pneumococcal polysaccharide vaccine (Pneumovax II®). In order to estimate the concentration of antibodies 007sp antibody concentrations were defined through bridging to the previously established standard 89SF

Anti-Toxoplasma IgG, Human (International Standard)

NIBSC code: 01/600

Anti-Toxoplasma IgG, Human (nibsc.org)

The standard contains specific IgG only and is intended for standardisation of the Sabin Feldman Dye test and can be used in various immunoassays for the diagnosis of toxoplasmosis.

Anti-Parvovirus B19, plasma, human (2nd International Standard)

NIBSC code: 01/602

2nd International Standard for Anti-Parvovirus B 19 plasma, human (nibsc.org)

This material will serve as the primary biological standard for antibodies to Parvovirus B19. In the collaborative study in which this material was calibrated in International Units against the First International Standard for Anti-Parvovirus B19 serum, only assay kits which contained VP2 were used in the assignment of potency to this material.

Tetanus Toxoid for Flocculation Test (2nd International Standard)

NIBSC code: 16/302

WHO International Standard 3rd International Standard for Tetanus Toxoid for use in Flocculation Test

The 3rd International Standard for Tetanus Toxoid for use in Flocculation Test (16/302) was established by the Expert Committee on Biological Standardization of the World Health Organisation in October 2019 and replaces the 2nd IS coded 04/150. The material is intended to be used for standardization of the flocculation test to determine the Lf content of tetanus toxoid or toxin.

Syphilitic plasma IgG (human) (1st International Standard)

NIBSC code: 05/122

1st IS for human syphilitic plasma IgG

The standard can be used to calibrate the *Treponema pallidum* passive particle agglutination assay (TPPA). In addition, the standard can be used as a positive control in the Fluorescent Treponema Antibody assay and IgG enzyme immunoassays

Anti-human papillomavirus type 16 serum (1st International Standard)

NIBSC code: 05/134

05-134 HPV 16 antibodies (nibsc.org)

This material will serve as the primary biological standard for antibodies to HPV 16. This material may be used in immunoassays utilising virus-like particles and pseudovirion neutralisation tests of adequate sensitivity

Pertussis Antiserum (human)1st IS-WHO international Standard

NIBSC code: 06/140

Pertussis Antiserum (Human) 1st IS (nibsc.org)

The freeze-dried anti-serum (ampoule code 06/140) was prepared from sera kindly donated by Dr Carl Heinz Wirsing von König, Institut für Infektiologie Krefeld GmbH (IIK), Krefeld, Germany. On behalf of WHO and in collaboration with members from CBER, FDA, USA and Institut für Infektiologie Krefeld GmbH, Germany a collaborative study to compare the candidate material with the US reference preparations lot 3, lot 4 and lot 5 was organized by NIBSC in 2007. Twenty-two laboratories from 15 countries participated in this study. ELISAs for IgG anti-PT, anti-FHA and anti-69kDa were carried out by the participants. Data from the study showed that estimates of the antibody activity of preparation 06/140 in terms of the relevant US reference lot were in good agreement among laboratories

Human Plasma Anti-Influenza H5N1 (1st International Standard)

NIBSC code: 07/150

International Standard for antibody to influenza H5N1 virus (nibsc.org)

This material has been prepared from plasma of human recipients of A/Vietnam/1194/2004 (H5N1) (NIBRG-14) vaccine. This is the first International Standard for antibody to influenza H5N1 virus. This material will serve as the primary biological standard for antibodies to A/Vietnam/2004 H5N1 Clade 1 virus. In the collaborative study in which this material was calibrated in International Units, it was shown that this material is not suitable as a biological standard for antibodies to H5N1 viruses from other genetic clades.

Anti-hepatitis B surface antigen (Anti-HBs) immunoglobulin (2nd International Standard)

NIBSC code: 07/164

Second International Standard for anti-hepatitis B surface antigen (anti-HBs) immunoglobulin, human (nibsc.org)

Hepatitis B immunoglobulins are produced in many countries and the minimum potency requirements and potencies of individual batches are expressed in International Units (IU). This material is also used in diagnostic assays to determine the antibody content of sera from naturally infected individuals and vaccinees. The unitage indicative of sero-conversion and sero-protection and the unitage indicative of long-term protection is also given in IU.

Chagas (anti-*Trypanosoma cruzi* II) antibody in human plasma (1st International Standard)

NIBSC code: 09/186

1st International Standard for Chagas (anti-Trypanosoma cruzi II) antibody in human plasma (nibsc.org)

Freeze-dried preparation 09/186 contains anti-*Trypanosoma cruzi* antibodies and is representative for seropositive samples from autochthonous individuals living in Brazil, the region where *T. cruzi* II is endemic. However, the parasite could not be isolated from blood of individual donors who are in the chronic stage of disease. Thus the *T. cruzi* genotype could not be confirmed. The preparation has been assessed in a collaborative study for its suitability for use in various enzyme linked immunosorbent assays, immunofluorescence assays, agglutination assays, lateral flow assays or rapid immunochromatographic assays, western blots and a radioimmunoprecipitation assay. The collaborative study report contains full details on the reactivity of 09/186. The preparation can be used to assess the analytical sensitivity of the tests for detection of antibodies to *T.cruzi*. 09/186 is one of two standards that make up the 1st WHO Anti *Trypanosoma cruzi* I and II Antibody Reference Panel and should be used concurrently with standard 09/188.

Chagas (anti-*Trypanosoma cruzi* I) antibody in Human Plasma (1st International Standard)

NIBSC code: 09/188

1st International Standard for Chagas (anti-Trypanosoma cruzi I) antibody in Human Plasma (nibsc.org)

Freeze-dried preparation 09/188 contains anti-*Trypanosoma cruzi* antibodies and consists of seropositive samples from autochthonous individuals living in Mexcio, the region where *T. cruzi* I is endemic. However, the parasite could not be isolated from blood of individual donors who are in the chronic stage of disease. Thus the *T. cruzi* genotyope could not be confirmed. The preparation has been assessed in a collaborative study for its suitability for use in various enzyme linked immunosorbent assays, immunofluorescence assays, agglutination assays, lateral flow assays/rapid immunographic assays, western blots and a radioimmunoprecipitation assay.

Anti-human papillomavirus [HPV] 18 serum (1st International Standard)

NIBSC code: 10/140

1st WHO International Standard for anti-human papillomavirus 18 serum (nibsc.org)

This material will serve as the primary biological standard for antibodies to HPV 18. This material may be used in immunoassays utilising virus-like particles and pseudovirion neutralisation tests of adequate sensitivity

Influenza antibody (Human) to A/California/7/2009 'like' (H1N1v) virus (2nd International standard)

NIBSC code: 10/202

WHO 2nd International Standard for antibody to influenza H1N1pdm virus (nibsc.org)

This material has been prepared from plasma of human recipients of A/California/7/2009 (H1N1) (NYMC X179A) vaccine. This is the second International Standard for antibody to influenza H1N1pdm virus. This material will serve as the primary biological standard for antibodies to A/California/7/2009 like virus

Diphtheria Antitoxin Human IgG (1st International Standard)

NIBSC code: 10/262

1st International Standard for Diphtheria Antitoxin Human (nibsc.org)

The 1st International Standard for Diphtheria Antitoxin Human is intended for use as a reference preparation in assays to determine the concentration of diphtheria antibody in human serum samples. The results of an international collaborative study suggest that this standard is suitable for use as a reference preparation in toxin neutralization tests and in vitro immunoassays (including ELISA), and that the standard showed comparable behaviour to native human serum samples in the majority of different assay methods used

HBsAg (HBV genotype B4, HBsAg subtypes ayw1/adw2) (3rd International Standard)

NIBSC code: 12/226

Third International Standard for HBsAg (HBV genotype B4, HBsAg subtypes ayw1/adw2) (nibsc.org)

This preparation contains inactivated HBsAg (HBV genotype B4, HBsAg subtypes ayw1/adw2) and has been calibrated in International Units (IU) in an international collaborative study. It was calibrated against the 2nd international standard (IS) for HBsAg (A2, adw2) along with additional study samples representing different HBV genotypes. The 3rd WHO IS for HBsAg is intended to be used for the determination of analytical sensitivity of HBsAg assays and for the calibration of secondary references for HBsAg

Antibodies, Human, to Toxoplasma gondii (4th International Standard)

NIBSC code: 13/132

4th IS for Antibodies, Human, to Toxoplasma gondii (nibsc.org)

A collaborative study compared the potency of 13/132, the 4th International Standard (IS) for Antibodies, Human, to Toxoplasma gondii in the Sabin Feldman dye test relative to TOXM (3rd IS for anti-Toxoplasma Serum, Human; 1000 IU per ampoule) and 01/600 (1st IS for anti-Toxoplasma IgG, Human; 20 IU per ampoule). 13/132 is suitable for use in Enzyme Linked Fluorescence Assays and Enzyme Linked Immuno Sorbent Assays for Ig, IgA, IgM, IgG and IgG avidity, and for agglutination assays, Immuno Fluorescence Assays and Immunoblot assays to detect IgG and IgM. 13/132 reacted strongly positive for Ig, IgA, IgG and IgM in all these assays. The avidity of IgG from 13/132 is low and similar to TOXM. In terms of antibody potency, IS 13/132 falls between TOXM and 0/1600.

Diphtheria Toxoid for use in Flocculation Test (3rd International Standard)

NIBSC code: 13/212

3rd IS for Diphtheria Toxoid for use in Flocculation Test (nibsc.org)

The 3rd International Standard for Diphtheria Toxoid for use in Flocculation Test (13/212) was established by the Expert Committee on Biological Standardization of the World Health Organisation in October 2015. The material is intended to be used for standardization of flocculation assay to determine the Lf content of diphtheria toxoid. 13/212 may also be suitable as a reference preparation in other methods used to measure the Lf content of diphtheria toxoid, such as ELISA or SRD. Anti-Brucella abortus Serum, Bovine (International Standard) (2BADS)

Anti-Poliovirus serum types 1,2 and 3 (3rd International Standard)

NIBSC code: 82/585

3rd International Standard Anti-Poliovirus serum Types 1, 2 and 3 (nibsc.org)

The 3rd International Standard serum is the primary material for the assay of human serum containing antibodies to the three poliovirus serotypes. The methodology used to assess the standard was a neutralizing antibody assay, with constant virus-varying serum (WHO/VSQ97.94)

HIV-1 P24 Antigen WHO International Standard

NIBSC code: 90/636

HIV-1 P24 Antigen (nibsc.org)

This material is intended for the calibration of secondary reference materials

Anti-HBc (Anti-hepatitis B core antigen) WHO International Standard

NIBSC code: 95/522

First International Standard for anti-Hepatitis B core antigen (anti-HBc), plasma, human (nibsc.org)

A WHO Collaborative Study organised by the Paul Ehrlich Institute was undertaken to assess the suitability of a candidate reference material (NIBSC code 95/522) for detection of antibodies to hepatitis B core antigen (anti-HBc) in diagnostic assays. Thirteen laboratories from 10 countries tested the above-described materials using 20 different anti-HBc assays

Anti-Hepatitis A Immunoglobulin, Human (W1041) WHO International Standard

NIBSC code: 97/646

2nd International Standard for Anti-Hepatitis A, Immunoglobulin, Human (nibsc.org)

The Second International Standard for Anti Hepatitis A, immunoglobulin, human, was established in 1998 by the Expert Committee for Biological Standardisation and will serve as a biological reference preparation for antibodies to hepatitis A virus. This material was calibrated against the First International Standard for Anti-Hepatitis A Virus (Ferguson et al., 2001).

Anti-Measles serum (3rd International Standard)

NIBSC code: 97/648

3rd International Standard for Anti-Measles (nibsc.org)

The dual International Standard for anti-measles and anti-polio sera (2nd International Standard Anti-Measles serum (Human)/2nd International Standard for anti-poliovirus serum types 1, 2, and 3: NIBSC Code: 66/202) was established by the Expert Committee on Biological Standardization of the World Health Organization in 1991 (WHO, 1992) Stocks of the above standard are now exhausted and collaborative study was run in 2005/06 to establish a replacement. The 3rd International Standard was established by ECBS in 2006 and is available from NIBSC.

Anti-rubella Immunoglobulin, Human WHO International Standard

NIBSC code: RUBI-1-94

Anti Rubella Immunoglobulin, Human (nibsc.org)

This material has been prepared and characterised by the Statens Serum Institut (SSI), Copenhagen, Denmark. With effect from 1st July 1997, the National Institute for Biological Standards and Control (NIBSC), Potters Bar, UK is the custodian and distributor of this material. The package insert from SSI is shown on the last two pages of this document. IVD manufacturers, regulators and assay users should be made aware of RUBI-1-94 potential lack of commutability when used as a calibrant. This was highlighted by the WHO Expert Committee on Biological Standardization in TRS 68th Report (Section 3.3.4: 2018). RELEVANT INFORMATION : For details of this International Standard, please refer to the enclosed package insert from the Statens Serum Institut (prepared in 1995) which describes the preparation as the proposed 3rd international standard preparation for anti-rubella serum, human. The unpublished report of the 47th meeting of the WHO Expert Committee on Biological Standardization in 1996 contains the following explanation: "The Committee noted the confused nomenclature that had arisen when the first International Reference Preparation of Anti-Rubella Serum, Human (a serum) was replaced by the second (a preparation of normal immunoglobulin) and decided to take the necessary corrective action on replacement of the latter. In view of the results of the collaborative study, the Committee established the preparation coded RUBI-1-94 as the first International Standard for Anti-Rubella Immunoglobulin, Human and assigned a potency of 1600 International Units to the contents of each vial". The first sentence of section 3 (Use of the Standard) in the enclosed package insert from the Statens Serum Institute should be disregarded

Anti-Varicella zoster immunoglobulin WHO International Standard

NIBSC code: W1044

The First International Standard for varicella zoster immunoglobulin (1987) (nibsc.org)

This material was established as the First International Standard for varicella zoster immunoglobulin and is suitable for use in EIAs. This material was previously distributed by Sanquin Diagnostic Services, the Netherlands. With effect from 1st February 2007, the National Institute for Biological Standards and Control (NIBSC), Potters Bar, UK, is the custodian and distributor of this material.

Consultation on the WHO International Standard for antiRubella

WHO headquarters, room M205, Geneva, Switzerland
30 June 2017

AGENDA

09:00	Introduction	S. Hill, M. Nübling
	History of the WHO IS antiRubella, claims, collaborative study	S. Kempster
09:30	Standardization status of current Rubella assays, discrepancies, clinical implications	W. Dimech
10:15	How is “antiRubella positive” defined and is “low positive” safe enough? variability of immune response	L. Grangeot-Keros
10:45	Coffee break	
11:00	Regulatory experience with evaluation of antiRubella assays and threshold (IU/ml) for immunity	S. Nick
11:45	Perspective from vaccine immunization	J. Icenogle
12:30	Lunch break	
13:30	Views of IVD manufacturers	
	Roche	R. Bollhagen
	Biokit	J. Serra
	Abbott	A. Vockel
	bioMerieux	G. Bouchard
14:30	Coffee break	
14:45	Discussion; statement resulting from this consultation (for e.g. ECBS, WER)	all
16:00	Conclusions, next steps	M. Mulders, M. Nübling

Medicines & Healthcare products Regulatory Agency

NIBSC Confidence in Biological Medicines

WHO International Standard
Anti Rubella Immunoglobulin, Human
NIBSC code: RUBI-1-94
Instructions for use
(Version 9.0, Dated 04/05/2020)

1. INTENDED USE

This material has been prepared and characterised by the Statens Serum Institut (SSI), Copenhagen, Denmark. With effect from 1st July 1997, the National Institute for Biological Standards and Control (NIBSC), Potters Bar, UK is the custodian and distributor of this material. The package insert from SSI is shown on the last two pages of this document.
IVD manufacturers, regulators and assay users should be made aware of RUBI-1-94 potential lack of commutability when used as a calibrant. This was highlighted by the WHO Expert Committee on Biological Standardization in TRS 68th Report (Section 3.3.4: 2018).

RELEVANT INFORMATION

For details of this International Standard, please refer to the enclosed package insert from the Statens Serum Institut (prepared in 1995) which describes the preparation as the proposed 3rd international standard preparation for anti-rubella serum, human.

The unpublished report of the 47th meeting of the WHO Expert Committee on Biological Standardization in 1996 contains the following explanation: "The Committee noted the confused nomenclature that had arisen when the first International Reference Preparation of Anti-Rubella Serum, Human (a serum) was replaced by the second (a preparation of normal immunoglobulin) and decided to take the necessary corrective action on replacement of the latter. In view of the results of the collaborative study, the Committee established the preparation coded RUBI-1-94 as the first International Standard for Anti-Rubella Immunoglobulin, Human and assigned a potency of 1600 International Units to the contents of each vial".

The first sentence of section 3 (Use of the Standard) in the enclosed package insert from the Statens Serum Institute should be disregarded

2. CAUTION

This preparation is not for administration to humans or animals in the human food chain.

The preparation contains material of human origin, and either the final product or the source materials, from which it is derived, have been tested and found negative for HBsAg, anti-HIV and HCV RNA.
The ampouled material has been tested and found negative fo HBsAg and anti-HIV, and HCV RNA not detectable by PCR. However, As with all materials of biological origin, this preparation should be regarded as potentially hazardous to health. It should be used and discarded according to your own laboratory's safety procedures. Such safety procedures should include the wearing of protective gloves and avoiding the generation of aerosols. Care should be exercised in opening ampoules or vials, to avoid cuts.

3. UNITAGE

1600 International Units (IU) per vial

4. CONTENTS

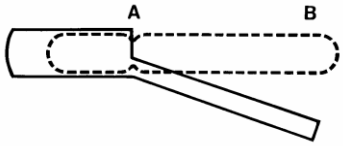
Country of origin of biological material: Denmark.

5. STORAGE

Unopened vials should be stored at -20°C or below
Please note: because of the inherent stability of lyophilized material, NIBSC may ship these materials at ambient temperature.

6. DIRECTIONS FOR OPENING

Tap the ampoule gently to collect the material at the bottom (labelled) end. Ensure ampoule is scored all round at the narrow part of the neck, with a diamond or tungsten carbide tipped glass knife file or other suitable implement before attempting to open. Place the ampoule in the ampoule opener, positioning the score at position 'A'; shown in the diagram below. Surround the ampoule with cloth or layers of tissue paper. Grip the ampoule and holder in the hand and squeeze at point 'B'. The ampoule will snap open. Take care to avoid cuts and projectile glass fragments that enter eyes. Take care that no material is lost from the ampoule and that no glass falls into the ampoule.



Side view of ampoule opening device containing an ampoule positioned ready to open. 'A' is the score mark and 'B' the point of applied pressure.

7. USE OF MATERIAL

No attempt should be made to weigh out any portion of the freeze-dried material prior to reconstitution
Dissolve the total contents of the ampoule in a known volume of physiological saline or any other suitable diluent.The solution will contain 1600 International Units per ampoule.

8. STABILITY

I
Reference materials are held at NIBSC within assured, temperature-controlled storage facilities. Reference Materials should be stored on receipt as indicated on the label.

NIBSC follows the policy of WHO with respect to its reference materials.

9. REFERENCES

Not applicable

10. ACKNOWLEDGEMENTS

Not applicable

11. FURTHER INFORMATION

Further information can be obtained as follows;
This material: enquiries@nibsc.org
WHO Biological Standards:
http://www.who.int/biologicals/en/
JCTLM Higher order reference materials:
http://www.bipm.org/en/committees/jc/jctlm/
Derivation of International Units:
http://www.nibsc.org/standardisation/international_standards.aspx
Ordering standards from NIBSC:
http://www.nibsc.org/products/ordering.aspx
NIBSC Terms & Conditions:
http://www.nibsc.org/terms_and_conditions.aspx

12. CUSTOMER FEEDBACK

Customers are encouraged to provide feedback on the suitability or use of the material provided or other aspects of our service. Please send any comments to enquiries@nibsc.org

National Institute for Biological Standards and Control,
Potters Bar, Hertfordshire, EN6 3QG. T +44 (0)1707 641000, nibsc.org
WHO International Laboratory for Biological Standards,
UK Official Medicines Control Laboratory

World Health Organization

13. CITATION

In all publications, including data sheets, in which this material is referenced, it is important that the preparation's title, its status, the NIBSC code number, and the name and address of NIBSC are cited and cited correctly.

14. MATERIAL SAFETY SHEET

Classification in accordance with Directive 2000/54/EC, Regulation (EC) No 1272/2008: Not applicable or not classified

Physical and Chemical properties	
Physical appearance: Freeze-dried powder	Corrosive: No
Stable: Yes	Oxidising: No
Hygroscopic: No	Irritant: No
Flammable: No	Handling: See caution, Section 2
Other (specify):	Contains dried material of human origin
Toxicological properties	
Effects of inhalation:	Not established. Avoid ingestion
Effects of ingestion:	Not established. Avoid ingestion
Effects of skin absorption:	Not established. Avoid ingestion
Suggested First Aid	
Inhalation:	Seek medical advice
Ingestion:	Seek medical advice
Contact with eyes:	Wash with copious amounts of water. Seek medical advice
Contact with skin:	Wash thoroughly with water.
Action on Spillage and Method of Disposal	
Spillage of ampoule contents should be taken up with absorbent material wetted with an appropriate disinfectant. Rinse area with an appropriate disinfectant followed by water. Absorbent materials used to treat spillage should be treated as biological waste.	

15. LIABILITY AND LOSS

In the event that this document is translated into another language, the English language version shall prevail in the event of any inconsistencies between the documents.

Unless expressly stated otherwise by NIBSC, NIBSC's Standard Terms and Conditions for the Supply of Materials (available at http://www.nibsc.org/About_Us/Terms_and_Conditions.aspx or upon request by the Recipient) ("Conditions") apply to the exclusion of all other terms and are hereby incorporated into this document by reference. The Recipient's attention is drawn in particular to the provisions of clause 11 of the Conditions.

16. INFORMATION FOR CUSTOMS USE ONLY

Country of origin for customs purposes*: Denmark * Defined as the country where the goods have been produced and/or sufficiently processed to be classed as originating from the country of supply, for example a change of state such as freeze-drying.
Net weight: 0.1g
Toxicity Statement: Non-toxic
Veterinary certificate or other statement if applicable**.** **Attached:** No

17. CERTIFICATE OF ANALYSIS

NIBSC does not provide a Certificate of Analysis for WHO Biological Reference Materials because they are internationally recognised primary reference materials fully described in the instructions for use. The reference materials are established according to the WHO Recommendations for the preparation, characterization and establishment of international and other biological reference standards http://www.who.int/bloodproducts/publications/TRS932Annex2_Inter_biolefstandardsrev2004.pdf (revised 2004). They are officially endorsed by the WHO Expert Committee on Biological Standardization (ECBS) based on the report of the international collaborative study which established their suitability for the intended use.

National Institute for Biological Standards and Control,
Potters Bar, Hertfordshire, EN6 3QG. T +44 (0)1707 641000, nibsc.org
WHO International Laboratory for Biological Standards,
UK Official Medicines Control Laboratory

WHO International Laboratory for Biological Standards

Phone: +45 32 68 34 66 (G.A. Hansen direct)
Telefax: +45 32 68 31 50 (Laboratory direct)
E-mail (Internet): GAHANSEN@STANDARD.SSI.DK

STATENS SERUM INSTITUT
prevention and control of infectious diseases and congenital disorders

Proposed

INTERNATIONAL STANDARD
for
ANTI-RUBELLA SERUM, HUMAN
(proposed 3rd international standard preparation)

1. THE STANDARD PREPARATION

The second International Standard for Anti-Rubella Serum, Human was established in 1970. Despite of the name this preparation consisted of ampoules each containing the freeze-dried residues of 2 ml of a mixture of normal human immunoglobulin and an equal volume of saline. An International Unit for Anti-Rubella Serum was defined as the activity contained in 0.14595 mg of the dry material in the ampoule.

The demand for the second International Standard for Anti-Rubella Serum, Human has particularly in later years, been high and the stocks are almost depleted. An interim stop for the distribution of this preparation has therefore been in force since primo 1995. The remaining ampoules will be needed for arranging the replacement and for future reference purposes.

In the interim period until a replacement preparation has been finally established, laboratories may, if needed, obtain the proposed replacement for calibration purposes.

The WHO Expert Committee for Biological Standardization noted at its 45th meeting in 1995 that "following the recognition of the need for a replacement for the second International Reference Preparation of Anti-Rubella Serum, Human, which was prepared from normal human immunoglobulin (WHO Technical Report Series No. 848, 1994 p.10), a new preparation of normal human immunoglobulin had been obtained (BS/94.1762). The Committee was informed, that this preparation will be proposed as a replacement preparation for the second International Reference Preparation of Anti-Rubella Serum, Human, and that a limited collaborative assay will be arranged".

The proposed replacement preparation (code RUBI-1-94) consist of vacuum sealed vials each containing the freeze-dried residues of 2 ml of a mixture of normal human immunoglobulin and an equal volume of sterile distilled water. Each vial is after freeze-drying controlled for vacuum by high frequency testing and individually packed in vacuum sealed plastic bags.

2. AMPOULE CONTENTS

Preliminary estimates of potency and stability of the proposed replacement preparation RUBI-1-94 have indicated the preparation to be a suitable replacement.

Statens Serum Institut
5 Artillerivej
2300 Copenhagen S
Denmark

Tel.: +45 3268 3268
Fax: +45 3268 3868
E-mail: serum@ssi.d

An international collaborative study to calibrate the proposed replacement (RUBI-1-94) against the second International Standard for Rubella Serum, Human has been carried out in 11 laboratories in 7 countries using both Immunoassays methods and Hemagglutination Inhition Test. This study has almost been completed. Based on results obtained from 9 laboratories in 7 countries using both methods potency of RUBI-1-94 is estimated to 1600 International Units (IU) of Anti-Rubella Serum, Human for the total content of an ampoule.

3. USE OF THE STANDARD

A solution of the total contents of an ampoule will contain 1700 IU in the total volume. The ampoule might be reconstituted in physiological saline or any other suitable diluent. After reconstitution the solution of the standard must preferably be used at once for calibrating working standards unless a validated procedure for storage is used.

Many laboratories have, however, experienced difficulties by calibrating their own working standards. Thus the proposed third International Standard for Anti-Rubella Serum,Human can be requested also as working standard.

The use of an immunoglobulin preparation as a reference material for diagnostic assays of human sera in clinical diagnostic laboratories is not an ideal solution. Therefore work has been going on for some time to prepare reference materials both from acute phase sera with high contents of rubella IgM antibodies and low avidity rubella IgG antibodies as well as from convalescent sera with high avidity IgG antibodies. These two new reference preparations for rubella IgM antibodies and rubella IgG antibodies are intended to replace the immunoglobulin preparation in few years.

5. WARNING

International reference materials of human origin might constitute a risk in regard to transmission of blood borne infections. All samples should therefore be treated as if capable of causing infection.
It should also be noted that distribution of international reference materials is being performed as a public service. Consequently, by accepting delivery and proceeding to use the reference material concerned, the recipient agrees not to hold the Statens Seruminstitut, Copenhagen or WHO liable for the consequences of any injury or illness attributable to infection acquired from the reference material.
If a recipient does not agree to this condition on use of the reference materials, he must immediately return the reference material to the Statens Seruminstitut, Copenhagen.

4. GENERAL REMARKS ABOUT INTERNATIONAL REFERENCE MATERIALS

International biological standards and international biological reference reagents provide a means of ensuring uniformity throughout the world in the designation of the potency or activity of preparations used in the prophylaxis, therapy, or diagnosis of disease, where this cannot be expressed in terms of physical or chemical quantities. The International Units are units of quantities of "effective constituent"

The standard is the material as it exists in the ampoules; the "material" thus includes the effective constituents together with all the other constituents that may be present (moisture, carrier, buffer, salt etc., according to the form in which the standard is available).
International biological reference materials are intended for use in the calibration of the contents of "effective constituent" in national or working standard preparations and for the expression of these contents in the respective International Units. For the routine use in the laboratory the national or working standards should preferable be used. A handling charge will be claimed for distribution of internation reference materials to other than national control laboratories.

National Institute for Biological Standards and Control,
Potters Bar, Hertfordshire, EN6 3QG. T +44 (0)1707 641000, nibsc.org
WHO International Laboratory for Biological Standards,
UK Official Medicines Control Laboratory