EVALUATION OF THE DIAGNOSTIC ACCURACY AND UTILITY OF A NEW DENGUE RAPID DIAGNOSTIC TEST IN MALAYSIA

CHONG ZHUO LIN

FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

EVALUATION OF THE DIAGNOSTIC ACCURACY AND UTILITY OF A NEW DENGUE RAPID DIAGNOSTIC TEST IN MALAYSIA

CHONG ZHUO LIN

THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PUBLIC HEALTH

FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Chong Zhuo Lin

Matric No: MWA 160009

Name of Degree: Doctor of Public Health

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Evaluation of the Diagnostic Accuracy and Utility of a New Dengue Rapid Diagnostic Test in Malaysia

Field of Study: Dengue Diagnostics

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work, I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

EVALUATION OF THE DIAGNOSTIC ACCURACY AND UTILITY OF A NEW DENGUE RAPID DIAGNOSTIC TEST IN MALAYSIA

ABSTRACT

Dengue is an emerging arboviral disease endemic to many tropical countries including Malaysia. Due to its unspecific symptoms, dengue diagnosis can only be confirmed through diagnostic tests. Existing laboratory-based dengue diagnostics like enzyme-linked immunosorbent assay (ELISA) is accurate in detecting dengue non-structural protein-1 (NS1), and anti-dengue immunoglobulin M (IgM) and G (IgG); but is also more complicated and time-consuming. Commercially available dengue rapid immunochromatographic test (RIT) is efficient but less sensitive. A new dengue rapid diagnostic rest (RDT) using biosensors may be both efficient and accurate in diagnosing acute dengue. This study evaluated this new dengue NS1-based RDT and compare its diagnostic accuracy and utility with a NS1/IgM/IgG combo RIT using a phase III diagnostic evaluation study design. It also examined the effect of various study characteristics on test performance. Firstly, a systematic review and meta-analysis was performed to review the performance of various RITs for the diagnosis of acute dengue. The effect of various study characteristics on test performance was explored through subgroup and sensitivity analyses. Secondly, after a case-control pre-test of the new RDT in a laboratory on archived samples found it to be useful; it was evaluated together with a RIT in a cross-sectional study in a public clinic on prospectively and consecutively recruited patients for the diagnosis of acute dengue infection, using established laboratory tests as reference standard. Subgroup analyses was also performed parallel to that of the systematic review. The systematic review and meta-analysis identified SD Bioline Dengue Duo as the most evaluated RIT with rather good performance. Study design, virus serotype, previous dengue exposure, disease phase, RIT brand, and reference test, were found to modify test accuracy. In the primary study, the sensitivity (SN) and

specificity (SP) of the new RDT – ViroTrack Dengue Acute NS1, were 62.3% (95%CI 55.6-68.7) and 95.0% (95%CI 91.7- 97.3), as compared to that of SD NS1/IgM/IgG combo RIT at 82.4% (95%CI 76.8-87.1) and 87.4% (95%CI 82.8-91.2) (both $p<0.001$), respectively. In addition, NS1/IgM/IgG combination would produce less false negatives compared to NS1-only test. The pattern by which different study characteristics affected test performance was found to be similar in both the review and primary study. Although ViroTrack had rather good diagnostic performance and utility, it had significantly lower SN compared to SD NS1/IgM/IgG RIT. Hence, the latter was superior to ViroTrack as it would miss less dengue patients. Detection of more dengue patients allows clinical management and vector control activities to be administered where they are needed, thus reducing disease mortality and morbidity. Future combination of serology components to ViroTrack has to be evaluated anew together with currently available RIT using phase III cross-sectional design with consecutive sampling, and proven more accurate, before it can be considered better alternatives to the existing ones.

Keywords: dengue, rapid diagnostic test, biosensors, evaluation, diagnostic performance

PENILAIAN KETEPATAN DAN KEBERGUNAAN DIAGNOSTIK SATU UJIAN DIAGNOSTIK DENGGI SEGERA DI MALAYSIA

ABSTRAK

Denggi adalah penyakit arbovirus yang kian berleluasa di banyak negara tropika termasuk Malaysia. Disebabkan kepelbagaian gejalanya yang tidak khusus, diagnosis denggi hanya boleh disahkan melalui ujian diagnostik. Diagnostik denggi berasaskan makmal yang sedia ada seperti asai imunoserapan terangkai enzim (ELISA) tepat dalam mengesan protin tidak berstruktur-1 (NS1) denggi, imunoglobulin anti-denggi M (IgM) dan G (IgG). Namun ia juga adalah lebih rumit dan memakan masa. Ujian immunokromatografik denggi segera (RIT) yang dikomersialkan adalah cekap tetapi kurang peka. Satu ujian diagnostik denggi segera (RDT) baru yang berasaskan biopenderia mungkin mampu mendiagnosis denggi akut dengan cekap dan tepat. Kajian ini merupakan kajian penilaian diagnostik fasa III yang menilai RDT NS1 denggi baru ini dan membandingkan ketepatan dan kebergunaan diagnostiknya dengan satu RIT kombo NS1/IgM/IgG. Ia juga mengkaji pengaruh pelbagai ciri kajian terhadap prestasi ujian. Pertama sekali, satu ulasan sistematik dan meta-analisis dilakukan untuk mengkaji prestasi pelbagai RIT untuk diagnosis denggi akut. Pengaruh pelbagai ciri kajian terhadap prestasi ujian diterokai melalui analisis subkumpulan dan kepekaan. Kedua, setelah RDT baru tersebut didapati berguna dalam satu pra-kajian kes-kawalan yang dijalankan di makmal mengunakan sampel serum yang diarkibkan, ia dinilai bersama satu RIT dalam satu kajian keratan rentas. Kajian ini dijalankan di klinik kesihatan atas pesakit yang disyaki menjangkiti denggi. Mereka disampel secara prospektif dan berturutan. Ujian berasaskan makmal yang mantap digunakan sebagai piawai rujukan, Analisis subkumpulan juga dilakukan selari dengan ulasan sistematik. Ulasan sistematik dan meta-analisis mengenal pasti SD Bioline Dengue Duo sebagai RIT yang paling banyak dinilai dengan prestasi yang agak baik. Reka bentuk kajian, serotip virus, jangkitan

denggi yang lepas, fasa penyakit, jenama RIT dan ujian piawai rujukan didapati mengubah ketepatan ujian. Dalam kajian utama, kepekaan (SN) dan kekhususan (SP) untuk RDT baru – ViroTrack Dengue Acute NS1, adalah masing-masing 62.3% (95%CI 55.6-68.7) dan 95.0% (95%CI 91.7- 97.3), berbanding dengan 82.4% (95%CI 76.8-87.1) dan 87.4% (95%CI 82.8-91.2) bagi RIT kombo NS1/IgM/IgG SD ($p<0.001$ bagi kedua-dua). Di samping itu, kombinasi NS1/IgM/IgG akan menghasilkan negatif palsu yang lebih rendah berbanding dengan ujian NS1 tunggal. Corak di mana ciri kajian yang berbeza mempengaruhi prestasi ujian didapati hampir serupa dalam ulasan sistematik dan kajian utama. Walaupun ViroTrack mempunyai ketepatan dan kebergunaan diagnostik yang agak baik, ia mencatatkan kepekaan yang jauh lebih rendah berbanding RIT NS1/IgM/IgG jenama SD. Oleh itu, bilangan pesakit denggi yang RIT NS1/IgM/IgG ini bakal terlepas pandang adalah kurang sedikit, sekaligus menjadikannya lebih unggul daripada ViroTrack. Pengesanan lebih ramai pesakit denggi membolehkan penjagaan klinikal dan aktiviti kawalan vektor dijalankan di mana mereka diperlukan, sekaligus mengurangkan kadar mortaliti dan morbiditi denggi. Pada masa hadapan, kombinasi komponen serologi dengan ujian NS1 jenama ViroTrack harus dinilai semula bersama dengan RIT yang sedia ada melalui kajian keratan rentas fasa III dengan pensampelan berturutan. Hanya sekiranya ia terbukti lebih tepat berbanding RIT yang sedia ada bolehlah ia dianggap sebagai pilihan alternatif yang lebih baik.

Kata kunci: denggi, ujian diagnostik segera, biopenderia, penilaian, prestasi diagnostik

ACKNOWLEDGEMENTS

Five years ago, in 2014, I was denied the federal scholarship for the post-graduate Public Health program in University of Malaya. In my reluctance to retry for the following year, I prayed in Jesus' name that unless You would alleviate my three worries, let me fail the second attempt. Against all odds, I was granted the scholarship and saw my first two prayers answered with overflow. Now that I am getting close to graduation, You obviously remember my final petition. Looking back now, I cannot be grateful enough to You, my Faithful God, who continues to love me and empower me, despite my lack of faith and many weaknesses. In Your divine arrangement, countless people contributed to shaping me into who I am, and journeyed with me this past four years. This thesis is dedicated to You, the Amazing God, and to all you wonderful people.

To my parents, Chong Chey Yean and Eng Gaik Sim, for your unconditional love. Your selfless sacrifice and provision made it possible for me to be where I am and do what I do now. I love you too.

To my wife, Yi Ping, for your unwavering trust, comforting support, and moving motivation. You cheered me up even when I doubted my own self. Please continue to do so. I need it for the rest of my life.

To my brothers and sisters in Christ in DUMC and EMCLC, for your relentless prayers and encouraging words when obstacles and challenges overwhelmed me. Please keep praying. God does listen. I can testify to that.

To my classmates, especially those of you (you know who you are), for your kind concern and company. Just knowing that you have more life commitments yet suffered as much as I did kept me going. I am amazed by you!

To my supervisors, Chiu Wan and Sanjay, for your insightful guidance and confidence in me. You were always there when I most needed you. You might not be aware that sometimes by just sitting in your room calmed my urge to cry, sometimes the opposite was true. Your importance to me cannot be overstated. For all you have done I thank you from the bottom of my heart.

To the people of Malaysia who sponsored my study, and to those who were instrumental in making this doctoral research project possible, the university staff of the Department of Social and Preventive Medicine, the university laboratory experts from the Department of Medical Microbiology, the Ministry of Health collaborators from Petaling district health office and its clinic, the funder - BluSense Diagnostics, as well as to many others who helped, I express my heartfelt appreciation.

Finally, to those of you who are reading this thesis, and even this part of it, thank you for taking it off the shelf. Writing this thesis was the hardest stunt I have ever pulled off in my entire life. Your attention makes my effort worthwhile. I hope you find it useful. Please, as the Malay nursery rhyme "Rasa Sayang" goes, '*saya budak baru belajar, kalau salah tolong tunjukkan*' (I'm just a kid new in learning, enlighten me if I'm mistaken). Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ADE	:	antibody-dependence enhancement
Ae.	:	Aedes
AF	:	Thailand Armed Forces Research Institute of Medical Science
AIDS	:	acquired immunodeficiency syndrome
AUC	:	area under curve
BD	:	BluSense Diagnostics
CDC	:	United States Centers for Disease Control and Prevention
CFR	:	case fatality rate
CI	:	confidence interval
DENV	:	dengue virus
DF	:	dengue fever
DHF	:	dengue haemorrhagic fever
DOR	:	diagnostic odd ratio
DSS	:	dengue shock syndrome
ELISA	:	enzyme-linked immunosorbent assay
EQ	:	equivocal
ESS	:	effective sample size
FBC	:	full blood count
FN	:	false negative
FP	:	false positive
HI	:	haemagglutination inhibition
HSROC	:	hierarchical summary receiver operating characteristics
IFA	:	immunofluorescence assay
IgG	:	immunoglobulin G

IgM	:	immunoglobulin M
IMR	:	Institute for Medical Research
IR	:	incidence rate
JE	:	Japanese encephalitis
KKS7	:	Shah Alam Section 7 Health Clinic
lnDOR	:	diagnostic log odd ratio
LR-	:	negative likelihood ratio
LR+	:	positive likelihood ratio
MAC-ELISA	:	IgM antibody capture enzyme-linked immunosorbent assay
MDGs	:	Millennium Development Goals
MNPs	:	magnetic nanoparticles
MOH	:	Ministry of Health
NAAT	:	nucleic acid amplification tests
NPHL	:	National Public Health Laboratories
NPV	:	negative predictive value
NS1	:	non-structural (antigen/protein)-1
NTDs	:	neglected tropical diseases
Panbio IC	:	Panbio Immunochromatographic Card Test
PCR	:	polymerase chain reaction
PPV	:	positive predictive value
PRNT	:	plaque reduction neutralization test
QUADAS-2	:	Quality Assessment of Diagnostic Accuracy Studies-2 (checklist)
RDT(s)	:	rapid diagnostic test(s)
RIT(s)	:	rapid immunochromatographic test(s)
RNA	:	ribonucleic acid
RT-PCR	:	reverse-transcription polymerase chain reaction

SD	:	Standard Diagnostics
SDGs	:	Sustainable Development Goals
SEA	:	Southeast Asia
SN	:	sensitivity
SP	:	specificity
SROC	:	summary receiver operating characteristic
STARD	:	Standards for the Reporting of Diagnostic Accuracy Studies (checklist)
TN	:	true negative
TP	:	true positive
UM	:	University of Malaya
UMMC	:	University Malaya Medical Centre
US	:	United States (of America)
VI	:	virus isolation
WHO	:	World Health Organization
-ve	:	negative
+ve	:	positive

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Introduction

The rapid development in science and technology over the past decades has brought mankind betterment of health and increased life expectancy, accompanied by epidemiological shift from communicable to non-communicable diseases. Nevertheless, some communicable diseases are emerging, one of which is dengue fever (DF). According to the World Health Organization (WHO), the incidence rate of dengue has risen 30 times over the past 50 years. It is now endemic in more than 100 countries including Malaysia with up to 100 million infections happening yearly worldwide (World Health Organization, 2012). A more recent disease distribution model gave a very high estimate of 390 million dengue infections worldwide in 2010, of which around one million happened in Malaysia (Bhatt et al., 2013). The resulting dengue economic burden was estimated to be US$ 8.9 billion globally in 2013 (Donald S. Shepard, Undurraga, Halasa, & Stanaway, 2016). These cost estimations would be higher if the global yearly loss of up to 20000 human lives and 264 disability-adjusted life years per million population were included (World Health Organization, 2012).

WHO in its latest *Global Strategy for Dengue Prevention and Control* implored its member countries, among others, to improve outbreak prediction and detection by establishing a coordinated epidemiological and entomological surveillance to reduce dengue morbidity, and to implement early detection and management of dengue cases by reorienting health services and training personnel at primary care level to reduce dengue mortality (World Health Organization, 2012). A key to the success of both the above recommendations is an early and accurate diagnosis for every dengue patient that is made within the short duration of healthcare utilisation. Without that, on a patient level, misdiagnosis or late diagnosis of dengue could lead to death from severe complications. On a macro level, the dengue surveillance data gathered would be misreported, thus

leading to ineffective disease prevention and control. Since diagnosis of dengue based solely on clinical symptoms is difficult due to its unspecific symptoms that often resemble other diseases, a rapid yet accurate diagnostic tool that can provide early dengue diagnosis is essential (Peeling et al., 2010; World Health Organization, 2009a).

The focus of this thesis is on dengue diagnostics, in particular on the diagnostic performance of a new dengue rapid diagnostic test (RDT) as compared to a commercially available dengue RDT, or more specifically, a rapid immunochromatographic test (RIT), for the diagnosis of acute dengue infection among Malaysian population; and the public health implications of their diagnostic performance. Dengue RDTs, in general, are easy to perform and quick to obtain result, as opposed to laboratory-based dengue diagnostics. If proven to have comparable good performance, they can provide an early and accurate dengue diagnosis at the point-of-care to facilitate not just timely secondary prevention to reduce mortality, but also primary prevention to reduce morbidity. In this introductory chapter, Section 1.1 provides a general overview of dengue and the importance of dengue RDT as a key to its prevention and control, Section 1.2 presents the study motivation, Section 1.3 contains the objectives of the study, Section 1.4 describes the significance of the study, and Section 1.5 lays out the structure of the entire thesis.

1.2 Study motivation

Dengue is one of the neglected tropical diseases (NTDs) characterised by their close association with poverty and proliferation in tropical climates, where most of the world population resides (World Health Organization, 2013). Despite their negative impact to a huge number of human lives, due to the lack of political voice and representation of these populations, and more subtle yet crucial, the lack of profitable return of investment into these NTDs, they were previously not regarded as important and not given much attention internationally.

This neglect was evident following the launch of the eight Millennium Development Goals (MDGs), when the sixth goal that aimed to combat acquired immunodeficiency syndrome (AIDS), malaria and other diseases was successful in drawing the world's attention to fight them and even tuberculosis, but left behind these other diseases. As such, the NTDs conceptual framework was formulated to define these other diseases and provide strategies for their control and elimination. In essence, it was a branding that gave a human face to the people suffering these other diseases, highlighted their plight, and raised awareness towards the importance of NTDs control and elimination (Institute of Medicine (US) Forum on Microbial Threats, 2011). This strategy was proven successful when NTDs were formally incorporated into the third goal of a total of 17 Sustainable Development Goals (SDGs) that replaced the MDGs after year 2015, which aims to end their epidemics among other things (United Nations, 2015).

Fortunately, in contrast to other NTDs, the attention given to dengue began much earlier, perhaps mainly due to the rapid economic development, rise of middle-class, and emergence of consumer market with great potential in some of the populous countries and regions it is endemic to, such as Brazil, India, Mexico, and Southeast Asia (SEA) (Goldstein Research, 2018; Kaddar, Milstien, & Schmitt, 2014). Compounding on the

above factors was the spread of dengue to richer and more developed temperate regions due to increased international travel and commerce, with many cases reported in the United States of America (US), Southern Europe, Japan, and Korea since the advent of 21st century (Arima et al., 2014; Park & Lee, 2012; Rezza, 2014). Coupled with climate change that saw increasingly hotter summer in the seasoned countries that favoured breeding of its vector, dengue fever has become one of the most widely distributed and impactful mosquito-borne viral disease affecting mankind with potential of more widespread outbreaks (Bowman, Donegan, & McCall, 2016; Faraji & Unlu, 2016; Goubert, Minard, Vieira, & Boulesteix, 2016).

Happening alongside these global trends were the development and advancement of dengue laboratory diagnostics tools. For many years the laboratory diagnostic tests for dengue have been characterised by the complexity in their methodology, long duration, requirement of well-equipped laboratory and skilled personnel, and high costs. These factors became the barriers of their wide adoption especially in low-resource settings such as rural areas, primary care, and emergency department. In these settings, diagnosis and management were solely based on clinical symptoms with or without full blood count (FBC). This was suboptimal as dengue symptoms are not specific, and changes in FBC might not be apparent in the early phase of illness (Peeling et al., 2010; World Health Organization, 2009a).

In view of that, the appearance of dengue RDT in the market around two decades ago, majority of which were RIT, was seen as a solution to the diagnostic gap in low-resource settings (Miller & Sikes, 2015). Being mostly card-based tests like widely used urine pregnancy tests, these RITs are intuitive to users and simple to perform, at the same time more affordable compared to conventional laboratory-based dengue diagnostics. They can provide clinicians results within the duration of a patient's visit, and therefore, are

very suitable for point-of-care diagnosis of acute dengue infection. Early diagnosis of dengue infection allows timely clinical intervention to be given to patients and prevents mortality due to misdiagnosis or missed and late diagnosis. In view of that, healthcare providers in many countries have adopted RIT into their dengue diagnostic algorithm since then (Bisordi et al., 2011; Huang et al., 2013; Huits et al., 2017; Shih et al., 2016).

Unfortunately, the efficiency of these RITs usually came with a sacrifice of their test accuracy, i.e. test sensitivity and specificity. In fact, their manufacturers often claimed good performance, but their accuracy as evaluated by independent researchers was almost always lower than the official figures (S. D. Blacksell et al., 2011; Hunsperger et al., 2014; Standard Diagnostics Inc., 2008). This may be due to the differences in the diagnostic landscape that was unique for each location and setting, such as patient populations, dominating serotype, predominance of primary or secondary dengue, preference of healthcare seeking, and also partly due to systemic bias. Apart from these extrinsic factors, intrinsic factors specifically related to RIT built and usage, such as their qualitative nature and subjective interpretation, may have played a role too (P. M. Bossuyt et al., 2015; Shamala, 2015).

In addition to the above, evaluation studies of dengue RDT may produce differing results due to the difference in their original context. Due to the absence of widely accepted international consensus on standard and guidelines of diagnostic evaluation studies, they were often performed, reported, and interpreted without having a framework in mind. In other words, evaluation studies differed in many ways. Blind comparison and generalization of their results is scientifically unsound (Leeflang, 2014).

In view of this, when interpreted out of context, the results of dengue RDT evaluation studies can deceive health policy makers and physicians. RDTs that actually performed poorly could have been taken as a good one and purchased for use. Not only would this

affect the accuracy of the diagnosis and subsequent management for individual patients, it would also upset the accuracy of the disease surveillance systems and effectiveness of subsequent dengue prevention and control activities that rely on it. If the underperformance goes undetected, poorly performed but affordable dengue RDTs may even put good but more expensive RDTs out of competition. Therefore, it is important to always evaluate a dengue RDT, whether RIT or not, before it is being used for the first time, before using it on another population, on a different type of sample, or for another purpose other than the one it was originally intended for and validated on (P. M. Bossuyt et al., 2015; Leeflang, 2014; Shamala, 2015).

Moreover, evaluation studies should be designed and performed to fit a framework according to established guidelines and recommendations, for quality assurance and to guide results interpretation. Four phases of diagnostic test evaluation studies proposed previously are listed in Table 1.1.

Table 1.1 Four phases of diagnostic test evaluation framework

Phase	Description
I (Discovery)	Establishment of technical parameters, algorithms, and diagnostic criteria
II (Introductory)	Early quantification of test performance (or accuracy) in clinical settings
III (Mature)	Comparison to other tests in prospective, typically multi-institutional studies (efficacy)
IV (Disseminated)	Assessment of the test as utilised in the community at large (effectiveness)

(adapted from Patient-Centered Outcomes Research Institute (PCORI), 2012)

Phase I studies aim to determine the normal range of values for a diagnostic marker by observing healthy people, while phase II studies are usually case-control studies that determine diagnostic accuracy by also including people with known disease as determined by diagnostic standard. Phase III studies are randomised trials conducted prior to the licensing of a test; while phase IV studies evaluate a licensed test used in practice on large cohorts of consecutive participants to examine whether its diagnostic accuracy corresponds to that found in phase III (Gluud & Gluud, 2005). Apart from diagnostic accuracy, both phase III and IV studies also aim to determine test diagnostic utility, i.e. clinical consequences or effects of introducing a new diagnostic test into clinical practice (P. M. M. Bossuyt, Reitsma, Linnet, & Moons, 2012; Leeflang, 2014). In view of this, phase III study should be the bare minimum prior to the licensing of a dengue RDT, which, however, is not yet made mandatory in many countries (Committee for Medicinal Products for Human Use, 2009).

As of now, efforts are still ongoing to improve the diagnostic accuracy of dengue RDT, particularly in the field of biosensors. This new technology has the potential to overcome some shortcomings of the currently used dengue RIT. One of the latest candidates is the ViroTrack immuno-magnetic agglutination assay developed by BluSense Diagnostics (BD), Denmark. It is believed to be able to improve test sensitivity by employing magnetic nanoparticles (MNPs) capable of forming sandwich agglutination with target analytes in just one drop of blood. Using microfluidic (lab-on-a-chip), the whole process of blood plasma separation, metering, mixing, resuspension of MNPs, and sample analysis happen inside a portable lightweight opto-magnetic reader that produce result within 15 minutes. This result is quantifiable, making its interpretation objective in nature, and may be more accurate compared to RIT.

In view of the above-mentioned difference in context, diagnostic landscape between studies, and potential of bias, this new dengue RDT should be evaluated by independent researchers, together with other established dengue diagnostic tests including dengue RIT that have been proven accurate, to provide a scientifically valid comparison for its diagnostic performance (Antunes et al., 2015; Shamala, 2015). Furthermore, this evaluation study should be a phase III study that also demonstrate the diagnostic utility or potential impact to clinical outcomes as a result of the application of this new RDT and its comparators (Committee for Medicinal Products for Human Use, 2009; Gluud & Gluud, 2005; Rutten, Moons, & Hoes, 2006).

1.3 Study objectives

The aim of this study was to evaluate the diagnostic accuracy and diagnostic utility of a new dengue rapid diagnostic test for the diagnosis of acute dengue infection in Malaysia using phase III diagnostic evaluation study design.

The specific objectives of this study were:

1) To review the diagnostic accuracy of commercially available dengue rapid immunochromatographic tests for the diagnosis of acute dengue infection:

   a. To determine the actual diagnostic accuracy of commercially available dengue rapid immunochromatographic tests for the diagnosis of acute dengue infection;

   b. To identify the source of heterogeneity that modify the diagnostic accuracy of commercially available dengue rapid immunochromatographic tests for the diagnosis of acute dengue infection;

   c. To identify the commercially available dengue rapid immunochromatographic test with the best diagnostic accuracy for the diagnosis of acute dengue infection.

2) To evaluate and compare the diagnostic accuracy and diagnostic utility of a new dengue rapid diagnostic test and a commercially available dengue rapid immunochromatographic test for the diagnosis of acute dengue infection in a primary care setting in Malaysia.

1.4 Study significance

The evaluation of the diagnostic performance of a new dengue RDT in Malaysia is important and the outcomes can improve patient care and management, inform health policy making, increase competition in the RDT market, and improve healthcare and population health.

1.4.1 Improve patient care and management

The most immediate benefit of this study is that it can help physicians seeing patients with suspected dengue in similar settings make sense of dengue RDT results. Ideally, physicians want a dengue RDT that can capture all dengue cases without missing one (highly sensitive), yet able to exclude other diseases with similar clinical symptoms (highly specific). Unfortunately, the test results are also often interpreted in the same manner, where positive result means presence of dengue infection and negative – absence of it (Peeling et al., 2010). Confusion arises when the test result contradicts physicians' provisional diagnosis. In some cases, test results are taken as truth; in other cases, where confirmation bias is stronger, physicians may choose to follow own instinct and manage according to that (Parmley, 2006).

The truth is, there exists no perfect diagnostic test. For dengue RDT, the matter is complicated by the progression of disease that produces changing level of target biomarkers that may or may not be captured by a test (Stuart D. Blacksell, 2012). The outcome of this study can provide physicians the actual performance of this new dengue RDT. Even if they do not have access to it, the performance of other dengue diagnostic tests evaluated for comparison including that of the RIT can inform them on the possible actual performance of other dengue RITs they are using. It is also the intention of this study to help them put into perspective the test results and their interpretation, with the

hope that the diagnosis and subsequent care and management of patients with suspected dengue can be improved (Florkowski, 2008; McGee, 2002).

1.4.2 Inform health policy making

In many developing countries, the sales and usage of dengue RDT are not strictly regulated. In Malaysia, only dengue RDT that has the potential to be procured and used by the Ministry of Health (MOH) is required to undergo a phase II evaluation by the Institute for Medical Research (IMR). Even then this applies to those used in primary care setting (Insitute for Medical Research, 2012; Thayan, 2018). Hospital pathologists have the authority and liberty to evaluate and procure their own tests. And most of the time, the initiative for evaluation comes from diagnostic companies or distributors that intend to sell their kits to the government (Ministry of Health, 2010). In private sector, the sales and usage of dengue RDT are self-regulated and more subject to the influence of market force. Even when a dengue RDT was evaluated internally, like in the public sector, the results were rarely published.

This asymmetry of information due to the lack of transparency and information collaboration restricts health policy makers from making an informed decision in the selection of dengue RDT in one setting or another. In the public sector, better dengue RDTs available in the market may not have a fair chance of being selected. The private sector, on the other hand, need to rely more on the diagnostic performance reported by the manufacturers in decision making. As a result, there is always an uncertainty that dengue RDTs were procured and used without sound evidence of its effectiveness (Peeling et al., 2010; World Health Organization, 2009b).

An evaluation study like this by an independent third party provides a fair assessment to the diagnostic performance of the evaluated dengue RDT. Its phase III design gives health policy makers its actual performance, as well as that of the comparator diagnostic tests, in a specific setting and helps them make an informed decision in the selection of dengue RDT for their own settings. On top of that, it may also highlight the sources of discrepancy in the self-reported and independently evaluated diagnostic performance for these tests. This may point to a need of a more coordinated effort in setting research priority and policy making for the application of dengue RDT in Malaysia (P. M. Bossuyt et al., 2015; Miller & Sikes, 2015; Peeling et al., 2010).

1.4.3 Increase competition in dengue rapid diagnostic tests market

Introduction of a new dengue RDT into the market increases the competition between existing ones by providing more product variety. Furthermore, if this new dengue RDT employs new technology that has the potential to outperform existing RDTs, i.e. improves product quality, the competition becomes tighter. In both situations, price reduction and/or product innovation may happen as a result of this increased competition since they are necessary for a dengue RDT manufacturer to remain competitive in the market. Innovation, in turn, leads to product with higher quality (Gaynor, 2007).

In some cases, however, competition may lead to reduced product quality through unhealthy price reduction coming especially from low quality dengue RDTs that compete on price instead of effectiveness. This price reduction may erode the market share and profit margin of other manufacturers, thus disincentivises them from producing high quality tests that are more costly to make. Fortunately, this issue can be overcome through regulation imposed on this market such as the abovementioned coordinated plan of actions that may include mandatory evaluation by one or a few independent third parties

prior to the approval of sales and usage of any dengue RDT in Malaysia (Gaynor, 2007; Miller & Sikes, 2015; Peeling et al., 2010).

As such, a study that evaluates new dengue RDT and compares it with existing tests may reduce the price of the latter and improve their quality through additional competition. More similar studies such as this, in the long run, will help eliminate unfit low-quality dengue RDTs still available in the market. The nett effect is better dengue RDTs at lower price, which benefits health policy makers when it comes to their selection and application, and eventually increases their access for the patients who need them (Miller & Sikes, 2015).

1.4.4 Improve healthcare and population health

When health policy makers and the populations they serve have access to high quality dengue RDTs at a lower price, the healthcare system becomes more cost-effective and efficient. As these tests are designed for point-of-care dengue diagnosis, settings like primary care and even emergency care do not need to rely on laboratory-based dengue diagnostics that require more resources. Access to high quality RDTs that are accurate on-site means physicians misdiagnose dengue less. This means fewer patients with dengue are missed and those detected are managed appropriately, thus reducing complications, shorten hospitalisation, and prevent loss of lives. On the other hand, non-dengue patients with non-life threatening diseases are managed accordingly and excluded from unnecessary follow-up and further interventions (Ministry of Health, 2015).

Up until this point, the benefits that were discussed concentrated mainly on the improvement of dengue clinical outcomes, i.e. reduction in mortality. But the impact study like this brings goes beyond that. In the long run, it may bring as much benefits if

not more to the public health aspect of dengue. The first and foremost significance is the improvement of dengue surveillance. Dengue is a notifiable disease in Malaysia. Upon diagnosis, it is required to be notified within 24 hours with dengue RDT results. Hence, improvement in dengue RDT diagnostic accuracy also means improvement in the accuracy of dengue surveillance system through reduction in misreporting, i.e. increase in the notification of actual dengue and decrease in the notification of non-dengue cases (Imai, Dorigatti, Cauchemez, & Ferguson, 2015).

A more accurate dengue surveillance system provides health policy makers more accurate estimation of the actual burden of dengue in Malaysia. This may translate into more informed health planning, more efficient allocation of human and financial resources and so on. One good example is dengue vector control activities conducted by the health authority, which depend very much on the surveillance system. At most places in Malaysia especially urban areas, where vector control units are already overwhelmed with gluts of daily dengue notification, improvement in their accuracy means these activities can be targeted at locations that really need it. As a result, the conduct of these activities also become more efficient and effective, which can in turn reduce the transmission of dengue virus (DENV) leading to less dengue infection among the population, i.e. decreased dengue burden or morbidity (World Health Organization, 2012).

As evident above, this study has the potential to improve the accuracy of dengue diagnosis and has scores of other chain benefits, which eventually reduce both dengue morbidity and mortality. Either way reduces wastage and results in a more efficient and cost-effective healthcare system, where additional resources can be relocated to other neglected areas. All these, together with decrease in burden of dengue, will translate into a healthier population and a stronger nation.

1.5 Thesis layout

This thesis consists of eight chapters and is arranged in the following manner. The current Chapter 1 is the introduction, followed by two chapters of literature review. Chapter 2 describes briefly the epidemiology of dengue and then focuses on challenges of diagnosing acute dengue infection. Chapter 3 shifts the attention to the development and current dengue situation in Malaysia.

Subsequent two chapters each answers a specific objective. Chapter 4 reviews the diagnostic performance of commercially available dengue rapid immunochromatographic tests for the diagnosis of acute dengue infection. Chapter 5 evaluates the diagnostic performance of a new dengue rapid diagnostic test for the diagnosis of acute dengue infection in a primary care setting in Malaysia.

The final two chapters of this thesis are Chapter 7 that discusses the findings from this study, and Chapter 8 that concludes this thesis with policy recommendations. Following the final chapter are references, list of publications and presentations, and appendix.

Last but not least, to avoid confusion, the usage of certain terms in this thesis is explained here. Dengue RIT is a subset of dengue RDT, and is used in a narrow sense. However, dengue RDT may be used specifically to indicate the new test, or broadly to include all rapid tests. Diagnostic performance and diagnostic accuracy may be used interchangeably, but the former also includes diagnostic utility that is more applied and practical. In whichever situation, the context in which these terms are used is important as it determines their meaning.

CHAPTER 2: DIAGNOSING DENGUE FEVER AND ITS CHALLENGES

2.1 Introduction

Dengue has coexisted with humanity for a long time, even before the rise of germ theory. Although many earlier accounts reported illnesses similar to dengue, this infection was first diagnosed and described in detail in 1789 by a physician among the US founding fathers - Benjamin Rush, as a 'bilious remitting fever' (Gubler, 1998; Liu, Liu, & Cheng, 2016). Interestingly, this infection was also named and is still called currently in some cultures according to its manifestation, the most pictorial and self-explaining being break-bone fever. Other names include dandy, three-day, seven-day, and giraffe fever. In the 19th century, the term 'dengue fever' slowly became popular and widely accepted (Liu et al., 2016; Lo & Perng, 2016). However, it was only until the 20th century that its pathogen DENV was discovered (Gubler, 2004; Scitable, 2014). Since then, mankind has been advancing in the knowledge of dengue. Nevertheless, diagnosing it remains a challenging task today.

Section 2.2 of this chapter reviews briefly the epidemiology of dengue, followed by human host body response towards dengue infection in section 2.3. Section 2.4 reviews commonly used dengue laboratory diagnostics currently and discusses their challenges. Section 2.5 focuses on commercially available dengue RDTs and their issues. Section 2.6 summarises the whole chapter on dengue fever diagnosis.

2.2 Epidemiology of dengue

Dengue fever is an arthropod-borne infectious disease caused by DENV with four serotypes (DENV-1 to DENV-4) from the genus Flavivirus (family Flaviviridae), which also includes other antigenically closely related sylvatic viruses such as yellow fever virus, Japanese encephalitis (JE) virus, West Nile virus, and Zika virus. DENV originally transmitted between non-human primates and arboreal mosquitoes but became endemic among human over the last few centuries due to population growth and increasing human activities in the wild (Lambrechts & Lequime, 2016; Moi, Takasaki, & Kurane, 2016; Tambo, Chuisseu, Ngogang, & Khater, 2016).

Among humans, DENV is transmitted by female Aedes (Ae.) mosquitoes from subgenus Stegomyia, primarily Ae. aegypti and Ae. albopictus. The life cycle of an Aedes mosquito is around eight to 10 days at room temperature, which comprises of two phases – aquatic (eggs, larvae, and pupae) and terrestrial (adults). Nutrients from human blood are needed for egg production. Upon an infectious blood meal, the digestive tract of an adult female mosquito is infected. After replication in its midgut, DENV is transferred to the haemocoel followed by its systemic dissemination to all secondary organs including salivary glands, which release DENV into the blood stream of a healthy human host during subsequent bite. The virus is also vertically transmitted to the eggs, making the offspring infectious from the very beginning (Khetarpal & Khanna, 2016; Lambrechts & Lequime, 2016).

Aedes mosquito is a container-breeder that thrives in urban and suburban areas. It oviposits in fresh water deposited anywhere in the peridomestic environment, even waste items. Initially distributed mainly in developing tropical and subtropical regions, it has now spread to developed temperate regions due to international travel and commerce, as well as global warming. This spread is mainly accomplished by the extremely invasive

Ae. albopictus, also termed Asian tiger mosquito due to its resemblance. As a result, dengue fever has become one of the most widely distributed and impactful mosquito-borne viral disease affecting mankind with outbreaks reported in more than 100 countries (Bowman et al., 2016; Faraji & Unlu, 2016; Goubert et al., 2016).

2.3 Human host body response towards dengue infection

In human host, DENV causes a broad spectrum of pathological conditions ranging from self-limiting asymptomatic infection and mild undifferentiated DF; to more severe forms of dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS), according to the WHO 1997 dengue case definition (World Health Organization, 1997). Even then, due to the lack of representativeness of the population the symptoms were based upon, ineffectiveness of tourniquet test to differentiate between DF and DHF, the unintended emphasis the term DHF had on haemorrhage instead of more life threatening plasma leakage, research findings suggesting the three seemingly distinct yet overlapping conditions as a continuum, and its failure in identifying many severe patients, the 1997 classification was replaced by another one in 2009. This latest classification groups patients into dengue with and without warning signs, and severe dengue (Figure 2.1). It was proven to be broader and more sensitive in detecting both dengue cases and severe dengue (Gan et al., 2013; Hadinegoro, 2012; World Health Organization, 2009a).

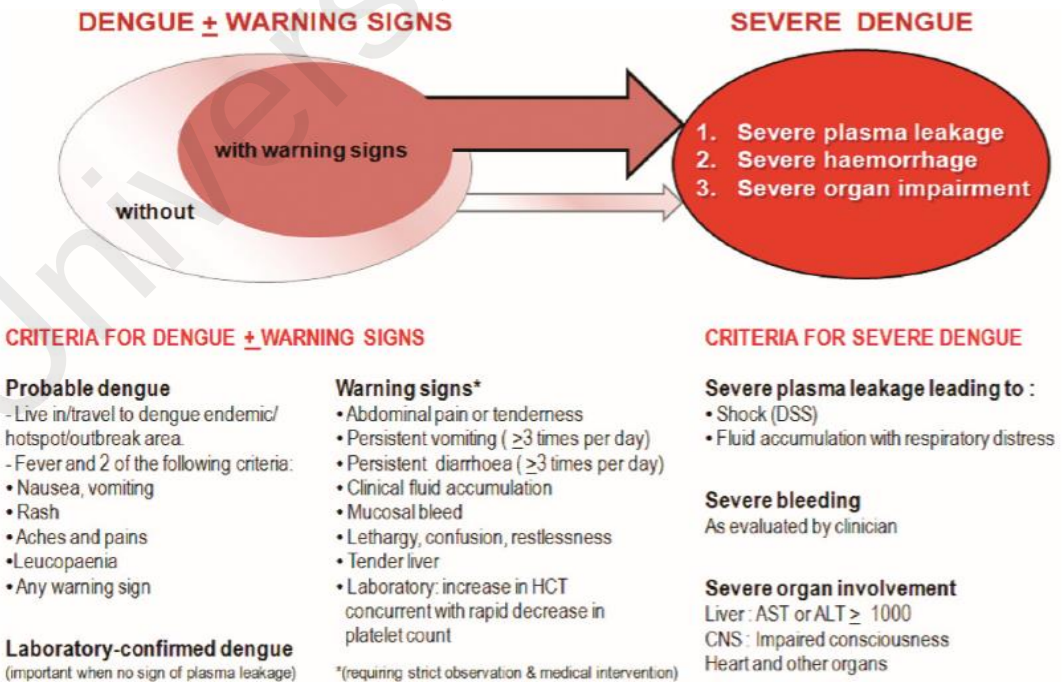


Figure 2.1 World Health Organization 2009 dengue case classification

(adapted from 2015 dengue clinical practice guidelines, MOH, Malaysia)

The variation in dengue clinical manifestation is due to the paradoxical role of the host immune response to consecutive infections by two different DENV serotypes. A primary dengue infection results in serotype-specific and serotype cross-reactive immunity. The latter provides protection against other serotypes but slowly wanes. A secondary heterotypic dengue infection leads to antibody-dependence enhancement (ADE) - activation of homotypic serotype-specific immunity that is directed against the primary infection and triggers an immune cascade that induces severe symptoms. Recovery from secondary infection is believed to confer serotype cross-reactive immunity against all four DENV serotypes. But asymptomatic or mild tertiary and quaternary infections may still occur in hyperendemic area (Endy et al., 2011; Guzman, Alvarez, & Halstead, 2013; Moi et al., 2016; Olkowski et al., 2013; Sabin, 1952).

A dengue infection usually starts with viraemia even before the onset of symptoms. It lasts from two to 12 days from the start of illness that usually coincides with the start of fever, followed by human host immunological response that is different depending on previous exposure to DENV. In primary dengue infection, immunoglobulin M (IgM) can be detected as early as fifth day of fever (or day 5) and remains for up to three months, while immunoglobulin G (IgG) is detectable as early as day 7 onwards for life (Figure 2.2). In secondary infection, IgM may appear earlier or at the same time as in the primary, but at lower titres; while IgG titre increases rapidly from its usual level that is present from previously (Figure 2.3). This dynamics of dengue infection has implications for laboratory diagnosis as discussed in the next section (Stuart D. Blacksell, 2012; World Health Organization, 2009a).

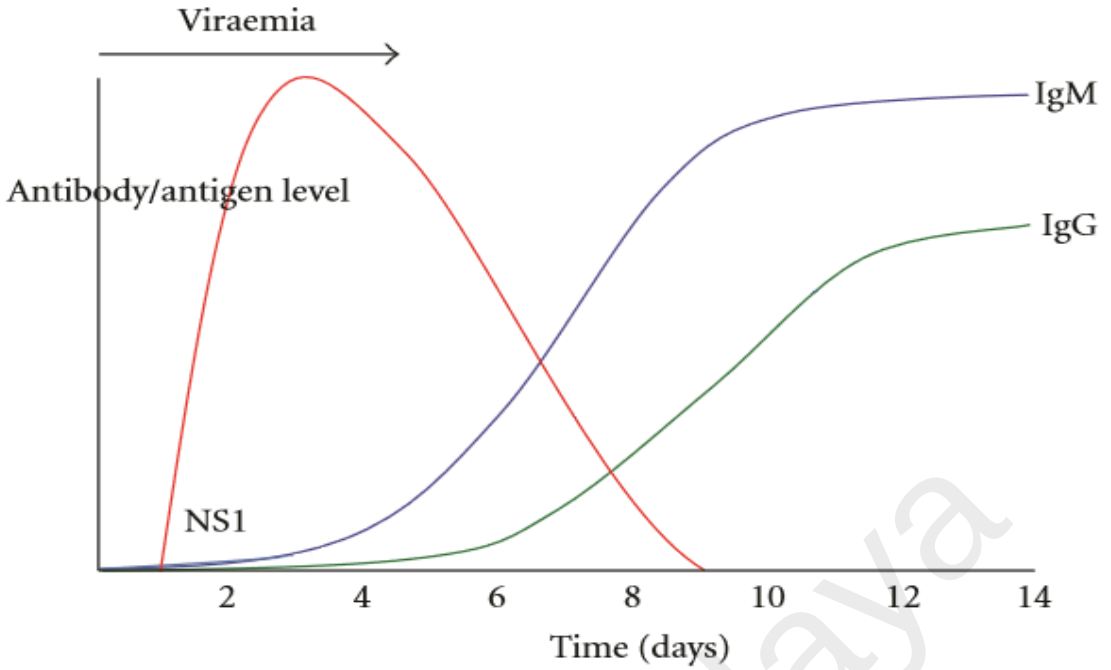


Figure 2.2 The kinetics of dengue NS1 antigen, IgM and IgG antibodies in a primary infection

(adapted with modification from Blacksell et al., 2012)

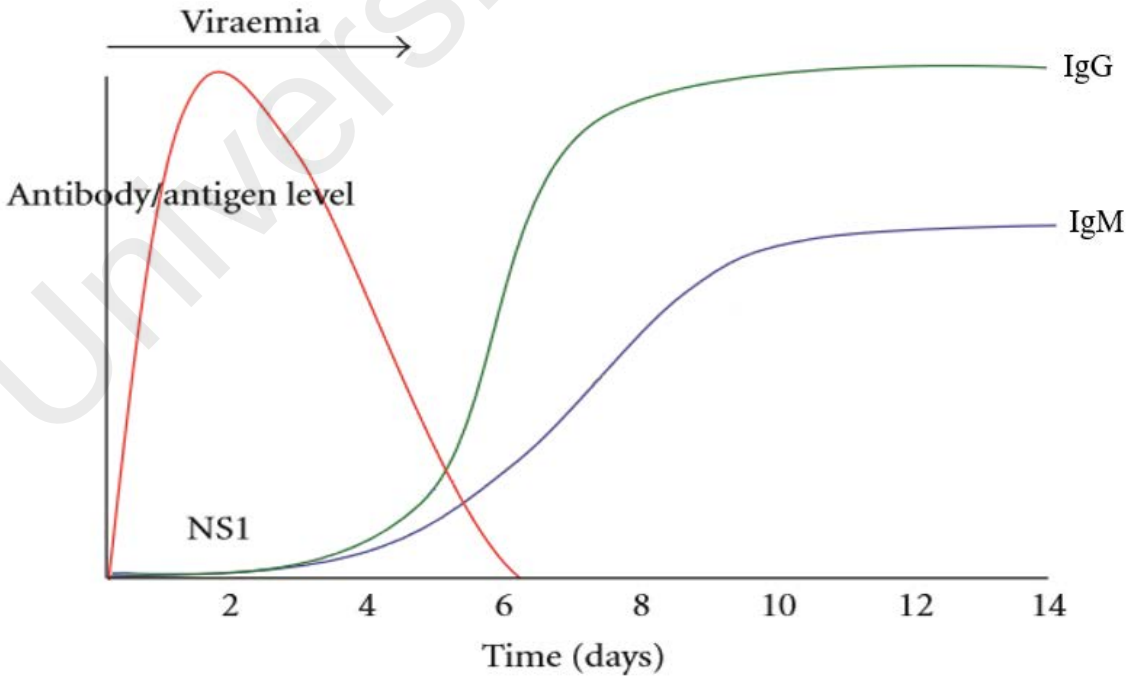


Figure 2.3 The kinetics of dengue NS1 antigen, IgM and IgG antibodies in a secondary infection

(adapted with modification from Blacksell et al., 2012)

2.4 Laboratory diagnosis of dengue infection and its challenges

Owing to the diversity and non-specificity in its clinical manifestation, making a dengue diagnosis based solely on clinical symptoms is difficult. Since DENV or its ribonucleic acid (RNA) and antigens can be detected in the blood of an infected patient within the first few days from the onset of symptoms or fever, methods for their detection are the direct confirmation of acute dengue infection. On the other hand, the detection of anti-dengue IgM and IgG is an indirect confirmation of dengue diagnosis (Peeling et al., 2010; Shamala, 2015).

The detection of disease-causing pathogen has always been the gold standard in the diagnosis of any infectious diseases (Table 2.1). Virus isolation (VI) and RNA detection are, therefore, very specific tests. However, VI requires mosquito or cell culture facilities, seven to 12 days to perform, and depends on virus survival in samples; making it less sensitive compared to nucleic acid amplification tests (NAAT), among which reverse-transcription polymerase chain reaction (RT-PCR or just PCR) is the most commonly performed. Nevertheless, they are still labour-intensive, time-consuming, and require well-resourced laboratory facilities, even though progress has been made in simplifying RT-PCR. As such, enzyme-linked immunosorbent assay (ELISA) capable to detect DENV particles such as non-structural protein 1 (NS1 protein or NS1) is preferred due to its simplicity, rapidity, and affordability. It is also quite specific although not serotype-specific. However, it still requires a few hours to obtain result and is not suitable for low-resource settings. Furthermore, its sensitivity may be compromised by pre-existing anti-dengue IgG in secondary infection (Peeling et al., 2010; A. Rathakrishnan & Sekaran, 2013; Sekaran & Soe, 2017).

Indirect serological tests available are haemagglutination inhibition (HI), plaque reduction neutralization test (PRNT), and ELISA. HI is one of the earliest developed and has been the gold standard among indirect methods (Table 2.1). Although sensitive, it cross-reacts with other flaviviruses, making it less specific. PRNT is capable of differentiating between DENV serotypes, although the neutralizing antibodies also tend to cross-react with other serotypes in hyperendemic areas where all serotypes are present. As both HI and PRNT are extremely tedious and require a few days to perform; ELISA has become the most preferred method as it takes only a few hours to obtain results, with acceptable level of test performance. However, as antibody response varies from person to person and depends on previous exposure to DENV, confirming dengue diagnosis using indirect methods requires paired samples collected a few days apart (Sekaran & Soe, 2017).

Table 2.1 Summary of laboratory dengue diagnostic methods

Diagnostic methods	Certainty of diagnosis	Time to results	Specimen types	Timing for specimen collection after onset of illness	Facilities requirement and costs
Virus Isolation	Confirmed	1-2 weeks	Whole blood, serum, tissues	Day 1-5	+++ / +++
RNA detection	Confirmed	1-2 days	Tissues, whole blood, serum, plasma	Day 1-5	++ / +++
Antigen detection	Not yet determined	1 day	Serum	Day 1-6	+ / ++
	Confirmed	>1 day	Tissue for immuno-chemistry	Not available	+ / +++
IgM ELISA	Probable	1-2 days	Serum, plasma, whole blood	After day 5	+ / +
IgG by ELISA, HI or PRNT	Confirmed	7 days or more	Serum, plasma, whole blood (paired samples)	Acute: day 1-5; Convalescent: after day 15	+ / +

(adapted with modification from WHO 2009 dengue guidelines)

In summary, laboratory-based dengue diagnostic methods generally require well-equipped facilities and trained personnel, are more tedious to perform, cost more, need more time to obtain results, and are not suitable for point-of-care diagnosis, especially in primary care, emergency department, and rural settings where resources are scarce (Peeling et al., 2010). Nevertheless, they are also more established and accurate for the diagnosis of acute dengue infection and have been widely accepted as diagnostic criteria in WHO 2009 dengue guidelines (Table 2.2) (Ministry of Health, 2017; World Health Organization, 2009a).

Table 2.2 Dengue diagnostic criteria

Confirmed/Laboratory-confirmed dengue	Highly suggestive/Presumptive
One of the following: 1) PCR + 2) Virus culture + 3) IgM seroconversion in paired samples 4) IgG seroconversion in paired samples or fourfold increase in HI IgG titre in paired samples	One of the following: 1) IgM + in a single sample 2) IgG + in a single sample or a HI titre of 1280 or greater

(adapted with modification from WHO 2009 dengue guidelines)

2.5 Issues with commercially available rapid dengue diagnostic tests

In response to the limitations of laboratory-based dengue diagnostics came the dengue RDT in the late 20th century, primarily in the form of RIT. RIT is designed usually in the shape of a lateral flow cassette hosting a nitrocellulose strip, where a defined volume of patient's specimen is applied (either whole blood, serum or plasma) to one end with or without a washing buffer. This sample bolus frees mobile antibodies coupled with reporter species, usually gold nanoparticles, which are placed at the same end. The mixture then migrates to the other end drawn by capillary action, crossing a test line and a control line in between. The test line and control line are impregnated with immobilised capture antibodies and control antibodies, respectively. Corresponding dengue target biomarkers or analytes, either NS1 antigen, IgM, or IgG, if present, bind with both the reporter antibodies as they travel together, and later also with the immobilised capture antibodies on the test line. Excess reporter antibodies continue along the strip and bind to the immobilised control antibodies. The concentration of captured reported antibodies on these lines turns them into maroon colour. Without the target biomarkers, reporter antibodies can only bind with immobilised control antibodies. As a maroon control line indicates that the reporter species has migrated and maintains binding ability, and that patient sample has also been drawn across the test line, a valid test must always be accompanied by a maroon control line. The whole process takes usually 15 to 30 minutes depending on manufacturer after the application of specimen (Stuart D. Blacksell, 2012; Miller & Sikes, 2015).

As such, RDT is deemed capable to provide healthcare practitioners and patients with rapid dengue diagnosis at the point-of-care that is timely and actionable. Their arrival can close the diagnostic gap for dengue in settings with low resources, as they appear to fulfil the characteristics of an ideal diagnostic test as defined by ASSURED criteria: i) Affordable (by those at risk of infection), ii) Sensitive (few false-negatives), iii) Specific

(few false-positives), iv) User-friendly (simple to perform and requiring minimal training), v) Rapid (to enable treatment at first visit) and Robust (not requiring refrigerated storage) vi) Equipment-free, and vii) Delivered to those who need it. With that dengue RDT, in particular RIT proliferated in the market and became almost the only type of RDT used for point-of-care dengue diagnosis (Stuart D. Blacksell, 2012; Miller & Sikes, 2015).

However, in reality, dengue RIT may not be as accurate, i.e. as sensitive and specific as the manufacturers claimed them to be. In fact, there is a substantial discrepancy between the close-to-perfect diagnostic performance reported in any dengue RIT product sheet and those evaluated by independent researchers. Take for example, an established dengue RIT that is widely available in the world including Malaysia, the SD Bioline Dengue Duo (Standard Diagnostics, Korea), a combo dengue RIT with all NS1, IgM, and IgG components. Its product insert stated NS1 assay sensitivity and specificity of 92.8% and 98.4%, respectively. The combined IgM/IgG assay performance was equally good at 99.4% for sensitivity and 93.0% for specificity (Standard Diagnostics Inc., 2008). However, these estimates as evaluated by independent researchers varied from study to study. Its NS1 assay was reported to have point estimates for sensitivity at 44.4-94.9% and specificity at 70.9-100% (Andries et al., 2012; Pal et al., 2014; Shih et al., 2016); while IgM had sensitivity and specificity at 10.0-98.0% and 66.0-100.0% (Hunsperger et al., 2014; Shih et al., 2016; Wang & Sekaran, 2010a), and IgG - 38.8-90.1% and 92.5-100.0%, respectively (Krishnananthasivam et al., 2015; Sanchez-Vargas, Sanchez-Marce, & Vivanco-Cid, 2014; Vickers et al., 2015).

Even when the between-study heterogeneity was removed by assessing different dengue RITs side-by-side, their diagnostic performance also varied widely. Using two different studies Blacksell et al conducted in year 2006 and 2007 as examples, where

eight different dengue RITs for the detection of IgM were evaluated. The point estimates of their sensitivity and specificity in the 2006 retrospective study were 6.4-65.3% and 69.1-100.0%, respectively (Stuart D. Blacksell et al., 2006). In the 2007 prospective study, these point estimates were 2.9-33.3% for sensitivity and 74.4-100.0% for specificity (Stuart D. Blacksell et al., 2007).

From the above estimates, it can be noticed that, although both diagnostic accuracy parameters varied for all dengue RITs, the sensitivity had wider range and was lower compared to specificity. This apparent underperformance in sensitivity may be mainly due to three interrelated intrinsic factors related to the built and usage of RIT, namely insensitivity to low diagnostic biomarker load, inability to quantify these diagnostic biomarkers, and subjective interpretation of test result. These shortcomings become problematic if a patient presents at a phase when the target biomarker is borderline positive, the resulting test line would be very faint. The interpretation of this qualitative result is subjective and may be read as negative by some people, making the test less sensitive. Another less probable reason contributing to RIT's failure to detect dengue cases may be its susceptibility to heat and humidity that can be easily overcome. On the contrary, threats to specificity are only test cross-reactivity with other antigens or antibodies that are closely-related to target biomarkers, observed mainly in IgM- and IgG-based assays; and time-dependence of signal development, when a negative test line becomes positive after the recommended interpretation time due to non-specific adsorption of other elements to the test line (Miller & Sikes, 2015).

As such, the between- and within-study differences in the diagnostic performance of dengue RIT demonstrated that they are far from being accurate, their performance varies between manufacturers or brands, and these performance parameters can be modified by the original purpose and context that influenced the study design, as well as other sources

of heterogeneity, including intrinsic factors related to their built and interpretation. Therefore, their diagnostic performance in a particular setting and population may not be generalizable. Whenever possible, their application to a new population and setting should be evaluated. In the same way, newly developed dengue RDT should also be evaluated together with existing RIT and even laboratory-based dengue diagnostics to benchmark their diagnostic performance for a proper comparison (Leeflang, 2014; Miller & Sikes, 2015; Peeling et al., 2010).

2.6 Summary

Making a dengue diagnosis solely from its undifferentiated non-specific clinical manifestation is difficult. However, the viraemia following the introduction of DENV into human body and subsequent immunological response provide a venue for its detection through direct and indirect methods. Direct detection of DENV and its genome remains the gold standard in the diagnosis of acute dengue infection. Confirmation of diagnosis using indirect methods requires paired samples to observe the changes in the level of anti-dengue antibodies.

Laboratory-based diagnostic tests are more accurate for the diagnosis of dengue fever. However, they are also more costly as they require well-equipped laboratories, skilled personnel, and longer time to perform. Therefore, they are not suitable for low-resource settings such as primary care, emergency department, and rural area. Dengue RDT is simple to perform and are able to provide rapid diagnosis at point-of-care, but may not be accurate. Their diagnostic performance varies between settings and populations, and can be influenced by other factors.

Therefore, generalising diagnostic accuracy of dengue RDT blindly is not scientifically correct. The application of any existing dengue RDT to another population or setting, and the introduction of any new dengue RDT to a population or setting, should be evaluated together with existing diagnostic methods to facilitate a fair comparison of their diagnostic performance.

CHAPTER 3: THE IMPORTANCE OF DENGUE DIAGNOSTICS IN MALAYSIA

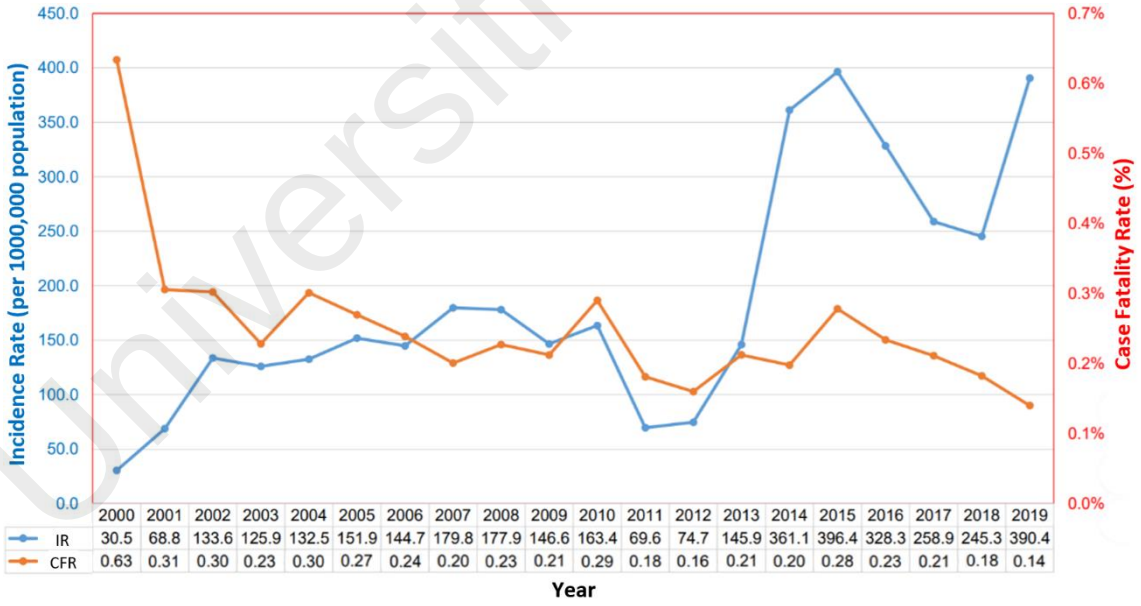
3.1 Introduction

In Malaysia, dengue fever was first reported prior to its independence from British colonial rule. It was in the last month of 1901, on the island of Penang, a dengue outbreak developed and persisted for four months, infecting mainly its local population (Skae, 1902). Sporadic cases were reported following that but it was only in 1962 that a major epidemic occurred again involving most of the states in the country (Ministry of Health, 2003). Since then, dengue fever has become a public health problem with gradual increase in incidence and mortality (Wallace et al., 1980). With the country developing to be more populous and urbanised, coupling with other interrelated factors such as population movement, host immunity, tropical climate, vector capacity, circulating DENV etc., this increasing trend in dengue has continued on. It has become endemic to Malaysia since early 1990s, when dengue outbreaks started to happen more frequently, with greater magnitude, and lasted longer (Hii, Zaki, Aghamohammadi, & Rocklov, 2016; Mohd-Zaki, Brett, Ismail, & L'Azou, 2014).

In this chapter, the current situation of dengue infection in Malaysia is explored, with special focus on the need for dengue diagnostics, in particular the dengue RDTs, to curb the spread of the disease and to assist clinical management of those infected. Section 3.2 describes the burden of dengue infection in Malaysia, while Section 3.3 reviews its national dengue surveillance system, in particular the dengue notification system. Section 3.4 describes the development of laboratory-based dengue diagnostics and their current usage in Malaysia. The selection and application of dengue RDT in Malaysia and related issues are discussed in Section 3.5 before Section 3.6 provides a summary of the chapter.

3.2 Burden of dengue infection in Malaysia

In Malaysia, the national dengue surveillance system recorded a total of 1428 cases or 8.4 cases per 100,000 population in 1988. Only 10 years later in 1998, this number grew to 27381 cases, or equivalent to 123.4 cases/100,000 population (Ministry of Health, 2003). Although the incidence rate (IR) has reduced to 30.5 cases/100,000 population in 2000, it has since risen more than ten-fold and reaching 396.4 cases/100,000 population in year 2015, before gradually reducing to 245.3 cases/100,000 population in 2018, and subsequently climbed back up to 2015-level in year 2019 (Figure 3.1). On the other hand, the case fatality rate (CFR) in 2000 was 0.62% but halved to 0.31% in the next year. Since then, it has been fluctuating within 0.2-0.3%, and registered its lowest point of 0.14% in 2019 (Figure 3.1) (idengue, 2020).



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
IR	30.5	68.8	133.6	125.9	132.5	151.9	144.7	179.8	177.9	146.6	163.4	69.6	74.7	145.9	361.1	396.4	328.3	258.9	245.3	390.4
CFR	0.63	0.31	0.30	0.23	0.30	0.27	0.24	0.20	0.23	0.21	0.29	0.18	0.16	0.21	0.20	0.28	0.23	0.21	0.18	0.14

Figure 3.1 Dengue incidence rate and case fatality rate in Malaysia from year 2000-2019

(adapted with modification from idengue, 2020)

From year 2000 to 2010, the economic burden dengue incurred on Malaysia was estimated to be at an average of US$ 128 million annually. At US$ 4.73 per capita, Malaysia ranked second among the countries in SEA in terms of dengue economic burden, only after neighbouring Singapore (D. S. Shepard, Undurraga, & Halasa, 2013). However, this cost estimation only included direct and indirect costs incurred as a result of dengue infection, and did not take into account the costs incurred in the prevention of dengue infection, primarily in the form of vector control. An estimation for the year 2010 revealed that Malaysia spent US$ 73.5 million on its National Dengue Vector Control Program. This was an additional US$ 2.68 per capita above the above-mentioned figure. As such, the estimated dengue economic burden in Malaysia in 2010 could have been US$ 201.5 million or 0.08% of its gross domestic product (Packierisamy et al., 2015).

However, due to the ability of the disease to cause asymptomatic and mild undifferentiated fever, many dengue patients may not even have accessed a healthcare facility for treatment. For the same reason, even if a dengue patient does get consultation from a physician, a dengue diagnosis might not be made and notified. The consequence is the underreporting of dengue infection in the national dengue surveillance system (Chew et al., 2016; Imai, Dorigatti, Cauchemez, & Ferguson, 2016; Mohd-Zaki et al., 2014). Although underreporting was taken into account when estimating dengue economic burden in Malaysia, the actual burden of dengue infection could still be higher (D. S. Shepard et al., 2013). In the next section, Malaysia national dengue surveillance system and its underreporting is explored.

3.3 National dengue surveillance system

Following the first dengue outbreak in modern Malaysia in 1962, dengue fever has officially become a notifiable disease in 1971. Subsequent enactment of the Prevention and Control of Infectious Disease Act 1988 made it a legal requirement for any dengue infection to be notified within 24 hours upon diagnosis to the nearest district health office by the attending physician (Ministry of Health, 2003). Initially, a notification could be made even if the diagnosis of dengue was only a provisional one based solely on clinical symptoms that fit the case definition. The conduct of dengue laboratory diagnostic test or its confirmation was not required at all. This early notification was encouraged to facilitate quick activation of dengue prevention and control activities after verification and registration of cases. Despite that, and the legal consequences in case of failure to notify, dengue infection remained underreported or was reported late in our passive dengue surveillance system (Ministry of Health, 2009; Mohd-Zaki et al., 2014). This underreporting became apparent in 2014, when the criteria for case registration was modified.

As mentioned above, prior to 2014, all suspected dengue that fit the case definition with or without a diagnostic test could be notified and registered as cases. Despite the presence of underreporting, the number of dengue cases was still high and its IR rose from 69.6 in 2011 to 145.9 cases/100,000 population in 2013 (Figure 3.1) (idengue, 2020). As only half of these cases had a dengue IgM performed, and around 45-99% of these tests were positive, two suspicions were raised. The first was that some of the dengue notifications might not be dengue cases, i.e. over-reporting. The second one was an inclination of notification only after the release of serology test result, which in turn delay dengue prevention activities. This was evident in the fact that, among the total dengue notification, only 4% were contributed by public primary care clinics, and only 33% were infection within three days from onset. On the other hand, only 40% of prevention

activities were carried out within five days from the onset of disease (Ministry of Health, 2009).

To overcome the presumed over-reporting, and also to facilitate early diagnosis and swift targeted prevention, starting from 2014, in order for a suspected dengue case to be notified, a dengue diagnostic test, whether laboratory-based or RDT, should also be performed. And only cases with a positive test result would be registered as dengue in the surveillance system (Ministry of Health, 2016, 2017). In conjunction with that, all public primary care clinics in Malaysia were supplied with dengue RDTs since 2014. The expected outcome was increased in number and timing of notification, but reduction in registration of actual dengue cases, which would facilitate earlier and more precise, and therefore, more effective vector control activities. However, despite having a stricter set of case registration criteria, the registered dengue cases in 2014 more than doubled itself from the year before, with an IR of 361.2 cases/100,000 population. The spike after the modification in the criteria was contrary to the expected result and hinted a substantial underreporting in the years before. Following that, private clinics were given an allocation of RM 30 million in the 2015 budget through the MOH for the distribution of 55,000 units of dengue RDTs (Bernama, 2014). In addition, they were also encouraged to source for additional dengue RDTs from two companies at extremely affordable price (Director General of Health, 2015). Subsequently, for the year 2015, the IR of dengue further increased to 396.4 cases/100,000 population (Figure 3.1).

The increase in registered dengue cases or incidence after the change in case registration criteria was not discussed in any publication. It may still be the result of the widely accepted shift in dominating dengue serotype, increasing urbanisation, and weather condition favourable to vector breeding. However, these factors would usually lead to a gradual change in the dengue burden as captured by the surveillance system

along the years. A sudden increase in burden of close to 250% within a year clearly pointed to vast underreporting prior to that. In the past, when dengue RDT was not available in public primary care clinics, they contributed to only 4% of total dengue notification made. Most notifications were made by hospitals for cases requiring hospital care. After dengue RDT became available, public clinics contribution to dengue notification and registration rose to 60%. And proportion of cases within three days of infection also rose from 33% to 49%. These changes clearly revealed the vast underreporting in the dengue surveillance system in the years prior to 2014, particularly from the primary care clinics. However, despite the increase in early diagnosis and notification, the percentage of prevention activities carried out within five days remained unchanged at 40% (Ministry of Health, 2009, 2016).

Although dengue underreporting in the previous years may have contributed to ineffective dengue vector control, increased in reporting witnessed above would only be useful if the underlying diagnostic tests used were accurate. Unfortunately, their actual diagnostic performance was unknown. Therefore, it is important for any dengue diagnostic test to be evaluated before its application in any setting, and its selection should be based on comparable results of different tests. When more accurate tests are used in the diagnosis of dengue, not only the overall accuracy of the passive surveillance is increased, the extent of underreporting may be estimated too. In the next two sections, laboratory-based dengue diagnostic tests and dengue RDT used in Malaysia are discussed.

3.4 Laboratory-based dengue diagnostic tests in Malaysia

In Malaysia, the IMR was the first to establish a unit for virus research including DENV in 1953, which later seconded in 1964 into an arbovirus research unit housed under University of Malaya (UM) for the development of a virus diagnostic and research laboratory that would focus on DENV, as well as the training of undergraduate students. However, it was only in 1969 that this laboratory was opened upon the completion of then University Hospital, which would later become University Malaya Medical Centre (UMMC). Subsequently, virology laboratories were opened in other universities in Malaysia. Together with IMR, they provided dengue diagnostic services for hospitals in many states. Only in 2003, the National Public Health Laboratories (NPHL) were expanded to relieve IMR of the burden of routine infectious diseases diagnostics including dengue. In the course of time, several major hospitals were also equipped with the capacity to provide dengue diagnostic services to their patients (Chua, 2009).

Initially, most of these laboratories except the hospitals were capable of performing direct methods such as VI, molecular diagnostics, antigen assays; and indirect serological methods such as HI. However, as technology advances, VI became less preferred compared to more sensitive and specific molecular diagnostics for the detection of DENV or simpler and more affordable dengue NS1 ELISA, and HI was replaced by in-house and subsequently commercial dengue IgM and IgG ELISA. As of year 2018, laboratory-based dengue diagnostic tests available for the diagnosis of dengue in the public healthcare system is found only in the major hospitals and are limited to commercial NS1, IgM, and IgG ELISA, the selection of which is dependent on each individual hospital following existing procedures. Only for severe patients, additional sample is required to be sent to NPHL for serotyping. In hospitals attached to some public universities, all these diagnostic services are provided under the same setting. Private hospitals utilise their own laboratories for dengue diagnosis and are likely subject to the same limitations

like public hospitals in terms of choices, but they are allowed to utilise NPHL or other private laboratories for more advanced dengue diagnostics at a fee. Apart from diagnostic services, NPHL also replaced IMR in providing the MOH diagnostic services for outbreak investigations and laboratory surveillance such as serotype monitoring and dengue seroprevalence study (Chew et al., 2016; Chua, 2009; Ministry of Health, 2010; National Public Health Laboratory, 2018; Thayan, 2016).

As such, for most of the time, dengue diagnostic tests were laboratory-based and were not available to many parts of Malaysia, especially in low-resource settings such as primary care, emergency department, and rural areas. This diagnostic gap was filled with the incorporation of dengue RDT into the diagnostic algorithm from 2014, which is discussed in the next section.

3.5 Dengue rapid diagnostic tests in Malaysia

In the previous sections, the burden of dengue as captured by the passive national dengue surveillance system was described and its underreporting and late reporting highlighted. This underreporting and late reporting happened despite the fact that notification based solely on clinical symptoms that fit the case definition was encouraged. It is suspected that these cases of underreporting and late reporting had been primarily contributed by primary care doctors. In order to overcome the above issues and also the possibility of over-reporting in the case of notification without a test result, dengue RDT was introduced to mainly the primary care setting in the form of combo tests that included NS1, IgM, and IgG assays (Ministry of Health, 2016).

Prior to that, some hospitals have already started utilising dengue IgM RDTs in their emergency departments, but only in a very small scale, and limited to patients with suspected dengue after day 5 of fever. The discrepancy in adoption timing was due to the public hospitals having relatively higher autonomy in selecting and providing diagnostic services needed, including RDT for dengue, provided that the quality is ensured (Ministry of Health, 2010). For primary care, the selection and application of dengue RDT come under the purview of the MOH public health division. And the rule is that companies selling dengue RDT that are interested to receive government tenders must submit their products to IMR for laboratory-based evaluation on archived clinical samples. However, this rule as well as the criteria for selection are not published and not freely available to the public. Nevertheless, as disclosed recently, in order for a dengue RDT to be selected and bulk-procured by the MOH, it has to have both sensitivity and specificity of more than 90% as evaluated by IMR, and preferably produced by a local company (Thayan, 2018).

As such, there was a lack of transparency in terms of the selection of a particular dengue RDT for the usage in the public primary clinics in Malaysia. It was also unsure whether the decision making took into account the result of an earlier health technology assessment on point-of-care tests that was unfavourable to dengue RDT (Ministry of Health, 2007). Based on the two known criteria, it was unlikely that the dengue RDT selected would be accurate. First and foremost, the diagnostic landscape of a primary care setting may not be replicable in laboratory-based evaluation studies. Field evaluation of any diagnostic test often produces lower diagnostic performance than in laboratory (Kohn, Carpenter, & Newman, 2013). This became evident after the first dengue RDT, a RIT called Acco Rapid Test Dengue IgG/IgM-NS1 Combo (Accobiotech, Malaysia), was purchased and used in Malaysia public clinics. It was reported to produce a lot of false negative tests over the years, i.e. not sensitive. However, it was unclear how this issue was discovered and whether proper evaluation was carried out or not (Kan, 2018). As a result, in 2016, the third year after this RDT was implemented, it was replaced with RVR Dengue Combo NS1-IgG/IgM produced by another company called RVR diagnostics, which was acquired by its American partner, Chembio Diagnostics, at the end of the same year (Chembio Diagnostics, 2016).

As a result, in the end, the other known criterion in the selection of dengue RDT, the protectionism shown towards a local company also failed to work out. But the main issues in the whole process of selection and application of dengue RDT that transpired in Malaysia was that no practice-based evaluation study was conducted to inform the decision making, and that laboratory-based evaluation studies performed lacked transparency. It appeared that priority was given more to cost of a dengue RDT over its actual diagnostic accuracy on the field. This was evident in MOH recommendation of another two dengue RDTs with untested accuracy to private clinics earlier in 2015 (Director General of Health, 2015). Although affordability should be an important

consideration in the selection of a dengue RDT, but in the absence of effectiveness, even low cost is not justifiable.

In summary, the most widely used dengue RDT in Malaysia was not evaluated in the place it is currently used. The only published data was performed in laboratory on archived samples, two years after its adoption across Malaysia (Ainulkhir et al., 2018). Therefore, its actual diagnostic performance in clinical setting is not known. The implications are misdiagnosis or late diagnosis of dengue patients that may lead to unfavourable outcomes; misreporting to the local health authority leading to inefficient deployment of prevention and control activities due to wastage as a result of false positive and missed opportunity as a result of false negative; and misreporting to the national dengue surveillance system causing more difficulties in the estimation of the actual burden of dengue from the incidence data.

3.6 Summary

Dengue has been reported in Malaysia even before its independence. It has currently reached an epidemic level in this country. However, the actual burden of dengue infection in Malaysia may be more severe. This is due to the underreporting in its national dengue surveillance system, in particular the passive dengue notification system. Simple notification requirement based solely on clinical symptoms in keeping with the case definition without the need of a diagnostic test, and even legal consequences in the event of neglect, failed to address this underreporting. Together with late reporting and potential over-reporting, dengue prevention activities were misinformed and became less effective, which might have contributed to the spread of dengue in Malaysia.

Among the efforts to address the late-reporting and potential over-reporting was the modification in the notification process that required a diagnostic test to be performed for every notified case, and only case tested positive on any assay would be registered, started in 2014. However, laboratory-based dengue diagnostic tests are available only in major hospitals and some specialised institutions in Malaysia. Resource-scarce settings such as public primary care clinics that cater for majority of Malaysian population received only dengue RDT. This new stricter criteria for case notification and registration unexpectedly and substantially increased not only early reporting, but also the overall number of cases, which indicated vast underreporting in the previous years.

However, due to the lack of transparency in the selection process, the absence of field evaluation prior to adoption, and the prioritisation of other factors over the more important diagnostic performance, the dengue RDT incorporated into the diagnostic algorithm of public clinics in Malaysia may not be accurate. As such, dengue diagnosis in this setting, dengue prevention and control activities, and passive dengue surveillance system and its underreporting, remained misinformed.

CHAPTER 4: SYSTEMATIC REVIEW AND META-ANALYSIS OF THE DIAGNOSTIC ACCURACY OF COMMERCIALLY AVAILABLE DENGUE RAPID IMMUNOCHROMATOGRAPHIC TESTS

4.1 Introduction

Early dengue diagnosis requires diagnostic test that is simple to perform, rapid, yet accurate. Many dengue RITs have been developed and marketed to be that. Their manufacturers often claim good diagnostic accuracy. Unfortunately, much discrepancies exist between the official figures and those evaluated by independent researchers. Moreover, the performance of any individual test varied widely between different studies (Hunsperger et al., 2009; Hunsperger et al., 2014; Peeling et al., 2010; Shamala, 2015).

Several prior systematic reviews were conducted to determine the diagnostic accuracy of various dengue laboratory tests, but none catered specifically for RIT and few meta-analysed them (Alagarasu, Walimbe, Jadhav, & Deoshatwar, 2016; Costa, Marques-Silva, & Moreli, 2014; Shamala, 2015; Shan et al., 2015; H. Zhang et al., 2014). In view of that, a systematic review and meta-analysis was conducted with the objectives: a) to determine the actual diagnostic accuracy of commercially available RITs for the diagnosis of acute dengue infection in exposed population; b) to identify the source of heterogeneity that modify the diagnostic accuracy of commercially available dengue rapid immunochromatographic tests for the diagnosis of acute dengue infection; and c) to identify the commercially available dengue rapid immunochromatographic test with the best diagnostic accuracy for the diagnosis of acute dengue infection.

This chapter details out the methodology of this systematic review and meta-analysis in Section 4.2, presents the results in Section 4.3. Section 4.4 is the discussion and Section 4.5 - the conclusion.

4.2 Methodology

The protocol of this meta-analysis was registered with PROSPERO (Protocol No: CRD42017071252) (Appendix A). The study selection, data extraction, and quality assessment were performed by two independent reviewers. A third reviewer was consulted when disagreement occurred.

4.2.1 Search strategy

There were three stages to the search strategy. Initial searches in MEDLINE were conducted using the index terms dengue and diagnostics, followed by exploration of the titles and abstracts to identify more key words. Additional key words were also identified from MeSH term library.

In the second stage, MEDLINE Complete, Scopus, CINAHL Complete, Science Direct, and Web of Science were searched for English articles that evaluated the accuracy of dengue RIT using the identified keywords: (dengue OR dengue virus OR dengue fever OR dengue infection OR dengue disease) AND (diagnostics OR diagnosis OR detection OR assay OR rapid diagnostic test OR rapid diagnosis OR point-of-care OR commercially available OR commercial). The search was last updated on 31st July 2018 (Table 4.1).

In the third stage, the reference lists of identified articles were searched for additional studies. Authors were contacted for full text if not available. Grey literature was not searched as this systematic review and meta-analysis intended to include only published studies that were peer-reviewed.

Table 4.1 Systematic review and meta-analysis search strategy

Search Settings	Search Terms
MEDLINE Complete (via EBSCOhost)	
Search Field: AB abstract Search Modes: Boolean/Phrase Expanders (default): apply related words, also search within the full text of the articles Limiters: From 1950, English, Human	"dengue *" AND "diagnos*" OR "detection" OR "assay" OR "rapid diagnos*" OR "point-of-care" OR "commercial*"
Scopus	
Search Field: Article titles, abstract, keywords Limiters: All years, English	"dengue *" AND "diagnos*" OR "detection" OR "assay" OR "rapid diagnos*" OR "point-of-care" OR "commercial*"
CINAHL Complete (via EBSCOhost)	
Search Field: AB abstract Search Modes: Boolean/Phrase Expanders (default): apply related words, also search within the full text of the articles Limiters: From 1950, English, Human	"dengue *" AND "diagnos*" OR "detection" OR "assay" OR "rapid diagnos*" OR "point-of-care" OR "commercial*"
Science Direct (up to 27/2/2018)	
Search Field: Title, Abstract, Keywords Filters: From 1950	(dengue*) AND (diagnos* OR detection OR assay OR {rapid diagnos*} OR {point-of-care} OR commercial*)
Science Direct (27/2/2018 onwards) +	
Search Field: Title, Abstract, Keywords Filters: From 1950	(dengue) AND (diagnosis OR diagnostic OR diagnostics OR detection OR assay); (dengue) AND ({rapid diagnosis} OR {rapid diagnostic} OR {point-of-care} OR commercial OR commercially)
Web of Science	
Field tags: TS Topic Limiters: All years, Science Citation Index Expanded, Emerging Sources Citation Index	dengue* AND diagnos* OR detection OR assay OR “rapid diagnos*” OR “point-of-care” OR commercial*

4.2.2 Study selection

All identified articles were exported to EndNote X7 citation management software, where duplicates were identified and removed. Subsequently, two independent reviewers screened through the title and abstract of these articles. Included articles were those that: 1) evaluated the sensitivity and specificity of commercially available dengue RIT for the diagnosis of dengue infection; and 2) used archived human samples or samples collected from patients with suspected dengue with or without control; and 3) used any established diagnostic method as reference; and 4) were published in English in peer-review journals. Articles with insufficient information to compute 2X2 table, duplicate data, as well as review, animal studies, meeting abstract, and comment, were excluded. Any disagreement between the first two reviewers was resolved with the help of a third reviewer. The whole process was summarised into a flow diagram (Figure 4.3).

4.2.3 Data extraction

The Standards for the Reporting of Diagnostic Accuracy Studies (STARD) (P. M. Bossuyt et al., 2015) was referred to ascertain data for extraction. They included main author, publication year; methodology such as study design, direction, setting, location, duration, inclusion and exclusion criteria; participant demographics including sample type, sampling frequency and timing, disease severity and characteristics; description of index and reference tests including their performers and interpreters, criteria for positive test; 2x2 tables with true positive (TP), false negative (FN), true negative (TN), and false positive (FP) values (Table 4.2); as well as presence of study flow chart and source of funding. The above information was adapted into the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) checklist (Whiting, Rutjes, Westwood, & et al., 2011) and used for data collection and quality assessment by two independent reviewers (Appendix B). Any disagreement between the reviewers was resolved with the help of a third reviewer.

Table 4.2 Components of a 2x2 table

	Disease Status (based on reference standard)		
Index Test Results	Diseased (D+)	Non-diseased (D-)	Total
Index test positive (T+)	***True positives (TP)***	***False positives (FP)***	Test positives (=TP+FP)
Index test negative (T-)	***False negatives (FN)***	***True negatives (TN)***	Test negatives (=TN+FN)
Total	Disease positives (=TP+FN)	Disease negatives (=TN+FP)	N (=TP+FN+TN+FP)

(adapted with modification from Macaskill et al., 2010)

4.2.4 Quality assessment

The quality of the included articles was assessed using QUADAS-2 checklist. All four domains of this tool, namely patient selection process, index test conduct and interpretation, reference standard conduct and interpretation, and flow and timing of the conduct of the index and reference tests, were assessed by 11 signalling questions, to which the responses can be "yes" (low risk of bias), "no" (high risk of bias), or "unclear" (unclear risk of bias) (Table 4.3). These signalling questions were rather straightforward. For the third and last questions on inclusion and exclusion of study participants or samples, loss of less than 10% was deemed acceptable and taken as fulfilment of criteria, i.e. answered yes. Appropriate interval in eighth question was defined as up to six months with storage in freezer. The only question that was not applicable to this review was the fifth question that asked whether index test threshold was pre-specified or not, as all dengue RITs were constructed with pre-specified threshold. Instead, it was used to assess whether the interpretation of index test was subject to potentially biased interpretation by only one interpreter. Only interpretation by two independent researchers, and a third that would decide in the event of disagreement between the first two, was taken as fulfilment of this criterion.

Each domain was then assigned a certain level of risk of bias and concerns regarding applicability to the review question, whether "low", "high" or "unclear". The risk of bias would be taken as low only if all corresponding signalling questions were given a "yes" answer; and "unclear" if all answers were "unclear". If any question was given a "no" answer, the risk level was taken as 'high' (Table 4.3). These responses were added into the dataset.

All studies were included into the review regardless of the quality to allow for subgroup or sensitivity analyses based on quality categories. However, studies with high

or unclear concerns regarding applicability to the review question were excluded. The outcome of the quality assessment for each signalling question and domain for all included articles were summarised in tabular form and also presented in proportional stack bar charts (Appendix E, Figure 4.4 & Figure 4.5).

Table 4.3 Components of QUADAS-2 checklist

Domain	Patient Selection	Index Test	Reference Standard	Flow and Timing
Description	Describe methods of patient selection Describe included patients (previous testing, presentation, intended use of index test, and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index tests or reference standard or who were excluded from the 2 X 2 table (refer to flow diagram) Describe the interval and any interventions between index tests and the reference standard
Signalling Questions (yes, no, or unclear)	Q1. Was a consecutive or random sample of patients enrolled? Q2. Was a case–control design avoided? Q3. Did the study avoid inappropriate exclusions?	Q4. Were the index test results interpreted without knowledge of the results of the reference standard? Q5. If a threshold was used, was it pre-specified?	Q6. Is the reference standard likely to correctly classify the target condition? Q7. Were the reference standard results interpreted without knowledge of the results of the index test?	Q8. Was there an appropriate interval between index tests and reference standard? Q9. Did all patients receive a reference standard? Q10. Did all patients receive the same reference standard? Q11. Were all patients included in the analysis?

(Continued in the following page)

Table 4.3, continued

Domain	Patient Selection	Index Test	Reference Standard	Flow and Timing
Risk of bias (high, low, or unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns about applicability (high, low, or unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or its interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

(adapted with modification from Whiting et al., 2011)

4.2.5 Data analysis

4.2.5.1 Selection of statistical model and software

The meta-analysis of diagnostic accuracy evaluation studies differs from that of the traditional therapeutic or interventional studies (Table 4.4). Firstly, diagnostic accuracy of any test is usually measured by a pair of summary parameters, i.e. sensitivity and specificity. Secondly, most tests are qualitative that provide results in binary form, such as positive or negative to the target disease, which is based on a certain test positivity threshold or diagnostic threshold that may be different between tests and studies. Its variation results in threshold effect that affects the sensitivity and specificity in the opposite ways, i.e. lower value increases sensitivity but decreases specificity, and vice versa. In other word, the pair of summary parameters is inversely correlated and cannot be pooled individually. Thirdly, due to this interrelation of the sensitivity and specificity, in addition to the ever-present differences in other study characteristics, the between-study heterogeneity of diagnostic accuracy studies is inevitably larger than in therapeutic or interventional studies (Juneyoung Lee, Kim, Choi, Huh, & Park, 2015; Macaskill, Gatsonis, Deeks, Harbord, & Takwoingi, 2010).

Therefore, for the meta-analysis of this systematic review, heterogeneity was presumed to be present and random effects model was used by default. Only two hierarchical methods can concurrently handle heterogeneity, correlation between sensitivity and specificity, and threshold effect, namely bivariate and hierarchical summary receiver operating characteristics (HSROC) models. The difference between them is that bivariate model caters for studies with common diagnostic threshold, while HSROC model caters for studies with different thresholds (Table 4.5). For this study, the HSROC model was selected as dengue RITs from multiple manufacturers were included. In addition, the qualitative and subjective nature of dengue RIT interpretation in different

studies, even for the same brand, would inevitably subject to threshold-like effect (Juneyoung Lee et al., 2015; Macaskill et al., 2010).

In view of the above, STATA version 12 (StataCorp, TX, US) with MIDAS and METANDI commands capable of hierarchical modelling was selected for the meta-analysis in this study.

Table 4.4 Comparison of the meta-analysis between therapeutic or interventional studies and diagnostic accuracy evaluation studies

	Therapeutic or Interventional Study	Diagnostic Accuracy Evaluation Study
Number of outcome variables	Single outcome	Pair of outcomes, sensitivity and specificity, which generally inversely correlated
Analysis of heterogeneity between studies	Chi-square test (Cochrane Q statistic): $p < 0.1$ generally indicates significant heterogeneity Higgins' I^2 statistic: 0% to 25% - might not be important; 25% to 50% - may represent low heterogeneity; 50% to 75% - may represent moderate heterogeneity; 75% to 100% - high heterogeneity	Cochrane Q or Higgins' I^2 statistics alone may not be informative as they do not consider threshold effect Visual evaluation of coupled forest plot or SROC plot to find threshold effect Spearman correlation analysis between sensitivity and false positive rate: $r \geq 0.6$ generally indicates considerable threshold effect
Meta-analytic summary	Summary point and its 95% CI obtained with: Fixed-effects model: when study heterogeneity does not exist Random-effects model: when existence of study heterogeneity is suspected	Summary point Summary sensitivity and specificity and their 95% CI is obtained with bivariate model: recommended Summary plot (SROC curve) Moses-Littenberg model: not recommended HSROC curve: recommended

CI = confidence interval, HSROC = hierarchical summary receiver operating characteristic, SROC = summary receiver operating characteristic (adapted with modification from Lee et al., 2015)

Table 4.5 Comparison of statistical methods for the meta-analysis of diagnostic accuracy evaluation studies

Method	Summary Measures	Weighting	Recommendations
Summary point			
Separate pooling	Summary sensitivity, specificity, LR+, LR-, and DOR	Fixed effects or random effects	Not recommended: Conducts separate meta-analyses for each summary point Ignores threshold effect as well as correlation between sensitivity and specificity
Hierarchical methods (bivariate/HSROC model)	Summary sensitivity, specificity, LR+, LR-, and DOR	Random effects	Recommended: Accounts for correlation between sensitivity and specificity For practical reasons, bivariate model is preferred for computing summary points, while HSROC model is preferred for constructing HSROC curve
Summary line (SROC analysis)			
Moses-Littenberg model	SROC curve, AUC, and Q*	Similar to fixed effects	Not recommended: Does not account for variability between studies Does not weight studies optimally Ignores correlation between sensitivity and specificity
Hierarchical model	HSROC curve, AUC, confidence region, and prediction region	Random effects	Recommended: Accounts for within- and between-study heterogeneity Accounts for correlation between sensitivity and specificity

AUC = area under the ROC curve, DOR = diagnostic odds ratio, HSROC = hierarchical summary receiver operating characteristic, LR+/LR- = positive/negative likelihood ratio, SROC = summary receiver operating characteristic, Q* = the point where SROC intersects the downward diagonal

(adapted with modification from Lee et al., 2015)

4.2.5.2 Tabular and graphical presentation of summary diagnostic accuracy parameters

The results of included primary evaluation studies were summarised and presented in graphical ways in conjunction with different summary parameters such as sensitivity (SN), specificity (SP), positive and negative likelihood ratios (LR+, LR-), diagnostic odd ratio (DOR), and area under curve (AUC) with their 95% confidence interval (CI). These summary parameters were pooled from their corresponding estimates calculated from the TP, FN, TN, and FP for each individual RIT in the primary studies using commonly used formula. They were selected to be presented as they are not influenced by the prevalence of the disease, and may perhaps be applicable to other similar settings (Table 4.2 & Table 4.6) (Harbord & Whiting, 2009; Macaskill et al., 2010; Šimundić, 2009):

Table 4.6 Formulas for the measures of diagnostic accuracy

Measure	Definition	Formula	Remarks
Sensitivity (SN)	the probability of index test positive in a diseased	TP/TP+FN	Negative rules out
Specificity (SP)	the probability of index test negative in a non-diseased	TN/TN+FP	Positive rules in
Positive Likelihood Ratio (LR+)	the ratio of the probability of index test positive in a diseased to the probability of the same in a non-diseased	SN/(1-SP)	Rule-in measure, 1 means test is indiscriminative, higher is better
Negative Likelihood Ratio (LR-)	the ratio of the probability of index test negative in a diseased to the probability of the same in a non-diseased	(1-SN)/SP	Rule-out measure, 1 means test is indiscriminative, lower is better
Diagnostic Odd Ratio (DOR)	the ratio of the odds of index test positive in diseased to the odds of the same in non-diseased	LR+/LR- or $\frac{(TP/FN)}{(FP/TN)}$	Global measure, 1=indiscriminate, higher is better
Area Under Curve (AUC)	the probability that the index test will correctly classify a randomly chosen diseased above a randomly chosen non-diseased	(SN+SP)/2	Global measure, 0.5=indiscriminate, 1=perfect test

For the graphical presentation, the HSROC graph was used instead of coupled forest plots as the former summarises test accuracy and allows for visual measurement of both heterogeneity and threshold effects, while only heterogeneity is observable in the latter (Macaskill et al., 2010). A HSROC graph plots sensitivity on the y-axis and 1-specificity (or specificity in reverse) on the x-axis, where each primary study is represented by a circle, with the size of the circle indicating the sample size (Figure 4.1). The curve is the best fitted line and is the expected summary or average ROC curve across studies with different positivity threshold levels, with the assumption that the shape of the true ROC curve underlying each individual study is the same. In other word, it shows how test SN and SP change as threshold changes across different studies (JBI, 2015; Leeflang, 2014).

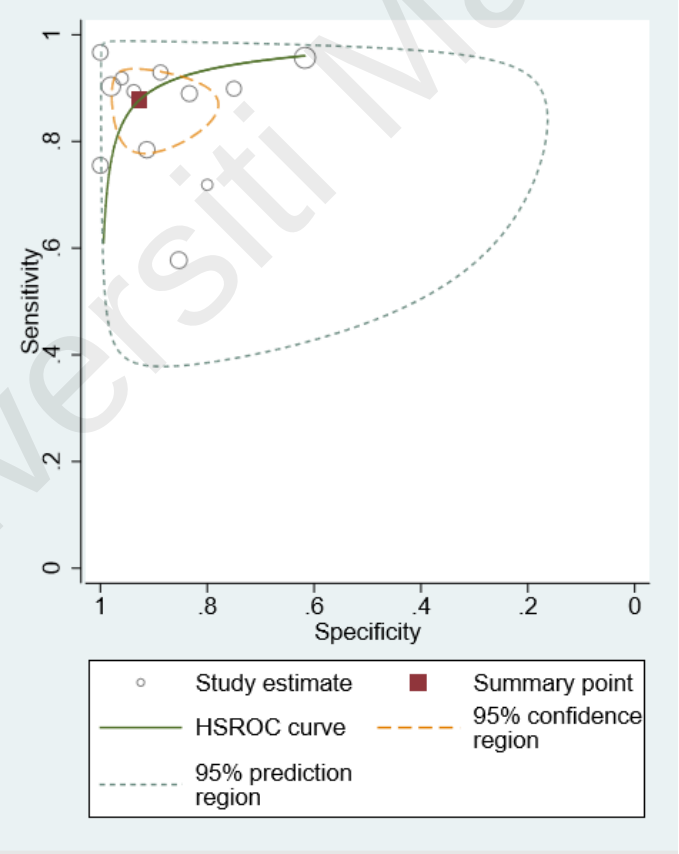


Figure 4.1 A HSROC graph

The summary point is the average sensitivity and specificity of all included tests, while the 95% confidence region reflects the precision of this average accuracy estimate and also measures within-study variability. It means that there is a 95% chance that the summary point will fall within the confidence region should the included individual studies be repeated again. On the other hand, between-study heterogeneity including that introduced by threshold effect can be gauged from the scatter of the circles representing primary studies and the 95% prediction region. The larger the heterogeneity, the more scattered are the circles and the further the prediction ellipse is away from the confidence region. Otherwise, 95% prediction region also means that there is a 95% chance that a future study will yield a summary point within its boundary. Finally, the variation in threshold across studies is depicted by the position of the circles relative to the HSROC curve, the closer they are to the curve, the lesser the variation in threshold, and vice versa (Figure 4.1) (P. Bossuyt et al., 2013; Macaskill et al., 2010).

Apart from the above, HSROC curve is able to depict test accuracy by its shape and position, and is related to DOR and AUC (Figure 4.2). If the test is uninformative or indiscriminative, it is a straight line that lies exactly on the upward diagonal of the plot and divides it into equal left upper and right lower halves, and corresponds to DOR of 1 and AUC of 0.5. An informative test always have a SROC that lies above the upward diagonal, with more discriminative test possessing higher DOR and AUC appearing more curvy and locating further away from it (P. Bossuyt et al., 2013). On a symmetric curve, DOR is constant along its whole length. However, this is not the case if the curve is asymmetric, which occurs when the variability of the test measurement distribution is unequal between diseased and non-diseased. Increased variability happens if a disease cause a biomarker to rise erratically, and vice versa (Macaskill et al., 2010).

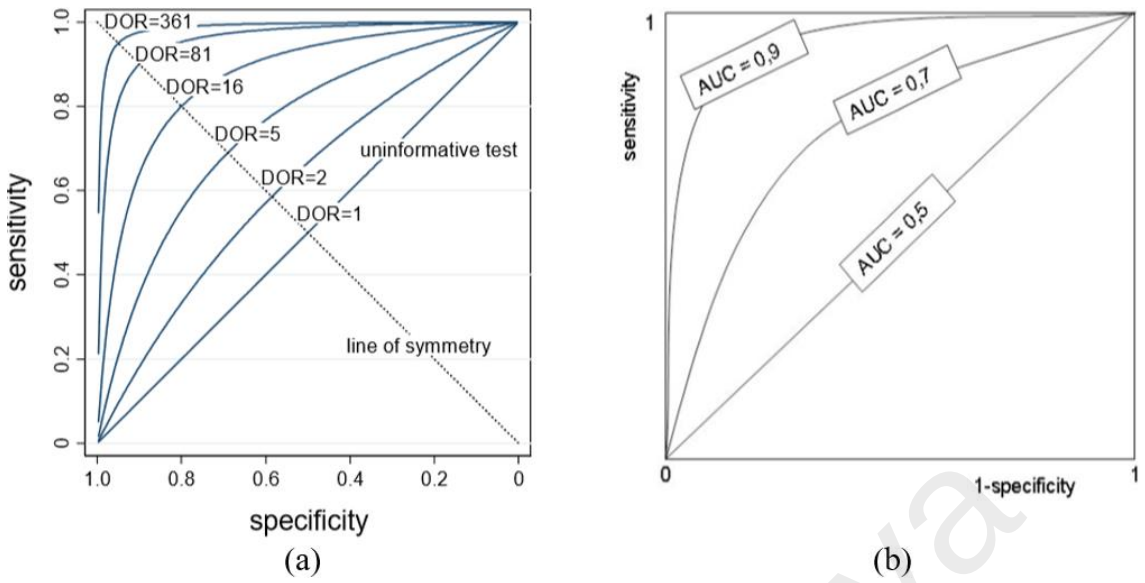


Figure 4.2 SROC curve and its relationship with DOR (a) and AUC (b)

(adapted from Macaskill et al., 2010 (a), and Šimundić, 2009 (b))

As evident from the description above, HSROC graph is superior to coupled forest plot and is preferred for graphical presentation of the individual and summary results. However, it is not useful in itself standing alone. Two or more curves need to be put overlapping each other on the same graph for meaningful comparison. However, doing that clutters the graph and is not feasible technically. Therefore, in this review, an individual HSROC curve was generated for each analysis, and comparable graphs were placed side by side for comparison. They can be interpreted with the aid of summary parameters in numbers presented in respective sections of the tables. The row for a particular analysis contained the total number of dengue RIT included, the total number of diseased and non-diseased participants or test samples as defined by reference standard, followed by the AUC, pooled SN, SP, DOR, LR+, LR-, I^2 statistic and their 95% CI, and p-value. In addition, the range of SN and SP of the underlying individual tests would be presented alongside in tables, if corresponding HSROC graphs were placed apart.

All the above parameters can be produced by MIDAS command in STATA, but only up to 2 decimals. Since all outputs come in the form of fraction, SN and SP values lost

its precision in MIDAS after conversion to percentage. Its HSROC curve also lacks information on study size and looks more cluttered. Therefore, whenever possible, all outputs that can be generated by METANDI, namely SN, SP, DOR, LR+ and LR-, were used in the tables in the result section, as they were more precise. This applied to HSROC curve too. All the other parameters were from MIDAS. Only in the situation when METANDI failed to pool that MIDAS outputs were used exclusively. If both failed to pool, outliers were identified and removed. Understandably, in the case when only MIDAS failed to pool, only five parameters and HSROC curves were presented, together with numbers of RITs and manually calculated numbers of diseased and non-diseased.

4.2.5.3 Interpretation of summary diagnostic accuracy parameters

Having the above in mind, in this review, the summary measures and HSROC were interpreted in the following manner. First of all, comparable HSROC curves were evaluated visually for heterogeneity (scatter of circles and prediction regions), contribution of variation in threshold to heterogeneity (circles position relative to curve), study number and size (circles quantity and size), and test accuracy (curve position, shape, summary point and its confidence region). Secondly, the corresponding tables were referred to for the comparison of summary measures and their 95% CI in light of the graphs. In this review, DOR was taken as the most important parameter to look for in order to decide if one test was better than the other. However, it should be noted that this approach has its own limitation, if one of the curves were not symmetrical. In this case, the DOR would not be constant along it, and should be interpreted with caution. Although AUC is also a global measure of accuracy, having generated from MIDAS, its lack of precision allowed it to be taken only as an indicator of test usefulness in this review. Nevertheless, it was definitely better in this compared to DOR as it is a more meaningful

indicator with a fixed range, and can be used alone. DOR can only be used to compare between two tests and is not meaningful alone (Macaskill et al., 2010). In this review, test with AUC of <0.5 was taken as not useful, 0.5-0.6 – poor, 0.6-0.7 – moderate, 0.7-0.8 good, 0.8-0.9 – very good, 0.9-1.0 – excellent (Šimundić, 2009).

If two or more tests were significantly different from each other, the SN and SP would be referred to estimate their portion of contribution towards the difference. If necessary, LR+ and LR- would also be referred to for clinical relevance of a test, i.e. its ability to increase or decrease the pre-test probability of disease with its result (Table 4.7). When it came the quantification of heterogeneity, Higgins I^2 statistic of more than 50% and Cochran-Q chi-square p-value less than 0.05 that indicate significant difference between the included studies can be helpful (Dwamena, 2014). Again, these measures have to be interpreted in light of the HSROC curve and should not be used alone, as they are not able to isolate the threshold effect (Macaskill et al., 2010). On the other hand, if no significant difference was found between tests based on HSROC curves and DORs, the reference pool size could help to deduce if it was due to the actual absence of effect size or the inadequacy of sample size.

Table 4.7 Likelihood ratios and approximate changes in probability of disease

Likelihood Ratio	Approximate Change in Probability (in absolute %)	Likelihood Ratio	Approximate Change in Probability (in absolute %)
Values below 1 decrease the probability of disease		Values above 1 increase the probability of disease	
0.1	- 45	2	+ 15
0.2	- 30	3	+ 20
0.3	- 25	4	+ 25
0.4	- 20	5	+ 30
0.5	- 15	6	+ 35
Value of 1 is uninformative		8	+ 40
1	0	10	+ 45

(adapted from McGee, 2002)

4.2.5.4 Pooling of dengue RITs diagnostic accuracy

In this review, the assessment of publication bias and meta-analysis were first performed according to dengue RIT individual and combined assays as presented in the included articles, namely IgM, IgG, IgM/IgG, NS1, NS1/IgM, and NS1/IgM/IgG. For combined assays like IgM/IgG, NS1/IgM, and NS1/IgM/IgG, testing positive on any one component in the respective combination was ascertained as diseased.

The assessment of publication bias was done using Deek's funnel plot asymmetry test for each of the individual and combined assays stated above. However, instead of the standard test that plots standard error of odd ratio against its log scale, the test was performed in MIDAS for association between the effective sample size (ESS) and diagnostic log odd ratio (lnDOR), with ESS as a function of the number of diseased and non-diseased samples. The reason was because the traditional method would produce incorrect result due to low power, as diagnostic studies tended to have large odd ratios or DOR. Nevertheless, the Cochrane collaboration advised caution in the interpretation of the results as this method would still suffer from low power if there was heterogeneity in the DOR (Deeks, Macaskill, & Irwig, 2005; Macaskill et al., 2010).

As heterogeneity would be inevitable, within each of the individual and combined assays, univariable meta-regression and subgroup analyses were performed to identify sources of heterogeneity. However, only results of important subgroups were presented, namely study design, serotype, previous exposure to dengue, disease phase, commercial brand, and reference test. Other subgroups would be mentioned if found significant. As MIDAS and METANDI require at least four studies or tests to pool, in the absence of enough samples for a subgroup, sensitivity analyses were used to assess the effect of test exclusion on the outcome. In this situation, the SN, SP, and AUC of the unpooled test or tests in the opposite subgroup were calculated using STATA and presented for

comparison. However, their comparison with the overall result of the corresponding assay should be made with caution, in light of the unpooled test or tests, as the distribution of the said subgroup characteristic may not be known in all the studies contributing towards the overall pool. Only the distribution of study design and commercial brand was known in all the studies. Information on serotype, previous exposure to dengue infection, and disease phase was not available in most studies. As for reference test, most studies used a combination of a few tests and only some also presented stratified results by different reference tests. Therefore, the actual distribution each individual reference test used in all the underlying was also unknown. Since sensitivity analysis may not be reliable if the distribution of characteristic of interest was unknown in all the studies, the direction and magnitude of bias were unpredictable, and can only be deduced with caution as mentioned above, in combination with scientific reasoning.

Multivariable analysis was not possible due to the absence of command that can take the correlation of SN and SP into consideration. MIDAS and METANDI are only capable of univariable analysis. Although multivariable analysis using METAREG based on AUC was attempted, it failed to generate meaningful results due to loss of power from small cell size, in addition to the global nature of AUC. Therefore, no result from multivariable analysis was presented.

Last but not least, the meta-analysis was performed on all dengue RIT as presented in the included articles. However, RIT with zero or missing value for SN or SP was excluded from the meta-analysis to avoid error in analysis. Commonly used Haldane correction, where zero or missing values are replaced by a small number, was not adopted to avoid bias. As some articles contained more than one study, and all studies presented results of more than one RIT with variable subgroups, the number of RIT results was more than the number of included articles.

4.3 Results

The results section is presented in the following subsections: search and selection, characteristics of included studies, quality assessment, publication bias, and pooling of dengue RITs diagnostic accuracy. The findings from the pooling were summarised and presented last to facilitate the discussion section that follows.

4.3.1 Search and selection

The study flow was as summarised in Figure 4.3. The search strategy as described in Table 4.1 yielded 18519 citations throughout the five databases. Out of all, Scopus provided the most at 8188 citations, followed by Web of Science, MEDLINE, and Science Direct. CINAHL only contributed 182 citations. After the removal of 5750 duplicates, titles and abstracts of 12769 citations were screened, of which 12654 obviously irrelevant ones were removed. Full text of the remaining 115 citations were retrieved for eligibility assessment. Out of that, 46 citations were further excluded with insufficient information to compute 2x2 table as the most common reason. Eventually 69 articles were included in the systematic review. 3 articles contained 2 studies each with different characteristics, making the total studies 72. In total, they contributed to 179 dengue RIT results or data points, each with its own 2x2 tables. However, 10 results were excluded due to zero cell or missing value for specificity. Out of the remaining 169 test results included for meta-analysis, 58 were IgM, followed by 56 from NS1, while IgG assay had only 13. Combination of assays each had only below 20 data points.

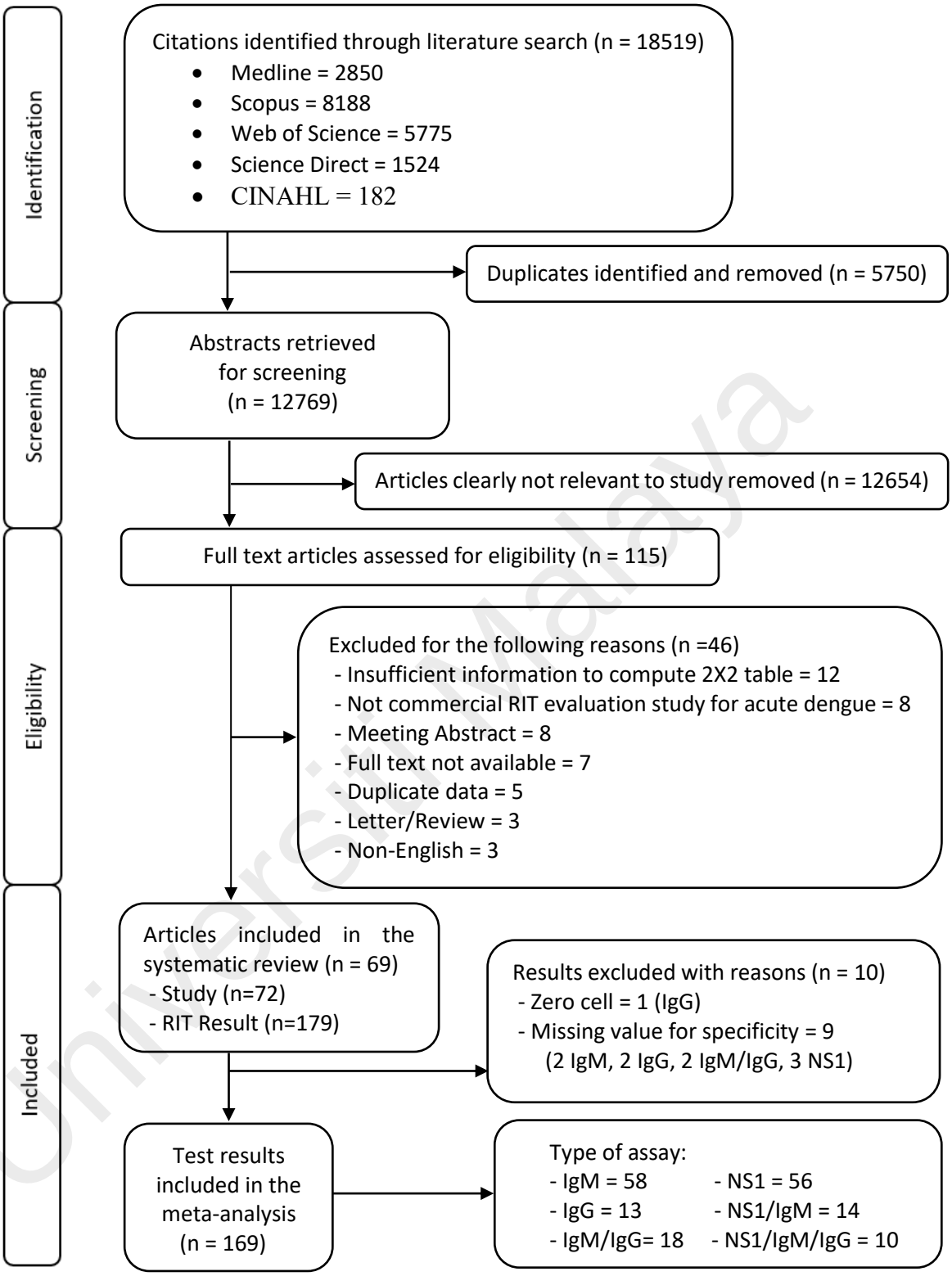


Figure 4.3 Systematic review and meta-analysis study flow chart

4.3.2 Characteristics of included studies

The characteristics of the 69 articles included in this review were summarised in Table 4.8 and detailed out in Table 4.9, which included author and year, study design, direction, setting and location, total sample size, dengue RIT brand and assays evaluated, and reference tests used. Another list with full title of the included articles together with their respective main author name and year of publication was attached in Appendix C. The 46 excluded ones followed in Appendix D together with the reason for exclusion. The earliest Dengue RIT evaluation study was published in the year of 1998. Among the articles included in this review, most years contributed two to three articles, five to six articles came from 2009-2010 and 2014-2017, while nine were published in 2011. On the other hand, only one paper each was included from 2005 and 2008, and no paper was identified or included in 2003-2004.

The first articles published in 1998 evaluated IgM and IgG dengue RITs. NS1 assay was only started to be reported since 2008. Up until then, majority of the articles reported results on IgM, IgG, and IgM/IgG, and only a few reported on IgM alone. After the appearance of NS1 assay that often come in a standalone kit, results of individual assays especially NS1 were reported more frequently compared to that of combo results, i.e. NS1/IgM and NS1/IgM/IgG, apart from the previously mentioned combination.

A total of 23 brands of dengue RIT were identified. As many as 59 studies evaluated only one brand name, while the rest had more than one. The most evaluated brand was Panbio dengue RIT. However, out of the 30 times it were evaluated, 11 of it was on its out-of-use first generation Panbio Immunochromatographic Card Test (Panbio IC) (Panbio, Brisbane), the remaining ones were divided between its standalone NS1 assay – Panbio Dengue Early, and IgM/IgG combo kit – Panbio Dengue Duo, that were sold separately. SD Bioline Dengue Duo (Standard Diagnostics, Korea), in its original name

or that for the Latin America – the SD Bioeasy, was evaluated for 28 times, mostly in its combo kit that includes all three assays and occasionally separately. Following that was Biorad NS1 STRIP (Biorad, France), a standalone NS1 assay that was evaluated 18 times. Most of the 20 other identified brands were only evaluated once, with some evaluated twice, and only two – thrice.

In total, 72 studies were presented in these 69 included articles, as there were three articles that reported two individual studies each (Andries et al., 2012; Fry et al., 2011; Pal et al., 2015). When it came to study design, most studies employed case-control design followed by cross-sectional, at 33 versus 31, respectively. However, case-control studies were observed more frequent initially, and cross-sectional ones were seen more frequently in recent decade. One study used a mix of case-control and cross-sectional, two were case only, and five did not provide adequate information on study design. As all case-control studies and some cross-sectional were also retrospective, in terms of study direction, retrospective studies were majority at 47. Only 20 cross-sectional studies were conducted prospectively. Four were unclear on the direction, and again, the only one mix design mentioned above also had a mix of study direction.

Up to 40 out of the 72 studies were conducted in laboratory setting, most of which were case-control studies. In fact, case-control studies were exclusively laboratory-based. 20 studies were conducted in hospital setting, while three were carried out in some unspecified medical care facilities. One study was conducted in both laboratory and a unspecified medical care (Pal et al., 2015). Only one study was performed in clinic setting. One study in Japan was conducted in a community setting. It used clinical data of health screening on inbound travelers with suspected dengue in an airport (Sugimoto, Haseyama, Ishida, Yoshida, & Kamiya, 2011). There were four studies that used information from

patient database for evaluation of dengue RITs, and two studies were unclear on their study setting.

When it came to study location, Asia was the largest contributor with 42 studies. Among them, 24 were from SEA countries, namely Thailand (n=6), Malaysia (n=5), Singapore (n=4), Vietnam (n=4), Cambodia (n=3), Laos (n=1), and a mix of several countries (n=1). 14 studies came from South Asia, i.e. India (n=8), Sri Lanka (n=5), and Pakistan (n=1). Only four studies came from East Asia, two from Taiwan, one each from China and Japan. The American continent came in second by contributing 17 studies, mostly from South America, such as Brazil (n=6), French Guiana (n=2), and one each from Columbia, Venezuela, and Peru in combination with Honduras. Five studies came from the Caribbean islands, with Jamaica giving three and one each from French Caribbean and Barbados. One study was conducted in Mexico. Finally, there were six studies that were conducted in multiple countries worldwide. Three European countries, namely Belgium, France, and Netherlands contributed one study each. Two more studies were conducted on Oceania or Western Pacific Islands. And another two studies were unclear on their study location.

Most studies used a combination of different reference tests to ascertain dengue infection. Only 21 studies used single test as reference. Commonly used direct methods were VI, PCR, and NS1 ELISA. Most used serological methods were HI and ELISA. The latter included those developed by Thailand Armed Forces Research Institute of Medical Science (AF), United States Centers for Disease Control and Prevention (CDC), other in-house IgM antibody capture ELISAs (MAC-ELISA), and commercial IgM and IgG ELISA. One study also used immunofluorescence assay (IFA) and blot. Four studies used clinical diagnosis as standard, with three using it exclusively. Three studies even included dengue RIT into their reference standard definition.

An evolution of preferred reference tests was observed over the years. Indirect methods such HI and various in-house ELISAs were more frequently used initially, and subsequently replaced gradually by commercial IgM and IgG ELISAs. Prior to 2008 before the introduction of NS1-based dengue RIT, either HI or in-house ELISA was used in at least a third of the studies. The former disappeared after year 2012, while in-house ELISA was no longer seen after 2015. Direct methods were more frequently used since 2008, with PCR usage increased more substantially from presence in only a quarter of studies to at least two-third, and less so for VI that was used in about a quarter before and a third of studies after NS1 dengue RIT appearance. Commercial NS1 ELISA was first used in 2014. As such, studies from the recent years mainly used PCR and NS1, IgM and IgG ELISA as their reference standard.

Finally, in total, the 72 studies contributed to 169 tests results that were included in the meta-analysis (Figure 4.3), with as many as 38800 of dengue RIT tests performed on collected patient samples. Generally, most of the early studies had smaller sample size, below 100 samples or participants. The study size gradually grew bigger since 2006 and were mostly above 200 counts, although smaller studies can still be found occasionally (Table 4.9).

Table 4.8 Summarised characteristics of included studies

Characteristics	N (%) *
Publication Year*	
1998-2004	13 (18.9) *
2005-2011	27 (39.1) *
2012-2018	29 (42.0) *
Study design	
Case-control	33 (45.8)
Cross-sectional	31 (43.1)
Case-only	2 (2.8)
Mix	1 (1.4)
Unclear	5 (6.9)
Study direction	
Retrospective	47 (65.3)
Prospective	20 (27.8)
Mix	1 (1.4)
Unclear	4 (5.5)
Study setting	
Laboratory	40 (55.5)
Hospital	20 (27.8)
Unspecified Medical Care	3 (4.2)
Database	4 (5.5)
Clinic	1 (1.4)
Community	1 (1.4)
Mix	1 (1.4)
Unclear	2 (2.8)
Study location	
Asia	42 (58.3)
America	17 (23.6)
Worldwide	6 (8.3)
Europe	3 (4.2)
Oceania	2 (2.8)
Unclear	2 (2.8)
Sample size	
Below 100	13 (18.1)
100-199	16 (22.2)
200-499	34 (47.2)
500 and above	9 (12.5)
RIT brand evaluated	
Single	59 (81.9)
Multiple	13 (18.1)
Reference test used	
Single	21 (29.2)
Multiple	51 (70.8)

*****Denominator for all % was 72 studies, except for publication year – 69 articles. Three articles presented results from two different studies each.

Table 4.9 Characteristics of included studies in the systematic review and meta-analysis

No.	Author	Year	Study Design	Study Direction	Study Setting	Study Location	Sample Size	Brand	Assay	Reference
1	Lam	1998	Case-control	Retrospective	Laboratory	Unclear	130	Panbio IC	IgM, IgG	MAC-ELISA+HI
2	Sang	1998	Case-control	Retrospective	Laboratory	Singapore	92	Panbio IC	IgM, IgG	HI
3	Vaughn	1998	Case-control	Retrospective	Laboratory	SEA	202	Panbio IC	IgM, IgG	AF & CDC ELISA
4	Branch	1999	Case-control	Retrospective	Laboratory	Barbados	92	Panbio IC	IgM, IgG	VI/Clinical/Serology
5	Palmer	1999	Case-control	Retrospective	Laboratory	Jamaica	100	Panbio IC	IgM, IgG	VI/ELISA/HI
6	Porter	1999	Case-control	Retrospective	Laboratory	Unclear	97	Panbio IC	IgM, IgG	VI/PCR/HI
7*	Chakravarti	2000	Cross-section	Prospective	Hospital	India	71	Panbio IC	IgM, IgG	Clinical
8	Groen	2000	Case-control	Retrospective	Laboratory	Netherlands	128	Panbio IC	IgM, IgG	ELISA/IFA/RIT/Blot
9	Wu	2000	Case-control	Retrospective	Laboratory	Worldwide	164	Panbio IC	IgM	VI/ELISA
10	Cuzzubo	2001	Case-control	Retrospective	Laboratory	Thailand	179	Panbio Dengue Duo	IgM, IgG	VI/PCR/MAC-ELISA/HI*
11	Vajpayee	2001	Case-control	Retrospective	Laboratory	India	58	Panbio IC	IgM, IgG	AF & CDC ELISA
12	Charrel	2002	Cross-section	Unclear	Hospital	France	37	Panbio Dengue Duo	IgM	ELISA
13*	Kittigul	2002	Cross-section	Prospective	Hospital	Thailand	92	Panbio IC	IgM, IgG	HI
14	Berlioz-Arthaud	2005	Case-control	Retrospective	Laboratory	Oceania	85	Panbio Dengue Duo	IgM, IgG	PCR/ELISA
15	Kumarasamy	2006	Cross-section	Prospective	Laboratory	Malaysia	239	Acon Dengue Rapid Test Device	IgM	ELISA

*Included in systematic review but excluded from meta-analysis due to missing values

(Continued in the following page)

Table 4.9, continued

No.	Author	Year	Study Design	Study Direction	Study Setting	Study Location	Sample Size	Brand	Assay	Reference
16	Blacksell	2006	Case-control	Retrospective	Laboratory	Thailand	491	Core Dengue	IgM	AF ELISA+PCR
								Diazyme Combo Rapid Test		
								Globalmed Smartcheck		
								Minerva Vscan		
								Panbio Dengue Duo		
								SD Bioline		
								Teco DF Combo Test		
								Tulip Dengucheck-WB		
17	Blacksell	2007	Cross-section	Prospective	Hospital	Laos	151	Core Dengue	IgM	AF ELISA+PCR
								Diazyme Combo Rapid Test		
								Globalmed Smartcheck		
								Minerva Vscan		
								Panbio Dengue Duo		
								SD Bioline		
								Teco DF Combo Test		
								Tulip Dengucheck-WB		

(Continued in the following page)

Table 4.9, continued

No.	Author	Year	Study Design	Study Direction	Study Setting	Study Location	Sample Size	Brand	Assay	Reference
18	Cohen	2007	Cross-section	Prospective	Hospital	Thailand	723	Panbio Dengue Duo	IgM, IgG	AF ELISA+HI
19	Nga	2007	Case-control	Retrospective	Laboratory	Vietnam	200	Panbio Dengue Duo	IgM, IgG	ELISA
20	Dussart	2008	Case-control	Retrospective	Laboratory	French Guiana	320	Biorad NS1 STRIP	NS1	VI/PCR/MAC-ELISA
21	Chaiyaratana	2009	Case-control	Retrospective	Laboratory	Thailand	220	Biorad NS1 STRIP	NS1	ELISA
22	Hang	2009	Cross-section	Prospective	Hospital	Vietnam	263	Biorad NS1 STRIP	NS1	PCR/ELISA
23	Hunsperger	2009	Case-control	Retrospective	Laboratory	Worldwide	350	Panbio Dengue Duo	IgM	AF & CDC ELISA
								SD Bioline		
								Tulip Dengucheck-WB		
24	Moorthy	2009	Cross-section	Retrospective	Laboratory	India	136	Panbio Dengue Duo	IgM, IgG	ELISA
25	Ramirez	2009	Case-control	Retrospective	Laboratory	Venezuela	147	Biorad NS1 STRIP	NS1	VI/PCR/ELISA
26	Zainah	2009	Case-control	Retrospective	Laboratory	Malaysia	533	Biorad NS1 STRIP	NS1	VI/PCR/ELISA
27	Lima	2010	Case-control	Retrospective	Laboratory	Brazil	450	Biorad NS1 STRIP	NS1	VI/PCR/ELISA
28	Osorio	2010	Cross-section	Retrospective	Laboratory	Colombia	147	Biorad NS1 STRIP	NS1	VI/PCR/MAC-ELISA/HI
							310	SD Bioline	NS1, IgM, IgG	
29	Pok	2010	Case-control	Retrospective	Laboratory	Singapore	433	Biorad NS1 STRIP	NS1	PCR/ELISA
								Panbio Dengue Duo	IgM, IgG	

(Continued in the following page)

Table 4.9, continued

No.	Author	Year	Study Design	Study Direction	Study Setting	Study Location	Sample Size	Brand	Assay	Reference
30	Tricou	2010	Cross-section	Retrospective	Hospital	Vietnam	292	SD Bioline	NS1, IgM, IgG	PCR/ELISA
								Biorad NS1 STRIP	NS1	
31	Wang	2010	Case-control	Retrospective	Laboratory	Malaysia	420	SD Bioline	NS1, IgM, IgG	VI/PCR/HI/ NS1 ELISA/ELISA
32	Bisordi	2011	Cross-section	Retrospective	Database	Brazil	266	Biorad NS1 STRIP	NS1	VI
33	Blacksell	2011	Cross-section	Retrospective	Hospital	Sri Lanka	259	Biorad NS1 STRIP	NS1	AF ELISA+PCR
								Biosynex Immunoquick DF	IgM	
								Merlin DF Combo Device		
								Panbio Dengue Early/Duo	NS1, IgM	
								SD Bioline		
34	Chaterji	2011	Case-control	Retrospective	Laboratory	Singapore	354	Biorad NS1 STRIP	NS1	VI/PCR/ELISA
35	Fry	2011	Case-control	Retrospective	Laboratory	Vietnam	298	Panbio Dengue Early	NS1	PCR/ELISA
						Malaysia	293	Panbio Dengue Early/Duo	NS1, IgM, IgG	VI/PCR/MAC-ELISA/HI
36	Gunasekera	2011	Case-control	Retrospective	Laboratory	Sri Lanka	143	Panbio Dengue Duo	IgM, IgG	AF & CDC ELISA
37	Najioullah	2011	Cross-section	Prospective	Hospital	French Carribean	537	Biorad NS1 STRIP	NS1	PCR/NS1 ELISA/ AF+CDC ELISA
38	Shrivastava	2011	Unclear	Retrospective	Laboratory	India	91	SD Bioline	NS1	PCR/NS1 ELISA

(Continued in the following page)

Table 4.9, continued

No.	Author	Year	Study Design	Study Direction	Study Setting	Study Location	Sample Size	Brand	Assay	Reference
39	Sugimoto	2011	Cross-section	Retrospective	Community	Japan	23	Biorad NS1 STRIP	NS1	PCR/NS1 ELISA/ELISA
								SD Bioline	IgM	
40	Tontulawat	2011	Unclear	Retrospective	Laboratory	Thailand	237	SD Bioline	NS1, IgM	PCR/ELISA
41	Andries	2012	Case-control	Retrospective	Laboratory	Cambodia	286	SD Bioline	NS1, IgM, IgG	VI/PCR/MAC-ELISA/HI
			Cross-section	Prospective	Hospital		157			
42	Pun	2012	Cross-section	Prospective	Hospital	Sri Lanka	131	SD Bioline	IgM	ELISA
43	Ferraz	2013	Case-control	Retrospective	Laboratory	Brazil	77	SD Bioeasy	NS1	PCR/NS1 ELISA/ELISA
								Biorad NS1 STRIP		
								Panbio Dengue Early		
44*	Huang	2013	Case only	Retrospective	Database	Taiwan	375	Biorad NS1 STRIP	NS1	PCR/ELISA
45	Gan	2014	Cross-section	Prospective	Medical Care	Singapore	197	SD Bioline	NS1, IgM, IgG	VI/PCR/ELISA/ NS1 ELISA
46	Hunsperger	2014	Case-control	Retrospective	Laboratory	Worldwide	571	Biorad NS1 STRIP	NS1	AF/CDC ELISA+PCR/VI
							496	CTK Onsite Dengue		
							830	Abon Biopharma	IgM	
							835	CTK Onsite Dengue		
							755	Orgenics/Iverness ImmunoComb II		
47	Naz	2014	Unclear	Unclear	Unclear	Pakistan	184	Panbio Dengue Early/Duo	NS1, IgM, IgG	ELISA

*Included in systematic review but excluded from meta-analysis due to missing values

(Continued in the following page)

Table 4.9, continued

No.	Author	Year	Study Design	Study Direction	Study Setting	Study Location	Sample Size	Brand	Assay	Reference
48	Pal	2014	Case-control	Retrospective	Laboratory	Peru/ Honduras	193	Biorad NS1 STRIP	NS1	VI/ELISA
							106	Inbios Dengue NS1 Detect		
							193	Panbio Dengue Early		
							192	SD Bioline		
49	Sanchez-Vargas	2014	Case-control	Retrospective	Laboratory	Mexico	397	SD Bioline	NS1, IgM, IgG	NS1 ELISA/ELISA
50	Bibhas	2015	Unclear	Unclear	Hospital	India	940	Unclear	NS1, IgM, IgG	Clinical
51	Carter	2015	Cross-section	Retrospective	Clinic	Cambodia	328	SD Bioline	NS1, IgM	NS1 ELISA/ELISA
52	Krishna-nanthasivam	2015	Cross-section	Prospective	Hospital	Sri Lanka	143	SD Bioline	NS1, IgM, IgG	PCR/ELISA
53*	Lee	2015	Case only	Retrospective	Laboratory	Worldwide	42	SD Bioline	NS1	PCR
54	Pal	2015	Cross-section	Prospective	Hospital	Worldwide	213	Panbio Dengue Duo	IgM, IgG	AF/CDC ELISA+PCR/VI
			Mix	Mix	Laboratory & Medical Care		2067			
							1940	SD Bioline	NS1, IgM, IgG	
55	Vickers	2015	Cross-section	Retrospective	Laboratory	Jamaica	339	SD Bioline	NS1, IgM, IgG	NS1 ELISA/ELISA
56	Buonora	2016	Cross-section	Prospective	Hospital	Brazil	325	SD Bioeasy	NS1	PCR/ELISA
57	Chen	2016	Cross-section	Prospective	Hospital	China	294	Wondfo Biotech Diagnostic Kit	IgM	PCR/NS1 ELISA/RIT

*Included in systematic review but excluded from meta-analysis due to missing values

(Continued in the following page)

Table 4.9, continued

No.	Author	Year	Study Design	Study Direction	Study Setting	Study Location	Sample Size	Brand	Assay	Reference
58	Hunsperger	2016	Cross-section	Prospective	Medical Care	Western Pacific Islands	1678	SD Bioline	NS1, IgM	PCR/ELISA
59	Mitra	2016	Cross-section	Prospective	Hospital	India	276	J.Mitra Dengue Day 1 Test	IgM	Clinical
							280	Panbio Dengue Duo		
							276	Reckon Diagnostics		
							276	SD Bioline		
60	Shih	2016	Cross-section	Retrospective	Database	Taiwan	511	SD Bioline	NS1, IgM, IgG	PCR
61	Huits	2017	Cross-section	Retrospective	Database	Belgium	308	SD Bioline	NS1	PCR/ELISA
62	Mata	2017	Cross-section	Prospective	Hospital	Brazil	143	Bioeasy Dengue Eden NS1	NS1	PCR/NS1 ELISA
63	Shukla	2017	Cross-section	Prospective	Medical Care	India	249	J.Mitra Dengue Day 1 Test	NS1	PCR/NS1 ELISA
64	Simonnet	2017	Cross-section	Prospective	Hospital	French Guiana	3347	SD Bioline	NS1	PCR/NS1 ELISA
65	Vickers	2017	Case-control	Retrospective	Laboratory	Jamaica	339	OneStep, Diag. Automation	NS1, IgM, IgG	NS1 ELISA
66	Vivek	2017	Case-control	Retrospective	Laboratory	India	211	J.Mitra Dengue Day 1 Test	NS1, IgM	PCR/ELISA
67	Ainulkhir	2018	Case-control	Retrospective	Laboratory	Malaysia	98	RVR Dengue Combo	NS1, IgM	RIT/ELISA
68	Murugananthan	2018	Unclear	Unclear	Unclear	Sri Lanka	765	Cortez	IgM, IgG	ELISA
69	Prado	2018	Cross-section	Prospective	Hospital	Brazil	452	SD Bioeasy	NS1, IgM	VI/PCR/ELISA

4.3.3 Quality assessment

The quality assessment of the included 72 studies were summarised according to QUADAS-2 11 signalling questions (Figure 4.4) and four domains (Figure 4.5). The detailed scoring of this assessment for each of the studies were attached in Appendix E. Revisit Table 4.3 again for detailed explanation of questions and domains, and their roles in assessment of bias.

First of all, in terms of sampling method, at least 33 (45.8%) studies employed either consecutive or random sampling, eight (11.1%) used neither one, and 31 (43.1%) were unclear about that (Figure 4.4). 31 (43.1%) studies avoided case-control design, while exactly half of them did not. Only nine (12.5%) studies avoided inappropriate exclusions at sampling stage, but most studies did not give a clue. Almost half of the studies had blinded index test interpretation, but only a quarter of them pre-specified its process in detail (used as a replacement of threshold). 57 (79.2%) studies had accurate reference standard, and 36 (50.0%) studies also blinded their interpretation. Around two-third of

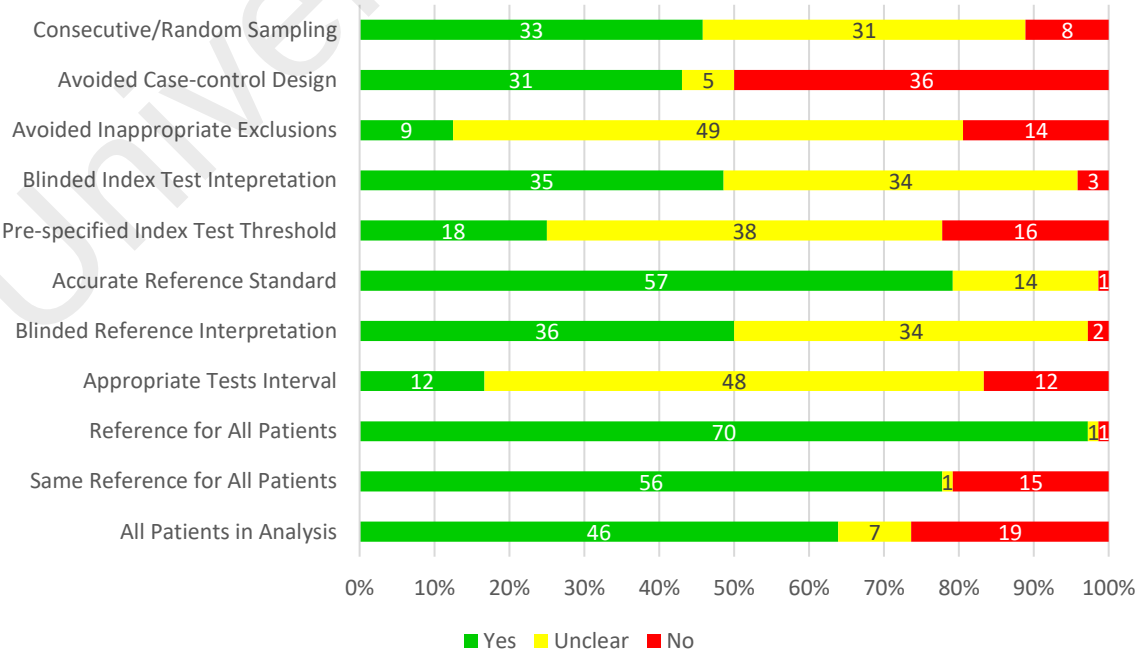


Figure 4.4 Quality assessment of included studies by QUADAS-2 signalling questions

studies failed to report the interval between index and reference tests, while only 12 (16.7%) had an appropriate interval. Almost all studies had at least a reference test carried out for all patients or samples, with one study each (1.4%) that did not or was unclear about that. However, only 56 (77.8%) studies performed the same reference for all patients, while as many as 15 (20.8%) had different reference tests for different patients. Lastly, 46 (63.9%) studies included all patients in analysis, while 19 (26.4%) did not (Figure 4.4).

The biases as assessed according to domain using the above signalling questions were as below (Figure 4.5). Only 13 (18.1%) studies had low patient selection bias, while as many as 42 (58.3%) were highly biased in that. Exactly two-third of the studies were unclear on index test conduct and its bias, compared to only seven (9.7%) studies that did not raise a concern in that. Almost the same number of studies were also unclear on bias in the conduct of reference test, but as many as 29 (40.3%) studies had low bias in this domain. Finally, for flow and timing bias, 15 (20.8%) studies were graded low and 23 (31.9%) were graded high, while almost half of the studies did not provide adequate information for assessment.

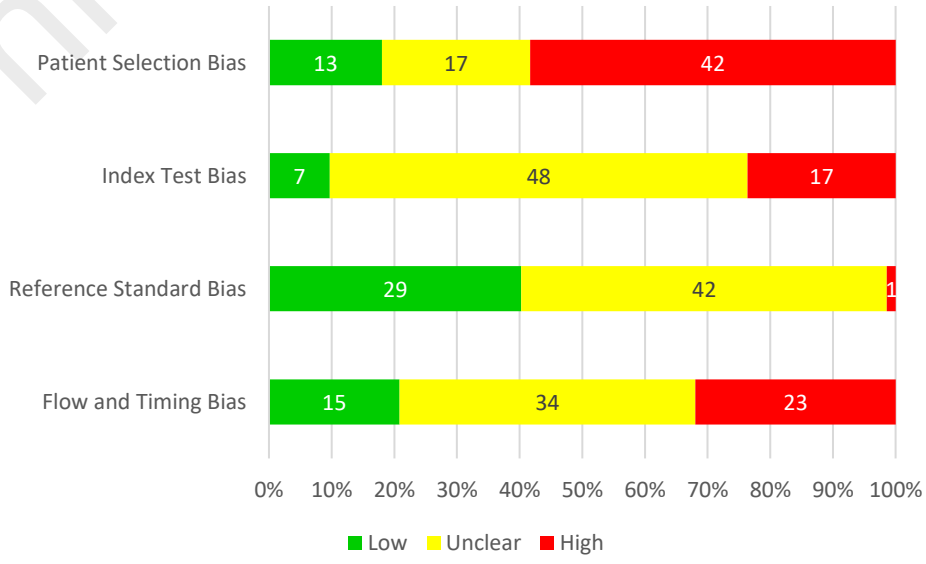


Figure 4.5 Quality assessment of included studies by QUADAS-2 domains

4.3.4 Publication bias

For the assessment of publication bias, one funnel plot was produced for each individual and combined assay (Appendix F). Visually, only the plots of IgG, IgM/IgG, and NS1/IgM/IgG had a resemblance of a funnel. Other plots especially IgM and NS1 assays did not. All plots appeared to be more or less symmetrical, except for IgM/IgG and NS1/IgM/IgG assays. However, among all six assays, only the asymmetry test for IgM/IgG assay generated a p-value of <0.05 (Figure 4.6), indicating that publication bias may be present. Even then, they were more studies that reported results with lower DOR or poorer performance, pulling the regression line towards the left.

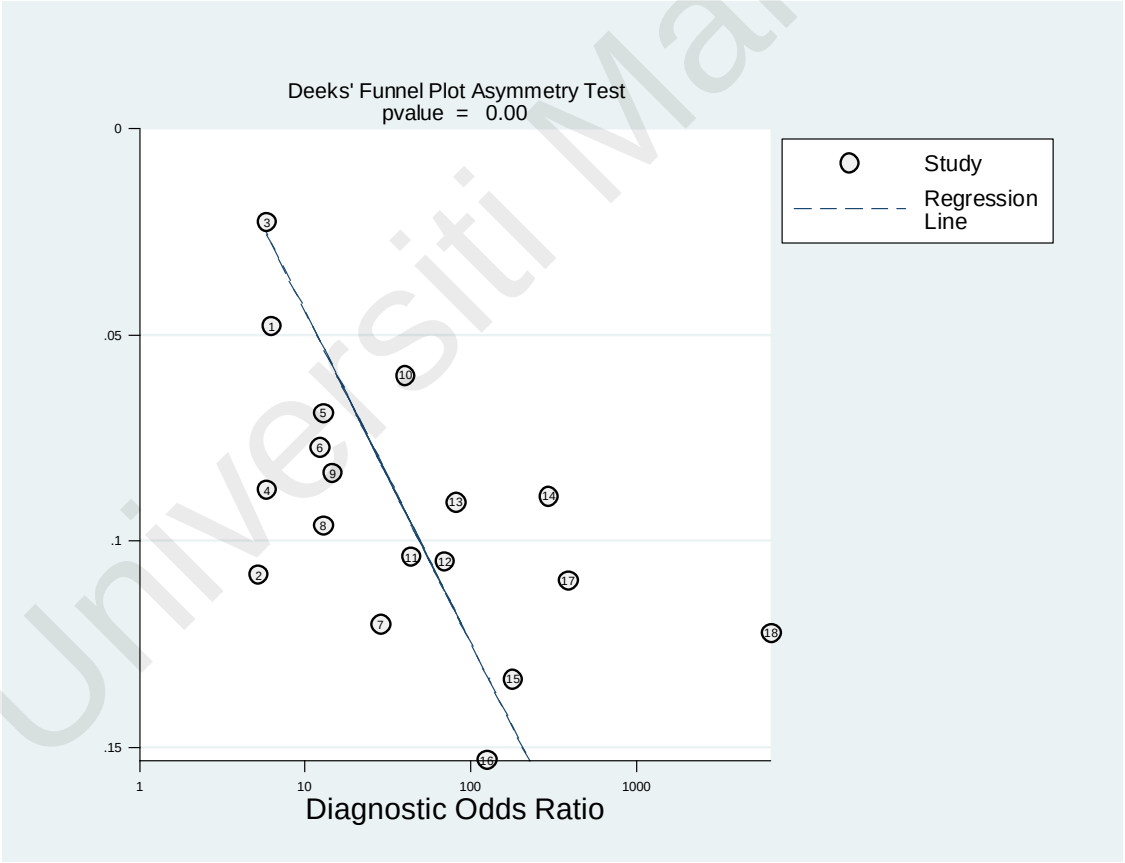


Figure 4.6 Funnel plot for dengue RIT IgM/IgG assay

4.3.5 Pooling of dengue RIT diagnostic accuracy

The overall diagnostic accuracy of dengue RIT was first pooled according to the type of assay. Subsequently, within each assay type, they were pooled by study design, serotype, previous exposure to dengue, disease phase, commercial brand, reference test etc.; for the exploration of sources of heterogeneity.

4.3.5.1 Overall according assay type

The results were as summarised in Figure 4.7 and corresponding Table 4.10. IgM assay had 58 tests pooled, followed by NS1 at 56 tests. The rest of them did not exceed 18, with NS1/IgM/IgG assay having just 10. NS1 assay had the biggest diseased and non-diseased pools with size of slightly more than 10,000 each, while IgM had slightly below that figure for each pool. All the other assays had around 1500 to 2800 samples in each pool. But only those for the above-mentioned and IgM/IgG assays were similar in size. All the other assays had substantially less samples in theirs non-diseased pool.

Between-study heterogeneity was huge, especially for IgM, IgM/IgG, and NS1. NS1/IgM/IgG was least heterogeneous, followed by a tie between NS1/IgM and IgG. The most heterogeneous was IgM assay. It could not be judged from the result whether IgM/IgG or NS1 was more heterogeneous, as it would be the former based on prediction region, the latter based on Higgin's I^2, and a tie based on dispersion of the circles. The extend of the contribution of threshold effect to heterogeneity followed the sequence of the latter, except that IgG appeared to have the least threshold effect based on the closeness to the HSROC curve of higher proportion of its studies.

The AUC (95% CI) of both IgM and IgG assays were above 0.8, indicating very good tests. For NS1 and other combined assays, the numbers were 0.9 and above, especially

for those with NS1 component, making them excellent tests. If AUC were to be compared, all assays with NS1 component and even IgM/IgG were significantly better compared to both IgM and IgG assays. However, according to the HSROC curves and their 95% confidence regions, and the DOR (95%CI), only NS1, NS1/IgM, and NS1/IgM/IgG performed significantly better than standalone IgM assay.

The superior performance of NS1 assay was achieved through both higher SN and SP. In fact, its pooled SP was significantly higher compared to IgM, IgM/IgG and NS1/IgM/IgG. While the pooled SP was more or less uniformed across assays with lower bound of CI always above 80%, the pooled SN varied more widely. The observation made from HSROC curves and DOR mentioned above stood true when it came to pooled SN, but here, IgM/IgG also had significantly better performance compared to IgM alone (Table 4.10*), but was masked in DOR due to poor SP or high FP. The same was observed for the significantly higher SN of both NS1/IgM and NS1/IgM/IgG as compared to IgG (Table 4.10^). On the other hand, combination with IgM alone or both IgM and IgG yielded significantly better SN compared to NS1 alone (Table 4.10+). Finally, combination with NS1 almost produced significantly better SN compared to combination of only serology (Table 4.10#).

Similar to SP, LR+ for NS1 was highest and significantly higher than that of IgM, IgM/IgG, and NS1/IgM/IgG assays. On the other hand, NS1/IgM/IgG assay had the lowest and best LR-, and together with that of NS1/IgM, was significantly lower than the estimates of all standalone assays (Table 4.10~). In addition, NS1 and IgM/IgG assays also had significant lower LR- compared to IgM (Table 4.10!). As opposed to SN and SP, the comparison for likelihood ratios can only be made through the tables.

Despite these interesting findings, due to high between-study heterogeneity, none of the pooled values can be taken as true. These pooled values were the average of all the

included studies, SN and SP of which varied widely as evident visually in Figure 4.7. For IgM assay, these SN ranged from close to 0 to 100%, while SP ranged from around 55 to 100%. IgG assay had SN of 40-90% and SP of 65-100%. The SN and SP of IgM/IgG combination were in between that registered by its individual components as shown above, ranging between 20-100% and 60-100%, respectively. For NS1 assay, the SN were 25-100%, and the SP were 70-100%. The combination of NS1 with serology generally had better underlying SN and SP. NS1/IgM assay had SN ranging from just below 60% to slightly below 100%, and SP from 60 to 100%. Lastly, the combination of all assays had SN of 80-100% and SP of 55-100%.

As such, subgroup/sensitivity analyses followed below to investigate the source of heterogeneity.

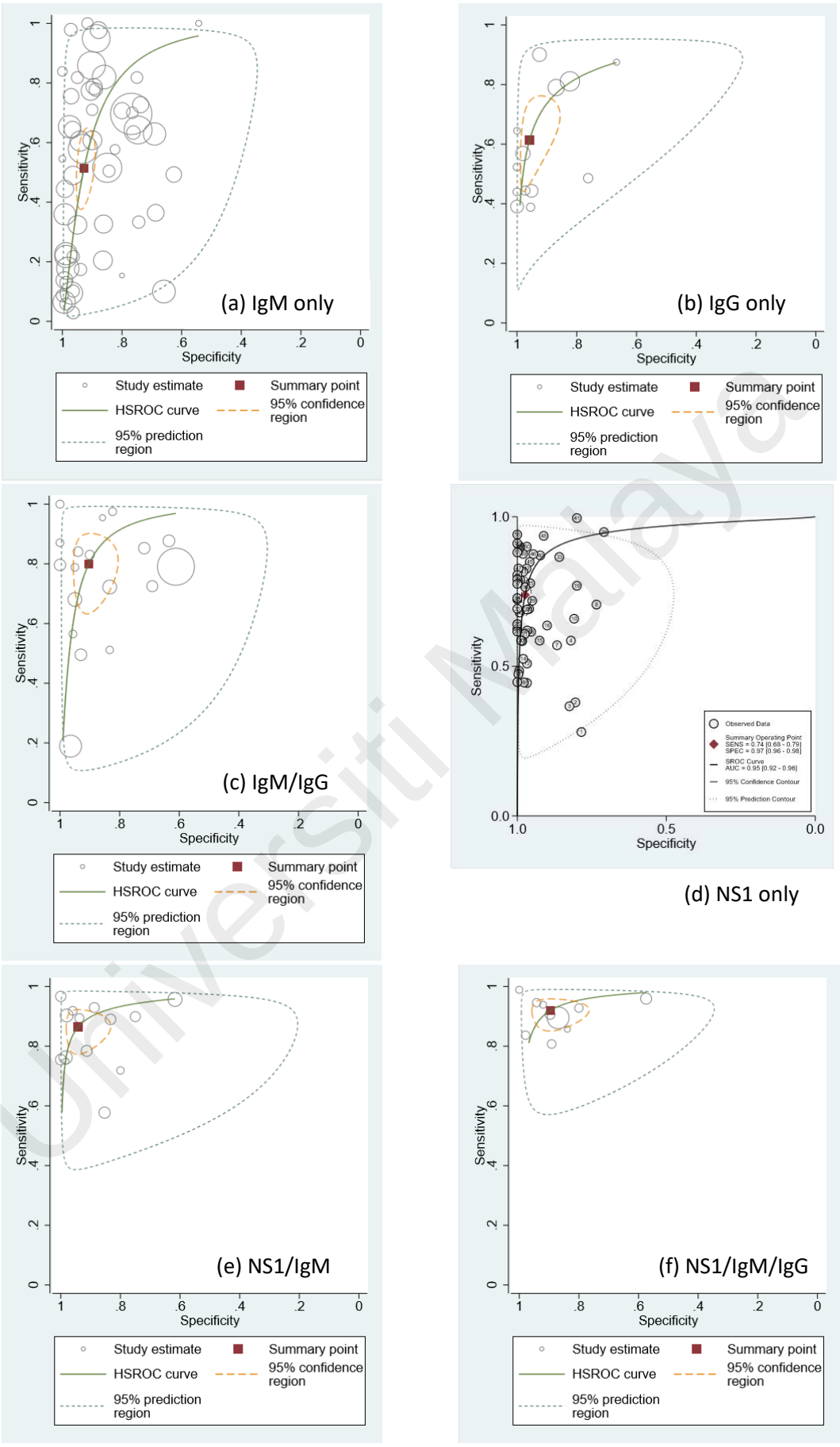


Figure 4.7 Pooling of overall dengue RIT diagnostic accuracy according to assay type (HSROC graphs)

Table 4.10 Pooling of overall dengue RIT diagnostic accuracy according to assay type (table)

Assay	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve								
IgM	58	9830	9852	0.87 (0.84-0.90)	**51.00 (41.00-62.00)***	**93.00 (90.00-95.00)**	**14.00 (9.00-22.00)**	**7.50 (5.30-10.50)**	**0.52 (0.43-0.65)~!**	100 (100-100)	<0.001
IgG	13	2166	1575	0.88 (0.85-0.90)	**61.38 (49.41-72.12)^**	95.78 (89.94-98.30)	36.10 (16.22-80.32)	14.55 (6.46-32.81)	**0.40 (0.31-0.53)~**	99 (98-99)	<0.001
IgM/IgG	18	2491	2775	0.93 (0.90-0.95)	79.91 (68.39-87.97)*#	**90.35 (83.76-94.44)**	37.24 (16.46-84.24)	**8.28 (4.81-14.24)**	0.22 (0.14-0.36)~!	100 (99-100)	<0.001
NS1	56	10359	10135	0.95 (0.92-0.96)	74.00 (68.00-79.00)*+	***97.00 (96.00-98.00)***	***106.00 (62.00-180.00)***	***28.30 (17.90-44.80)***	0.27 (0.22-0.33)~!	100 (100-100)	<0.001
NS1/IgM	14	2506	1540	0.95 (0.93-0.97)	***86.40 (80.34-90.81)*^+***	94.18 (87.56-97.38)	***102.80 (40.55-260.65)***	14.84 (6.77-32.52)	***0.14 (0.10-0.21)~***	99 (98-99)	<0.001
NS1/IgM/IgG	10	2770	1823	0.96 (0.94-0.98)	***91.98 (87.88-94.78)*^+#***	**89.54 (81.97-94.15)**	***98.21 (45.32-212.85)***	**8.79 (4.99-15.50)**	***0.09 (0.06-0.14)~***	97 (96-99)	<0.001

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only. For comparison between more than one sets of assays, symbol (*, ^, +, #, ~, !) indicates relevant set. Only symbols will be highlighted if any estimate was both inferior and superior at the same time. Apply for all following tables with pooling estimates.

4.3.5.2 Subgroup/Sensitivity analyses

The between-study heterogeneity in the overall estimates pooled by assay type were explored within each assay type. Results of subgroup or sensitivity analyses performed according to study design, serotype, previous exposure to dengue, disease phase, commercial brand, reference test, and any other subgroups that were found to be significant; were presented in tables together with texts, while their corresponding HSROC graphs were attached in Appendix G.

(a) IgM assay

The results of subgroup analysis for IgM assay were summarised in Table 4.11 and Appendix G (I). None of the subgroup could satisfactorily explain away the heterogeneity found in the overall pool. Although most analyses were able to slightly reduce the scatter of circles and the size of their prediction regions compared to the overall HSROC graphs, their I^2 remained above 50%, and their p-values <0.05. Among all, Panbio IC had the lowest I^2 as well as the lowest number of studies. The heterogeneity found in all studies was partly due to threshold effect except in the pool for DENV-1 and DENV-4, where almost all circles were located on the curve.

Among all subgroup analyses, only the distribution of study design and commercial brand was known in all the underlying studies. And their diseased and non-diseased pools were exclusive from each other, as opposed to other subgroups that may have common non-diseased pools, such as serotype, previous exposure to dengue infection, and reference test. Here, it is important to note that case-control studies had equal diseased and non-diseased pool size, and cross-sectional studies had bigger non-diseased pools size.

Coincidentally, the categories within both these subgroups with known distribution also had significant difference between some of the performance parameters. IgM RIT evaluated in case-control studies had significantly higher DOR and LR+ as compared to cross-sectional studies, due to the combination of insignificantly higher SN and SP. On the other hand, Panbio IC and Panbio Dengue Duo IgM component (Panbio in Table 4.11) had significantly better SN and LR-, and the DOR of the former was also significantly higher compared to the combination of all other small brands. However, only SD had larger non-diseased pool pointing to a substantial presence of cross-sectional studies.

In addition to the above, significant differences were found between the performance of IgM RIT as evaluated by serology-based and virology-based methods. The SN and corresponding LR- were significantly better, but the SP was significantly lower, when serology was used as reference test.

Among those subgroups that did not produce significant results within their categories, it is important to note that IgM RIT was less sensitive in detecting DENV-4, secondary and acute infection, although its SP was comparable to the rest. However, these analyses were underpowered.

Table 4.11 Subgroup/sensitivity analyses for IgM assay

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
Overall IgM (a)	58	9830	9852	0.87 (0.84-0.90)	51.00 (41.00-62.00) 2.90-100.00^	93.00 (90.00-95.00) 46.25-100.00^	14.00 (9.00-22.00)	7.50 (5.30-10.50)	0.52 (0.43-0.65)	100 (100-100)	<0.001
By study design											
Case-control (b)	26	5984	4303	0.93 (0.91-0.95)	58.15 (42.03-72.69) 6.44-100.00^	95.31 (92.59-97.06) 69.09-100.00^	***28.21 (14.87-53.51)***	***12.39 (8.02-19.14)***	0.44 (0.30-0.64)	100 (100-100)	<0.001
Cross-section (c)	28	2936	4723	0.80 (0.76-0.83)	41.00 (28.00-56.00) 2.90-100.00^	91.00 (85.00-95.00) 46.25-100.00^	**7.00 (4.00-13.00)**	**4.60 (2.80-7.30)**	0.65 (0.52-0.81)	100 (100-100)	<0.001
By serotype											
DENV-1 (d)	10	1430	1439	0.97 (0.95-0.98)	44.14 (17.54-74.60) 10.56-100.00^	97.84 (95.83-98.89) 82.35-100.00^	35.84 (12.36-103.94)	20.46 (10.83-38.66)	0.57 (0.32-1.01)	97 (95-99)	<0.001
DENV-2 (e)	16	662	1764	0.94 (0.92-0.96)	26.64 (10.25-53.58) 1.28-100.00^	97.41 (95.16-98.63) 74.39-100.00^	13.65 (4.71-39.54)	10.28 (4.53-23.32)	0.75 (0.56-1.02)	98 (96-99)	<0.001
DENV-3 (f)	12	223	1405	0.94 (0.92-0.96)	37.00 (13.00-69.00) 3.57-100.00^	97.00 (94.00-98.00) 74.39-100.00^	18.00 (6.00-56.00)	11.60 (5.40-25.00)	0.65 (0.40-1.06)	97 (95-99)	<0.001

(Continued in the following page)

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section I) IgM assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

Table 4.11, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By serotype (continued)											
DENV-4 (g)	8	253	1077	0.96 (0.94-0.98)	19.47 (3.01-65.34) 2.50-100.00^	97.84 (95.78-98.91) 82.35-100.00^	10.96 (2.20-54.58)	9.02 (2.62-31.00)	0.82 (0.56-1.21)	96 (93-99)	<0.001
By previous exposure to dengue infection											
Primary (h)	9	576	1909	0.94 (0.92-0.96)	79.49 (61.91-90.24) 30.00-95.71^	92.56 (85.97-96.19) 76.19-100.00^	48.20 (14.70-158.05)	10.68 (5.30-21.51)	0.22 (0.11-0.45)	97 (95-99)	<0.001
Secondary (i)	9	1567	1909	0.92 (0.89-0.94)	66.70 (42.56-84.40) 10.00-97.39^	92.62 (85.96-96.26) 76.19-100.00^	25.15 (10.04-62.96)	9.04 (5.01-16.32)	0.36 (0.19-0.68)	99 (98-99)	<0.001
By disease phase											
Acute (<=5 days) (j)	20	2656	4907	0.87 (0.84-0.90)	40.00 (24.00-60.00) 3.07-94.64^	95.00 (90.00-97.00) 65.96-100.00^	13.00 (6.00-28.00)	8.00 (4.20-15.20)	0.63 (0.46-0.85)	100 (100-100)	<0.001
Convalescent (>=6 days) (k)	20	3059	3629	0.95 (0.93-0.97)	66.83 (47.25-81.92) 9.82-100.00^	96.15 (92.58-98.04) 69.09-100.00^	50.30 (23.10-109.53)	17.35 (9.63-31.29)	0.35 (0.20-0.59)	100 (100-100)	<0.001

(Continued in the following page)

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section I) IgM assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

Table 4.11, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By commercial brand											
SD (l)	17	2932	4005	0.85 (0.82-0.88)	**49.78 (36.21-63.38)**	93.24 (88.25-96.20)	13.67 (5.95-31.37)	7.36 (3.90-13.90)	0.54 (0.41-0.71)	99 (99-100)	<0.001
					10.00-85.90^	65.96-100.00^					
Panbio (m)	10	1111	1099	0.91 (0.88-0.93)	***77.33 (60.23-88.48)***	88.12 (79.68-93.34)	25.29 (11.37-56.29)	6.51 (3.89-10.88)	***0.26 (0.14-0.47)***	98 (97-99)	<0.001
					21.74-100.00^	54.29-97.58^					
Panbio IC (n)	5	379	230	0.96 (0.93-0.97)	***89.81 (65.56-97.61)***	93.10 (86.66-96.55)	***118.80 (17.59-802.36)***	13.01 (6.16-27.48)	***0.11 (0.03-0.44)***	71 (37-100)	0.015
					57.69-100.00^	82.35-100.00^					
Others (o)	26	5408	4518	0.78 (0.74-0.82)	**33.00 (21.00-46.00)**	94.00 (90.00-97.00)	**8.00 (5.00-14.00)**	5.90 (3.60-9.90)	**0.71 (0.60-0.84)**	100 (100-100)	<0.001
					2.90-94.95^	46.25-100.00^					

(Continued in the following page)

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section I) IgM assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

Table 4.11, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By reference test											
Serology (p)	27	4529	8683	0.89 (0.85-0.91)	***64.78 (53.54-74.59)***	**91.11 (86.58-94.22)**	18.86 (10.24-34.72)	7.29 (4.74-11.22)	***0.39 (0.29-0.52)***	100 (100-100)	<0.001
					14.00-100.00^	54.29-100.00^					
Virology (q)	19	2565	2057	0.93 (0.91-0.95)	**30.73 (17.27-48.53)**	***98.28 (96.53-99.16)***	25.42 (8.57-75.37)	17.91 (7.17-44.76)	**0.70 (0.56-0.89)**	97 (94-99)	<0.001
					6.44-100.00^	65.96-100.00^					
Clinical (r)	6	778	918	0.92 (0.90-0.94)	65.33 (31.51-88.53)	94.37 (81.78-98.43)	31.60 (6.65-150.11)	11.61 (3.61-37.28)	0.37 (0.15-0.90)	99 (98-99)	<0.001
					13.85-97.73^	68.71-100.00^					

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section I) IgM assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

(b) IgG assay

The results of subgroup analysis for IgG assay were summarised in Table 4.12 and Appendix G (II). Similar to IgM, none of the analyses could explained the heterogeneity. Having a narrow prediction region, the HSROC graph of case-control studies looked promising initially, but the SN still ranged from around 40% to 90%, and the I^2 remained high. Only acute had the lowest I^2, but also had least studies. Threshold effect was present in all poolable groups but was minimal in case-control and SD.

Subgroup analysis was possible only for disease phase. IgG RIT appeared to be more sensitive in convalescent phase. The 95%CI of both its SN and corresponding LR- overlapped only slightly with that of the acute phase. However, the analyses were underpowered due to small sample size. For previous exposure to dengue infection, subgroup analysis was possible although primary could not be pooled due to presence of only three studies. The SN of IgG assay in secondary infections was significantly higher than that of the acute infections in all three individual studies. As for reference test, the only study using clinical picture as reference test had higher SN compared to the rest, but with corresponding lowest SP.

Sensitivity analysis for stratification by study design and by commercial brand was possible by directly comparing the poolable categories with the overall pool. Case-control studies generally had better performance estimates across the board compared to the overall estimates, except SN and LR-. However, majority of the unpooled individual cross-sectional studies had lower SN and SP with one outlying study that generated very high SN and low SP. The SD IgG RIT also had generally better estimates compared to the overall, except SN and LR-. The unpooled SN and SP of its contenders variably fell on both sides of its estimates. Nonetheless, all analyses for IgG assay were underpowered to detect meaningful difference, and still too heterogeneous even if there were.

Table 4.12 Subgroup/sensitivity analyses for IgG assay

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
Overall IgG (a)	13	2166	1575	0.88 (0.85-0.90)	61.38 (49.41-72.12) 38.84-90.06^	95.78 (89.94-98.30) 66.67-100.00^	36.10 (16.22-80.32)	14.55 (6.46-32.81)	0.40 (0.31-0.53)	99 (98-99)	<0.001
By study design											
Case-control (b)	7	1047	634	0.96 (0.94-0.98)	58.46 (43.63-71.91) 44.00-90.06^	***97.16 (94.77-98.48)*** 92.48-100.00^	48.22 (25.90-89.76)	20.61 (12.01-35.38)	0.43 (0.30-0.60)	96 (92-99)	<0.001
Cross-section (unpooled)	3	121	22	0.67 (0.61-0.73)	**38.80 (30.10-48.10)**	95.50 (77.20-99.90)	-	-	-	-	-
		256	83	0.70 (0.67-0.73)	**39.10 (33.00-45.30)**	100.00 (95.70-100.00)	-	-	-	-	-
		40	39	0.77 (0.68-0.86)	87.50 (73.20-95.80)	**66.70 (49.80-80.90)**	-	-	-	-	-
By previous exposure to dengue infection											
Primary (unpooled)	3	46	107	0.52 (0.48-0.56)	**6.52 (1.37-17.90)**	97.20 (92.00-99.40)	-	-	-	-	-
		20	100	0.59 (0.50-0.68)	**20.00 (5.73-43.70)**	98.00 (93.00-99.80)	-	-	-	-	-
		45	42	0.59 (0.49-0.69)	**42.20 (27.70-57.80)**	76.20 (60.50-87.90)	-	-	-	-	-
Secondary (c)	4	173	279	N.A.	***86.21 (62.85-95.85)*** 51.55-100.00^	96.30 (85.48-99.14) 76.19-100.00^	162.85 (12.02-2207.09)	23.31 (4.85-112.06)	0.14 (0.04-0.46)	N.A.	N.A.

(Continued in the following page)

Table 4.12, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By disease phase											
Acute (<=5 days) (d)	4	174	279	0.50 (0.46-0.54)	43.78 (35.03-52.93) 38.89-50.00^	96.46 (84.83-99.25) 76.19-100.00^	21.19 (4.55-98.70)	12.35 (2.78-54.90)	0.58 (0.50-0.68)	85 (70-100)	0.001
Convalescent (>=6 days) (e)	6	282	759	0.81 (0.77-0.84)	62.05 (51.16-71.85) 50.00-79.00^	96.23 (86.62-99.02) 76.19-100.00^	41.77 (10.40-167.75)	16.47 (4.44-61.12)	0.39 (0.30-0.52)	93 (86-99)	<0.001
By commercial brand											
SD (f)	4	868	431	0.96 (0.94-0.98)	59.56 (33.98-80.82) 38.84-90.06^	97.89 (93.36-99.35) 92.48-100.00^	68.28 (26.33-177.04)	28.21 (10.91-72.96)	0.41 (0.22-0.76)	96 (93-99)	<0.001
Panbio (unpooled)	3	142	42	0.62 (0.55-0.70)	48.60 (40.10-57.10)	**76.20 (60.50-87.90)**	-	-	-	-	-
		72	107	0.71 (0.65-0.77)	44.40 (32.70-56.60)	97.20 (92.00-99.40)	-	-	-	-	-
		40	39	0.77 (0.68-0.86)	***87.50 (73.20-95.80)***	**66.70 (49.80-80.90)**	-	-	-	-	-
Panbio IC (unpooled)	3	100	30	0.72 (0.67-0.77)	**44.00 (34.10-54.30)**	***100.00 (88.40-100.00)***	-	-	-	-	-
		65	59	0.76 (0.70-0.82)	**52.30 (39.50-64.90)**	***100.00 (93.90-100.00)***	-	-	-	-	-
		62	30	0.82 (0.76-0.88)	64.50 (51.30-76.30)	***100.00 (88.40-100.00)***	-	-	-	-	-

(Continued in the following page)

Table 4.12, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By reference test											
Serology (g)	10	1662	1058	0.90 (0.87-0.92)	61.42 (43.27-76.87) 18.00-90.06^	95.29 (87.80-98.27) 66.67-100.00^	32.20 (12.25-84.63)	13.04 (5.32-31.93)	0.40 (0.26-0.62)	99 (98-99)	<0.001
Virology (unpooled)	3	30	100	0.64 (0.56-0.73)	**30.00 (14.70-49.40)**	***98.00 (93.00-99.80)***	-	-	-	-	-
		50	100	0.68 (0.61- 0.75)	**38.00 (24.70-52.80)**	***98.00 (93.00-99.80)***	-	-	-	-	-
		50	100	0.86 (0.80-0.92)	74.00 (59.70-85.40)	***98.00 (93.00-99.80)***	-	-	-	-	-
Clinical (single study)	1	100	450	0.83 (0.79-0.87)	***79.00 (69.70-86.50)***	**86.90 (83.40-89.90)**	-	-	-	-	-

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section II) IgG assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

(c) IgM/IgG assay

The results of subgroup analysis for IgG assay were summarised in Table 4.13 and Appendix G (III). In a similar fashion to the above assays, the subgroup categories remained heterogeneous from the HSROC graphs and the table. Threshold effect was present in all the pools.

The results of the subgroup analyses were expectedly the effect of both IgM and IgG assays. It was neutralised in previous exposure to dengue infection, where no difference between primary and secondary was found; but was compounded in disease phase, where IgM and IgG combined had significantly better SN and corresponding LR-. Although IgM/IgG assay appeared to also have higher SN when serology was used rather than virology, the difference was not statistically significant.

On the other hand, assay performance was better across the board when evaluated by case-control studies as compared to the overall estimates, as well as that of the majority individual cross-sectional studies. Similar to the assays above, Panbio IC had significantly higher SN and DOR.

Table 4.13 Subgroup/sensitivity analyses for IgM/IgG assay

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
Overall IgM/IgG (a)	18	2491	2775	0.93 (0.90-0.95)	79.91 (68.39-87.97) 18.94-100.00^	90.35 (83.76-94.44) 61.11-100.00^	37.24 (16.46-84.24)	8.28 (4.81-14.24)	0.22 (0.14-0.36)	100 (99-100)	<0.001
By study design											
Case-control (b)	13	1378	609	0.95 (0.93-0.97)	***83.30 (70.83-91.11)*** 49.54-100.00^	***92.44 (87.08-95.69)*** 71.74-100.00^	60.97 (23.17-160.44)	11.02 (6.22-19.52)	0.18 (0.10-0.33)	95 (91-99)	<0.001
Cross-section (unpooled)	3	132	591	0.58 (0.54-0.61)	**18.90 (12.60-26.70)**	96.40 (94.60-97.80)	-	-	-	-	-
		57	156	0.76 (0.70-0.81)	87.70 (76.30-94.90)	**63.50 (55.40-71.00)**	-	-	-	-	-
		52	40	0.87 (0.80-0.94)	78.80 (65.30-88.90)	95.00 (83.10-99.40)	-	-	-	-	-
By previous exposure to dengue infection											
Primary (c)	8	278	330	0.93 (0.90-0.95)	82.49 (60.41-93.56) 42.11-100.00^	88.92 (77.88-94.82) 69.05-100.00^	37.81 (13.73-104.15)	7.45 (3.86-14.37)	0.20 (0.08-0.48)	97 (94-99)	<0.001
Secondary (d)	8	450	330	0.94 (0.92-0.96)	86.89 (76.39-93.14) 65.71-100.00^	89.68 (79.45-95.13) 69.05-100.00^	57.57 (17.59-188.45)	8.42 (3.97-17.84)	0.15 (0.08-0.28)	77 (51-100)	0.006

(Continued in the following page)

Table 4.13, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By disease phase											
Acute (<=5 days) (e)	7	809	1354	0.83 (0.79-0.86)	**61.79 (43.82-77.03)**	89.16 (73.63-96.03)	13.30 (3.81-46.45)	5.70 (2.14-15.18)	**0.43 (0.27-0.68)**	99 (98-99)	<0.001
					18.94-85.00^	63.53-100.00^					
Convalescent (>=6 days) (f)	10	808	1403	0.96 (0.94-0.97)	***95.81 (85.43-98.89)***	85.84 (73.46-92.99)	138.51 (32.12-597.28)	6.76 (3.48-13.17)	***0.05 (0.01-0.18)***	99 (99-100)	<0.001
					59.09-100.00^	58.35-100.00^					
By commercial brand											
SD (single study)	1	166	120	0.82 (0.78-0.86)	**68.10 (60.40-75.10)**	95.00 (89.40-98.10)	-	-	-	-	-
Panbio (g)	9	1682	2435	0.85 (0.82-0.88)	67.85 (52.30-80.25)	86.92 (75.68-93.41)	**14.02 (6.72-29.27)**	5.19 (2.87-9.37)	0.37 (0.25-0.56)	100 (99-100)	<0.001
					18.94-87.72^	61.11-100.00^					
Panbio IC (h)	8	643	220	0.97 (0.95-0.98)	***90.49 (77.98-96.24)***	92.63 (83.25-96.95)	***119.70 (30.19-474.54)***	12.28 (5.16-29.25)	0.10 (0.04-0.25)	89 (79-100)	<0.001
					56.52-100.00^	71.74-100.00^					

(Continued in the following page)

Table 4.13, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By reference test											
Serology (i)	9	831	935	0.92 (0.89-0.94)	74.43 (58.01-85.99) 18.94-95.45^	90.91 (82.55-95.48) 69.05-100.00^	29.12 (12.27-69.09)	8.19 (4.30-15.58)	0.28 (0.16-0.48)	99 (98-99)	<0.001
Virology (unpooled)	2	43	42	0.67 (0.58-0.77)	51.20 (35.50-66.70)	83.30 (68.60-93.00)	-	-	-	-	-
		145	120	0.80 (0.76-0.84)	64.80 (56.50-72.60)	95.00 (89.40-98.10)	-	-	-	-	-

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section III) IgM/IgG assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

(d) NS1 assay

The results of subgroup analysis for IgG assay were summarised in Table 4.14 and Appendix G (IV). According to the HSROC graphs, most analyses were able to reduce the scatter of the circle and prediction region, especially all the serotype pools. However, only the pooling by DENV-4 infections and serology had I^2 of <50 and p>=0.05. This was in spite of scatter of the circles across SN-axis and moderate size of prediction region for serology. For the others, there remained unexplained between-study heterogeneity, with partial contribution of threshold effect in all pools.

Unfortunately, for DENV-4 and serology pools that could explain the heterogeneity, it was indicating NS1 assay significantly worse SN and LR- in both of these situations. However, the SP of NS1 RIT was significantly higher when serology was used as reference test compared to clinical pictures only, but so was when virology was used.

Just like IgM assay, the non-diseased pool used for NS1 RIT evaluation was also smaller in case-control studies, but bigger in cross-sectional. However, the results for both designs were similar. Despite that, NS1 assay was obviously better in detecting primary and acute dengue infections, evident by its significantly better SN and corresponding LR-, as compared to secondary infections and those in convalescent phase.

All commercial brands did not demonstrate statistical difference in SN, but Biorad NS1 STRIP had significantly higher SP and corresponding LR+ compared to that of all other small brands combined. However, only SD had bigger non-diseased pool indicating larger proportion of cross-sectional design in the underlying studies.

Table 4.14 Subgroup/sensitivity analyses for NS1 assay

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
Overall NS1 (a)	56	10359	10135	0.95 (0.92-0.96)	74.00 (68.00-79.00) 28.04-99.54^	97.00 (96.00-98.00) 70.92-100.00^	106.00 (62.00-180.00)	28.30 (17.90-44.80)	0.27 (0.22-0.33)	100 (100-100)	<0.001
By study design											
Case-control (b)	27	4959	2901	0.95 (0.93-0.97)	73.51 (66.07-79.81) 28.04-94.03^	97.79 (95.80-98.84) 78.50-100.00^	122.52 (49.74-301.80)	33.19 (16.47-66.91)	0.27 (0.21-0.35)	94 (90-99)	<0.001
Cross-section (c)	25	4588	6594	0.96 (0.93-0.97)	74.39 (65.06-81.92) 44.44-99.54^	97.37 (95.19-98.57) 70.92-100.00^	107.42 (58.47-197.34)	28.26 (15.93-50.13)	0.26 (0.19-0.36)	100 (99-100)	<0.001
By serotype											
DENV-1 (d)	17	998	1644	0.98 (0.96-0.99)	***86.06 (80.81-90.06)*** 61.40-98.00^	98.34 (96.87-99.13) 91.13-100.00^	366.58 (174.44-770.39)	51.95 (27.35-98.68)	***0.14 (0.10-0.20)***	89 (77-100)	<0.001
DENV-2 (e)	18	1384	4529	0.98 (0.97-0.99)	***79.98 (71.63-86.35)*** 47.35-98.00^	98.56 (97.34-99.23) 91.13-100.00^	273.67 (137.24-545.69)	55.58 (30.34-101.82)	***0.20 (0.14-0.29)***	97 (94-99)	<0.001
DENV-3 (f)	14	528	1321	0.92 (0.89-0.94)	***79.85 (75.05-83.93)*** 61.29-90.24^	98.65 (96.89-99.42) 91.13-100.00^	290.68 (114.75-736.36)	59.36 (25.31-139.20)	***0.20 (0.16-0.26)***	84 (67-100)	0.001
DENV-4 (g)	12	413	944	0.96 (0.94-0.98)	**60.41 (49.04-70.76)** 33.33-100.00^	98.40 (96.93-99.17) 95.00-100.00^	94.03 (36.49-242.26)	37.83 (17.93-79.80)	**0.40 (0.30-0.53)**	0 (0-100)	0.500

(Continued in the following page)

Table 4.14, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By previous exposure to dengue infection											
Primary (h)	18	1069	2341	0.95 (0.92-0.96)	***78.60 (70.90-84.71)***	98.24 (95.65-99.30)	204.88 (60.54-693.35)	44.63 (16.95-117.53)	***0.22 (0.16-0.30)***	95 (90-99)	<0.001
					38.46-95.00^	78.50-100.00^					
Secondary (i)	18	2264	2341	0.84 (0.80-0.87)	**52.00 (42.00-61.00)**	98.00 (96.00-99.00)	68.00 (20.00-234.00)	33.10 (10.90-100.80)	**0.49 (0.40-0.60)**	98 (97-99)	<0.001
					5.00-92.86^	78.50-100.00^					
By disease phase											
Acute (<=5 days) (j)	41	6558	7031	0.94 (0.92-0.96)	***76.18 (70.58-81.00)***	97.28 (95.06-98.51)	114.21 (58.66-222.38)	27.97 (15.34-51.01)	***0.25 (0.20-0.31)***	99 (99-100)	<0.001
					40.00-99.54^	70.92-100.00^					
Convalescent (>=6 days) (k)	20	1112	1559	0.91 (0.88-0.93)	**55.00 (42.00-68.00)**	98.00 (95.00-99.00)	57.00 (16.00-201.00)	26.10 (9.20-74.30)	**0.46 (0.34-0.63)**	95 (91-99)	<0.001
					12.35-100.00^	78.50-100.00^					
By commercial brand											
SD (l)	23	4507	6504	0.93 (0.91-0.95)	72.03 (63.91-78.93)	96.83 (94.11-98.31)	78.55 (38.48-160.36)	22.69 (12.15-42.37)	0.29 (0.22-0.38)	99 (99-100)	<0.001
					44.44-94.86^	70.92-100.00^					
Panbio (m)	7	1291	582	N.A.	65.28 (53.55-75.40)	95.49 (89.77-98.08)	39.85 (10.97-144.66)	14.49 (5.46-38.44)	0.36 (0.26-0.52)	N.A.	N.A.
					37.94-88.06^	80.50-100.00^					

(Continued in the following page)

Table 4.14, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By commercial brand (continued)											
Biorad (n)	17	2917	1948	0.97 (0.95-0.98)	77.38 (66.64-85.42)	***99.10 (97.11-99.72)***	375.26 (107.82-1306.09)	***85.65 (26.72-274.49)***	0.23 (0.15-0.35)	98 (97-99)	<0.001
					36.66-99.54^	80.00-100.00^					
Others (o)	7	1326	1056	0.94 (0.92-0.96)	80.58 (64.21-90.56)	**92.15 (85.92-95.76)**	48.66 (13.55-174.72)	**10.26 (5.13-20.51)**	0.21 (0.11-0.43)	97 (94-99)	<0.001
					28.04-93.60^	78.50-97.37^					
By reference test											
Serology (p)	9	987	1105	0.95 (0.93-0.97)	**63.00 (48.00-75.00)**	***99.00 (97.00-100.00)***	190.00 (37.00-988.00)	71.50 (18.10-282.70)	0.38 (0.26-0.55)	46 (0-100)	0.079
					28.00-86.00^	83.33-100.00^					
Virology (q)	43	5258	9133	0.97 (0.95-0.98)	***83.00 (78.00-87.00)***	***98.00 (97.00-99.00)***	265.00 (146.00-480.00)	44.60 (26.60-75.10)	0.17 (0.13-0.22)	100 (99-100)	<0.001
					44.00-99.54^	70.92-100.00^					
Clinical (single study)	1	550	390	0.86 (0.84-0.89)	***86.50 (83.40-89.30)***	**85.90 (82.00-89.20)**	-	-	-	-	-

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section IV) NS1 assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

(e) NS1/IgM assay

The results of subgroup analysis for IgG assay were summarised in Table 4.15 and Appendix G (V). Only case-control and virology were able to reduce the heterogeneity as assessed visually. And truly, case-control studies had low I^2, but virology still had it high. That said, case-control also had small number of studies as compared to the latter. In a similar way, both the above pools did not suffer from threshold effect like the rest.

Due to small number of studies, only study design had two pooled subcategories. As mentioned above, pooling of NS1/IgM RIT accuracy by case-control studies was not heterogeneous, and the estimates were too underpowered to be concluded as better than those generated by cross-sectional design, which was rather similar to that found for NS1.

Otherwise, the combination of NS1/IgM appeared to be more sensitive in detecting primary and convalescent infections. SD had higher SP compared to Panbio. And the SN and SP were not significantly lower when virology was used as reference test. However, the underlying heterogeneity of all groups except one would undermine the validity of most findings here.

Table 4.15 Subgroup/sensitivity analyses for NS1/IgM assay

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
Overall NS1/IgM (a)	14	2506	1540	0.95 (0.93-0.97)	86.40 (80.34-90.81) 57.75-96.69^	94.18 (87.56-97.38) 61.70-100.00^	102.80 (40.55-260.65)	14.84 (6.77-32.52)	0.14 (0.10-0.21)	99 (98-99)	<0.001
By study design											
Case-control (b)	4	802	220	0.90 (0.87-0.92)	89.00 (86.00-91.00) 75.00-90.31^	96.00 (88.00-98.00) 83.33-98.28^	173.00 (58.00-512.00)	20.10 (7.20-56.10)	0.12 (0.10-0.14)	18 (0-100)	0.147
Cross-section (c)	10	1704	1320	0.95 (0.93-0.96)	86.17 (77.37-91.90) 57.75-96.69^	93.63 (83.19-97.76) 61.70-100.00^	91.60 (26.14-320.94)	13.53 (4.86-37.70)	0.15 (0.09-0.25)	99 (98-99)	<0.001
By previous exposure to dengue infection											
Primary (unpooled)	3	110	32	***0.95 (0.90-0.99)***	***95.50 (89.70-98.50)***	93.80 (79.20-99.20)	-	-	-	-	-
		66	47	0.92 (0.87-0.96)	83.30 (72.10-91.40)	100.00 (92.50-100.00)	-	-	-	-	-
		20	100	***0.99 (0.98-1.00)***	***100.00 (83.20-100.00)***	98.00 (93.00-99.80)	-	-	-	-	-
Secondary (unpooled)	3	53	32	**0.86 (0.78-0.93)**	**77.40 (63.80-87.70)**	93.80 (79.20-99.20)	-	-	-	-	-
		176	47	**0.86 (0.83-0.90)**	**72.70 (65.50-79.20)**	100.00 (92.50-100.00)	-	-	-	-	-
		20	100	0.94 (0.87-1.00)	90.00 (68.30-98.80)	98.00 (93.00-99.80)	-	-	-	-	-

(Continued in the following page)

Table 4.15, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By disease phase											
Acute (<=5 days) (d)	6	877	462	0.95 (0.93-0.97)	86.50 (77.35-92.32) 73.08-95.68^	94.59 (83.42-98.38) 61.70-100.00^	112.03 (39.15-320.59)	15.99 (5.14-49.75)	0.14 (0.09-0.24)	98 (97-99)	<0.001
Convalescent (>=6 days) (unpooled)	3	97	50	0.95 (0.91-0.99)	93.80 (87.00-97.70)	96.00 (86.30-99.50)	-	-	-	-	-
		20	100	0.97 (0.91-1.00)	95.00 (75.10-99.90)	98.00 (93.00-99.80)	-	-	-	-	-
		20	32	0.97 (0.93-1.00)	100.00 (83.20-100.00)	93.80 (79.20-99.20)	-	-	-	-	-
By commercial brand											
SD (e)	10	1925	1260	0.96 (0.93-0.97)	86.00 (78.00-92.00) 57.75-96.69^	***95.00 (87.00-98.00)*** 61.70-100.00^	127.00 (34.00-466.00)	18.30 (6.30-52.80)	0.14 (0.09-0.24)	99 (98-99)	<0.001
Panbio (unpooled)	2	99	160	0.82 (0.78-0.87)	89.90 (82.20-95.00)	**75.00 (67.60-81.50)**	-	-	-	-	-
		263	30	0.86 (0.79-0.93)	89.00 (84.50-92.50)	83.30 (65.30-94.40)	-	-	-	-	-

(Continued in the following page)

Table 4.15, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
By reference test											
Serology (unpooled)	2	50	100	0.95 (0.91-0.99)	92.00 (80.80-97.80)	98.00 (93.00-99.80)	-	-	-	-	-
		100	100	0.96 (0.93-0.98)	93.00 (86.10-97.10)	98.00 (93.00-99.80)	-	-	-	-	-
Virology (f)	6	776	523	0.96 (0.94-0.98)	90.39 (84.51-94.19)	95.45 (86.42-98.58)	197.46 (76.65-508.67)	19.88 (6.57-60.13)	0.10 (0.06-0.16)	98 (96-99)	<0.001
					70.00-95.68^	61.70-98.00^					

'Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section V) NS1/IgM assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

(f) NS1/IgM/IgG assay

The results of subgroup analysis for IgG assay were summarised in Table 4.16 and Appendix G (VI). Acute had least scatter of circles and smallest prediction region, but I^2 remained high like all the other pools. Threshold effect was present in all analyses.

Due to the smallest number of studies involved, none had directly comparable pooled estimates within subgroup. Only pooled SN for acute was generally lower than that of the convalescent. And unpooled virology-based reference test gave NS1 assays better SN when compared to overall estimates. SN and SP of NS1 assay for primary and secondary infections were comparable. The same was true for commercial brand except that SP from the single study evaluating Panbio was insignificantly lower than that of the SD pooled estimate. But, again, in the presence of heterogeneity, none of the above results were definitive.

Table 4.16 Subgroup/sensitivity analyses for NS1/IgM/IgG assay

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve		Unpooled Range^	Unpooled Range^					
Overall NS1/IgM/IgG (a)	10	2770	1823	0.96 (0.94-0.98)	91.98 (87.88-94.78) 80.73-98.90^	89.54 (81.97-94.15) 57.45-100.00^	98.21 (45.32-212.85)	8.79 (4.99-15.50)	0.09 (0.06-0.14)	97 (96-99)	<0.001
By study design											
Case-control (unpooled)	3	263	30	0.86 (0.79-0.94)	92.80 (88.90-95.60)	80.00 (61.40-92.30)	-	-	-	-	-
		310	87	0.90 (0.87-0.94)	90.60 (86.80-93.60)	89.70 (81.30-95.20)	-	-	-	-	-
		166	120	0.94 (0.92-0.97)	94.60 (90.00-97.50)	94.20 (88.40-97.60)	-	-	-	-	-
Cross-section (b)	6	1287	390	0.97 (0.95-0.98)	92.28 (84.25-96.39) 80.73-98.90^	91.36 (75.99-97.24) 57.45-100.00^	126.33 (27.95-570.97)	10.68 (3.52-32.36)	0.09 (0.04-0.18)	98 (96-99)	<0.001
By previous exposure to dengue infection											
Primary (unpooled)	3	66	47	0.91 (0.86-0.96)	**83.30 (72.10-91.40)**	97.90 (88.70- 99.90)	-	-	-	-	-
		139	87	0.88 (0.84-0.93)	87.10 (80.30-92.10)	89.70 (81.30- 95.20)	-	-	-	-	-
		19	120	0.97 (0.95-0.99)	100.00 (82.40-100.00)	94.20 (88.40-97.60)	-	-	-	-	-
Secondary (unpooled)	3	176	47	0.91 (0.88-0.94)	84.10 (77.80-89.20)	97.90 (88.70-99.90)	-	-	-	-	-
		171	87	0.92 (0.88-0.95)	93.60 (88.80-96.70)	89.70 (81.30-95.20)	-	-	-	-	-
		83	120	*0.96 (0.93-0.99)*	***97.60 (91.60-99.70)***	94.20 (88.40-97.60)	-	-	-	-	-

(Continued in the following page)

Table 4.16, continued

Characteristic (HSROC graph)*	No of Index Test	Reference Pool Size		Area Under Curve (95% CI)	Sensitivity,% (95% CI)	Specificity,% (95% CI)	Diagnostic Odd Ratio (95% CI)	Likelihood Ratio + (95% CI)	Likelihood Ratio - (95% CI)	I^2 Statistic (95% CI)	p-value
		Dengue +ve	Dengue -ve								
By disease phase											
Acute (<=5 days) (c)	6	1081	847	0.92 (0.89-0.94)	**85.50 (77.37-91.04)**	85.47 (76.08-91.58)	34.68 (25.81-46.60)	5.89 (3.72-9.30)	0.17 (0.12-0.25)	98 (97-99)	<0.001
					75.41-95.95^	57.45-97.87^					
Convalescent (>=6 days) (unpooled)	3	247	568	0.91 (0.89-0.93)	93.90 (90.20-96.60)	87.10 (84.10-89.80)	-	-	-	-	-
		97	50	0.94 (0.90-0.98)	95.90 (89.80-98.90)	92.00 (80.80-97.80)	-	-	-	-	-
		70	25	0.95 (0.90-1.00)	***98.60 (92.30-100.00)***	92.00 (74.00-99.00)	-	-	-	-	-
By commercial brand											
SD (d)	9	2507	1793	0.96 (0.94-0.98)	91.95 (87.18-95.04)	90.41 (82.30-95.03)	107.71 (44.32-261.80)	9.59 (5.05-18.22)	0.09 (0.06-0.14)	97 (96-99)	<0.001
					80.73-98.90^	57.45-100.00^					
Panbio (single study)	1	263	30	0.86 (0.79-0.94)	92.80 (88.90-95.60)	80.00 (61.40-92.30)	-	-	-	-	-
By reference test											
Virology (unpooled)	3	370	141	0.77 (0.73-0.81)	95.90 (93.40-97.70)	57.40 (48.80-65.70)	-	-	-	-	-
		111	50	0.94 (0.90-0.98)	95.50 (89.80-98.50)	92.00 (80.80-97.80)	-	-	-	-	-
		145	120	0.94 (0.92-0.97)	94.50 (89.40-97.60)	94.20 (88.40-97.60)	-	-	-	-	-

Note: **Bold** font highlights inferior performance only. ***Bold*** and ***italic*** font highlights superior performance only.
*Alphabet in bracket provided for easy reference to corresponding HSROC graph in Appendix G section VI) NS1/IgM/IgG assay
^Lowest and highest point estimates of the respective parameter in the included individual studies

(g) Other subgroups within each assay type

Since none of the important subgroups mentioned above could fully explained the between-study heterogeneity, other subgroups such as publication year. study direction, study setting, location, selection method, age group, gender, type of test specimen, specimen sampling frequency, diseased and non-diseased severity, confidence level of dengue and non-dengue reference tests, presentation of study flow, source of funding, quality assessment outcome based on all 11 QUADAS-2 signalling questions, and risk of bias in all four QUADAS-2 domains; were also explored within each above-mentioned assay type.

However, still no single one of these subgroups could satisfactorily explain the heterogeneity found in the general pool (data not shown). Moreover, age and gender were also subjected to ecological fallacy due to the absence of individual-level data.

4.4 Discussion

4.4.1 Search and selection

This review identified as many as 18519 citations that dwarfed previously published four systematic reviews and meta-analyses on dengue diagnostics that found less than 2000 citations combined (Alagarasu et al., 2016; Costa et al., 2014; Shan et al., 2015; H. Zhang et al., 2014). The first reason was the use of five databases in the current review as compared to two to four used in the other studies. This could be seen from the high number and proportion of duplicates removed from this study. But the main reason was the very broad and comprehensive search terms used in this study that identified far too many irrelevant citations. However, such sensitive search ensured few to none relevant citations were missed. Nevertheless, these four previous reviews included only between nine to 30 articles due to their narrower objectives that focused only on two commercial brands of NS1 ELISA (Costa et al., 2014), NS1-based ELISA and RITs (H. Zhang et al., 2014), NS1-based ELISA and RITs in Asian populations (Shan et al., 2015), and rarely seen IgA-based tests (Alagarasu et al., 2016). They would have matched the number of articles or studies included in this review had their focuses been as broad.

With its objectives to review the diagnostic accuracy of dengue RIT without specifying the type of assay, the location of study, this review had to include all eligible citations. With some published articles presented the results of more than one study, and each study without fail contained the results of one or more standalone and combined assays in their overall as well stratified forms, the number of included results went from 69 articles to 72 studies to 169 included test results. Although this made the review more tedious and complicated to conduct and present, the effort was paid off with some findings that would likely remain undiscovered without the broad inclusion used here. These findings would be further elaborated and discussed below.

4.4.2 Changing trends in dengue RIT evaluation

This review identified some trends in the history of dengue RIT development and evaluation (Table 4.9). Over the years, they evolved slowly from mainly retrospective laboratory-based case-control studies to employ prospective care-based cross-sectional design, with the study size became increasingly bigger. The reference tests used also seen replacement of more complicated, time-consuming methods with those commercially available PCR, NS1, IgM, and IgG ELISA that have become increasingly more affordable and accurate. On top of that, there were also increasingly more evaluation studies on dengue RIT, especially after the introduction of NS1 dengue RIT. The focus of the studies was also shifted from just rapid diagnosis to include early diagnosis of acute dengue infection (Osorio, Ramirez, Bonelo, Villar, & Parra, 2010; Shrivastava et al., 2011; Wang & Sekaran, 2010a). It also was obvious that the market is dominated by several global names such as Panbio, Standard Diagnostics, and Biorad laboratories. It is likely that most of the small brands were no longer available in the market, with some only restricted to specific regions such as J. Mitra in India, Wondfo Biotech in China, and RVR in Malaysia (Chembio Diagnostics, 2016; J. Mitra, 2015; Wondfo Biotech, 2015).

4.4.3 Study quality and biases

The quality assessment of the included studies demonstrated two issues that plagued dengue RIT evaluation studies: study bias and incomplete reporting (Figure 4.4 & Figure 4.5). These studies were subjected to a multitude of biases. The most serious one was patient selection bias, where 42 studies (58.3%) had high risk of it. Most of it came from their case-control design and some from inappropriate exclusions of study subjects or samples. The second serious bias was flow and timing bias, where 23 studies (31.9%) had it, mainly from excluding the results of some patients or samples from the analysis for various reasons, followed by different reference tests for some patients or samples, and inappropriate tests interval. Although the conduct of tests, especially the reference tests were subjected to less bias, they were plagued by incomplete reporting in 58.3-66.7% of included studies. This left studies with low bias ranging from 40.3% in reference standard bias to only 9.7% in index tests bias. Fortunately, increasing awareness in the requirement for completeness in reporting as advocated by scientific community that developed the STARD checklist, and the formation of QUADAS and QUADAS-2 checklists for the quality assessment of these studies in systematic review and meta-analysis, more and more cross-sectional studies with higher quality that adhered to reporting guidelines were conducted (Figure 4.4, Figure 4.5, Table 4.8, Table 4.9 & Appendix E).

When it comes to publication bias, this review detected significant finding only for IgM/IgG assay (Figure 4.6). Logically, underrepresentation of studies with unfavorable outcomes was expected. However, in this review, studies that reported better assay performance were absent instead. Further analysis (data not shown) revealed that reporting of combined interpretation of IgM/IgG assay was at least 50% more commonly before the introduction of NS1, even though more studies were published after the introduction of NS1. The reason to this was perhaps prior to that, it was the only

combination available. The combination of NS1/IgM and/or NS1/IgM/IgG assay was more preferred later, when the focus shifted to early diagnosis, due to their higher SN (Andries et al., 2012; Fry et al., 2011).

On the other hand, even though the funnel plots for the other assays did not demonstrate asymmetry, it was more of a failure to show rather than actual absence of it (Appendix F). As mentioned above, this method would suffer from low power if there was heterogeneity in the DOR (Deeks et al., 2005; Macaskill et al., 2010). However, in the absence of better method currently, it was the only viable option. Still, the results should be interpreted with caution, and exploration of sources of heterogeneity should be carried out as it was in this review.

4.4.4 Appraisal of heterogeneity indicators and pooled accuracy estimates

For the assessment of heterogeneity, the four indicators used for its quantification in this review were the scatter of circles representing individual studies, the size of prediction region and its distance from the confidence region, the Higgin's I^2 statistics, and the p-value. The first two were given in the HSROC graphs and were recommended by the Cochrane collaboration. However, they were subjective in nature and were more difficult to interpret when the number of studies between subgroups were too different, as the presence of more circles cluttered the graphs and gave an impression of disorder that could be mistaken as scattered-ness, and smaller number of studies had huge prediction and confidence region that could point to a lack of precision more than heterogeneity. In other word, in analyses with smaller number of studies, it was the proportion of prediction region outside the confidence region that mattered for heterogeneity, not the size of the former. The same issue was also observed in the estimation of threshold effect between analyses, where the difference in proportion of circles that was located away from the HSROC curves was what researchers should look for, not the difference in number of circles. Simultaneously, the weight given to larger circles should also be taken into consideration. Therefore, although I^2 statistics and p-value were not recommended for the purpose of quantifying between-study heterogeneity in meta-analysis of diagnostic test accuracy, their objective nature still offered some value in it (P. Bossuyt et al., 2013; Macaskill et al., 2010). Nevertheless, they should be interpreted in light of the HSROC curves to compare between pools in subgroup or sensitivity analyses, and their numerical values should not be taken as an absolute to make inference on pool-ability of studies.

Among the parameters used in this review, it was noted that AUC generated by MIDAS was paradoxical in its precision. On one hand, it was less precise due to its lack of decimal. On the other hand, it was more precise compared to DOR with less degree of

variation in its 95% CI even in pools with few studies. And when compared to individual studies that were unpooled, pooled estimates appeared to be grossly overestimated. Therefore, it is not suitable to be used for comparison between pools or subgroups, but as an indication of test performance as suggested in this review. But given its known maximum value of 1 being the perfect test, and 0.5 being uninformative, as opposed to DOR that was less intuitive in interpretation, it's value in this review remained (Šimundić, 2009).

For the comparison of diagnostic test performance or accuracy between pools, DOR proved to be the best candidate. Although its precision was greatly dependent on the number of studies in the pool, making most analyses underpowered to detect the presence of effect, it's larger degree of change with relatively small changes in SN and SP made direction of bias detectable even in the absence of significant difference between analyses. However, this global measure of performance, together with AUC, could be identical with different combinations of SN and SP. Therefore, comparable DOR between two pools did not necessary indicate comparable underlying SN and SP. In fact, as demonstrated in this review, significant differences between SN or SP in different pools could be masked by overlapping DOR. On the other hand, when SN and SP in a pool differed in the same direction from that of the other pool, even insignificantly, could produce significant difference in DOR As such, it is a very powerful parameter in mete-analysis, even though in clinical practice it is of little relevance (Macaskill et al., 2010).

On the contrary, other accuracy estimates used in this review such as SN, SP, LR+, and LR- were clinically meaningful and relevant (Table 4.6). First of all, as demonstrated here and discussed in detail later, SN and SP were inversely correlated, pointing again to threshold effect. Secondly, SN and LR- would change for the better or worse simultaneously, with good value indicating exclusion of diagnosis if test was negative.

Here, it is also necessary to repeat again that, test with lower value of LR- is of higher value. On the other hand, SP and LR+ were the other pair that would change in the same direction at the same time, with high value indicating presence of disease of interest if test was positive. And last but not least, LR+ and LR- were better parameters for this meta-analysis compared to predictive values as they were not affected by the varying prevalence across studies. However, it should be reinforced again that all pooled accuracy estimates, as well as DOR, and the HSROC curves, were merely an observed average of all the included studies. They were not the true values or common effects of the included studies due to the ever-present between-study heterogeneity and cannot and should not tempt to be interpreted as such (JBI, 2015; Leeflang, 2014; Macaskill et al., 2010; Šimundić, 2009).

Before engaging in the discussion on the pooled accuracy of dengue RITs, one last observation from its process was worth mentioning. In the process of meta-analysis, during the pooling of results in different groups, it was mentioned above that MIDAS and METANDI occasionally failed to pool. This situation tended to happen in the presence of large number of study or very low number of studies. Whatever the situation, it was due to the presence of one or more outliers that threatened the stability of the model. In other word, failure to pool could provide a hint that the involved studies were too heterogeneous to begin with. However, this was not a reliable indicator as the ability to pool varied if STATA was allowed to analyse on random seed. In this review, although an identical random seed was set for all analyses, this indicator may still not be as reliable compared to other established parameters mentioned above due to the fact that ability to pool did not indicate absence of heterogeneity but merely stability of the model. Nevertheless, failure to pool could point to the presence of extreme outliers and relatively more heterogeneity in a given pool compared to another.

4.4.5 Results of pooling and exploration of heterogeneity

This review confirmed that RIT evaluation studies were characterized with heterogeneity. Almost all analyses were afflicted with high level of between-study heterogeneity, with analyses including higher number of studies suffering more from it. Apart from the difference in study characteristics, the threshold effect or the correlation between sensitivity and specificity, which was present in almost all analyses even of the same brand, was the main contributing factor in the study heterogeneity. It was also obvious that the variation was generally greater in sensitivity, instead of specificity that was usually high, indicating that the changes in positivity threshold affected the detection of diseased more than the exclusion of non-diseased. It also provided evidence of the above-mentioned low sensitivity issue common for dengue RIT (Miller & Sikes, 2015).

The most important implication for this review due to the persistence of between-study heterogeneity and inability to satisfactorily explain it in the exploration of heterogeneity is that most of the pooled accuracy estimates cannot be taken as the true or actual accuracy estimates. In other words, they cannot be used to reach definitive conclusions. Therefore, the objectives of this review may not be fully answered as a result. In addition, between-study heterogeneity has implication beyond this review. Variations between different evaluation studies imply that no two studies are comparable, unless they are identical. A conclusion that a diagnostic kit performs better than another is flawed, if it was based on the results of two different studies that evaluated two kits separately. The performance of two diagnostic kits can only be fairly compared if they were evaluated in the same study, on the same population, against the same reference standard, following the same study flow (Stuart D. Blacksell, 2012; Leeflang, 2014).

Even though the meta-analyses could not fully explain the heterogeneity because of its univariable nature, some valid observations could still be derived due to the fact that every

single diagnostic outcome was actually dependent on an individual patient underlying dengue biomarkers composition at the point of sample collection and their detection (Figure 2.2 & Figure 2.3). And the sum of all these individual diagnostic outcomes in a study would fill its 2X2 table, which would in turn determine the accuracy of an index test in relation to the reference standard. So, the effect seen in any analysis came primarily from the factors that could modify the composition of diagnostic markers in the patient or sample pool, i.e. previous exposure to dengue and disease phase; and then from factors that could modify the choice of reference standard used in relation to the index test. Therefore, in whatever analysis, the effect seen between subgroups of the characteristic of interest was eventually dependent on how they influenced the above factors differently, i.e. their impact on selection and measurement biases, and not their direct effect. In other words, although the meta-analyses could not provide the actual accuracy estimates with heterogeneity fully accounted for, it could still show us study characteristics that might modify accuracy estimates, with the help of scientific reasoning. These observations may not be definitive, but could provide us with adequate confidence especially if similar trend is observed across different analyses. They would provide guidance in the academic discourse that follows and the design of primary study,

So, from the pooling of the overall diagnostic accuracy according to type of assay, it was found that antigen-based tests performed better than serology tests especially IgM assay (Table 4.10). This finding was expected as serological tests for dengue tend to cross-react with other flaviviruses, leading to lower SP (Peeling et al., 2010; Shamala, 2015). On top of that, anti-dengue IgM could remain in human body for up to 90 days, while IgG – for life, meaning they are detectable in healthy individuals (World Health Organization, 2009a). Attempts to maintain useful SP for the diagnosis of acute dengue required raising the detection threshold higher for serological assays. In fact, for IgG assay, it could be set as high as the level of antibodies characteristic of secondary infection,

or equivalent to HI titre of >=1:2560 (Sang, Hoon, Cuzzubbo, & Devine, 1998). However, higher threshold would harm SN as observed in the analysis. On the contrary, dengue NS1 antigen is more specific and cross-reacts lesser (Hunsperger et al., 2014). However, it's higher average SN compared to serological ones may be due to recruitment of overwhelmingly more samples from acute phase than convalescent, in line with its focus for early diagnosis, contrary to more comparable pools seen for serology assays (Table 4.11-Table 4.14).

Likewise, similar observation was made from the insignificantly higher SN seen in IgG assay as compared to IgM (Table 4.10). Although theoretically speaking, the fact that IgG being a light-weight monomer can travel faster and reach the binding antigen first, allowing for its detection in test like HI better than a heavy pentamer like IgM that moves slower, can perhaps be used to relate to this finding here (Schroeder & Cavacini, 2010; World Health Organization, 1997, 2009a). The actual reason for higher SN for IgG assay seen in this review was due to its higher proportion of case-control studies that usually produced better accuracy estimates, which would be further discussed below. Here, it could be reasoned from IgG concurrently higher SN and SP compared to IgM, as opposed to the usual inverse relation between SN and SP. Although underpowered especially for SP, it remained valid and would be elaborated later.

Last but not least, the overall pooling according to assay also inform us that combined assays had better SN with corresponding decrease in SP compared to individual assays (Table 4.10). This was also expected as the window period for NS1 detection is up to around a week, while IgM can only be detected from day 5 and IgG from day 7, or earlier in secondary infection; combining assays increased test ability to detect disease at different timing. And this finding has to be valid as corresponding decrease in SP was present. However, the greater degree of change observed in the combination of purely

serological assays versus the lesser extent of change seen in the addition of serology to NS1 assay was due to the use of predominantly acute phase samples or patients in studies involving the latter, during which serological assays were not as useful (Table 4.15 & Table 4.16).

From the pooling by study design, it was observed that case-control studies always had bigger diseased pool and smaller non-diseased pool (Table 4.11-Table 4.16). This phenomenon was reversed for cross-sectional studies except in IgG, NS1/IgM and NS1/IgM/IgG assays. The reason was that in a study with case-control design, a panel of archived samples was usually compiled and tested retrospectively in a laboratory. Diseased and non-diseased pool comparable in size was necessary to ensure study efficiency (World Health Organization, 2009b). Whereas in cross-sectional studies that usually used consecutive sampling of all patients that came to a real clinical setting with dengue-like symptoms, more non-diseased patients would be recruited (Carter et al., 2015; Cohen et al., 2007; Simonnet et al., 2017). In view of this, the selection of samples in case-control studies was predominantly biased, which probably explained why case-control studies generally produced better test performance compared to cross-sectional studies (Groen, Koraka, Velzing, Copra, & Osterhaus, 2000; Kohn et al., 2013). Although the analysis for IgM/IgG was underpowered, and significant difference was detectable only for IgM, the fact that similar trend was observed across two assays and plausible explanations exist for the other assays that did not demonstrate the same trend still made this finding a legitimate one.

For NS1, IgG, NS1/IgM and all assays combined, there was no difference between study designs observed. However, this was likely an exclusion with explanation as test performance such as for NS1 assay even of the same brand could be higher in case-control studies compared to cross-sectional ones (Andries et al., 2012; Ferraz et al., 2013;

Najioullah et al., 2011; Zainah et al., 2009). The reason selection bias was masked in NS1 assay hid in the subgroup analyses by commercial brand for IgM and NS1 assays (Table 4.11 & Table 4.14). It can be seen there that SD Bioline/Bioeasy NS1 and IgM kits were evaluated mostly by cross-sectional studies as evident from the substantially larger non-diseased pool size, as opposed to all other kits that were evaluated mostly by case-control design. Despite that, SD NS1 assay performance was on par with all the rest, but its's serological assays especially that of IgM performed far worse than the other kits. This differential performance could aggravate the effect of study design for IgM, but would neutralise the same effect that was supposed to also be apparent in NS1 assay. For IgG, NS1/IgM, and NS1/IgM/IgG assays, the masking came from the supposedly less bias in term of selection cross-sectional studies that had smaller non-diseased pool just like its case-control counterparts, pointing to potentially comparable selection bias between both designs for these assays. On top of that, these assays also had smaller number of studies for pooling leading to underpowered analyses. Still, some hints on the upward direction of bias in case-control design could still be seen for IgG and NS1/IgM assays (Table 4.12 & Table 4.14).

Pooling by serotype carried out only for IgM and NS1 assays discovered that they were less sensitive in detection dengue infection caused by DENV-4 than by other serotypes (Table 4.11 & Table 4.14). Here, although the non-diseased pools for most analyses were larger, it did not mean that they were cross-sectional studies, as the same non-diseased samples in one or another individual study presenting data by serotype would contribute to the pooling of SP for all serotypes instead of being used just for one of the serotype. In other word, non-diseased pool was common to all serotypes and appeared in all analyses, inflating the size of the non-diseased pool. So, differential distribution of study design could not be used to explain lower SN found for DENV-4. In fact, all studies that

contributed to DENV-4 pooling for IgM were case-control, and only one was cross-section among those for NS1 (analysis not shown).

The fact that the same observation of lower SN for DENV-4 was seen in both NS1 and IgM assays, and similar finding across other reviews made this finding possibly true (Costa et al., 2014; Shan et al., 2015; H. Zhang et al., 2014). Although these meta-analyses were conducted primarily on NS1 ELISA, it could still be applicable to NS1 dengue RITs that were built to detect the same biomarker. This may be due to the possibilities of difference in the level of secreted NS1 between serotypes, presence of polymorphism in the DENV-4 NS1 gene, and higher proportion of secondary dengue caused by DENV-4 (Costa et al., 2014; Hunsperger et al., 2016; Shan et al., 2015). However, additional studies are needed to prove the above theories, and only the last could explained the same phenomenon in IgM assay. For that, a revisit of the primary study confirmed that all eight different brands of dengue IgM RIT undeniably had substantially lower SN for DENV-4 except one (Stuart D. Blacksell et al., 2006). The only other plausible explanations could be that the cocktail of dengue antigen used to manufacture most of the dengue IgM RITs in the primary study above for the detection of this antibody was underrepresented for DENV-4, or that the antigen-antibody binding affinity was lower for DENV-4 compared to that of the other serotypes (Cuzzubbo et al., 2001; Fry et al., 2011; Hunsperger et al., 2016; Jihoo Lee, Kim, Chong, & Song, 2015; Osorio et al., 2010; Shamala, 2015).

While study design influenced test performance of all assays in the same way, previous exposure to dengue and disease phase affected serology and antigen-based assays in different ways, with effect neutralised or masked in combined assays. Moreover, the effect was only found in SN, and SP between subgroups in all assays were rather comparable. For previous exposure to dengue, lesser variability in SP could also be due

to the same reason of common non-diseased pools explained above for serotype. Anyway, the difference in effect for serology and antigen-based tests were expected due to the kinetics of dengue NS1, IgM, and IgG in human body that differ across time and between numbers of previous infection (Figure 2.2 & Figure 2.3). Therefore, they were the actual determinants of test performance of all assays, and other study characteristics were generally simply effect modifiers. Last but not least, it is important to note that, although the effect was generally nullified for both the above patient characteristics for combined assays with NS1 component, i.e. NS1/IgM and NS1/IgM/IgG; their SN and SP, especially the SN, were generally higher compared to other individual assays. Despite the difference between assays was underpowered to detect, this high SN had important clinical implication. Because, in clinical practice, it is more important for a kit to make a dengue diagnosis as early as possible and not to miss a case. In other word, ability to diagnose in both primary and secondary dengue in any day of presentation was more crucial compared to ability to differentiate between them (Andries et al., 2012; Fry et al., 2011).

In the pooling by previous exposure to dengue, IgM assay was found to do better in the detection of primary dengue, IgG – secondary, and the effect was nullified when both assays were combined (Table 4.11-Table 4.13). NS1 was more sensitive for the detection of primary dengue, and remained so with the addition of IgM assay with the same effect, but did not differ with the further addition of IgG (Table 4.14-Table 4.16). As mentioned above in section 2.3, a dengue infection would start with viraemia when antigen could be detected in the human host, followed with immunological response in the secretion of antibodies, first IgM, then IgG. In primary infection, the lead in IgM secretion over that of IgG is larger in time and concentration. In secondary infection, however, the lead in secretion time for IgM is minimal. Instead, it trails IgG in terms of concentration. This made IgM assay better in the detection of primary dengue, and IgG – secondary. The observation that NS1 assay performed worse in secondary infection was not due to the

absence of this antigen, but merely its complexing with neutralising antibodies and the inability of NS1 assay to detect complexed antigen (Stuart D. Blacksell, 2012).

On the other hand, disease phase influenced test performance in a slightly different manner compared to previous exposure (Table 4.11-Table 4.16). Here, the SN for both serology assays tended to be higher in infection more than five days from onset; while NS1 assay was better in detecting infection up to five days from onset. Addition of IgM and IgG to NS1 assay seemed to nullify the effect with a hint of upward bias towards the SN for the detection of dengue in convalescence that was too underpowered to prove. It is not difficult to explain this finding given NS1 formation in the early phase of disease followed by antibodies secretion that neutralise the former (Figure 2.2 & Figure 2.3) (Stuart D. Blacksell, 2012).

Next, pooling by commercial brand produced some interesting findings, some of which were briefly mentioned above. For IgM assays, Panbio IC had the best performance mainly due to higher SN, followed by Panbio Dengue Duo, while SD and all the other brands had it worse. The similar trend was also observed in IgM/IgG pools but not in IgG due to the preference of primary studies to report SD IgG results alone and results for both Panbio kits in the form of IgM/IgG (Table 4.11-Table 4.13). However, it could not be concluded that Panbio IC was the best because of the underlying difference in other study characteristics such as study design. SD worse performance was most probably due to its predominantly cross-sectional design. The fact that Panbio IC with more impressive performance was replaced by its second generation counterpart in the market casted serious doubt in the comparability of the published results of this first generation kit (Charrel & de Lamballerie, 2002). Now, although it could not be assessed if SD or Panbio Dengue Duo serology assays were better, all the other brands did genuinely perform much worse in spite of their predominantly case-control design.

On the other hand, for NS1 assay, SD had comparable accuracy estimates with all the rest of commercial brands despite its predominantly cross-sectional design, pointing to a genuinely better performance. Panbio Dengue Early performed worse despite its majority case-control design and could be genuinely lacking in the accuracy of this standalone NS1 assay. Biorad NS1 STRIP and the others' better performance may be also largely due to their primarily case-control design. The combination of NS1/IgM and NS1/IgM/IgG assays for both SD and Panbio was comparable (Table 4.14-Table 4.16). As such, pooling by commercial brand could have inform about the best performing one in the market currently. But due to the modifying effect of other underlying study characteristics, it may not be appropriate to jump to a definitive conclusion for most assay other than NS1, where SD was likely better specifically due to its higher SN. In addition, this exercise also revealed the most and better evaluated brand – SD Bioline/Bioeasy Dengue Duo.

Up till now, the discussion focused on study characteristics that could modify the levels of diagnostic biomarkers in patients, or that contributed to selection bias. Reference tests were the only study characteristic that could modify the outcome due to measurement bias. Here, evaluating serology assays using only serology tests as reference almost always produced better SN, at the expense SP (Table 4.11-Table 4.13). Similarly, using virology tests only as reference yielded better SN for NS1 assay (Table 4.14). The effect was neutralised when NS1 was combined with IgM assay (Table 4.15). No comparable subgroup was available for all assays combined (Table 4.16). These results were expected as virology and serology tests detect biomarkers that appear at different time after the onset of illness (Table 2.1, Figure 2.2 & Figure 2.3). Evaluating an index test similar to the reference would result in higher performance, and vice versa. On the other hand, individual assays with a subgroup using only clinical diagnosis as reference produced either comparable or higher SN at the expense of SP, hinting a potential presence of incorporation bias, where a positive index test on a patient played a

role in classifying this very patient as a diseased in the reference, i.e. index tests was incorporated into the reference. This happened when clinicians were not blinded to the index test results when ascertaining diseased and non-diseased (Kohn et al., 2013).

The influence of each study characteristic towards diagnostic accuracy, in particular SN, as discussed above, is summarised in Table 4.17.

Table 4.17 Diagnostic accuracy of different individual assays and combinations stratified by study characteristics

Assay	Better accuracy in					
	Design	Serotype	Exposure	Phase	Reference	Brand
IgM	**Case-control**	All poor especially DENV-4	Primary	Convalescent	**Serology**	**Panbio IC**
IgG	**Case-control**	-	**Secondary**	*Convalescent*	**Clinical**	Incon-clusive
IgM/ IgG	**Case-control**	-	Both	**Convalescent**	Serology	**Panbio IC**
NS1	Both case-control & cross-sectional	**All except DENV-4**	**Primary**	**Acute**	**Virology & clinical**	Incon-clusive
NS1/ IgM		-	**Primary**	Both	Both serology & virology	Incon-clusive (SD or Panbio)
NS1/ IgM/ IgG		-	Both	Both	-	

Note: **Bold** font indicates at least one accuracy estimate was significantly different within the subgroup. *Italic* font indicates at least one accuracy estimate within the subgroup was close to demonstrating statistically significant difference.

4.4.6 Study strengths and limitations

The strength of this systematic review and meta-analysis was its broad objective that include both serology and antigen-based assays and their combinations, which led to an abundance of observations from their multidimensional comparison. On top of that, it amassed a huge number of studies for pooling to a larger sample size. Third, the statistical methods used were appropriate and rigorous. Nevertheless, there was a lack of tool to adequately address the publication bias. And also due to the lack of a tool for multivariable analysis and problems with small cell size beyond two levels of stratified analysis, the exploration of heterogeneity could only be univariable up to second level, i.e. on different study or sample characteristics within individual assays. However, it's sheer volume and allowance of multidimensional comparison, coupled with scientific reasoning and comparison to other studies and reviews, made possible to produce plausible and scientifically valid observations with implication for clinical practice and research.

For clinical practice, this review informed that the actual performance of dengue RITs would always be lower in real life compared to the published figures. The closest to reality diagnostic performance dengue RIT could be found in the pooled estimates of cross-sectional studies, which had less bias in the selection of patient characteristics such as previous exposure and disease phase. However, the important thing should not be classifying patients into this or that categories, but to make a dengue diagnosis as early as possible regardless of the situation. For that, a positive test to any of NS1, IgM, or IgG assay in a combo kit should raise doctor's suspicion and the probability of dengue diagnosis (Andries et al., 2012; Fry et al., 2011). But, since there exists no perfect test, a negative test cannot totally exclude dengue, and should be handled on case-by-case basis. Although there may be false positives, in the absence of alternative diagnosis that is more life-threatening, doctor may choose to err for the benefit of the patient, since dengue

management, even if not beneficial, is not more harmful. On the reverse, discharging an actual dengue patient that was mistaken as FN may put the patient at risk. Lastly, it is also important that the accuracy estimates generated in this review cannot be used as the true estimates, especially in relation to commercial brand. Therefore, it was intended that the more clinically relevant LR+ and LR- estimates were not over-interpreted in terms of their ability to increase and decrease post-test probability of disease. All accuracy estimates were but the average of the same in the underlying primary studies. Generalisation of the results of any primary study to any clinical setting should be made with careful consideration of their similarities and differences, which bring us into the next point that is more relevant to this thesis - the implication for research.

In view of the limitation of generalisability of the results of this meta-analysis to other clinical settings, there is a necessity to evaluate the actual performance of any dengue RIT or RDT before its application. Whenever possible, this evaluation study should be cross-sectional, prospective, and care-based, to minimise selection bias and to ensure capture of more patients with diverse characteristic such as primary and secondary, early or late presentation, and so on. Case-control, retrospective, laboratory-based evaluation study on archived samples has its value as pre-test of the dengue kit of interest (Peeling et al., 2010). And regardless of the design, the limitation in generalisation of evaluation study result made it almost mandatory to have other comparator diagnostic kit with known diagnostic performance such as those identified in this review. In that case, based on the outcome of the current study, the relative performance of any new dengue RDT can be estimated in all primary studies that have previously evaluated the comparator kit.

4.5 Conclusion

The conclusion of this review was that, due to the presence and persistence of heterogeneity, there was not a true diagnostic performance for any dengue RIT that was applicable to all situations. The pooled accuracy estimates were merely average corresponding estimates of the underlying primary studies, which were very different from each other and generally biased.

Nevertheless, the objectives of this review were more or less answered. For the first objective, that the performance of combined assays with NS1 component was probably better compared to all the others, and should be preferred in clinical practice since it would produce least FN. Secondly, type of assay, study design, dengue serotype, patients' previous exposure to dengue and disease phase, commercial brand, and type of reference test, appeared to be the main sources of heterogeneity, although none of them could fully explain it away. Lastly, for the third objective, this review identified SD Bioline Dengue Duo as the most evaluated dengue RIT with good performance.

The absence of true values implied that the pooled performance estimates cannot be blindly generalized to another clinical setting, even if estimates from cross-sectional studies were more applicable to real life situation. For that, estimates from primary cross-sectional evaluation studies with similar background to the setting of interest may be a better reference. However, the most scientifically valid estimates can only be produced by directly evaluating new dengue test in the setting it will be applied to. Moreover, this evaluation study has to be a prospective cross-sectional phase III study with consecutive sampling independent of disease status to minimise selection bias; and composite reference standard to minimise measurement bias. Case-control can be used for pre-test purposes or as a phase II study. Finally, another established dengue diagnostic test should be evaluated together with the new test for direct comparison of diagnostic performance.

CHAPTER 5: CLINIC-BASED EVALUATION OF A NEW DENGUE RAPID DIAGNOSTIC TEST WITH A COMMERCIALLY AVAILABLE DENGUE RAPID IMMUNOCHROMATOGRAPHIC TEST: A PHASE III STUDY

5.1 Introduction

Dengue is an emerging infectious disease endemic to more than 100 tropical countries (World Health Organization, 2012). In Malaysia, dengue incidence stood as high as 396.4 cases/100,000 population with case fatality rate ranging from 0.2-0.28% in recent years. In 2015 alone, dengue infected up to 120,000 people and caused 336 deaths (idengue, 2020). Despite absence of medical treatment to date, early disease recognition and timely intervention with proper fluid management and supportive care can prevent mortality due to dengue (Ministry of Health, 2015).

The obstacle to early dengue diagnosis lies in its unspecific clinical symptoms that resemble other diseases, which leads to delay in health-seeking and misdiagnosis. Laboratory tests such as VI, PCR, HI, and ELISA for the detection of dengue NS1 antigens and antibodies, can help in diagnosis but are time-consuming (Peeling et al., 2010). This led to growing demand for point-of-care diagnostics for early diagnosis of dengue gave rise to many dengue RDT that flooded the market in the past two decades, majority of which was RIT.

However, their performance varied widely especially for sensitivity (S. D. Blacksell et al., 2011; Hunsperger et al., 2014; Shamala, 2015). One main reason for this among others is their interpretation that is qualitative in nature. Most RIT manufacturers treat the appearance of any faint line at the test region of a valid test as positive. This interpretation

is subjective and vague lines may not be detected by naked eye, thus reducing the sensitivity (Stuart D. Blacksell, 2012; Miller & Sikes, 2015).

New development in biosensors that can quantify and amplify the immunological reaction between test reagent and target analyte into objective interpretable result may provide a solution to the dilemma of having to choose either rapidity or accuracy in the early diagnosis of dengue infection (B. Zhang, Salieb-Beugelaar, Nigo, Weidmann, & Hunziker, 2015). This study aimed to evaluate the diagnostic accuracy and diagnostic utility of a new dengue RDT for the diagnosis of acute dengue infection in a primary care setting in Malaysia, and compare them with that of a commercially available dengue RIT.

In this chapter, Section 5.2 describes the methodology. Section 5.3 presents the results. Section 5.4 discusses the results and Section 5.5 concludes the chapter.

5.2 Methods

5.2.1 Study design

This phase III diagnostic evaluation study employed a prospective cross-sectional study design. Research participants were patients with suspected dengue presented to a public clinic. Blood samples were drawn upon presentation and tested for acute dengue using two dengue RDTs. Excess blood samples were subsequently sent to a virology laboratory for acute dengue case ascertainment using gold standard reference tests.

5.2.2 Ethical statement and reporting standard

This study adhered to the principles of the revised 2013 Declaration of Helsinki (WMA, 2013) and obtained ethical approval from University Malaya Medical Centre (MRECID.NO: 2017426-5171) (Appendix H) and National Institute of Health, Malaysia (NMRR-17-853-34393) (Appendix I). The reporting of this study followed STARD guidelines to ensure completeness in reporting (P. M. Bossuyt et al., 2015).

5.2.3 Population

The study site was Shah Alam Section 7 Health Clinic (KKS7), a public clinic with the highest number of dengue patients located in the dengue-endemic district of Petaling, Selangor state, Malaysia. Petaling district, in turn, was the district with heaviest burden of dengue infection in Malaysia (Gill, 2012; idengue, 2020). The study population was all patients with suspected dengue that visited KKS7 for consultation. It was also the sampling frame from which study participants were selected.

The inclusion criteria were febrile patients aged 9 months and above with symptoms fulfilling WHO 2009 criteria for suspected dengue (Ministry of Health, 2015; van Panhuis

et al., 2011; World Health Organization, 2009a). Patients in need of emergency care or with pre-existing conditions that were prone to complications from blood sampling were excluded. The detailed versions of inclusion and exclusion criteria were summarised in Table 5.1.

Table 5.1 Study inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
1) 9 months old and above, AND 2) Fever, AND 3) Two or more of the following signs: a) aches and pains (retroorbital eye pain OR headache OR generalised body ache OR arthralgia OR myalgia), AND/OR b) rash, AND/OR c) nausea OR vomiting OR anorexia, AND/OR d) leukopenia, AND/OR e) any dengue fever warning signs: i) abdominal pain OR tenderness; OR ii) persistent vomiting (>= 3 times per day); OR iii) persistent diarrhoea (>= 3 times per day); OR iv) clinical fluid accumulation; OR v) mucosal bleed; OR vi) lethargy OR confusion OR restlessness; OR vii) tender liver; OR	1) dengue shock syndrome, OR 2) respiratory distress, OR 3) severe bleeding, OR 4) any blood coagulation disorder, OR 5) taking blood thinning medication for medical reason; 6) immunocompromised; 7) any condition that is contraindicated for blood sampling.

The sample size required for this study was calculated using the formula for sample size estimation in a matched-groups diagnostic study, where two diagnostic methods were tested on the same subjects (Beam, 1992):

$$n = \frac{[SLF * W^{0.5} + PF * (W - d^2)^{0.5}]^2}{d^2}$$

where, W = probability of disagreement between the two diagnostic tests;

d = difference between SN (or SP) of these two diagnostic tests;

SLF = significant level factor (=1.645 for a one-tailed test with 95% confidence) PF = power factor (=0.84 for power of 80%).

The number of subjects needed were calculated separately for diseased and non-diseased groups using the SN and SP of two diagnostic tests, respectively. For each of them, the highest probability of disagreement between the SN (and SP) of these two tests was first calculated as:

lower SN/SP * 0.05 + 0.20 * higher SN/SP.

According to the systematic review, the expected SN and SP for a NS1/IgM/IgG RIT assay were around 92.0% and 90.0% (Table 4.16). If the new RDT were able to achieve the lower bound of its 95%CI for SN (80.8%) and SP (83.5%) based on the pre-test performed prior to this study as described below, 113 dengue positive and 379 dengue negative patients were required to detect the effect between its performance and that of the RIT with 95% confidence level and 80% power. In total, 492 patients would need to be recruited for this study with an expected prevalence or probability of dengue at 23.0%. These participants were enrolled using universal consecutive sampling method during the study period from 13th November 2017 to 30th March 2018 between 8 am to 5 pm, which were the working hours of KKS7.

5.2.4 Data collection

In KKS7, patients would undergo vital signs screening immediately after registration. Due to high dengue burden in the area, patients with recorded temperature of ≥38˚c or those who complained of having fever for three days and above were routinely sent for FBC by nurses at the screening counter. These patients were identified by the research team member prior to blood taking and were screened for selection criteria (Table 5.1). Those who do not fulfil the criteria were released back to the clinic's original patient flow, i.e. blood taking by clinic staff. Those who fulfilled the selection criteria and consented to participate in the study were recruited. Blood specimens for this research and routine FBC required by clinic would be drawn from them only by researchers to avoid double poking. Written informed consent was obtained from each participant. For participants below 18 years old, minor assent was obtained where possible on top of consent from guardian.

The socio-demographic background and clinical history of the participants were captured using content- and face-validated structured questionnaire through face-to-face interview by a medically trained research team member (Appendix J). Diagnosis of acute dengue was made in-situ using two dengue RDTs under evaluation, termed index tests, to compare with the reference standard. Both capillary and venous blood samples were collected immediately from each patient using EDTA tubes to run on the index tests. The new RDT was also tested with serum extracted from another venous sample collected in plain tube, that was chilled and meant to be sent to the laboratory at the end of each day for reference tests that constituted the reference standard. Blood sampling was performed by experienced medical personnel. The data collection was only conducted once for each patient. No follow-up blood sample was drawn. The conduct of the index tests and reference tests were as described below. Results were not release to KKS7 doctors as they

have existing RIT at their disposal and manage patients according to their existing flow and protocol.

5.2.4.1 Index tests

A new dengue RDT – the ViroTrack Dengue Acute (BluSense Diagnostics, Denmark), and a commercially available dengue RDT/RIT - the SD Bioline Dengue Duo NS1 Ag + IgG/IgM (Standard Diagnostics, Korea), were evaluated in this study for comparison. The latter was selected to be the comparator as it was extensively evaluated previously and found to have good diagnostic performance. As for ViroTrack, a pre-test was performed from June to August 2017 in a virology laboratory of Department of Medical Microbiology, Faculty of Medicine, University of Malaya; using 91 archived serum or plasma samples consisted of 50 dengue cases and 41 controls previously collected in a prospective study from patients aged 14 and above admitted to two tertiary public hospitals in Malaysia with suspected dengue infection from June 2010 to April 2011 (Anusyah Rathakrishnan et al., 2014). The SN and SP of ViroTrack Dengue Acute in this pre-test were demonstrated to be 92.0% (95%CI 80.8-97.8) and 95.1% (95%CI 83.5-99.4), respectively.

Both ViroTrack Dengue Acute and SD Bioline Dengue Duo were intended for point-of-care use. They were performed in KKS7 upon patient recruitment using fresh samples. Both were tested on both capillary and venous blood samples by a medically trained research team member. Additionally, ViroTrack Acute Dengue was also tested with serum sample extracted from the plain tube.

- *ViroTrack Dengue Acute*

ViroTrack Dengue Acute is a biosensor-based semi-quantitative immune-magnetic agglutination assay packed in a polymer centrifugal microfluidic cartridge. Its diagnostic mechanism was detailed out previously (Antunes et al., 2015). Briefly, for each test, a ViroTrack microfluidic loaded with 30 mcl of blood sample was inserted into a portable opto-magnetic reader – the BluBox. The sample was centrifuged, metered, and mixed with magnetic nanoparticles (MNPs) pre-coated with anti-dengue antibodies located within the cartridge. Dengue NS1 protein, if present, formed sandwich agglutination with these MNPs and were forced to rotate under an oscillating magnetic field, which modulated the intensity of a laser beam passing through them. A photodetector with a Blu-ray optical pickup unit would then measure the phase difference between the modulated light transmission and the applied field, which corresponded to the level of dengue NS1 protein. This measurement was presented in a relative unit and interpreted by the BluBox according to a pre-set threshold value, where positive was defined as >=27, negative if <23, and equivocal (EQ) if 23-26.9 unit. The whole process after the insertion of the microfluidic was automatic and the result was ready in less than 15 minutes. The results of ViroTrack Dengue Acute were recorded by one research assistant and verified by three others independently. For analysis, a patient was considered tested positive for ViroTrack Dengue Acute if either capillary or venous sample was positive, EQ if both were EQ, and negative for all other combinations. A patient tested positive on ViroTrack was considered to have acute dengue infection.

- *SD Bioline Dengue Duo*

This is a commercially available dengue RIT used widely in Malaysia. It comes in a combo of two joint cassettes, one for NS1 and another for IgM/IgG. Only 100 mcl blood sample was needed for NS1 assay, while serology required 10 mcl followed by assay diluent. Around 15-20 minutes after the application of the specimen to the cassettes, the results were interpreted according to manufacturer's instruction, where the appearance of a test line was considered positive in the presence of a control line. Presence of only control line was considered negative. Since the interpretation was subjective in nature, two research assistants blinded to the clinical information of the patient performed the read-out independently. Discrepancies between them were resolved with the help of a third interpreter. For analysis, a patient was considered tested positive to an assay on SD Bioline Dengue Duo if either capillary or venous sample was found positive, and negative if both were negative. For the analysis, a patient tested positive to any assay of SD Bioline Dengue Duo, whether NS1, IgM, or IgG, was considered an acute dengue case. However, additional analyses for individual components and their other combination of this index test, as well as analysis with additional unmatched-to-ViroTrack samples for all assays combined, were presented in Appendix K.

5.2.4.2 Reference standard

The reference tests that constituted the reference standard were validated and commercially available iTaq Universal SYBR Green One-Step real-time RT-PCR (Bio-Rad Laboratories, Hercules, CA), Panbio Dengue Early ELISA, and SD Dengue IgM and IgG capture ELISA (Standard Diagnostics, Korea). They were performed and interpreted according to the manufacturers' instructions in the same virology laboratory mentioned above by trained laboratory personnel blinded to the clinical information and results of the point-of-care index tests. Specimens collected were tested on reference tests not later than one month from the date of collection. The procedures of these reference tests were more complicated and are freely available for reference in previous publications (Bessoff, Delorey, Sun, & Hunsperger, 2008; Shu et al., 2003; World Health Organization, 2009b; Yong, Thayan, Chong, Tan, & Sekaran, 2007).

Laboratory-confirmed and presumptive dengue were defined according the criteria used for the study that contributed to the archived samples used in the pre-test, with reference to WHO 2009 guidelines (Table 2.2). A laboratory-confirmed dengue was defined as 1) RT-PCR positive, or 2) Panbio NS1 ELISA positive; while a presumptive dengue tested negative for both the above but positive for IgM ELISA (Anusyah Rathakrishnan et al., 2014). Both laboratory-confirmed and presumptive dengue were included in the analysis as dengue positive. On top of that, a combination of "Recife" method and IgM/IgG ratio from ELISA was used to classify these dengue positive patients into primary and secondary dengue, whereby primary was defined as IgG negative with positive on either IgM, NS1 or PCR; while secondary – positive IgG with negative IgM and positive on either PCR or NS1. If both IgM and IgG were present, IgM/IgG ratio >=1.2 was a primary dengue, while <1.2 was considered secondary (Cordeiro, Braga-Neto, Nogueira, & Marques, 2009; Shu et al., 2003; Vickers et al., 2015; World Health Organization, 2009a).

5.2.5 Data entry and analysis

A structured questionnaire was created using TeleForm version 10.2 (Cardiff Software, CA, US). Each questionnaire consisted of variables with letter or check boxes to be filled in accordingly. The completeness of all questionnaire was examined upon completion of data collection for each participant. All completed questionnaires were scanned using TeleForm scanner. Data entry was performed automatically by the software and verified box by box upon scanning by a research team member. Errors were immediately corrected with reference to the original questionnaire.

After the data entry, each data point in the verified dataset was cross-checked again prior to acceptance. The accepted dataset was explored and cleaned. The original questionnaires were referred to again for missing, invalid, and extreme value for clarification and correction. Categories with small numbers were merged, while numerical variables were categorised accordingly or left for analyses as they are.

Descriptive analysis was used to describe the sociodemographic and clinical characteristics of the participants. The results were presented first as total, and then divided by dengue and non-dengue patients. The difference between the latter in these characteristics was tested using independent t-test for continuous variables, or Wilcoxon-Mann-Whitney test if the data were not normally distributed. Difference for categorical variables was tested using two-sample tests of proportions.

The interrater agreement between the first and second interpreters for each assay in SD Bioline Dengue Duo was assessed using Kappa statistics (k). It was also computed for the agreement of test results between capillary and venous samples for this combo, while ViroTrack Dengue Acute also have additional results for capillary-serum and venous-serum. The analysis for ViroTrack was weighted in view of additional EQ category, where EQ-positive and EQ-negative were weighted at 0.5, and positive-

negative was given no weight. Agreement was interpreted as poor if k was <0, slight if 0-0.2, fair if 0.2-0.4, moderate if 0.4-0.6, good if 0.6-0.8, and excellent if 0.8-1.0 (Osorio et al., 2010).

The TP, FN, FP, and TN of each index test as compared to the reference standard were used to calculate various diagnostic accuracy parameters and their 95% confidence intervals (95% CI) using standard formulae as below (Florkowski, 2008; Šimundić, 2009): SN = TP/(TP+FN); SP = TN / (TN+FP); positive predictive value (PPV) = TP / (TP + FP); negative predictive value (NPV) = TN / (TN + FN); LR+ = SN / (1-SP); LR- = (1-SN) / SP; AUC = (SN+SP) / 2; and DOR = (TP/FN) / (FP/TN).

All accuracy parameters were compared between both index tests using their confidence intervals; as well as p-value estimated using McNemar's test for binary matched-pairs data for SN and SP (Fagerland, Lydersen, & Laake, 2013), two-sample test of proportions for PPV and NPV, and test of equality of ROC areas for AUC. Likelihood ratios and DOR were compared between both tests using indirect comparison of their confidence intervals as p-values were not estimable.

Additionally, subgroup analyses by exposure (serotype, disease phase, previous exposure to dengue infection) and outcome (lab-confirmed vs presumptive dengue) were also performed to compare SN estimates. SP was not compared like this as it was computed from non-dengue patients. The comparison between the SN of both tests in the subgroup analyses was performed using their confidence intervals and p-value estimated from McNemar's test for binary matched-pairs data.

For diagnostic utility, the overall SN and SP with their 95%CI for both tests were applied to 10000 patients with suspected dengue that would have presented to KKS7 or another similar setting, with the prevalence set according to the figure found in this study

but rounded to the closest double digit (50%), to demonstrate the outcome of the application of these index tests from a macro view.

Data analysis was performed using STATA version 12 (StataCorp, TX, US). All inconclusive and missing test results, whether of reference standard or index tests, were excluded from the analysis.

5.3 Results

5.3.1 Description of the population

Out of the 504 potentially eligible patients who attended the clinic over the study period, 494 (98.0%) agreed to participate in the study. All 494 recruited patients had either capillary and/or venous sample tested on both point-of-care index tests. The flow of participants for the index tests and their results was presented using STARD diagrams (Figure 5.1-Figure 5.2).

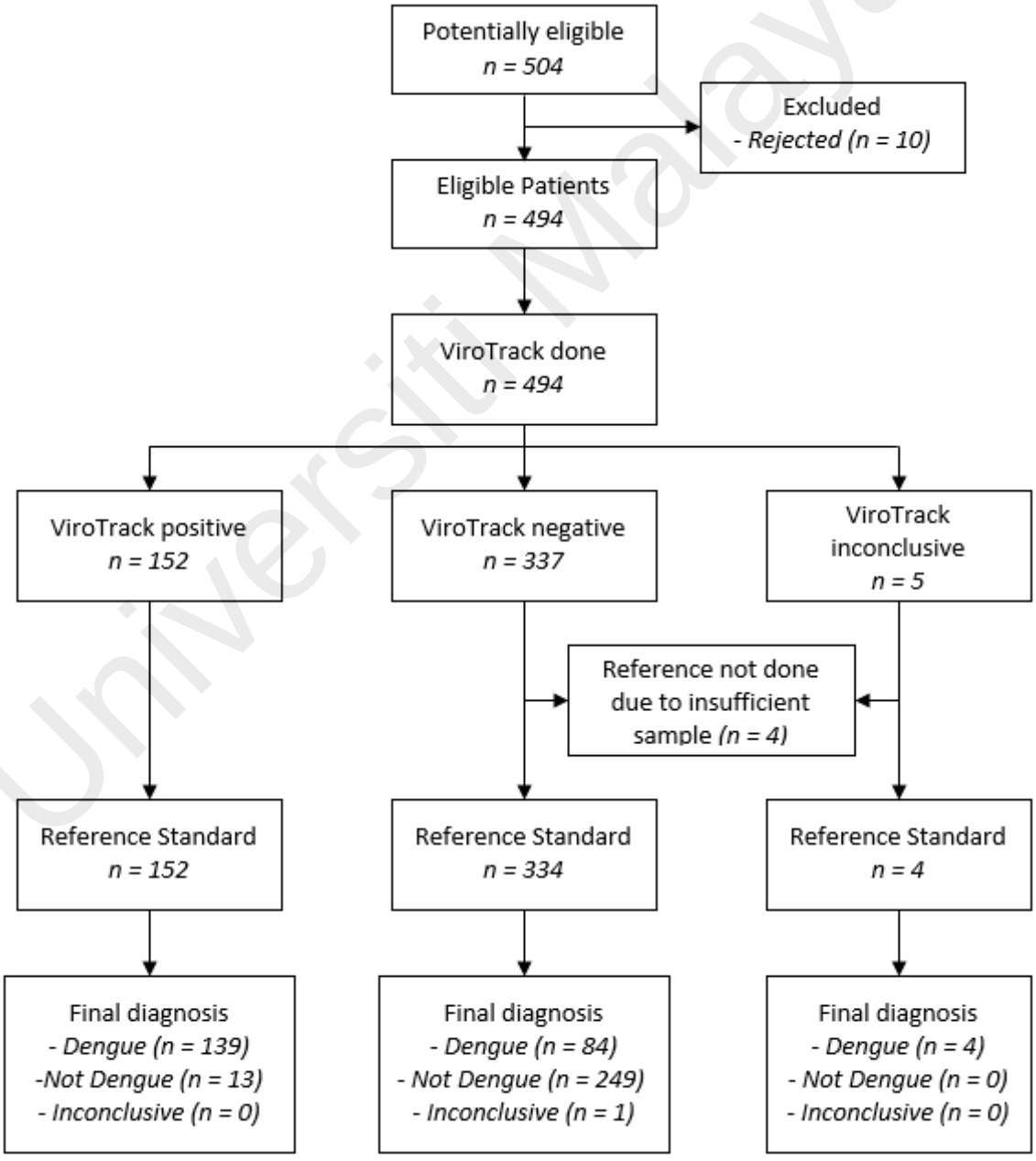


Figure 5.1 STARD flow diagram for ViroTrack Dengue Acute

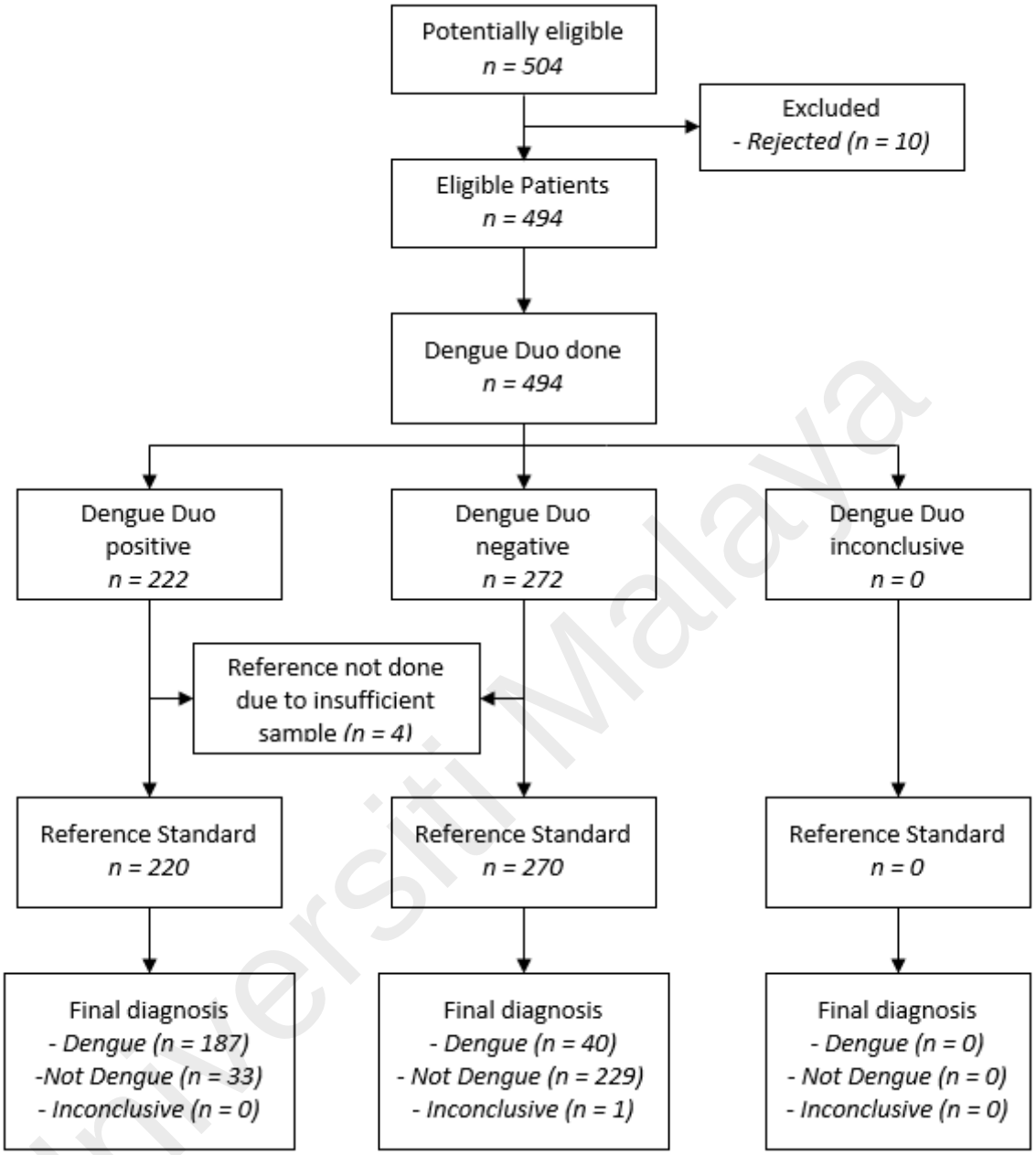


Figure 5.2 STARD flow diagram for SD Bioline Dengue Duo

Out of the 494 recruited patients, only 490 had sufficient sample volume for the reference tests in laboratory, of which one had inconclusive dengue diagnosis. An additional four patients had inconclusive index test result (Figure 5.1 & Figure 5.2). All five of them were excluded from the final analysis. Hence, the final number of patients included in the final analysis was 485, making the actual response rate of 96.2%. Among them, 223 (46.0%) were dengue positive and 262 (54.0%) were negative (Table 5.2 & Table 5.3).

The age of these 485 patients ranged from one to 71 years with a mean of 27.2 years (SD 11.8). 278 (57.3%) were male. Majority of the patients were Malay (n=387, 79.8%), followed by foreigners (n=36, 7.4%). Most were educated up to secondary (n=199, 41.0%) and diploma level (n=134, 27.6%). Their household income level was mainly RM 3000 and above (n=173, 35.7%), with the others rather equally divided between the two lower income categories (Table 5.2).

When comparing the sociodemographic characteristics between dengue and non-dengue patients, statistically significant differences were found in age, patients of foreign origin, and patients with no formal education. The mean age of dengue patients was significantly higher (p=0.01) at 28.5 (SD 10.6) years, as compared to that of non-dengue patients at 26.1 (SD 12.6) years. The proportion of foreigners with dengue was significantly higher (p=0.03) at 23 out of 223 (10.3%), versus 13 out of 262 (5.0%) for those without dengue. Lastly, in the reverse fashion, the proportion of non-dengue patients without formal education was significantly higher (p<0.01) at 18 out of 262 (6.9%) compared to 3 out of 223 (1.4%) dengue patients (Table 5.2).

Table 5.2 Sociodemographic characteristics of included patients

Characteristic	Mean (SD) or N (%)*			p-value
	Total (N=485)	Dengue (N=223)	Non-dengue (N=262)	
Age (years)	27.2 (11.8)	28.5 (10.6)	26.1 (12.6)	**0.01**
Gender				
Male	278 (57.3)	131 (58.7)	147 (56.1)	0.56
Female	207 (42.7)	92 (41.3)	115 (43.9)	0.56
Ethnicity				
Malay	387 (79.8)	175 (78.5)	212 (80.9)	0.50
Chinese	4 (0.8)	3 (1.4)	1 (0.4)	0.24
Indian	46 (9.5)	15 (6.7)	31 (11.8)	0.06
Sabahan & Sarawakian	12 (2.5)	7 (2.1)	5 (1.9)	0.38
Foreigner	36 (7.4)	23 (10.3)	13 (5.0)	**0.03**
Education level				
No formal education	21 (4.3)	3 (1.4)	18 (6.9)	**<0.01**
Up to primary	45 (9.3)	20 (9.0)	25 (9.5)	0.83
Up to secondary	199 (41.0)	91 (40.8)	108 (41.2)	0.93
Up to diploma	134 (27.6)	67 (30.0)	67 (25.6)	0.27
Completed tertiary	84 (17.3)	42 (18.8)	42 (16.0)	0.42
Monthly household income level				
Less than RM 1500	155 (32.0)	70 (31.4)	85 (32.4)	0.80
RM 1500-2999	153 (31.6)	73 (32.7)	80 (30.5)	0.60
RM 3000 and above	173 (35.7)	79 (35.4)	94 (35.9)	0.92

*Denominator for % was corresponding total number of included patients within each column
Bold fonts indicate significant difference between dengue and non-dengue group

The clinical characteristics of the study cohort included in the analysis were summarised in Table 5.3. The mean day of fever upon recruitment was 4.2 (SD 1.9) days, with a range of 1 to 11 days.

Among all the clinical symptoms that could constitute probable dengue, rash was the least common one, reported only by 95 patients (19.6%). Almost all patients suffered from nausea, vomiting or anorexia (n=476, 98.1%), had at least one or more painful parts (n=466, 96.1%), or had at least one dengue warning sign (n=463, 95.5%). Among the warning signs, lethargy was complained by most (n=443, 91.3%), followed by abdominal pain (n=110, 22.7%). Liver tenderness (n=2, 0.4%) and fluid accumulation (n=1, 0.2%) were the least observed dengue fever warning sings (Table 5.3).

Statistically significant differences were found between dengue and non-dengue patients in day of fever at recruitment, rash, and abdominal pain. For dengue patients, the average day of fever upon recruitment into the study was 4.8 (SD 2.0) days, significantly longer than that of the non-dengue patients at 3.8 (SD 1.7) days, with a p-value of <0.001. Dengue patients were significantly more likely (p<0.001) to have rash (n=63, 28.3%), as compared to non-dengue patients (n=32, 12.2%). On the contrary, for abdominal pain, non-dengue patients were significantly more likely (p=0.01) to report it (n=71, 27.1%) compared to dengue patients (n=39, 17.5%) (Table 5.3).

Table 5.3 Clinical characteristics of included patients

Characteristic	Mean (SD) or N (%) *			p-value
	Total (N=485)	Dengue (N=223)	Non-dengue (N=262)	
Day of fever	4.2 (1.9)	4.8 (2.0)	3.8 (1.7)	**<0.001**
Symptoms				
i) Nausea, vomiting or anorexia	476 (98.1)	219 (98.2)	257 (98.1)	0.93
ii) Rash	95 (19.6)	63 (28.3)	32 (12.2)	**<0.001**
*iii) Aches and pain^*	466 (96.1)	217 (97.3)	249 (95.0)	0.20
Warning signs				
i) Presence of at least one	463 (95.5)	215 (96.4)	248 (94.7)	0.35
ii) Abdominal pain	110 (22.7)	39 (17.5)	71 (27.1)	**0.01**
iii) Persistent vomiting	76 (15.7)	32 (14.4)	44 (16.8)	0.46
iv) Persistent diarrhoea	66 (13.6)	27 (12.1)	39 (14.9)	0.37
v) Bleeding	17 (3.5)	11 (4.9)	6 (2.3)	0.11
vi) Lethargy	443 (91.3)	208 (93.3)	235 (89.7)	0.16
vii) Fluid accumulation	1 (0.2)	1 (0.5)	0 (0.0)	0.28
viii) Liver tenderness	2 (0.4)	2 (0.9)	0 (0.0)	0.12

*Denominator for % was corresponding total number of included patients within each column
^Aches and pain included retro-orbital pain, headache, body ache, and muscle and joint pain
Bold fonts indicate significant difference between dengue and non-dengue group

Among the 223 dengue positive patients, 71 (31.8%) were presumptive and 152 (68.2%) were laboratory-confirmed by definition. The latter comprised of 64 positives on both RT-PCR and Panbio NS1 ELISA, 27 positives only on RT-PCR, and 61 positives only on NS1 ELISA. Only eight (3.6%) had dengue without warning sign, while 215 (96.4%) had at least one warning sign according to WHO 2009 classification. 134 (60.1%) of them had primary while 89 (39.9%) had secondary dengue. Lastly, there were 31 (13.9%) DENV-1, 29 (13.0%) DENV-2, 30 (13.5%) DENV-3, and only one (0.5%) DENV-4 infections (Table 5.4).

Table 5.4 Stratification of dengue positive patients by different subgroups

Characteristics	N(%)*
By laboratory result	
Laboratory-confirmed	*152 (68.2)*
Panbio NS1 only	*61*
RT-PCR only	*27*
Both NS1 and PCR	*64*
Presumptive	*71 (31.8)*
By WHO 2009 classification	
Dengue without warning sign	*8 (3.6)*
Dengue with warning sign	*215 (96.4)*
By disease phase	
Acute (<=5 days)	*157 (70.4)*
Convalescent (>=6 days)	*66 (29.6)*
By previous exposure to dengue	
Primary	*134 (60.1)*
Secondary	*89 (39.9)*
By serotype	
DENV-1	*31 (13.9)*
DENV-2	*29 (13.0)*
DENV-3	*30 (13.5)*
DENV-4	*1 (0.5)*

*****Denominator for all % was 223 dengue positive patients

5.3.2 Agreement between test interpreters and different blood specimens

For SD Bioline NS1 assay, 477 patients had both samples tested and 12 patients had only venous results; while 485 had both results and 4 had only venous result for IgM/IgG assay. For ViroTrack Dengue Acute, 437 patients had both results, while 41 and 11 had either capillary or venous result, respectively. All 489 patients also had serum results for ViroTrack Dengue Acute. Comparison can only be made between results from different subgroups belonging to the same patient. The kappa and their 95% CI for all comparisons were more than 0.8, indicating excellent agreement (Table 5.5-Table 5.6).

For SD Bioline Dengue Duo, both interpreters almost completely agreed on NS1 assay tested on both capillary and venous samples with kappa of 0.99 (95%CI 0.90-1.00) and 0.99 (95%CI 0.91-1.00), respectively; while the results were almost comparable for serology tests with point estimates of k ranging from 0.96-0.98 (Table 5.5). When the final agreed test results between capillary and venous were compared, the kappa ranged from 0.96 (95%CI 0.87 -1.00) for IgG assay to 0.99 (95%CI 0.90-1.00) for NS1 (Table 5.6).

For ViroTrack Dengue Acute NS1 assay, the kappa stood at 0.95 (95%CI 0.86-1.00) for capillary-venous, 0.94 (95%CI 0.85-1.00) for capillary-serum, and 0.94 (95%CI 0.86 -1.00) for venous-serum (Table 5.6).

Table 5.5 Interrater agreements (95% CI) between two interpreters for capillary and venous specimens tested on different SD Bioline Dengue Duo assays

Assay Type	Kappa (95% CI)	
	Capillary	Venous
NS1	0.99 (0.90 - 1.00)	0.99 (0.91 - 1.00)
IgM	0.98 (0.89 - 1.00)	0.96 (0.88 - 1.00)
IgG	0.97 (0.88 - 1.00)	0.98 (0.89 - 1.00)

Table 5.6 Agreements (95% CI) between the results of different specimens tested on different SD Bioline Dengue Duo assays and ViroTrack Dengue Acute NS1 assay

Compare between	Kappa (95% CI)
Capillary-Venous on SD Bioline Dengue Duo	
NS1 assay	0.99 (0.90 - 1.00)
IgM assay	0.98 (0.89 - 1.00)
IgG assay	0.96 (0.87 - 1.00)
ViroTrack Dengue Acute NS1 assay	
Capillary-Venous	0.95 (0.86 - 1.00)
Capillary-Serum	0.94 (0.85 - 1.00)
Venous-Serum	0.94 (0.86 - 1.00)

5.3.3 Diagnostic accuracy of index tests as compared to reference standard

All the diagnostic accuracy parameters for both ViroTrack Dengue Acute NS1 and SD Bioline Dengue Duo all assays combined were presented in Table 5.7. Additional results for the latter individual components and their other combination, as well as that performed with additional unmatched samples for all assays combined, were presented in Appendix K. ViroTrack achieved SN of 62.3% (95%CI 55.6-68.7) and SP of 95.0% (95%CI 91.7-97.3), versus SD Bioline with SN of 82.5% (95%CI 76.9-87.3) and SP of 87.4% (95%CI 82.8-91.2). The 95%CI of both accuracy estimates did not overlap; and the p-value for both was <0.001. Both the above indicated that the SN and SP of both tests were significantly different from each other. More specifically, the ViroTrack had significantly lower SN, but significantly higher SP, compared to that of SD Bioline.

The low SN and high SP of ViroTrack corresponded to its high PPV and low NPV, at 91.4% (95%CI 85.8-95.4) and 74.8% (95%CI 69.8-79.4), respectively. Due to its rather comparable SN and SP, SD Bioline achieved similar PPV and NPV, at 84.8% (95%CI 79.3-89.3) and 85.4% (95%CI 80.6-89.4), respectively. The 95%CI of PPV for both tests overlapped, and the p-value equalled to 0.06, indicating absence of statistically significant difference between the PPV of both tests. On the contrary, the NPV of both tests did not overlap, and the p-value was <0.01, pointing to a statistically significant difference between them. In short, ViroTrack had insignificantly higher PPV, but significantly lower NPV, as compared to SD Bioline (Table 5.7).

The LR+, LR-, and DOR for ViroTrack were 12.6 (95%CI 7.3-21.5), 0.396 (95%CI 0.334-0.470), and 31.7 (95%CI 17.2-58.5), respectively. The same estimates for SD Bioline were 6.6 (95%CI 4.7-9.1), 0.200 (95%CI 0.150-0.267), and 32.7 (95%CI 19.8-54.0), respectively. Since the p-value was not estimable, comparison between them can only rely on their 95%CI. Out of all measures, only the 95%CI of LR- for both tests did

not overlap, and the figure was lower for SD Bioline. Hence, the LR- of SD Bioline was significantly lower than that of ViroTrack (Table 5.7).

Finally, ViroTrack had an AUC of 0.787 (95%CI 0.752-0.821), versus that of SD Bioline at 0.850 (95%CI 0.817-0.882). Although their 95%CI overlapped slightly, the test of equality of ROC areas for AUC produced a p-value of <0.001, indicating extremely significant difference between them. More specifically, this global accuracy measure signified that SD Bioline Dengue Duo performed generally better than ViroTrack Dengue Acute, even though their DOR were quite similar to each other (Table 5.7).

Table 5.7 Diagnostic accuracy estimates (95% CI) for ViroTrack Dengue Acute and SD Bioline Dengue Duo

Parameter	ViroTrack Dengue Acute NS1	SD Bioline NS1 or IgM or IgG	p-value
Sensitivity*, % (95% CI)	*139/223* 62.3 % (55.6 - 68.7)	*184/223* 82.5 % (76.9 - 87.3)	**<0.001**
Specificity*, % (95% CI)	*249/262* 95.0 % (91.7 - 97.3)	*229/262* 87.4 % (82.8 - 91.2)	**<0.001**
PPV*, % (95% CI)	*139/152* 91.4 % (85.8 - 95.4)	*184/217* 84.8 % (79.3 - 89.3)	0.06
NPV*, % (95% CI)	*249/333* 74.8 % (69.8 - 79.4)	*229/268* 85.4 %(80.6 - 89.4)	**<0.01**
LR + (95% CI)	12.6 (7.32 - 21.5)	6.6 (4.7 - 9.1)	N.A.
LR - (95% CI)	0.396 (0.334 - 0.470)	0.200 (0.150 - 0.267)	N.A.
AUC (95% CI)	0.787 (0.752 - 0.821)	0.850 (0.817 - 0.882)	**<0.001**
DOR (95% CI)	31.7 (17.2 - 58.5)	32.7 (19.8 - 54.0)	N.A.

*The italic numbers shown before the parameter estimates are number of correct tests over number of all tests for the corresponding parameters.
Bold fonts indicate significant difference between ViroTrack and SD Bioline.

5.3.4 Sensitivities of index tests in subgroup analyses

All SN estimates stratified according to different subgroups for both index tests were summarised in Table 5.8. Additional results for the individual components and their other combination of SD Bioline Dengue Duo, as well as that performed with additional unmatched samples for all assays combined, were presented in Appendix K. The SN of ViroTrack Dengue Acute was 64.5% (95%CI 45.4-80.8) in detection DENV-1 infection, 79.3% (95%CI 60.3-92.0) for DENV-2, 96.7 (95%CI 82.8 - 99.9) for DENV-3, and 0% (95%CI 0 - 97.5) for DENV-4. SD Bioline Dengue Duo had identical TP and all positives for all serotypes except DENV-2, making its SN identical to that of ViroTrack in all those categories. Its SN for the detection of DENV-2 infection was higher at 93.1% (95%CI 77.2-99.2) but overlapped with that of ViroTrack. Its p-value was also similar to all the rest at >0.05. As such, both index tests did not differ in the detection of dengue infection of different serotypes.

When stratified by disease phase or day of fever counted from the day of disease onset, ViroTrack achieved SN of 65.0% (95%CI 57.0-72.4) for acute infection five days and below, and 56.1% (95%CI 43.3-68.3) for convalescent infection five days and beyond. They were significantly lower (p-value for both <0.001) than the corresponding estimates for SD Bioline, which stood at 80.3% (95%CI 73.2-86.2) and 87.9% (95%CI 77.5-94.6) for acute and convalescent phases, respectively (Table 5.8).

Among those patients with primary infection, ViroTrack had a SN of 73.9% (95%CI 65.6-81.1). However, this SN dropped to 44.9% (95%CI 34.4-55.9) among those with secondary infection, significantly lower than that of the primary dengue. Together, both SN of ViroTrack were significantly lower than that of the corresponding estimate for SD Bioline, which remained almost constant for both primary and secondary infections, at 82.1% (95%CI 74.5-88.2) and 83.1% (95%CI 73.7-90.2), respectively. In fact, the

difference between the SN for these two index tests was more significant for secondary infection ($p<0.001$), than for primary infection ($p=0.03$) (Table 5.8).

Finally, when only lab-confirmed dengue patients, i.e. those tested positive on virology tests, whether RT-PCR or NS1 ELISA or both, were taken as dengue reference positive or disease positive; ViroTrack achieved a SN of 77.0% (69.5-83.4), as compared to 82.2% (95%CI 75.2-88.0) for SD Bioline. Both estimates were comparable without statistically significant difference ($p=0.07$). However, when only patients with presumptive dengue, i.e. those tested positive only on IgM ELISA, were taken as dengue positive; the SN of ViroTrack reduced to 31.0% (95%CI 20.5-43.1). This estimate was significantly lower than its own SN among lab-confirmed dengue only; as well as that of SD Bioline among presumptive dengue, which stood far higher at 83.1% (95%CI 72.3-91.0). In fact, the difference between the SN of both index tests among presumptive dengue was extremely significant with $p<0.001$ (Table 5.8).

Table 5.8 Sensitivities (95% CI) for ViroTrack Dengue Acute and SD Bioline Dengue Duo in different subgroups

Category	ViroTrack Dengue Acute NS1	SD Bioline NS1 or IgM or IgG	p-value
By serotype			
DENV-1	*20/31* 64.5 (45.4 – 80.8)	*20/31* 64.5 (45.4 – 80.8)	1.00
DENV-2	*23/29* 79.3 (60.3 – 92.0)	*27/29* 93.1 (77.2 – 99.2)	0.12
DENV-3	*29/30* 96.7 (82.8 – 99.9)	*29/30* 96.7 (82.8 – 99.9)	1.00
DENV-4	*0/1* 0 (0 - 97.5)		1.00
By disease phase/day of fever			
Acute (<=5 days)	*102/157* 65.0 (57.0 - 72.4)	*126/157* 80.3 (73.2 - 86.2)	**<0.001**
Convalescent (>=6 days)	*37/66* 56.1 (43.3 - 68.3)	*58/66* 87.9 (77.5 - 94.6)	**<0.001**
By previous exposure to dengue infection			
Primary	*99/134* 73.9 (65.6 - 81.1)	*110/134* 82.1 (74.5 - 88.2)	**0.03**
Secondary	*40/89* 44.9 (34.4 - 55.9)	*74/89* 83.1 (73.7 - 90.2)	**<0.001**
By reference definition			
Lab-confirmed	*117/152* 77.0 (69.5 - 83.4)	*125/152* 82.2 (75.2 - 88.0)	0.07
Presumptive	*22/71* 31.0 (20.5 - 43.1)	*59/71* 83.1 (72.3 - 91.0)	**<0.001**

i) The italic numbers shown before the sensitivity estimates are true positives over all positives for the respective assay.
ii) Bold fonts indicate significant difference between ViroTrack and SD Bioline.

5.3.5 Diagnostic utility: Application to clinical settings

The outcome of the application of the index tests evaluated in this study to a setting similar to the study location, on 10000 patients presented with suspected dengue according to WHO 2009 classification, among whom prevalence of dengue would be approximately 50% (rounded from the actual level of 46.3%), was summarised in Figure 5.3. Here, the number of patients actually infected with dengue would be equal with the number of patients without dengue, i.e. 5000 people in each group.

Out of 5000 patients with dengue, ViroTrack would detect 3115 (95%CI 2780-3435) of them, as compared to 4120 (95%CI 3840-4355) TP detected by SD Bioline Dengue Duo. However, at the same time, ViroTrack would have missed as many as 1885 (95%CI 1565-2220) out of these actual dengue patients by misclassifying them as negatives; while SD Bioline would have 880 (95%CI 645-1160) of these FN (Figure 5.3).

Among the other 5000 patients without dengue, ViroTrack would correctly exclude dengue in 4750 (95%CI 4585-4865) people, versus 4370 (95%CI 4140-4560) TN for SD Bioline. On the other hand, 630 (95%CI 440-860) non-dengue patients would be wrongly diagnosed as having dengue by SD Bioline. In contrast, ViroTrack would have only 250 (95%CI 135-415) FP (Figure 5.3). The differences between both index tests in all these parameters were statistically significant as none of their 95% CI crossed each other.

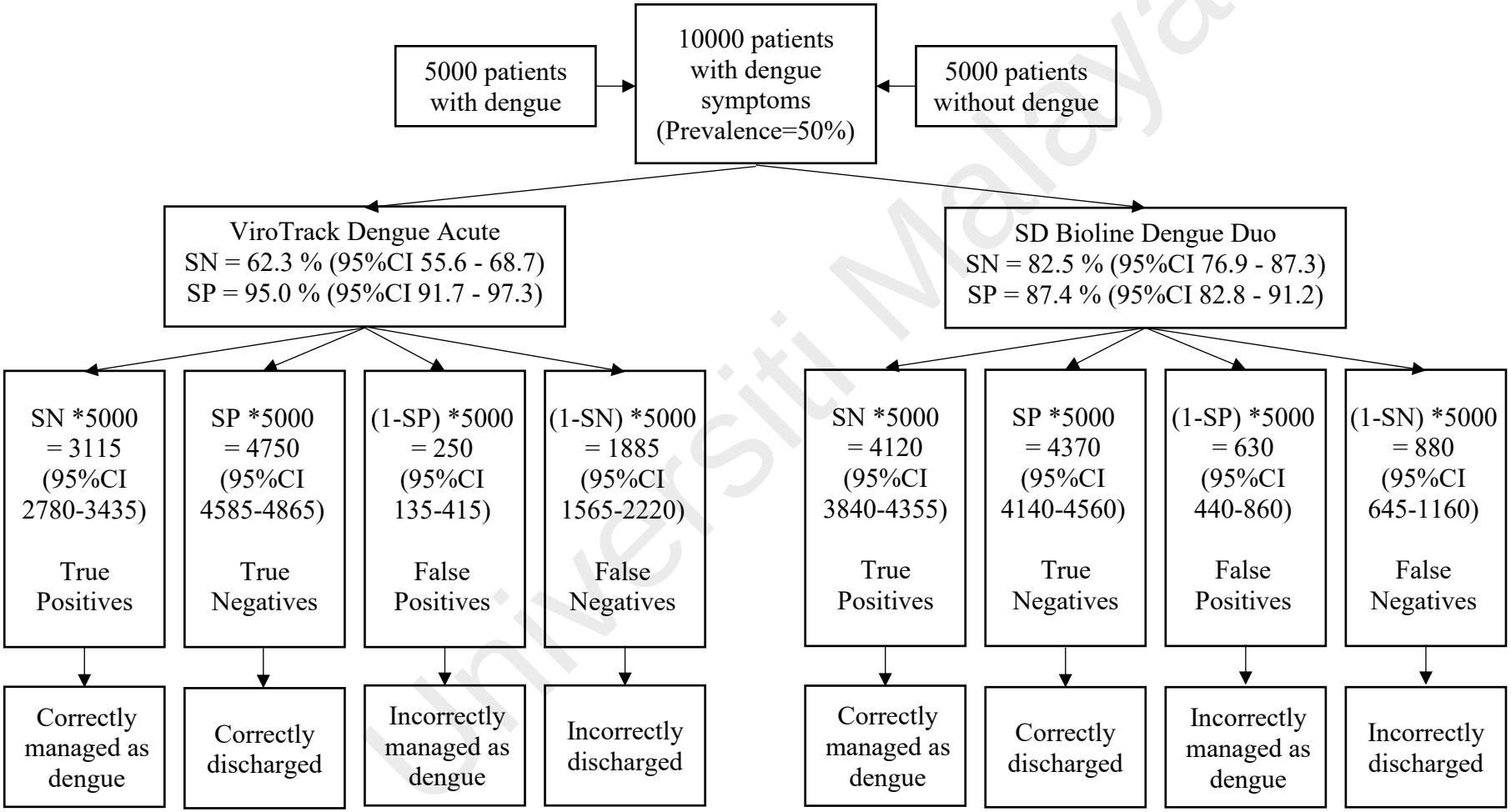


Figure 5.3 Outcomes of the application of two index tests for the diagnosis of acute dengue in 10000 patients presented with dengue symptoms to a setting similar to study location

5.4 Discussion

5.4.1 Sample size and response rate

This study achieved a large sample size with good response rate. Out of the 504 eligible patients, 485 (96.2%) were included in the final analysis. Response rate or retention rate was rarely if not never explicitly reported in previously studies (Stuart D. Blacksell et al., 2006; Lima, Nogueira, Schatzmayr, & dos Santos, 2010; Najioullah et al., 2011; Pok, Lai, Sng, & Ng, 2010; Wang & Sekaran, 2010a; Zainah et al., 2009). Manual calculation was possible only for some studies with clearer reporting and STARD flow cart. Some studies with similar sample size in the final analysis had very low retention rates of 33.4% (452/1353) and 511/1607 (31.8%) due to existing laboratory testing protocol at the study sites that did not require all samples to be tested with all reference tests (Prado et al., 2018; Shih et al., 2016). Other smaller studies had better retention rate at 86.7% (325/375), 82.2% (328/399), 80.1% (197/246), 91.9% (308/335), 96.6% (143/148), and 81.8% (211/258) (Buonora et al., 2016; Carter et al., 2015; Gan et al., 2014; Huits et al., 2017; Mata et al., 2017; Vivek et al., 2017). Exclusion of certain samples from the final analysis may be due to their intrinsic characteristics that could introduce difficulty in results interpretation, e.g. inconclusive or intermediate results (Osorio et al., 2010). This kind of sample attrition is differential and would upward bias the resulted accuracy estimates (Kohn et al., 2013). Although inconclusive reference and index tests results in this study were also excluded, they only constituted a negligible 1.0% (5/504) of total eligible participants.

5.4.2 Socio-demographic characteristics of study cohort

The study cohort recruited for this study had an average age of below 30, although toddlers and seniors were included as well (Table 5.2). Due to the location of the study site in a Malay-majority area, most of the study participants was of Malay root. However, there was also a sizable presence of foreigners working as labourers in surrounding factories, who lived in the affordable flats nearby. On the other hand, the observation of few participants of Chinese origin was probably reflective of the population structure at the locality, as well as their preference for private healthcare services (Atun, Berman, Hsiao, Myers, & Yap, 2016). In terms of education level, the participants appeared to be more or less normally distributed in the five categories with skew to the right, in line with the country's emphasis on higher education (Malaysia, 2015). However, the average household income demonstrated that majority of the participants belonged to the lower income group, as the actual percentage of households with an average income less than RM 3000 was only 5.2%, contrary to 63.6% found in this study (DOSM, 2016). Nevertheless, underreporting of income cannot be excluded. These sociodemographic characteristics were likely reflective of the underlying population that normally accessed the clinic for healthcare services, and may differ from place to place.

The significant differences between dengue and non-dengue patients in age, and proportion of foreigners and those without formal education informed us about two things (Table 5.2). First of all, dengue patients were significantly older than non-dengue patients. This observation was confirmed by the significantly lower proportion of dengue patients without formal education, i.e. children of pre-schooling age. This finding was in line with that of three other studies (Stuart D. Blacksell et al., 2007; Hang et al., 2009; Prado et al., 2018), but disagreed with one (Mitra et al., 2016). The main reason for this may be more frequent outdoor activities of adults increased their exposure to Aedes mosquito, as

compared to children who were more protected. Another plausible explanation for this underrepresentation of younger age groups among those with dengue may be higher proportion of asymptomatic or mild primary dengue in these groups (Rodriguez-Barraquer et al., 2015). Having said that, the higher proportion of adult foreign workers among dengue group may also be contributing to its higher average age. Their larger presence in dengue group, on the other hand, may be due to a mixture of their origin from dengue non-endemic countries with low dengue antibody level, lower socio-economic status, poor housing condition and hygiene practices, and health behaviours. However, further research is required to confirm any of these factors.

5.4.3 Clinical characteristics of study cohort

In this study, the mean day of fever upon recruitment of participants with dengue was significantly longer than that of non-dengue (Table 5.3). But it did not mean that dengue patients inclined to seek health later than non-dengue patients. The reason for this was purely operational, as patients diagnosed with dengue earlier under daily follow-up with the clinic were also recruited into the study. The actual mean day of fever for dengue patients upon first health contact may be shorter or may not differ from that of non-dengue patients (Andries et al., 2012; Carter et al., 2015; Mitra et al., 2016).

As mentioned previously, dengue is characterised by non-specific symptoms. Here, the dengue-probable symptoms and warning signs given in WHO 2009 classification did not differ in their frequencies between dengue and non-dengue patients, except for rash and abdominal pain (Table 5.3). More specifically, rash was significantly more frequently observed among dengue patients, while abdominal pain was significantly more frequently observed among non-dengue patients. The finding on rash was in line with

that reported by Blacksell et al among dengue patients in Laos (Stuart D. Blacksell et al., 2007). Other studies also found higher probability of rash among dengue patients compared to non-dengue, but could not demonstrate statistical significance (Buonora et al., 2016; Carter et al., 2015; Mitra et al., 2016).

On the other hand, for abdominal pain, Buonora et al also discovered its higher proportion among non-dengue patients, albeit not significantly higher (Buonora et al., 2016). In contrast, Carter et al found the exact opposite, that abdominal pain was significantly more likely to be observed in dengue patients. However, the latter conducted the study only among children (Carter et al., 2015). Anyway, the finding in this study can aid local physicians in diagnosing patients. However, while presence of rash in febrile patients should raise their suspicion for dengue; complain of abdominal pain should not be quickly interpreted as the absence of the disease, especially in view of the significance of this symptom among children, as a warning sign, and its subjective nature.

5.4.4 Dengue diagnosis among study cohort

When dengue positive patients were stratified, it was found that majority of them were diagnosed by virology tests (Table 5.4), in line with previous studies that used both virology and serology as reference (Andries et al., 2012; Hang et al., 2009; Osorio et al., 2010; Prado et al., 2018). This is understandable as virology tests are preferred for early diagnosis of dengue. In fact, in all these previous studies, as well as in this study, majority of dengue patients were in the acute phase when their blood sample was drawn.

When classified by WHO 2009 guidelines, it was found that almost all dengue patients in this study had warning sign (Table 5.4). Few studies reported this for their dengue positive cohort. Among three that reported, two classified them only according to WHO

1997 guidelines. Anyway, both of the latter studies reported more patients with DHF and/or DSS than uncomplicated dengue fever (Andries et al., 2012; Hang et al., 2009). Gan et al, in a study conducted primarily among patients from ambulatory setting in Singapore, classified dengue patients using both guidelines. When classified using old guidelines, only slightly less than 20% of dengue patients had DHF while the rest had less severe DF. New guidelines increased those with potentially more severe dengue to slightly more than 50% (Gan et al., 2014). While the higher proportion of more severe patients in the former two studies was probably due to their study settings in hospital (Table 4.8), where patients with more severe condition were admitted; the higher proportion of more severe patients when 2009 guidelines were used in the latter, as well as in this study, was most likely the result of higher sensitivity of the newer guidelines (Gan et al., 2013; Hadinegoro, 2012; World Health Organization, 2009a). The number of more severe patients in this study were specifically inflated by the warning sign "lethargy", due to its subjective nature and participants' tendency to agree to it during eligibility screening (Table 5.3).

In view of this, better assessment of dengue severity may be found in its stratification by previous exposure to dengue, where secondary may be more likely to be more severe compared to primary. This study discovered that around 2/3 of dengue positive patients had primary dengue, in line with what was expected of a primary care clinic. In contrast, all other hospital-based studies had primarily patients with secondary dengue (Buonora et al., 2016; Chen et al., 2016; Hang et al., 2009; Kittigul & Suankeow, 2002). The only previous study that was also conducted in clinic setting, but only among children, did not classified its dengue participants into primary and secondary infections. However, only around 1/5 of these patients were admitted to critical care unit, indicating the proportion of more severe patients was parallel to that found in this study (Carter et al., 2015).

Finally, stratification of dengue positive patients by serotype was only possible for those tested positive on RT-PCR. There was only around half of them among all positives in this study. The distribution of underlying dengue serotype among those infected here was different from all the other studies (Andries et al., 2012; Hang et al., 2009; Osorio et al., 2010; Tricou et al., 2010). This is understandable as each of these places has its own circulating and predominant serotypes at the moment the studies were carried out. The most important thing was that the distribution of dengue serotype found in this study agreed with previous finding for the study site (Mohd-Zaki et al., 2014)

5.4.5 Significance of index tests on different specimens

Most published dengue RDT evaluation studies used serum samples. Some studies also used whole blood specimen and only one used capillary blood (Carter et al., 2015; Gan et al., 2014; Mata et al., 2017; Pal et al., 2015). The excellent agreement between the index test results tested on capillary, venous, and/or serum samples in this study demonstrated that anyone of them can be used on RDT, provided that anticoagulant-coated tool is used for the collection of whole blood specimen. It is important to note that the kappa estimates of ViroTrack were generally lower than that of SD Bioline due to the additional EQ category, but agreement remained excellent (Brenner & Kliebsch, 1996). The validity of results from capillary blood has practical implication when minimal invasiveness and/or rapidity is required such as in young children or during massive dengue outbreak.

5.4.6 Comparison of index tests overall diagnostic performance

No study was published prior to this for the evaluation of ViroTrack Dengue Acute, while SD Bioline Dengue Duo had been extensively evaluated. The combination of NS1/IgM/IgG assay of the latter had point estimates of SN and SP for the diagnosis of acute dengue that ranged within 80.73-98.90% and 57.45-100.0%, respectively (Andries et al., 2012; Gan et al., 2014; Osorio et al., 2010; Pal et al., 2015; Sanchez-Vargas et al., 2014; Shih et al., 2016; Tricou et al., 2010; Vickers et al., 2015). On the other hand, the pooled SN and SP for SD Bioline all assays combined in the meta-analysis in Chapter 4 were correspondingly 91.95% (95%CI 87.18-95.04) and 90.41% (95%CI 82.30-95.03). The overall SN and SP estimates of this study fell within both the above ranges, although the lower bound of the SN was slightly below (Table 5.7). However, direct comparison like this, whether with primary studies or meta-analysis, may not be appropriate due to underlying difference in study design, patient population, definition of reference standard and other study characteristics (Leeflang, 2014; Whiting, Rutjes, Westwood, & et al., 2011).

Since difference in study characteristics modify the outcomes, apparent differences in diagnostic accuracy and utility parameters between two same or different tests evaluated in two different studies may be due to the difference between their study characteristics instead of the actual performance of the tests themselves. However, diagnostic tests evaluated within the same study on the same patient population under the same condition can be directly compared (Leeflang, 2014). In this study, the new RDT – ViroTrack Dengue Acute, had significantly lower SN, but significantly higher SP, as compared to SD Bioline Dengue Duo all assays combined. This means that, although the developer of ViroTrack claimed to be able to increase its NS1 test SN with the help of immuno-magnetic agglutination assay and objective interpretation of the results, this increase

could not outpace the various degree of increment in SN brought about by the combination of serology assays to NS1. However, similar to this study, such increase in SN in all these studies almost always compensated by corresponding decrease in SP, due to lowering of detection threshold (Andries et al., 2012; Gan et al., 2014; Osorio et al., 2010; Pal et al., 2015; Sanchez-Vargas et al., 2014; Shih et al., 2016; Tricou et al., 2010; Vickers et al., 2015).

Higher SN of SD Bioline Dengue Duo made its negative result significantly more useful clinically as compared to ViroTrack. Dengue diagnosis can be confidently excluded in 85.4% (95%CI 80.6 - 89.4) among that tested negative on this test. In reverse, only one to two out of 10 patients tested negative were FN, or there would only be 10-20% post-test probability of dengue in a negative patient, whichever way more understandable by clinicians. In contrast, post-test probability of dengue among that tested negative on ViroTrack would be at least 20-30%. On the other hand, higher SP for ViroTrack in this study did not make its positive results significantly more useful compared to that of SD Bioline, due to the former comparatively far lower SN (Table 5.7).

5.4.7 Comparison of index tests diagnostic performance by subgroups

The sensitivities of the index tests performed as expected in the subgroup analyses. Being NS1-only assay, ViroTrack performed significantly better in detecting dengue infection in the first 5 days versus 6 days and above as NS1 antigen is actively produced and secreted in the first week (Table 5.8, Figure 2.2 & Figure 2.3). However, SD Bioline Dengue Duo, due to the combination of serology assays that performed better in convalescent phase, not only improved its SN in this phase, but also in acute phase. This finding was in line with that found in the meta-analysis (Table 4.11, Table 4.12, Table

4.14 & Table 4.16). This improvement was so substantial that SD Bioline significantly outperformed ViroTrack in the detection of acute dengue in any phase of disease.

The same expected trend was also observed in the detection of primary versus secondary dengue, and laboratory-confirmed versus presumptive. Again, NS1-only ViroTrack had significantly higher SN in detecting primary and laboratory-confirmed dengue, as compared to that of secondary and presumptive dengue, evident from their non-overlapped 95%CI (Table 5.8). But its performance was lacklustre when compared to that of SD Bioline in all these categories and their subgroups. In fact, in all subgroups except lab-confirmed dengue, SD Bioline had significantly higher SN. Again, the reason for this lies predominantly in the definition of reference standard, where primary and laboratory-confirmed dengue were mostly those tested positive on RT-PCR and/or Panbio NS1 ELISA; while secondary and presumptive were more dependent on IgM and IgG capture ELISA for their definitions. These similar trends were in line with that found Chapter 4 (Table 4.11, Table 4.12, Table 4.14 & Table 4.16), and were also observed repeatedly in previous studies with some variations that can be attributed to difference in study characteristics (S. D. Blacksell et al., 2011; Stuart D. Blacksell et al., 2012; Hunsperger et al., 2014; Osorio et al., 2010; Pal et al., 2014; Pal et al., 2015; Sanchez-Vargas et al., 2014; Shih et al., 2016; Wang & Sekaran, 2010a).

Although the SN of both index tests differed significantly in almost all subgroups, the same was not observed when stratified by serotype. This was expected as classification by serotype was only possible for those tested positive on RT-PCR. Serology assay did not perform as well as NS1 assay when virology tests were used as reference (Table 4.11, Table 4.12 & Table 4.14). The SD Bioline SN estimates stratified by serotype were probably contributed largely by its NS1 component, making them rather similar with that of ViroTrack (Table 5.8). On the other hand, the failure of both tests to detect the sole

DENV-4 was probably due to chance, as this patient had secondary dengue (with low IgM level) on 5th day of illness, when the level of free and detectable IgG and NS1 happened to be too low after their union in vivo (Chaterji, Allen, Chow, Leo, & Ooi, 2011; Dussart et al., 2008; Hang et al., 2009; Osorio et al., 2010).

5.4.8 Recommendations

As such, the addition of serology assays to SD Bioline Dengue Duo was demonstrated to perform better than that of NS1-only ViroTrack. This trend was in line with previous findings (Andries et al., 2012; S. D. Blacksell et al., 2011; Fry et al., 2011; Gan et al., 2014; Krishnananthasivam et al., 2015; Osorio et al., 2010; Pal et al., 2015; Sanchez-Vargas et al., 2014; Shih et al., 2016; Tricou et al., 2010; Vivek et al., 2017; Wang & Sekaran, 2010b). The repercussion of this finding is that dengue combo test is always superior to RDT with only individual assay as the former has the ability to detect dengue infection regardless of the phase of illness and previous exposure to dengue infection (S. D. Blacksell et al., 2011; Fry et al., 2011).

This recommendation is further backed up by the PPV and NPV generated in this study as discussed above. But the advantages of combo test would be more obvious when the SN and SP of both index tests were applied to 10,000 patients that would visit a similar setting to the study site, among whom 50% were actually infected with dengue (Figure 5.3). Here, among 5000 patients without dengue, SD Bioline would be more than twice as likely compared to ViroTrack to wrongly diagnose as dengue, with number of FP at 630 (95%CI 440-860), versus 250 (95%CI 135-415) of the latter. On the contrary, among 5000 patients with dengue, ViroTrack would also be more than twice more likely compared to SD Bioline to miss the infection, with number of FN at 1885 (95%CI 1565-

2220), versus 880 (95%CI 645-1160) of the latter. While dengue FP may not be a big concern due to its small proportion and relatively less harmful supportive treatment unless in the case of misdiagnosis of other more severe diseases, high number of FN might lead to late diagnosis and delayed administration of required life-saving treatment for missed dengue patients. As such, in a clinical setting, especially in primary care, it is important for a dengue RDT to act as a screening tool that can detect more cases with minimal FN (World Health Organization, 2009a).

5.4.9 Study strengths and limitations

The strength of this study lies in its sound methodology and application. As mentioned above, it is difficult to directly compare diagnostic performance between evaluation studies due to different study characteristics (Leeflang, 2014). In the same way, it is also fundamentally incorrect to directly apply their results into daily practice or for policy making. Good performances reported in phase II case-control or laboratory-based studies may be due to biases instead of the discriminatory power of the evaluated tests (Peeling et al., 2010). In contrast, the cross-sectional prospective design in an actual primary care clinical setting seen in this phase III diagnostic evaluation study produced a more realistic set of diagnostic performance parameters that is true to other similar clinical settings, making the application of its results easier and more valid. Besides, it complied with STARD-guidelines for complete reporting and quality assurance (P. M. Bossuyt et al., 2015). In addition, diagnostic utility presented here is more intelligible to clinical practitioners and policy makers compared to the usually reported diagnostic accuracy. With simple calculation, the diagnostic utility of the two index tests evaluated in this study can be estimated for any clinical setting. However, it should be cautioned that this exercise took into account only dengue diagnostics without consideration of other

diseases. Practitioners should also use existing clinical reasoning for differential diagnosis.

In Malaysia, four previously conducted dengue RDT evaluation studies were exclusively laboratory-based and only one employed cross-sectional prospective design (Fry et al., 2011; Kumarasamy, Zuridah, Asmah Hani, Mariam, & Chua, 2007; Wang & Sekaran, 2010a; Zainah et al., 2009). This study was the first conducted prospectively among consecutively sampled patients in a primary care setting. It provides better and more updated insight into the application of dengue RDT in Malaysia. Furthermore, it was the first that evaluated a biosensor-based RDT in this setting and compared it with extensively used RDT for a more comprehensive understanding of their relative performance, which is absent in most other studies that evaluated only single RDT.

However, this study was not without its limitations. First of all, only single sample was collected from each patient, making the reference standard based on serology presumptive rather than conclusive (Anusyah Rathakrishnan et al., 2014; World Health Organization, 2009a). Secondly, the diagnostic utility calculated was based on the pre-test probability of disease or prevalence of dengue at around 50% as screened using the symptomatology of the WHO 2009 guidelines without considering of the haematological result. The latter would have increased this figure further, changing the predictive values in Table 5.7 and numbers in Figure 5.3. Nevertheless, these limitations perfectly reflect the actual situation faced by clinicians in the front line, making the study results more realistic and applicable to the real-world circumstances.

5.5 Conclusion

In conclusion, for the diagnosis of acute dengue, the new dengue RDT – ViroTrack Dengue Acute, achieved significantly lower SN but higher SP, as compared to SD Bioline Dengue Duo NS1/IgM/IgG combo test. The corresponding higher NPV of SD Bioline Dengue Duo that was statistically different from that of ViroTrack made the negative results of this combo test more clinically useful, as a patient with suspected dengue tested negative on it had lower post-test probability of dengue, or a diagnosis of dengue can be more confidently excluded. SD Bioline significantly higher SN and NPV relative to ViroTrack also means that it would produce less FN compared to the latter. Having less FN equals to missing less dengue patients, which would prevent loss of lives from late diagnosis and delayed administration of required life-saving treatment.

As such, the combination of NS1/IgM/IgG assays in SD Bioline Dengue Duo greatly enhanced its diagnostic accuracy and utility parameters beyond that of ViroTrack Dengue Acute NS1 assay, making the former a better point-of-care dengue diagnostic tool. ViroTrack Dengue Acute may be a potential alternative to existing RITs only if its combination with serology components match or outperform the latter in future phase III cross-sectional prospective evaluation studies conducted properly like the current one.

CHAPTER 6: DISCUSSION

This thesis comprised of two stages, namely: i) systematic review and meta-analysis of the diagnostic accuracy of commercially available dengue RIT; and ii) phase III clinic-based evaluation of a new dengue RDT. Each of the above components answered its specific objectives, with findings discussed in respective chapter. In this chapter, only the most important findings from each phase were discussed.

6.1 The value of systematic review and meta-analysis in the presence of heterogeneity

First of all, the systematic review identified numerous trends in the development and evaluation of dengue RIT (Table 4.9), starting from IgM and IgG kits that were first reported in 1998 with focus on just rapid diagnosis, to NS1 assay that followed a decade later that had additional focus on early diagnosis. Study design also shifted from predominantly laboratory-based case-control studies to more of prospective care-based cross-sectional ones with bigger sample size. There was also a simplification in the preferred reference tests over time. This wide array of differences between studies, together with other study characteristics such as the way RIT results were interpreted and so on, and also threshold effect, contributed to huge between-study heterogeneity in the meta-analysis. These differences between studies subjected evaluation of a specific same brand of RIT to threshold-like effect, even when threshold effect was not supposed to be present. Therefore, the results from the meta-analysis cannot be the true values of diagnostic performance, but merely average values of the underlying studies with one or another common characteristic.

As such, as opposed to the recommendation by Zhang et al., the pooled diagnostic performance of a particular RIT brand was not the best evidence of effectiveness and cannot be used to justify its purchase and application in another setting, where it was never evaluated (H. Zhang et al., 2014). Although, for the closest estimate of the true diagnostic accuracy of dengue RIT, cross-sectional studies that were usually conducted with less biases may still be referred to. Even then, the figures were the average of all brands in the underlying studies. Further meta-analyses according to brand and other relevant characteristics within the cross-sectional studies is needed to obtain a closer-to-the-truth and better estimates. But doing that would result in very few studies, making meta-analysis not possible and unnecessary. It would be more efficient to critically appraise the primary studies using STARD and QUADAS-2 checklists as proposed in this thesis to assess their applicability to one or another setting. But a better way is to properly evaluate a potential test before its application in a setting.

Although its results cannot be generalised directly, the meta-analysis did help in detecting factors or biases that could have affected the study outcomes, and identified the most evaluated dengue RIT with good performance. These findings provided guidance in the design and conduct of the main study presented in this thesis. It also demonstrated that the diagnostic performance of existing dengue RIT varied widely, and because they were affected by different study characteristics, they cannot be used to compare with each other blindly, as much as their pooled results cannot be taken as the true values of performance. The implication was that the evaluation of the new dengue RDT that was claimed to be better than existing dengue RITs cannot be conducted without the latter being evaluated together. Using another RIT as comparator would provide not only a comparison for the new RDT, but would also indicate the most likely performance of this new RDT in all the primary studies they were previously evaluated in.

Considering all the above, the best way to evaluate a new RDT was to conduct at least a phase III diagnostic evaluation study - a cross-sectional study in a clinic in a dengue-prone area, on prospectively recruited patients with suspected dengue, together with the most evaluated dengue RIT with good performance as identified in the systematic review. But before that, this new RDT had to be proven effective first to avoid wastage of research resources. A more efficient small-scale laboratory-based case-control study equivalent to phase II clinical trial sufficient for this purpose was conducted as pre-test. Therefore, the main study and its pre-test in this thesis were designed and conducted in such manner. Furthermore, the reference tests and standard used in the cross-sectional study mirrored as much as possible that of the pre-test to minimise the differences in measurement methods between both studies to make them more comparable for discussion.

6.2 The importance of sources of heterogeneity in understanding index test diagnostic performance

As mentioned above, the systematic review identified several sources of heterogeneity and the way they affect index tests diagnostic performance. These factors were study design, dengue serotype, previous exposure to dengue infection, disease phase, brand, and reference standard used. The conduct of this systematic review, the stratified analysis of the main study, and the pre-test performed before it, provided many opportunities to explore and confirm further the influence of these factors toward index tests diagnostic performance.

First of all, phase II case-control studies would almost always produce better performance estimates compared to phase III cross-sectional studies. This was most evident from the changes in ViroTrack Dengue Acute performance estimates compared between the pre-test and the actual study. Its SN and SP dropped from 92.0% (95%CI 80.8-97.8) and 95.1% (95%CI 83.5-99.4) to 62.3% (95%CI 55.6-68.7) and 95.0% (95%CI 91.7- 97.3), respectively. The difference was only apparent in SN. This indicated that, while selection bias was present in the pre-test, it was primarily contributed by the diseased pool. This can be explained by spectrum bias. For SN, it would be raised if the selection of the diseased was skewed towards those with higher severity compared to the more evenly distributed pattern seen in actual clinical practice, with the assumption that severe patients produced higher level of diagnostic markers. Having said that, it was unclear whether the original source of the dengue positive samples used in the pre-test could have contributed to the higher SN observed, as patients admitted to tertiary hospitals tended to be more severe. To prove this, further study is required. On the other hand, since non-diseased pools in both phases were sick patients and not healthy control,

spectrum bias was less likely to be present, resulting in comparable SP estimates (Kohn et al., 2013).

Although SD Bioline Dengue Duo was not present in the pre-test for comparison to its results in the main study, the pooled estimates in the systematic review and meta-analysis were a mix of both case-control and cross-sectional designs, and could provide an insight into the effect of case-control design to its performance (Table 4.16). SD Bioline all assay combined pooled SN and SP were 91.95% (87.18-95.04) and 90.41 (82.30-95.03), respectively. Its SN and SP in main study were 82.5 % (76.9 - 87.3) and 87.4 % (82.8 - 91.2), respectively (Table 5.7). Evidently, a mix with case-control studies produced better estimates for SD Bioline Dengue Duo. Similar direction of bias discovered for both index tests in relation to study design further confirmed that case-control studies were more biased compared to cross-sectional ones, and that evaluation studies with cross-sectional design would yield diagnostic performance closer to that of the actual clinical settings (Kohn et al., 2013).

As the sample size in the pre-test was not adequate to be analysed by subgroups, the comparison between subgroups were carried out only between the review and the main study, in particular between test sensitivities as only this parameter was computed for the subgroup analyses in clinic phase. In the review, it was found that all test performance parameters were comparable in detecting all serotypes except DENV-4, with none of the differences statistically significant (Table 4.11-Table 4.16). In the main study, no conclusion could be reached pertaining this due to lack of patient with DENV-4 infection (Table 5.8). The failure of both index tests to detect this single DENV-4 patient could be due to chance. Nevertheless, judging from the review findings, it was still possible that all index tests in the main study had lower accuracy when it comes to DENV-4 infection.

However, unless a shift in serotype occurs leading to a DENV-4 outbreak, the implication this finding has to a setting with little DENV-4 presence similar to this study is minimal.

In the review, both NS1 and IgM dengue RITs, as well as their combination were found to have higher SN in detecting primary dengue compared to secondary. The reverse was applicable for IgG and IgM/IgG assays, with the difference neutralised for NS1/IgM/IgG combination. However, statistically significant difference in SN was only found for NS1 and IgG assays (Table 4.11-Table 4.16). In the main study, similar finding was observed for ViroTrack, where its SN for the detection of primary dengue was significantly better than that for the detection of secondary dengue. On the other hand, the neutralisation of effect was observed for SD Bioline (Table 5.8). Both these findings were in line with the review. Furthermore, better performance in the detection of both primary and secondary dengue signified that combo test was more desirable than NS1-only assay (S. D. Blacksell et al., 2011; Fry et al., 2011).

When comparing between disease phases, the trend observed in the review and the main study was similar. NS1 assays performed better in the detection of dengue cases within five days of illness, while serological assays performed better in convalescent phase. The addition of IgM and IgG to NS1 assay nullified the difference but still pointed to an upward bias in the SN for the detection of cases in convalescence, which was underpowered to prove in both the review and main study. Furthermore, the SN estimates of all assays combined consistently demonstrated additional advantage over that of NS1-only assay not just in convalescent phase, but also in acute phase (Table 4.11-Table 4.16, Table 5.8). This finding indicated that, while NS1-only test is important in early dengue diagnosis, addition of serology assays would only add onto its value. Although this benefit was smaller for patients that presented early, it was huge and important for patients that present late to health facility, as the margin of diagnostic error and subsequent

window of opportunity for timely intervention were smaller for this group of patients (S. D. Blacksell et al., 2011; Fry et al., 2011)..

Lastly, changes in reference definition was revealed to influence the study outcome in both the review and the main study (Table 4.11-Table 4.16, Table 5.8). The implication of this finding was that the results in dengue RDT evaluation studies could be manipulated. Other factors apart, this one alone could be one of the biggest contributor to the inflations in the accuracy estimates as reported by RDT manufacturers compared to those evaluated by independent third-party researchers (Andries et al., 2012; Hunsperger et al., 2014; Krishnananthasivam et al., 2015; Pal et al., 2014; Sanchez-Vargas et al., 2014; Shih et al., 2016; Standard Diagnostics Inc., 2008; Vickers et al., 2015; Wang & Sekaran, 2010a).

In terms of diagnostic utility as assessed from population angle, the comparator dengue RIT used in the main study, the SD Bioline Dengue Duo, was the most suitable dengue screening test for application in a low-resource setting such as primary care where it was evaluated in, provided that positive test on any one of its individual component, whether NS1, IgM or IgG assay, is taken as confirmation dengue diagnosis and managed accordingly. The reason was because it would miss least FN compared to ViroTrack Dengue Acute (Figure 5.3). Furthermore, ViroTrack, being NS1-only assay, was less sensitive in the detection of patients that present later in their disease and those with secondary dengue. Thus, among those FN it may fail to diagnose, most would belong to both the above groups, which had higher probabilities of progressing to severe dengue and die from its complications. As such, the practical advantages of combo test over NS1-only test were further established with the help of knowledge on the sources of heterogeneity identified in the review.

6.3 Implications of study for patient care and management

The results of the study serve a reminder to practicing physicians dealing with patients with suspected dengue that there exists no perfect dengue diagnostic test, especially RDT. No test result can fully confirm or exclude a dengue diagnosis in a patient. Judging from the PPV and NPV of SD Bioline Dengue Duo all assays combined, there would be as much as 20% chances of the exact opposite happening, whether the test result was positive or negative (Table 5.7). This may be less important for those non-dengue patients that were wrongly diagnosed as dengue, unless the actual diagnosis was more life-threatening if missed. In the absence of other alternative diagnosis, non-dengue patients misdiagnosed as dengue could be managed as one since clinical management for dengue is mainly supportive in nature and would not harm them (Ministry of Health, 2015).

The more important aspect of dengue RDT misdiagnosis is the FN. Even with the most sensitive interpretation of the SD Bioline Dengue Duo all assays combined, there would be still as much as 23% chance of FN (Table 5.7). In view of this, even in the presence of a negative RDT result, physicians should not discharge patients without advising them to return if they do not recover, especially if warning signs occur. In dengue-endemic area, if a patient is really unwell at the moment of consultation; in the absence of other diagnosis, even in the face of negative RDT test results, the physician should still treat this patient as dengue. If any dengue warning sign is present, the patient should be referred to a hospital, where FBC and dengue RDT can be repeated, or dengue diagnosis can be confirmed with additional laboratory-based diagnostic tests (Ministry of Health, 2015).

In Malaysia, the most used dengue RDT is RVR Dengue Combo NS1-IgG/IgM, which is available in all public clinic (Chembio Diagnostics, 2016). Therefore, the results of this study may not be directly applicable for most physicians except those with access to

SD Bioline Dengue Duo. In the absence of RVR combo diagnostic performance in an actual clinical setting, and the lack of comparator for its only published phase II-equivalent case-control study conducted in a laboratory (Ainulkhir et al., 2018); even rough estimation of its accuracy and utility is not possible. Nevertheless, in view of the better performance of a combo RDT versus NS1-only assay as demonstrated in this study, RVR combo might also be better than the new RDT. Whatever it is, the same principles in dengue patient care and management as mentioned above in this section are applicable to RVR combo, that physicians using it should be vigilant when interpreting its result.

6.4 Implications of study for health policy making

The results of this study highlighted the discrepancies between dengue RDT diagnostic performance as reported by manufacturers and those evaluated by independent third-party researchers, as well as reduction in diagnostic performance in phase III cross-sectional studies when compared to phase II case-control laboratory-based studies. Health policy makers especially the MOH can benefit from this in several ways.

First of all, the results of this study demonstrated that dengue RDT all assays combined performed better than NS1-only assay. The latter produced too many FN and should not be encouraged to be used independently. Although the physicians in public clinics are currently also using dengue combo kit, many private practitioners specifically those in clinics may have NS1-only test due to higher cost of the former. Even then, most private general practitioners may not even use it to avoid additional financial burden to their patients (Loh, 2015). As such, the MOH should take effort to encourage the use of dengue RDT among general practitioners in the private sector, in particular in the form of combo test (Andries et al., 2012; Fry et al., 2011).

Secondly, the results of this study provided direct evidence on the good performance of the most evaluated dengue RDT – the SD Bioline Dengue Duo. When positive result on any of its individual components was taken as positive, it would be able to prevent most complications and deaths as a result of dengue. Health managers of clinical settings similar to the study site can utilise the evidence provided in this study to make an informed decision on the exact dengue RDT to be purchased and used in their settings, provided that they carefully appraised its results using the QUADAS-2 and STARD-guidelines as recommended above (P. M. Bossuyt et al., 2015; Whiting et al., 2011).

Thirdly, but most importantly, the health policy makers in Malaysia should formulate stricter regulations that mandate phase III evaluation of dengue RDTs prior to their approval for sales in Malaysia. Such research priority should be set by the MOH with the involvement and agreement of all relevant stakeholders such as the private healthcare providers, research institutions, and dengue RDT manufacturers themselves. If making phase III evaluation mandatory prior to the approval and sales of a RDT is too stringent and may stifle the growth of small brands with potentially good products, phase II evaluation performed together with established comparator should at least be available. Phase III or IV evaluation can be made mandatory within a certain period after the approval and sales of the RDT, also with a comparator, preferably the same one as in phase II for comparison (Committee for Medicinal Products for Human Use, 2009). Furthermore, in view of the superior performance of combo test, any new RDT that seeks to get evaluated for approval and sales in Malaysia should preferably be a combo test with all three NS1, IgM, and IgG components.

Finally, whether or not these regulations come into existence and get implemented, dengue RDT manufacturers should take the initiative to self-regulate to ensure the quality of both their products and the healthcare services their products would provide. For BluSense Diagnostics, this means adding IgM and IgG assays using the same technology to ViroTrack Dengue Acute NS1, and then evaluate this new combination anew with SD Bioline Dengue Duo. Having said that, dengue RIT currently in use that was never evaluated properly, such as RVR Dengue Combo NS1-IgG/IgM, should also be involved to produce directly comparable results to help health policy maker in decision making.

6.5 Implications of study for current dengue rapid diagnostic tests market

Evaluation studies, whether of new or existing dengue RDT, would increase the competition in current dengue RDT market. Evaluation of existing dengue RDT gives healthcare providers and patients assurance of its quality in diagnosing dengue accurately. Low quality dengue RDT or those that could not maintain its accuracy would be eliminated. The systematic review contained within this thesis have identified two pieces of evidence to support this.

Firstly, when dengue RDT first appeared in the market, the only option was IgM/IgG combined kit, so all evaluation studies would evaluate only this type of RDT. After the appearance of NS1 assay that could detect dengue earlier, only 11 out of 50 published articles evaluated IgM/IgG-only kit (S. D. Blacksell et al., 2011; Chen et al., 2016; Gunasekera, Senanayake, & Mendis, 2011; Hunsperger et al., 2009; Hunsperger et al., 2014; Mitra et al., 2016; Moorthy, Chandy, Selvaraj, & Abraham, 2009; Murugananthan, Coonghe, Ketheesan, & Noordeen, 2018; Pal et al., 2015; Pok et al., 2010; Pun, Shah, Gupta, Sherchand, & Pandey, 2012; Sugimoto et al., 2011).

Secondly, although the review identified a total of 23 brands of dengue RDT that were ever evaluated. Only several brands remained commercially available today in various parts of the world (Abbott, 2017; Chembio Diagnostics, 2016; J. Mitra, 2015; Wondfo Biotech, 2015).. Certain RDT was replaced by a better second generation kit, but most kits were no longer in production due to poor performance (Charrel & de Lamballerie, 2002; World Health Organization, 2009b).

On the other hand, evaluation of new dengue RDT informs health policy makers of upcoming new products that may be an alternative to the current ones. The incentives to change may be higher quality or lower price of the new RDT. For some manufacturers,

introduction of new RDT into the market would prompt them to lower the price (Director General of Health, 2015). For others, product innovation is unavoidable as it is the only long term solution to remain relevant in this field (Chembio Diagnostics, 2017; Teoh, 2018). Whether the strategy is price reduction or innovation, it would eventually lead to increased access to better dengue RDT for the patient populations who need it.

6.6 Implications of study for the healthcare system and population health

As a result of increased access to dengue RDT of higher quality at lower price, healthcare providers can provide more accurate point-of-care dengue diagnostics to patients with suspected dengue. More patients can be diagnosed earlier and managed appropriately at low-resource settings such as primary care and rural area, which would reduce the necessity to refer patients to secondary or tertiary hospitals for diagnosis, effectively decongesting their emergency departments for other patients who really need them. Early and accurate diagnosis in combination with timely intervention in primary care would also reduce dengue complication and the need for hospitalisation or prolonged hospital care, and eventually prevent unnecessary death from dengue. Such increment in effectiveness and efficiency would reduce dengue healthcare cost and eventually societal economic burden in general, which is in line with the aim of MOH to strengthen primary healthcare system in Malaysia (Tan, 2018).

On the other hand, the benefit of this study went beyond reducing mortality. Dengue RDT evaluation study would also benefit disease prevention. For dengue, the most used disease prevention method is vector control (Bowman et al., 2016). In Malaysia, vector control activities are carried out by district health office after investigating notifications from diagnosing physicians and confirming them to be dengue cases. In most cases, this confirmation of disease was based on the result of laboratory diagnosis (Packierisamy et al., 2015). Having dengue RDT with low accuracy in the arsenal would definitely produce inaccurate notification in the form of over-reporting due to FP and underreporting or absence of notification due to FN. The former would waste the resources of the district health office as vector control activities would be carried out where it was not needed. The latter would lead to further transmission of the disease due to the lack of vector control activities.

In the absence of phase III evaluation study of the dengue RDT currently in use, the extent of this over-reporting and underreporting cannot be estimated. Fortunately, the results of this study provided an avenue for their estimation. Assuming all cases were notified by diagnosing physicians and confirmed by the district health office based on the results of the dengue RDT, using ViroTrack Dengue Acute NS1 would produce as much as 8.3% over-reporting and 44.4% underreporting, while SD Bioline Dengue Duo all assays combined would produce 17.2% over-reporting and 23.2% underreporting (Figure 5.3). This means that the former RDT would misreport more than 50% of the patients with suspected dengue, as compared to around 40% of the latter. Moreover, although using SD Bioline Dengue Duo would waste twice as much resources to conduct vector control activities for FP cases; the ViroTrack would also miss doing that for twice as much of FN cases, which would be more detrimental to disease transmission.

Therefore, phase III dengue RDT evaluation study such as that presented in this thesis is important in assessing the accuracy of current passive dengue surveillance system. Combined with access to better dengue RDT at lower price, this disease surveillance system can be improved in its accuracy. Earlier and more accurate diagnosis made possible by dengue RDT would also improve the timeliness of the data captured by the surveillance. As a result, the vector control activities aimed at curbing the spread of dengue would become more efficient and effective, leading to reduction in DENV transmission and eventually dengue morbidity (Bowman et al., 2016; World Health Organization, 2012).

In summary, dengue RDT evaluation study conducted in this thesis contributes to the improvement in their quality and reduction in their price through competition, which increase patient access to them. Earlier and more accurate diagnosis for more patients in turn decreases dengue mortality and morbidity through case management and vector

control activities that are timely, efficient, and effective. At the same time, the accuracy of passive dengue surveillance system is also improved. Reduction in wastage allows reallocation of scarce healthcare resources to other areas that are more in need. The end result is a better healthcare system and healthier population.

6.7 Study strengths and limitations

The strength of this thesis was in its comprehensive systematic review and meta-analysis, as well as the sound methodology in its main study. The broad focus of the systematic review allowed the identification of findings not immediately obvious in the previously conducted primary studies included in the pooling, such as the modifying effect of study design on test diagnostic accuracy. The effect of factors such as previous exposure to dengue and disease phase was detected. The necessity to combine NS1, IgM, and IgG results together to improve SN, together with the most evaluated and likely better dengue RIT, were also identified. As such, the review helped in the proper and structured design of the main study and its pre-test presented in this thesis, which allowed them to in turn identify similar findings observed in the systematic review and meta-analysis.

Furthermore, instead of evaluating just the new dengue RDT, comparator dengue RIT was also evaluated side-by-side in the main study. This provided directly comparable actual performance of different dengue diagnostic tests. In fact, this phase III evaluation study presented in this thesis was the first and only one cross-sectional study conducted on prospectively and consecutively recruited patient in an actual clinical setting in Malaysia (Table 4.9). Previous studies were conducted exclusively in laboratory on collected samples, whatever their design and direction (Fry et al., 2011; Hunsperger et al., 2009; Hunsperger et al., 2014; Kumarasamy et al., 2007; Lam & Devine, 1998; Wang & Sekaran, 2010a; Zainah et al., 2009). In addition, the large study sample size in this study allowed the precision level of the overall test accuracy to fall around 5% from the point estimates. Therefore, its results were more likely to be representative of Malaysian population and generalizable to other similar clinical settings in this country.

The main limitation of the review and main study presented in this thesis was the use of only univariable analysis. However, this was unavoidable for the following reasons.

In the systematic review and meta-analysis, there was no statistical software capable of multivariable multivariate meta-regression. MIDAS and METANDI were only capable of univariable bivariate meta-regression, while METAREG - multivariable univariate analysis. Even for the latter, all analyses suffered from small cell size, making the primary studies unpoolable or underpowered to detect any meaningful difference. In the main study, the sample size was too small even for univariable analysis. But even though the clinic phase had adequate samples, odd ratio that can be adjusted using multiple logistic regression was less meaningful clinically than widely used accuracy parameters, and was used only in one previous study (Carter et al., 2015).

On top of that, multivariable analysis was most likely unnecessary, since the actual predictor of the diagnostic performance of any test was already known to be its ability to detect a disease biomarker as good as the reference. All other factors merely modified its performance indirectly through the biomarker level. The primary factors affecting the level of different diagnostic biomarkers in a dengue patient were logically deduced and proven to be previous exposure to dengue and disease phase. Although it was good to know which other factors remained significant after adjustment, doing so would not add any further practical value. This is because the purpose of the whole thesis was to evaluate the diagnostic performance of a new dengue RDT, where patients recruited should be as randomised as possible on all characteristics that could potentially modify the outcome variables. As such, the overall diagnostic accuracy and utility estimates for each index test as evaluated in the main study were adequate to answer the main objective. The various subgroup analyses, in this sense, served more as a proof of adequate randomisation through consecutive sampling, rather than a mere attempt to identify effect modifiers.

Finally, since the main purpose of a dengue RDT is to close the dengue diagnostic gap in a low-resource setting, by rapidly screening patients with suspected dengue and diagnosing it as early as possible for timely intervention, yet not missing patients that present late; an ideal dengue RDT should not have differential performance in detecting patients of diverse characteristics, be it timing of presentation, previous exposure to dengue, DENV-serotype and so on. This can only be achieved through the combination of antigen- and serology-based assays, of which SD Bioline Dengue Duo was the only one evaluated and proven to be most accurate if any positive test in this NS1/IgM/IgG combo was interpreted as dengue positive. This interpretation would produce least FN and consequently provide more benefit over risk to patients in general.

As such, notwithstanding the limitations of this study, its primary objective to evaluate the diagnostic performance of a new biosensors-based RDT – the ViroTrack Dengue Acute NS1, was adequately answered; that unless its combination with IgM and IgG assays using the same technology is proven to be better than that of the SD Bioline Dengue Duo, it cannot be a suitable alternative to this dengue RIT.

CHAPTER 7: CONCLUSION

In conclusion, the evaluation of the new dengue RDT - ViroTrack Dengue Acute NS1, demonstrated its rather good diagnostic performance and utility. However, the comparator RDT – SD Bioline Dengue Duo, due to its additional IgM and IgG components, when interpreted such that any positive result on any component for a patient was taken as dengue diagnosis, was significantly more sensitive but also significantly less specific as compared to the ViroTrack NS1 only kit. Although the SN of dengue RDT was found to be modified by study design, DENV serotype, disease phase, previous exposure to dengue infection, RDT brand, and reference tests used, in both the systematic review and the main study; SD Bioline Dengue Duo all assays combined demonstrated consistently better performance when compared to ViroTrack Dengue Acute.

In view of dengue RDT application mainly in low-resource settings for the purpose of early detection of acute dengue infection, the diagnostic utility of dengue RDT with higher SN is better as it would produce least FN. Missing less dengue patients allows timely clinical management to be administered to more of them, thus preventing more complications and deaths. Comparatively more vector control activities can also be carried out to prevent further transmission of disease. As such, notwithstanding ViroTrack Dengue Acute good performance and utility, due to the absence of serology components, its diagnostic performance and utility were inferior to that of SD Bioline Dengue Duo, despite the latter shortcomings being a RIT.

The implications of the findings in this thesis for health policy makers in Malaysia are as below. First of all, the ViroTrack Dengue Acute cannot yet be an alternative to a dengue combo kit like SD Bioline Dengue Duo, unless it is combined with IgM and IgG assays employing the same biosensors technology. Secondly, this ViroTrack combo kit,

if produced, should be evaluated anew in a clinical setting on the patient population it intends to help. Thirdly, this evaluation study should ideally be cross-sectional in design and conducted prospectively on consecutively recruited patients, with other dengue diagnostics such as SD Bioline Dengue Duo as comparator. Fourthly, it is advisable to re-evaluate any dengue RDT, especially RIT, which was not evaluated in such a way prior to its application to the current practice. And finally, in the future, new dengue RDTs should be properly evaluated with the results published prior to their application, especially when it involves subsidised public healthcare.

REFERENCES

Abbott. (2017). Dengue Fever. Retrieved from https://www.alere.com/en/home/products-services/infectious/dengue-fever.html

Ainulkhir, H., Fredolin, L., Azlan, M. A. N., Ahneez, A. H., Nazlina, I., & Ashraf, S. M. (2018). Evaluation of a commercial dengue combo rapid test kit for the detection of NS1 and IgM. *Trop Biomed, 35*(2), 511-520. Retrieved from <Go to ISI>://WOS:000435404000023

Alagarasu, K., Walimbe, A. M., Jadhav, S. M., & Deoshatwar, A. R. (2016). A meta-analysis of the diagnostic accuracy of dengue virus-specific IgA antibody-based tests for detection of dengue infection. *Epidemiol Infect, 144*(4), 876-886. doi:10.1017/S0950268815001922

Andries, A. C., Duong, V., Ngan, C., Ong, S., Huy, R., Sroin, K. K., . . . Buchy, P. (2012). Field evaluation and impact on clinical management of a rapid diagnostic kit that detects dengue NS1, IgM and IgG. *PLoS Negl Trop Dis, 6*(12), e1993. doi:10.1371/journal.pntd.0001993

Antunes, P., Watterson, D., Parmvi, M., Burger, R., Boisen, A., Young, P., . . . Donolato, M. (2015). Quantification of NS1 dengue biomarker in serum via optomagnetic nanocluster detection. *Scientific Reports, 5*, 16145. doi:10.1038/srep16145

http://www.nature.com/articles/srep16145#supplementary-information

Arima, Y., Matsui, T., Shimada, T., Ishikane, M., Kawabata, K., Sunagawa, T., . . . Oishi, K. (2014). Ongoing local transmission of dengue in Japan. *Western Pacific Surveillance and Response Journal, 5*(4), 27-29. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4318974/pdf/WPSAR.2014.5.4-027.pdf

Atun, R., Berman, P., Hsiao, W., Myers, E., & Yap, W. A. (2016). *MALAYSIA HEALTH SYSTEMS RESEARCH VOLUME I*. Malaysia Retrieved from http://www.moh.gov.my/penerbitan/Laporan/Vol%201_MHSR%20Contextual%20Analysis_2016.pdf

Beam, C. A. (1992). Strategies for improving power in diagnostic radiology research. *American Journal of Roentgenology, 159*(3), 631-637. doi:10.2214/ajr.159.3.1503041

Bernama. (2014). Budget: Dengue 'test kit' reduces clinic dependence on labs for blood tests. Retrieved from http://english.astroawani.com/budget2015-news/budget-dengue-test-kit-reduces-clinic-dependence-labs-blood-tests-45789

Bessoff, K., Delorey, M., Sun, W., & Hunsperger, E. (2008). Comparison of two commercially available dengue virus (DENV) NS1 capture enzyme-linked immunosorbent assays using a single clinical sample for diagnosis of acute DENV infection. *Clin Vaccine Immunol, 15*(10), 1513-1518. doi:10.1128/cvi.00140-08

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., . . . Hay, S. I. (2013). The global distribution and burden of dengue. *Nature, 496*, 504-507.

Bisordi, I., Rocco, I. M., Suzuki, A., Katz, G., Silveira, V. R., Maeda, A. Y., . . . Bessa, T. A. (2011). Evaluation of dengue NS1 antigen detection for diagnosis in public health laboratories, Sao Paulo State, 2009. *Rev Inst Med Trop Sao Paulo, 53*(6), 315-320. Retrieved from http://www.scielo.br/pdf/rimtsp/v53n6/a03v53n6.pdf

Blacksell, S. D. (2012). Commercial Dengue Rapid Diagnostic Tests for Point-of-Care Application: Recent Evaluations and Future Needs? *Journal of Biomedicine and Biotechnology, 2012*, 151967. doi:10.1155/2012/151967

Blacksell, S. D., Bell, D., Kelley, J., Mammen, M. P., Jr., Gibbons, R. V., Jarman, R. G., . . . Newton, P. N. (2007). Prospective study to determine accuracy of rapid serological assays for diagnosis of acute dengue virus infection in Laos. *Clinical And Vaccine Immunology: CVI, 14*(11), 1458-1464. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=17715330&site=ehost-live

Blacksell, S. D., Jarman, R. G., Bailey, M. S., Tanganuchitcharnchai, A., Jenjaroen, K., Gibbons, R. V., . . . Day, N. P. (2011). Evaluation of six commercial point-of-care tests for diagnosis of acute dengue infections: the need for combining NS1 antigen and IgM/IgG antibody detection to achieve acceptable levels of accuracy. *Clin Vaccine Immunol, 18*(12), 2095-2101. doi:10.1128/cvi.05285-11

Blacksell, S. D., Jarman, R. G., Gibbons, R. V., Tanganuchitcharnchai, A., Mammen, M. P., Nisalak, A., . . . Lalloo, D. G. (2012). Comparison of Seven Commercial Antigen and Antibody Enzyme-Linked Immunosorbent Assays for Detection of Acute Dengue Infection. *Clin Vaccine Immunol, 19*(5), 804-810. doi:10.1128/CVI.05717-11

Blacksell, S. D., Newton, P. N., Bell, D., Kelley, J., Mammen, M. P., Jr., Vaughn, D. W., . . . Day, N. P. J. (2006). The comparative accuracy of 8 commercial rapid immunochromatographic assays for the diagnosis of acute dengue virus infection. *Clinical Infectious Diseases: An Official Publication Of The Infectious Diseases Society Of America, 42*(8), 1127-1134. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=16575730&site=ehost-live

Bossuyt, P., Davenport, C., Deeks, J., Hyde, C., Leeflang, M., & Scholten, R. (2013). Chapter 11:Interpreting results and drawing conclusions. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.9.* Retrieved from https://methods.cochrane.org/sdt/handbook-dta-reviews

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., . . . Cohen, J. F. (2015). STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Clin Chem, 61*(12), 1446-1452. doi:10.1373/clinchem.2015.246280

Bossuyt, P. M. M., Reitsma, J. B., Linnet, K., & Moons, K. G. M. (2012). Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. *Clin Chem, 58*(12), 1636. doi:10.1373/clinchem.2012.182576

Bowman, L. R., Donegan, S., & McCall, P. J. (2016). Is Dengue Vector Control Deficient in Effectiveness or Evidence?: Systematic Review and Meta-analysis. *PLoS Negl Trop Dis, 10*(3), e0004551. doi:10.1371/journal.pntd.0004551

Brenner, H., & Kliebsch, U. (1996). Dependence of Weighted Kappa Coefficients on the Number of Categories. *Epidemiology, 7*(2), 199-202. Retrieved from http://www.jstor.org/stable/3703036

Buonora, S. N., Passos, S. R. L., do Carmo, C. N., Quintela, F. M., de Oliveira, D. N. R., dos Santos, F. B., . . . Daumas, R. P. (2016). Accuracy of clinical criteria and an immunochromatographic strip test for dengue diagnosis in a DENV-4 epidemic. *BMC Infect Dis, 16*, 37-37. doi:10.1186/s12879-016-1368-7

Carter, M. J., Emary, K. R., Moore, C. E., Parry, C. M., Sona, S., Putchhat, H., . . . Blacksell, S. D. (2015). Rapid Diagnostic Tests for Dengue Virus Infection in Febrile Cambodian Children: Diagnostic Accuracy and Incorporation into Diagnostic Algorithms. *Plos Neglected Tropical Diseases, 9*(2). doi:10.1371/journal.pntd.0003424

Charrel, R. N., & de Lamballerie, X. (2002). Low specificity of an immunochromatographic serological assay for diagnosis of dengue fever in travelers returning with malaria. *Clin Diagn Lab Immunol, 9*(6), 1400-1400. doi:10.1128/cdli.9.6.1400.2002

Chaterji, S., Allen, J. C., Jr., Chow, A., Leo, Y. S., & Ooi, E. E. (2011). Evaluation of the NS1 rapid test and the WHO dengue classification schemes for use as bedside diagnosis of acute dengue fever in adults. *Am J Trop Med Hyg, 84*(2), 224-228. doi:10.4269/ajtmh.2011.10-0316

Chembio Diagnostics. (2016). Chembio Agrees to Acquire Malaysia-based RVR Diagnostics and Establish Operations in Southeast Asia. Retrieved from https://globenewswire.com/news-release/2016/11/04/886679/0/en/Chembio-Agrees-to-Acquire-Malaysia-based-RVR-Diagnostics-and-Establish-Operations-in-Southeast-Asia.html

Chembio Diagnostics. (2017). About Us. Retrieved from http://chembio.com/about-us/

Chen, X. L., Chen, R., Gu, W. S., He, J., Cai, W. P., Li, J. J., . . . Yan, H. Y. (2016). Clinical evaluation of dengue RNA, NS1, and IgM for diagnosis of dengue in Southern China. *J Med Virol, 88*(1), 28-34. doi:10.1002/jmv.24314

Chew, C. H., Woon, Y. L., Amin, F., Adnan, T. H., Abdul Wahab, A. H., Ahmad, Z. E., . . . Lim, T. O. (2016). Rural-urban comparisons of dengue seroprevalence in Malaysia. *BMC Public Health, 16*(1), 824. doi:10.1186/s12889-016-3496-9

Chua, K. B. (2009). *Medical virology in Malaysia* (Vol. 24).

Cohen, A. L., Dowell, S. F., Nisalak, A., Mammen, M. P., Jr., Petkanchanapong, W., & Fisk, T. L. (2007). Rapid diagnostic tests for dengue and leptospirosis: antibody detection is insensitive at presentation. *Tropical Medicine & International Health: TM & IH, 12*(1), 47-51. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=17207147&site=ehost-live

Committee for Medicinal Products for Human Use. (2009). *Guideline on clinical evaluation of diagnostic agents*. European Medicines Agency Retrieved from https://www.ema.europa.eu/documents/scientific-guideline/guideline-clinical-evaluation-diagnostic-agents_en.pdf

Cordeiro, M. T., Braga-Neto, U., Nogueira, R. M., & Marques, E. T., Jr. (2009). Reliable classifier to differentiate primary and secondary acute dengue infection based on IgG ELISA. *PLoS One, 4*(4), e4945. doi:10.1371/journal.pone.0004945

Costa, V. G. d., Marques-Silva, A. C., & Moreli, M. L. (2014). A Meta-Analysis of the Diagnostic Accuracy of Two Commercial NS1 Antigen ELISA Tests for Early Dengue Virus Detection. *PLoS One, 9*(4), e94655. doi:10.1371/journal.pone.0094655

Cuzzubbo, A. J., Endy, T. P., Nisalak, A., Kalayanarooj, S., Vaughn, D. W., Ogata, S. A., . . . Devine, P. L. (2001). Use of recombinant envelope proteins for serological diagnosis of dengue virus infection in an immunochromatographic assay. *Clin Diagn Lab Immunol, 8*(6), 1150-1155. doi:10.1128/CDLI.8.6.1150-1155.2001

Deeks, J. J., Macaskill, P., & Irwig, L. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol, 58*(9), 882-893. doi:10.1016/j.jclinepi.2005.01.016

Director General of Health. (2015). Combo Rapid Test Kit Denggi dengan harga rendah untuk Pengamal Perubatan Swasta. Retrieved from https://kpkesihatan.com/2015/03/21/combo-rapid-test-kit-denggi-dengan-harga-rendah-untuk-pengamal-perubatan-swasta/

DOSM. (2016). *HOUSEHOLD INCOME AND BASIC AMENITIES SURVEY REPORT BY STATE AND ADMINISTRATIVE DISTRICT: SELANGOR*. Malaysia Retrieved from https://newss.statistics.gov.my/newss-portalx/ep/epFreeDownloadContentSearch.seam?cid=44915

Dussart, P., Petit, L., Labeau, B., Bremand, L., Leduc, A., Moua, D., . . . Baril, L. (2008). Evaluation of two new commercial tests for the diagnosis of acute dengue virus infection using NS1 antigen detection in human serum. *PLoS Negl Trop Dis, 2*(8), e280. doi:10.1371/journal.pntd.0000280

Dwamena, B. (2014). *MIDAS retouch regarding diagnostic accuracy meta-analysis.* Paper presented at the Stata Conference, Boston, MA

Endy, T. P., Anderson, K. B., Nisalak, A., Yoon, I.-K., Green, S., Rothman, A. L., . . . Gibbons, R. V. (2011). Determinants of inapparent and symptomatic dengue

infection in a prospective study of primary school children in Kamphaeng Phet, Thailand. *Plos Neglected Tropical Diseases, 5*(3), e975-e975. doi:10.1371/journal.pntd.0000975

Fagerland, M. W., Lydersen, S., & Laake, P. (2013). The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. *BMC medical research methodology, 13*(1), 91. doi:10.1186/1471-2288-13-91

Faraji, A., & Unlu, I. (2016). The Eye of the Tiger, the Thrill of the Fight: Effective Larval and Adult Control Measures Against the Asian Tiger Mosquito, Aedes albopictus (Diptera: Culicidae), in North America. *J Med Entomol*. doi:10.1093/jme/tjw096

Ferraz, F. O., Bomfim, M. R. Q., Totola, A. H., Ávila, T. V., Cisalpino, D., Pessanha, J. E. M., . . . Teixeira, M. M. (2013). Evaluation of laboratory tests for dengue diagnosis in clinical specimens from consecutive patients with suspected dengue in Belo Horizonte, Brazil. *Journal Of Clinical Virology: The Official Publication Of The Pan American Society For Clinical Virology, 58*(1), 41-46. doi:10.1016/j.jcv.2013.06.015

Florkowski, C. M. (2008). Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. *The Clinical Biochemist Reviews, 29*(Suppl 1), S83-S87. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2556590/

Fry, S. R., Meyer, M., Semple, M. G., Simmons, C. P., Sekaran, S. D., Huang, J. X., . . . Cooper, M. A. (2011). The diagnostic sensitivity of dengue rapid test assays is significantly enhanced by using a combined antigen and antibody testing approach. *Plos Neglected Tropical Diseases, 5*(6), e1199-e1199. doi:10.1371/journal.pntd.0001199

Gan, V. C., Lye, D. C., Thein, T. L., Dimatatac, F., Tan, A. S., & Leo, Y.-S. (2013). Implications of Discordance in World Health Organization 1997 and 2009 Dengue Classifications in Adult Dengue. *PLoS One, 8*(4), e60946. doi:10.1371/journal.pone.0060946

Gan, V. C., Tan, L.-K., Lye, D. C., Pok, K.-Y., Mok, S.-Q., Chua, R. C.-R., . . . Ng, L.-C. (2014). Diagnosing dengue at the point-of-care: utility of a rapid combined diagnostic kit in Singapore. *PLoS One, 9*(3), e90037-e90037. doi:10.1371/journal.pone.0090037

Gaynor, M. (2007). Competition and Quality in Health Care Markets. *Foundations and Trends® in Microeconomics, 2*(6), 441-508. doi:10.1561/0700000024

Gill, B. S. (2012). *Epidemiology of Dengue in Malaysia from 2005-2010 and Factors Contributing to Its Emergence.* (Doctor of Philosophy). University of Western Australia,

Gluud, C., & Gluud, L. L. (2005). Evidence based diagnostics. *Bmj, 330*(7493), 724-726. doi:10.1136/bmj.330.7493.724

Goldstein Research. (2018). Initiative Towards Healthy World: Global Dengue Vaccine Market. Retrieved from https://www.goldsteinresearch.com/pressrelease/initiative-towards-healthy-world-global-dengue-vaccine-market

Goubert, C., Minard, G., Vieira, C., & Boulesteix, M. (2016). Population genetics of the Asian tiger mosquito Aedes albopictus, an invasive vector of human diseases. *Heredity (Edinb), 117*(3), 125-134. doi:10.1038/hdy.2016.35

Groen, J., Koraka, P., Velzing, J., Copra, C., & Osterhaus, A. (2000). Evaluation of six immunoassays for detection of dengue virus-specific immunoglobulin M and G antibodies. *Clin Diagn Lab Immunol, 7*(6), 867-871. doi:10.1128/cdli.7.6.867-871.2000

Gubler, D. J. (1998). Dengue and Dengue Hemorrhagic Fever. *Clinical Microbiology Reviews, 11*(3), 480-496. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC88892/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC88892/pdf/cm000480.pdf

Gubler, D. J. (2004). COMMENTARY: Ashburn PM, Craig CF. Experimental Investigations Regarding the Etiology of Dengue. J Infect Dis 1907; 4:440–75. *J Infect Dis, 189*(9), 1744-1783. doi:10.1086/383418

Gunasekera, K. M., Senanayake, C. P., & Mendis, L. (2011). Dengue rapid strip test: A Sri Lankan experience. *Indian Journal of Pathology and Microbiology, 54*(4), 861-863. doi:10.4103/0377-4929.91537

Guzman, M., Alvarez, M., & Halstead, S. (2013). Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: An historical perspective and role of antibody-dependent enhancement of infection. *Archives Of Virology, 158*. doi:10.1007/s00705-013-1645-3

Hadinegoro, S. R. S. (2012). The revised WHO dengue case classification: does the system need to be modified? *Paediatrics and International Child Health, 32*(s1), 33-38. doi:10.1179/2046904712Z.00000000052

Hang, V. T., Nguyet, N. M., Trung, D. T., Tricou, V., Yoksan, S., Dung, N. M., . . . Simmons, C. P. (2009). Diagnostic accuracy of NS1 ELISA and lateral flow rapid tests for dengue sensitivity, specificity and relationship to viraemia and antibody responses. *Plos Neglected Tropical Diseases, 3*(1), e360-e360. doi:10.1371/journal.pntd.0000360

Harbord, R. M., & Whiting, P. (2009). metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *The Stata Journal, 9*(2), 211-229. Retrieved from http://www.stata-journal.com/sjpdf.html?articlenum=st0163

Hii, Y. L., Zaki, R. A., Aghamohammadi, N., & Rocklov, J. (2016). Research on Climate and Dengue in Malaysia: A Systematic Review. *Curr Environ Health Rep, 3*(1), 81-90. doi:10.1007/s40572-016-0078-z

Huang, C.-H., Kuo, L.-L., Yang, K. D., Lin, P.-S., Lu, P.-L., Lin, C.-C., . . . Wu, H.-S. (2013). Laboratory diagnostics of dengue fever: an emphasis on the role of commercial dengue virus nonstructural protein 1 antigen rapid test. *Journal Of Microbiology, Immunology, And Infection = Wei Mian Yu Gan Ran Za Zhi, 46*(5), 358-365. doi:10.1016/j.jmii.2012.07.011

Huits, R., Soentjens, P., Maniewski-Kelner, U., Theunissen, C., Van Den Broucke, S., Florence, E., . . . Bottieau, E. (2017). Clinical utility of the nonstructural 1 antigen rapid diagnostic test in the management of dengue in returning travelers with fever. *Open Forum Infect Dis, 4*(1). doi:10.1093/oid/ofw273

Hunsperger, E. A., Sharp, T. M., Lalita, P., Tikomaidraubuta, K., Cardoso, Y. R., Naivalu, T., . . . Margolis, H. S. (2016). Use of a rapid test for diagnosis of dengue during suspected dengue outbreaks in resource-limited regions. *Journal Of Clinical Microbiology, 54*(8), 2090-2095. doi:10.1128/JCM.00521-16

Hunsperger, E. A., Yoksan, S., Buchy, P., Nguyen, V. C., Sekaran, S. D., Enria, D. A., . . . Peeling, R. W. (2009). Evaluation of Commercially Available Anti-Dengue Virus Immunoglobulin M Tests. *Emerg Infect Dis, 15*(3), 436-440. doi:10.3201/eid1503.080923

Hunsperger, E. A., Yoksan, S., Buchy, P., Nguyen, V. C., Sekaran, S. D., Enria, D. A., . . . Margolis, H. S. (2014). Evaluation of Commercially Available Diagnostic Tests for the Detection of Dengue Virus NS1 Antigen and Anti-Dengue Virus IgM Antibody. *Plos Neglected Tropical Diseases, 8*(10). doi:10.1371/journal.pntd.0003171

idengue. (2020). MALAYSIA DENGUE INCIDENCE RATE & CASE FATALITY RATE FOR YEAR 2000-2019. Retrieved from http://idengue.arsm.gov.my/pdf/statistik.pdf#page=3

Imai, N., Dorigatti, I., Cauchemez, S., & Ferguson, N. M. (2015). Estimating dengue transmission intensity from sero-prevalence surveys in multiple countries. *PLoS Negl Trop Dis, 9*(4), e0003719. doi:10.1371/journal.pntd.0003719

Imai, N., Dorigatti, I., Cauchemez, S., & Ferguson, N. M. (2016). Estimating Dengue Transmission Intensity from Case-Notification Data from Multiple Countries. *PLoS Negl Trop Dis, 10*(7), e0004833. doi:10.1371/journal.pntd.0004833

Insitute for Medical Research. (2012). Objectives of Virology Unit. Retrieved from http://www.imr.gov.my/index.php/en/idrc/43-english-content/centre/about-idrc/virology-unit/136-objectives-of-virology-unit

Institute of Medicine (US) Forum on Microbial Threats. (2011). *The Causes and Impacts of Neglected Tropical and Zoonotic Diseases: Opportunities for Integrated Intervention Strategies*. doi: 10.17226/13087

J. Mitra. (2015). Company Profile. *J. Mitra & Co. Pvt. Ltd.* Retrieved from http://www.jmitra.co.in/home.aspx

JBI. (2015). *The systematic review of studies of diagnostic test accuracy*. Retrieved from Univeersity of Adelaide:

Kaddar, M., Milstien, J., & Schmitt, S. (2014). Impact of BRICS' investment in vaccine development on the global vaccine market. *Bulletin of the World Health Organization, 92*(6), 436-446. doi:10.2471/BLT.13.133298

Kan, W. T. (2018, 5 September 2018) *Dengue rapid tests used in primary clinic in Malaysia/Interviewer: Z. L. Chong*.

Khetarpal, N., & Khanna, I. (2016). Dengue Fever: Causes, Complications, and Vaccine Strategies. *J Immunol Res, 2016*, 6803098. doi:10.1155/2016/6803098

Kittigul, L., & Suankeow, K. (2002). Use of a rapid immunochromatographic test for early diagnosis of dengue virus infection. *European Journal of Clinical Microbiology and Infectious Diseases, 21*(3), 224-226. doi:10.1007/s10096-001-0691-z

Kohn, M. A., Carpenter, C. R., & Newman, T. B. (2013). Understanding the Direction of Bias in Studies of Diagnostic Test Accuracy. *Academic Emergency Medicine, 20*(1194-1206). doi: 10.1111/acem.12255

Krishnananthasivam, S., Fernando, A. N., Tippalagama, R., Tennekoon, R., De Man, J., Seneviratne, D., . . . De Silva, A. D. (2015). EVALUATION OF A COMMERCIAL RAPID TEST KIT FOR DETECTION OF ACUTE DENGUE INFECTION. *Southeast Asian J Trop Med Public Health, 46*(4), 602-610. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84979856721&partnerID=40&md5=cb940bff9c3baba4c1bb8c918839ba89

Kumarasamy, V., Zuridah, H., Asmah Hani, A. W., Mariam, M., & Chua, K. B. (2007). A comparative study on the performance of two commercial anti-dengue IgM assay kits. *Medical Journal of Malaysia, 62*(1), 85-86. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-34447127572&partnerID=40&md5=44a49367c70c52e4cb38044e9ea7c097

Lam, S. K., & Devine, P. (1998). Evaluation of capture ELISA and rapid immunochromatographic test for the determination of IgM and IgG antibodies produced during dengue infection. *Clin Diagn Virol, 10*(1), 75-81. doi:10.1016/s0928-0197(98)00002-6

Lambrechts, L., & Lequime, S. (2016). Evolutionary dynamics of dengue virus populations within the mosquito vector. *Curr Opin Virol, 21*, 47-53. doi:10.1016/j.coviro.2016.07.013

Lee, J., Kim, H.-Y., Chong, C.-K., & Song, H.-O. (2015). Development and clinical evaluation of a highly accurate dengue NS1 rapid test: from the preparation of a soluble NS1 antigen to the construction of an RDT. *Diagnostic Microbiology And Infectious Disease, 82*(2), 128-134. doi:10.1016/j.diagmicrobio.2015.03.009

Lee, J., Kim, K. W., Choi, S. H., Huh, J., & Park, S. H. (2015). Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical

Review for Clinical Researchers-Part II. Statistical Methods of Meta-Analysis. *Korean Journal of Radiology, 16*(6), 1188-1196. doi:10.3348/kjr.2015.16.6.1188

Leeflang, M. M. (2014). Systematic reviews and meta-analyses of diagnostic test accuracy. *Clin Microbiol Infect, 20*(2), 105-113. doi:10.1111/1469-0691.12474

Lijmer, J. G., Leeflang, M., & Bossuyt, P. M. (2009). *Proposals for a Phased Evaluation of Medical Tests*. In R. (MD) (Ed.), *Medical Tests-White Paper Series [Internet]*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK49467/

Lima, M. d. R. Q., Nogueira, R. M. R., Schatzmayr, H. G., & dos Santos, F. B. (2010). Comparison of three commercially available dengue NS1 antigen capture assays for acute diagnosis of dengue in Brazil. *Plos Neglected Tropical Diseases, 4*(7), e738-e738. doi:10.1371/journal.pntd.0000738

Liu, Y., Liu, J., & Cheng, G. (2016). Vaccines and immunization strategies for dengue prevention. *Emerg Microbes Infect, 5*(7), e77. doi:10.1038/emi.2016.74

Lo, Y. C., & Perng, G. C. (2016). Novel concept on antiviral strategies to dengue. *Curr Opin Virol, 18*, 97-108. doi:10.1016/j.coviro.2016.05.009

Loh, F. F. (2015). Don't blame us for not detecting dengue in patients earlier, say docs. *The Star*. Retrieved from https://www.thestar.com.my/news/nation/2015/02/01/gps-test-done-on-second-visit-dont-blame-us-for-not-detecting-dengue-in-patients-earlier-say-docs/

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). Chapter 10: Analysing and Presenting Results. *Cochrane*

Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. Retrieved from https://methods.cochrane.org/sdt/handbook-dta-reviews

Malaysia, M. o. E. (2015). *Executive Summary: Malaysia Education Blueprint 2015-2025 (Higher Education)*. Malaysia Retrieved from https://www.um.edu.my/docs/default-source/about-um_document/media-centre/um-magazine/4-executive-summary-pppm-2015-2025.pdf?sfvrsn=4

Mata, V. E., Passos, S. R. L., Hokerberg, Y. H. M., Berardinelli, G. M., dos Santos, M. A. B., Fukuoka, L. V. B., . . . de Oliveira, R. D. C. (2017). Accuracy and reliability of an NS1 rapid immunochromatographic test for DENV-1 diagnosis at the point of care and in the laboratory. *BMC Infect Dis, 17*. doi:10.1186/s12879-017-2679-z

McGee, S. (2002). Simplifying Likelihood Ratios. *Journal Of General Internal Medicine, 17*(8), 647-650. doi:10.1046/j.1525-1497.2002.10750.x

Miller, E., & Sikes, H. D. (2015). Addressing Barriers to the Development and Adoption of Rapid Diagnostic Tests in Global Health. *Nanobiomedicine, 2*, 6. doi:10.5772/61114

Ministry of Health. (2003). *Clinical Practice Guidelines: Dengue Infection in Adults*. Retrieved from http://www.acadmed.org.my/view_file.cfm?fileid=32

Ministry of Health. (2007). *Health Technology Assessment Report: Point of Care Testing (POCT)*. Retrieved from http://www.moh.gov.my/update2017/1361.pdf

Ministry of Health. (2009). *Pelan Strategik Pencegahan dan Kawalan Denggi 2009-2013*. Retrieved from http://www.moh.gov.my/images/gallery/Garispanduan/PELAN_DENGGI.pdf

Ministry of Health. (2010). *Departmental Policy of Pathology Services*. Retrieved from http://www.moh.gov.my/images/gallery/Polisi/201009-pathology.pdf

Ministry of Health. (2015). Clinical Practice Guidelines: Management of Dengue Infection in Adults (Third Edition). In. Putrajaya, Malaysia: Malaysia Health Technology Assessment Section.

Ministry of Health. (2016). *Pelan Strategik Denggi Kebangsaan 2016-2020*. Retrieved from https://books.google.com/books/about/MYCDCGP_PELAN_STRATEGIK_DENGGI_KEBANGSAA.html?id=pphaDwAAQBAJ

Ministry of Health. (2017). CASE DEFINITIONS FOR INFECTIOUS DISEASES IN MALAYSIA.

Mitra, S., Choudhari, R., Nori, H., Abhilash, K. P. P., Jayaseelan, V., Abraham, A. M., . . . Muliyil, J. (2016). Comparative evaluation of validity and cost-benefit analysis of rapid diagnostic test (RDT) kits in diagnosis of dengue infection using composite reference criteria: A cross-sectional study from south India. *J Vector Borne Dis, 53*(1), 30-36. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84962616297&partnerID=40&md5=649ff04be55057453e395bf964c9082f

Mohd-Zaki, A. H., Brett, J., Ismail, E., & L'Azou, M. (2014). Epidemiology of dengue disease in Malaysia (2000-2012): a systematic literature review. *PLoS Negl Trop Dis, 8*(11), e3159. doi:10.1371/journal.pntd.0003159

Moi, M. L., Takasaki, T., & Kurane, I. (2016). Human antibody response to dengue virus: implications for dengue vaccine design. *Trop Med Health, 44*, 1. doi:10.1186/s41182-016-0004-y

Moorthy, M., Chandy, S., Selvaraj, K., & Abraham, A. M. (2009). Evaluation of a rapid immunochromatographic device for the detection of IgM IgG antibodies to Dengue viruses (DENV) in a tertiary care hospital in South India. *Indian J Med Microbiol, 27*(3), 254-256. doi:10.4103/0255-0857.53210

Murugananthan, K., Coonghe, P. A. D., Ketheesan, N., & Noordeen, F. (2018). Comparison of a rapid immuno-chromatography assay with a standard ELISA for the detection of IgM and IgG antibodies against dengue viruses. *VirusDisease, 29*(2), 199-202. doi:10.1007/s13337-018-0440-x

Najioullah, F., Combet, E., Paturel, L., Martial, J., Koulmann, L., Thomas, L., . . . Cesaire, R. (2011). Prospective evaluation of nonstructural 1 enzyme-linked immunosorbent assay and rapid immunochromatographic tests to detect dengue virus in patients with acute febrile illness. *Diagnostic Microbiology And Infectious Disease, 69*(2), 172-178. doi:10.1016/j.diagmicrobio.2010.09.021

National Public Health Laboratory. (2018). Borang Permohonan Ujian. Retrieved from http://mkak.moh.gov.my/en/muat-turun/borang-dokumen-bahagian-penyakit/borang-permohonan-ujian.html

Olkowski, S., Forshey, B. M., Morrison, A. C., Rocha, C., Vilcarromero, S., Halsey, E. S., . . . Stoddard, S. T. (2013). Reduced risk of disease during postsecondary dengue virus infections. *J Infect Dis, 208*(6), 1026-1033. doi:10.1093/infdis/jit273

Osorio, L., Ramirez, M., Bonelo, A., Villar, L. A., & Parra, B. (2010). Comparison of the diagnostic accuracy of commercial NS1-based diagnostic tests for early dengue infection. *Virol J, 7*, 361-361. doi:10.1186/1743-422X-7-361

Packierisamy, P. R., Ng, C. W., Dahlui, M., Inbaraj, J., Balan, V. K., Halasa, Y. A., & Shepard, D. S. (2015). Cost of Dengue Vector Control Activities in Malaysia. *Am J Trop Med Hyg, 93*(5), 1020-1027. doi:10.4269/ajtmh.14-0667

Pal, S., Dauner, A. L., Mitra, I., Forshey, B. M., Garcia, P., Morrison, A. C., . . . Wu, S. J. (2014). Evaluation of dengue NS1 antigen rapid tests and ELISA kits using clinical samples. *PLoS One, 9*(11), e113411. doi:10.1371/journal.pone.0113411

Pal, S., Dauner, A. L., Valks, A., Forshey, B. M., Long, K. C., Thaisomboonsuk, B., . . . Wu, S. J. (2015). Multicountry prospective clinical evaluation of two enzyme-linked immunosorbent assays and two rapid diagnostic tests for diagnosing dengue fever. *J Clin Microbiol, 53*(4), 1092-1102. doi:10.1128/jcm.03042-14

Park, J.-H., & Lee, D.-W. (2012). Dengue Fever in South Korea, 2006–2010. *Emerg Infect Dis, 18*(9), 1525-1527. doi:10.3201/eid1809.111811

Parmley, M. C. (2006). *The Effects of the Confirmation Bias on Diagnostic Decision Making.* (PhD). Drexel University Retrieved from https://pdfs.semanticscholar.org/2570/f67e0d303b9c20269d408e666630509337 6a.pdf

PCORI. (2012). Standards in the Design, Conduct and Evaluation of Diagnostic Testing For Use in Patient Centered Outcomes Research In. Patient-Centered Outcomes Research Institute.

Peeling, R. W., Artsob, H., Pelegrino, J. L., Buchy, P., Cardosa, M. J., Devi, S., . . . Yoksan, S. (2010). Evaluation of diagnostic tests: dengue. *Nat Rev Microbiol, 8*(12 Suppl), S30-38. Retrieved from http://www.nature.com/nrmicro/journal/v8/n12_supp/pdf/nrmicro2459.pdf

Pok, K.-Y., Lai, Y.-L., Sng, J., & Ng, L.-C. (2010). Evaluation of nonstructural 1 antigen assays for the diagnosis and surveillance of dengue in Singapore. *Vector Borne*

And Zoonotic Diseases (Larchmont, N.Y.), 10(10), 1009-1016. doi:10.1089/vbz.2008.0176

Prado, P. S., Almeida, J. T. D., de Abreu, L. T., Silva, C. G., Souza, L. D., Gomes, M. C., . . . Romero, G. A. S. (2018). Validation and reliability of the rapid diagnostic test 'SD Bioeasy Dengue Duo' for dengue diagnosis in Brazil: a phase III study. *Mem Inst Oswaldo Cruz, 113*(8). doi:10.1590/0074-02760170433

Pun, R., Shah, Y., Gupta, G. P., Sherchand, S. P., & Pandey, B. D. (2012). Prognostic value of rapid test for diagnosis of dengue in Nepalese patients during 2010 epidemic. *Kathmandu University Medical Journal, 10*(37), 7-10. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84866748565&partnerID=40&md5=1e921783ffb42cd1b05ac8e49cc96cae

Rathakrishnan, A., Klekamp, B., Wang, S. M., Komarasamy, T. V., Natkunam, S. K., Sathar, J., . . . Sekaran, S. D. (2014). Clinical and Immunological Markers of Dengue Progression in a Study Cohort from a Hyperendemic Area in Malaysia. *PLoS One, 9*(3), e92021. doi:10.1371/journal.pone.0092021

Rathakrishnan, A., & Sekaran, S. D. (2013). New development in the diagnosis of dengue infections. *Expert Opin Med Diagn, 7*(1), 99-112. doi:10.1517/17530059.2012.718759

Rezza, G. (2014). Dengue and chikungunya: long-distance spread and outbreaks in naïve areas. *Pathog Glob Health, 108*(8), 349-355. doi:10.1179/2047773214Y.0000000163

Rodriguez-Barraquer, I., Solomon, S. S., Kuganantham, P., Srikrishnan, A. K., Vasudevan, C. K., Iqbal, S. H., . . . Cummings, D. A. (2015). The Hidden Burden of Dengue and Chikungunya in Chennai, India. *PLoS Negl Trop Dis, 9*(7), e0003906. doi:10.1371/journal.pntd.0003906

Rutten, F. H., Moons, K. G., & Hoes, A. W. (2006). Improving the quality and clinical relevance of diagnostic studies. *Bmj, 332*(7550), 1129. doi:10.1136/bmj.332.7550.1129

Sabin, A. B. (1952). Research on dengue during World War II. *Am J Trop Med Hyg, 1*(1), 30-50. doi:10.4269/ajtmh.1952.1.30

Sanchez-Vargas, L. A., Sanchez-Marce, E. E., & Vivanco-Cid, H. (2014). Evaluation of the SD BIOLINE Dengue Duo rapid test in the course of acute and convalescent dengue infections in a Mexican endemic region. *Diagn Microbiol Infect Dis, 78*(4), 368-372. doi:10.1016/j.diagmicrobio.2013.12.019

Sang, C. T., Hoon, L. S., Cuzzubbo, A., & Devine, P. (1998). Clinical evaluation of a rapid immunochromatographic test for the diagnosis of dengue virus infection. *Clin Diagn Lab Immunol, 5*(3), 407-409. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0031804098&partnerID=40&md5=82182798eb18f104303d3743e00e4bea

Schroeder, H. W., & Cavacini, L. (2010). Structure and Function of Immunoglobulins. *The Journal Of Allergy And Clinical Immunology, 125*(2 0 2), S41-S52. doi:10.1016/j.jaci.2009.09.046

Scitable. (2014). Dengue Viruses. Retrieved from https://www.nature.com/scitable/topicpage/dengue-viruses-22400925

Sekaran, S. D., & Soe, H. J. (2017). Issues in contemporary and potential future molecular diagnostics for dengue. *Expert Review of Molecular Diagnostics, 17*(3), 217-223. doi:10.1080/14737159.2017.1275963

Shamala, D. S. (2015). Laboratory Diagnosis of Dengue: A Review. *The International Medical Journal Malaysia, 14*(1).

Shan, X., Wang, X., Yuan, Q., Zheng, Y., Zhang, H., Wu, Y., & Yang, J. (2015). Evaluation of the diagnostic accuracy of nonstructural protein 1 Ag-based tests for dengue virus in Asian population: a meta-analysis. *BMC Infect Dis, 15*, 360. doi:10.1186/s12879-015-1088-4

Shepard, D. S., Undurraga, E. A., & Halasa, Y. A. (2013). Economic and disease burden of dengue in Southeast Asia. *PLoS Negl Trop Dis, 7*(2), e2055. doi:10.1371/journal.pntd.0002055

Shepard, D. S., Undurraga, E. A., Halasa, Y. A., & Stanaway, J. D. (2016). The global economic burden of dengue: a systematic analysis. *The Lancet Infectious Diseases, 16*(8), 935-941. doi:http://dx.doi.org/10.1016/S1473-3099(16)00146-8

Shih, H.-I., Hsu, H.-C., Wu, C.-J., Lin, C.-H., Chang, C.-M., Tu, Y.-F., . . . Sung, T.-C. (2016). Applications of a Rapid and Sensitive Dengue DUO Rapid Immunochromatographic Test Kit as a Diagnostic Strategy during a Dengue Type 2 Epidemic in an Urban City. *PLoS One, 11*(7), e0158437-e0158437. doi:10.1371/journal.pone.0158437

Shrivastava, A., Dash, P. K., Tripathi, N. K., Sahni, A. K., Gopalan, N., & Rao, P. V. L. (2011). Evaluation of a commercial Dengue NS1 enzyme-linked immunosorbent assay for early diagnosis of dengue infection. *Indian J Med Microbiol, 29*(1), 51-55. doi:10.4103/0255-0857.76525

Shu, P. Y., Chen, L. K., Chang, S. F., Yueh, Y. Y., Chow, L., Chien, L. J., . . . Huang, J. H. (2003). Comparison of capture immunoglobulin M (IgM) and IgG enzyme-linked immunosorbent assay (ELISA) and nonstructural protein NS1 serotype-specific IgG ELISA for differentiation of primary and secondary dengue virus infections. *Clin Diagn Lab Immunol, 10*(4), 622-630. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC164246/pdf/0008.pdf

Simonnet, C., Okandze, A., Matheus, S., Djossou, F., Nacher, M., & Mahamat, A. (2017). Prospective evaluation of the SD BIOLINE Dengue Duo rapid test during a dengue virus epidemic. *European Journal of Clinical Microbiology and Infectious Diseases, 36*(12), 2441-2447. doi:10.1007/s10096-017-3083-8

Šimundić, A.-M. (2009). Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC, 19*(4), 203-211. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4975285/

Skae, F. M. T. (1902). Dengue Fever in Penang. *Br Med J, 2*(2185), 1581-1582. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2402112/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2402112/pdf/brmedj08482-0009b.pdf

Standard Diagnostics Inc. (2008). SD Bioline Dengue Duo product insert. In S. D. Inc. (Ed.).

Sugimoto, M., Haseyama, M., Ishida, Y., Yoshida, S., & Kamiya, K. (2011). Evaluation of nonstructural 1 protein rapid test for dengue virus at the Narita Airport Quarantine Station, Japan. *Japanese Journal of Infectious Diseases, 64*(2), 169-170. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-79955600688&partnerID=40&md5=900bde1378d633fc1c85a978608b6b3e

Tambo, E., Chuisseu, P. D., Ngogang, J. Y., & Khater, E. I. (2016). Deciphering emerging Zika and dengue viral epidemics: Implications for global maternal-child health burden. *J Infect Public Health, 9*(3), 240-250. doi:10.1016/j.jiph.2016.02.005

Tan, S. C. (2018). Strengthening Malaysia's primary healthcare system. *The Star*. Retrieved from https://www.star2.com/health/wellness/2018/01/15/strengthening-malaysias-primary-healthcare-system/

Teoh, B. T. (2018, 5 Dec 2018) *Second Generation of SD Bioline Dengue Duo: Immunofluorescence Assay /Interviewer: Z. L. Chong*. Expert Forum: Innovation and Diagnostics - Pathway to Improve Dengue Patient Management in Malaysia.

Thayan, R. (2016). Molecular Diagnostics in Dengue. Retrieved from https://www.crc.gov.my/wp-content/uploads/2016/08/Molecular_Diagnostics_in_Dengue_CME_CRC_25_August_2016.pdf

Thayan, R. (2018, 14 July 2018). *National Laboratory Diagnostics - What test to use?* Paper presented at the One Day Dengue Symposium, MAHSA University.

Tricou, V., Vu, H. T. T., Quynh, N. V. N., Nguyen, C. V. V., Tran, H. T., Farrar, J., . . . Simmons, C. P. (2010). Comparison of two dengue NS1 rapid tests for sensitivity, specificity and relationship to viraemia and antibody responses. *BMC Infect Dis, 10*. doi:10.1186/1471-2334-10-142

United Nations. (2015). Sustainable Development Goal 3: Ensure healthy lives and promote well-being for all at all ages. Targets & Indicators. Retrieved from https://sustainabledevelopment.un.org/sdg3

van Panhuis, W. G., Luxemburger, C., Pengsaa, K., Limkittikul, K., Sabchareon, A., Lang, J., . . . Cummings, D. A. T. (2011). Decay and Persistence of Maternal Dengue

Antibodies among Infants in Bangkok. *Am J Trop Med Hyg, 85*(2), 355-362. doi:10.4269/ajtmh.2011.11-0125

Vickers, I. E., Harvey, K. M., Brown, M. G., Nelson, K., DuCasse, M. B., & Lindo, J. F. (2015). The performance of the SD BIOLINE Dengue DUO® rapid immunochromatographic test kit for the detection of NS1 antigen, IgM and IgG antibodies during a dengue type 1 epidemic in Jamaica. *J Biomed Sci, 22*(1). doi:10.1186/s12929-015-0164-9

Vivek, R., Ahamed, S. F., Kotabagi, S., Chandele, A., Khanna, I., Khanna, N., . . . Shet, A. (2017). Evaluation of a pan-serotype point-of-care rapid diagnostic assay for accurate detection of acute dengue infection. *Diagn Microbiol Infect Dis, 87*(3), 229-234. doi:10.1016/j.diagmicrobio.2016.09.020

Wallace, H. G., Lim, T. W., Rudnick, A., Knudsen, A. B., Cheong, W. H., & Chew, V. (1980). Dengue hemorrhagic fever in Malaysia: the 1973 epidemic. *Southeast Asian J Trop Med Public Health, 11*(1), 1-13.

Wang, S. M., & Sekaran, S. D. (2010a). Early diagnosis of dengue infection using a commercial dengue duo rapid test kit for the detection of NS1, IGM, and IGG. *American Journal of Tropical Medicine and Hygiene, 83*(3), 690-695. doi:10.4269/ajtmh.2010.10-0117

Wang, S. M., & Sekaran, S. D. (2010b). Evaluation of a commercial SD dengue virus NS1 antigen capture enzyme-linked immunosorbent assay kit for early diagnosis of dengue virus infection. *J Clin Microbiol, 48*(8), 2793-2797. doi:10.1128/jcm.02142-09

Whiting, P. F., Rutjes, A. S., Westwood, M. E., & et al. (2011). Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals Of Internal Medicine, 155*(8), 529-536. doi:10.7326/0003-4819-155-8-201110180-00009

WMA. (2013). Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. In. http://www.wma.net/en/30publications/10policies/b3/.

Wondfo Biotech. (2015). About us. Retrieved from https://en.wondfo.com.cn/about-us/

World Health Organization. (1997). *Dengue haemorrhagic fever : diagnosis, treatment, prevention and control — 2nd ed.* . Retrieved from

World Health Organization. (2009a). Dengue: guidelines for diagnosis, treatment, prevention and control - New edition.

World Health Organization. (2009b). *Evaluation of commercially available anti-dengue virus immunoglobulin M tests*. Retrieved from Philippines: http://www.who.int/tdr/publications/documents/diagnostics-evaluation-3.pdf?ua=1

World Health Organization. (2012). *Global strategy for dengue prevention and control 2012-2020*. Retrieved from World Health Organziation:

World Health Organization. (2013). *Neglected tropical diseases: Prevention, control, elimination and eradication. Report by the Secretariat* (A66/20). 66th World Health Assembly Retrieved from http://apps.who.int/gb/ebwha/pdf_files/WHA66/A66_20-en.pdf

Yong, Y. K., Thayan, R., Chong, H. T., Tan, C. T., & Sekaran, S. D. (2007). Rapid detection and serotyping of dengue virus by multiplex RT-PCR and real-time SYBR green RT-PCR. *Singapore Med J, 48*(7), 662-668.

Zainah, S., Wahab, A. H. A., Mariam, M., Fauziah, M. K., Khairul, A. H., Roslina, I., . . . Chua, K. B. (2009). Performance of a commercial rapid dengue NS1 antigen immunochromatography test with reference to dengue NS1 antigen-capture ELISA. *J Virol Methods, 155*(2), 157-160. doi:10.1016/j.jviromet.2008.10.016

Zhang, B., Salieb-Beugelaar, G. B., Nigo, M. M., Weidmann, M., & Hunziker, P. (2015). Diagnosing dengue virus infection: rapid tests and the role of micro/nanotechnologies. *Nanomedicine, 11*(7), 1745-1761. doi:10.1016/j.nano.2015.05.009

Zhang, H., Li, W., Wang, J., Peng, H., Che, X., Chen, X., & Zhou, Y. (2014). NS1-based tests with diagnostic utility for confirming dengue infection: a meta-analysis. *Int J Infect Dis, 26*, 57-66. doi:10.1016/j.ijid.2014.02.002

LIST OF PUBLICATIONS AND PAPERS PRESENTED

(a) Oral Presentation:

Chong, Z. L., Sekaran, S. D., Soe, H. J., Peramalah, D., Rampal, S., & Ng, C. W. (2018). The Diagnostic Accuracy of a Biosensor-Based Point-of-Care Test (POCT) and a Commercial Enzyme-Linked Immunosorbent Assay (ELISA) for the Diagnosis of Dengue Fever. Paper presented at the APRU 2018 Global Health Conference, University of Malaya, Kuala Lumpur, Malaysia.

(b) Poster Presentation:

Chong, Z. L., Sekaran, S. D., Soe, H. J., Peramalah, D., Rampal, S., & Ng, C. W. (2019). Diagnostic Accuracy of Three Dengue Diagnostic Tests for the Diagnosis of Acute Dengue in Malaysia. Paper presented at the 4th Asia Dengue Summit, Jakarta, Indonesia.

(c) Publication:

i. Zhuo Lin Chong, Shamala Devi Sekaran, Hui Jen Soe, Devi Peramalah, Sanjay Rampal Lekhraj Rampal, & Ng, C. W. (2020). Diagnostic Accuracy of Two Dengue NS1 Tests: New Biosensors-Based Rapid Diagnostic Test Versus Enzyme-linked Immunosorbent Assay. ASM Sc. J., 13 (Special Issue 5), 149-156. Retrieved from https://www.akademisains.gov.my/asmsj/article/diagnostic-accuracy-of-two-dengue-ns1-tests-new-biosensors-based/

ii. Chong, Z. L., Sekaran, S. D., Soe, H. J., Peramalah, D., Rampal, S., & Ng, C. W. (2020). Diagnostic accuracy and utility of three dengue diagnostic tests for the diagnosis of acute dengue infection in Malaysia. BMC Infect Dis, 20(1), 210. doi:10.1186/s12879-020-4911-5