Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations | Graduate School

2013

The Impact of Imprecision in HCV Viral Load Test Results on Clinicians' Therapeutic Management Decisions and on the Economic Value of the Test

Roberta M. Madej
Virginia Commonwealth University

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

Part of the Medicine and Health Sciences Commons

© The Author

Downloaded from
https://scholarscompass.vcu.edu/etd/3259

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

The Impact of Imprecision in HCV Viral Load Test Results on Clinicians' Therapeutic Management Decisions and on the Economic Value of the Test

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

Roberta Marie Madej
Bachelor of Science in Microbiology/ Medical Technology, San Jose State University, 1977
Master of Science, Clinical Laboratory Science, San Francisco State University, 1988
Master of Business Administration, Saint Mary's University, 1992

Director: William J. Korzun, Ph.D.
Associate Professor
Department of Clinical Laboratory Sciences

Virginia Commonwealth University
Richmond, Virginia
August 2013

Acknowledgement

This life-long venture would not have been completed without a village of support. The culmination of the research and dissertation presented here was directly influenced and supported by my academically diverse and insightful committee: David A. Holdford, PhD, William J. Korzun, PhD, Edward J. Moler, PhD, and Teresa S. Nadder, PhD. Special acknowledgement is given to my research director and advisor, Dr. Korzun for his willingness to consider a newer area for part of this work - health economics - and his uncompromising high standards for the final product.

The Health Related Sciences PhD program at VCU provides an alternative for aspiring students whose life situations exclude us from traditional doctoral programs. I am grateful to VCU for the establishment of an interdisciplinary curriculum that provides this opportunity while maintaining the academic rigor and expectations of traditional doctoral programs. Cecil B. Drain PhD, J. James Cotter, PhD, Paula K. Kupstas, PhD and Monica L. White are acknowledged and very much appreciated for their dedication to this program. Each scholar in my cohort of 2005, who in addition to their academic careers and mutual support, managed job changes, births, deaths, natural disasters, and other significant life events, were an inspiration to me.

Family, friends and colleagues re-energized me and motivated me. Too numerous to name are those in the molecular, diagnostic and clinical laboratory communities who patiently listened to my ideas and provided valuable insight, challenges and encouragement. Special acknowledgment is given to Giuseppe Colucci, MD for his substantial insight as a practicing

hepatologist and molecular science professional. Through the last eight years, several dear friends have supported me, and a very special group even prayed me through this. I thank them for their faith in me, even when mine wavered.

My ultimate gratitude is to my family. First, I thank my parents, Andrew and Beatrice Bianco for instilling in me an unfaltering love of learning and the disciplines of perseverance and hard work. To my daughter Annie, whose academic achievements inspire and challenge me. To Patrick, my son, whose fortitude encourages me, and to his wife Jenifer whose resilience heartens me. In an interesting reversal of roles, the anticipation of your pride in me for this accomplishment was exceptionally motivating. Finally, I give thanks to my husband, David. As young students in the 70's, our marriage vows - written by David- included the promise of supporting each other though school. He has steadfastly held to that promise for thirty-seven years though a bachelor's, two masters' and finally this doctoral program. Without him, none of these would have been possible. I am eternally grateful for his unconditional love and support in this aspect of our life together.

Table of Contents

List of Tables viii

List of Figures x

List of Abbreviations xi

Abstract xiii

Chapter One: Introduction 1

HCV Viral Load Testing: A Model 5

Statement of Problem 6

Research Questions 7

Research Hypotheses and Objectives 7

Null Hypotheses 7

The Value of Clinical Laboratory Test Information in Healthcare 8

Study Design 8

Rationale for this Study 9

Chapter Two: Literature Review 11

Clinical Laboratory Testing in Healthcare 11

Clinical Utility: How Clinicians Use and Value Laboratory Test Information 13

Health Economics and the Assessment of Value 20

Economic Assessments in Healthcare 24

Cost-benefit analysis 25

Cost-effectiveness analysis 27

Cost-utility analysis 28

Decision analysis and mathematical models 30

Hepatitis C and Viral Load Testing 44

The virus, transmission and course of disease 45

Treatment and patient management 46

Chronic HCV disease management and laboratory testing 48

Quantitative HCV RNA viral load testing 52

Justification for the Research 55

Contribution of the proposed research 58

Chapter Three: Methods 60

Survey Design 61

Selection of HCV VL test results and imprecision ranges 64

Selection of the Study Sample 66

Sample size 66

Variables in the Survey 67

Testing a Draft Version of the Survey 67

Final survey version 68

The Decision Analytic Model 70

Parameters for the development model 71

Testing the development model 71

Conclusions of development model testing and final model 73

Hypothesis Testing 79

Chapter Four: Results 81

Impact of Test Imprecision on Clinician Decisions 81

Sample 82

Hypothesis testing 84

H1 null hypothesis 84

H2 null hypothesis 84

Analysis of decisions from all respondents independent of order seen 87

Analysis of responses between two groups based on order of presentation of results with and without CI 90

Analysis based on order of the presentation of results, with or without CI 94

Further analysis of the decision outcomes for a VL of 560 IU/ml with CI 95

Impact of presentation order, with and without CI, on changes in decisions from the first to the second set of results 98

H1 and H2 null hypothesis findings 99

H3 null hypothesis 100

Other analyses for clinical utility 102

Decision to repeat the test 102

Summary of survey analyses 104

Determining Value and the Impact of Clinicians' Use of Test Information on that Value 106

Analyses to evaluate the economic impact of the clinicians' decisions 106

The value of HCV VL test information 111

Analysis of a representative population with several VLs 114

Impact of repeat-testing on test value 114

Analyzing model- outputs' sensitivity to variation in parameters 115

Section summary: determining value..... 118

Summary of Results 119

Chapter five: Discussion and Conclusions 120

Summary of the Study 120

Discussion of Survey Findings and Implications..... 121

Limitations of the survey research 126

Discussion of Economic Model Findings and Implications 128

Limitations of the health economic research 129

Conclusions and Recommendations for Future Research 130

List of References 136

Appendix A 151

Vita..... 159

List of Tables

1. Types of Health Economic Analyses 24

2. Three Studies with Published Intra-laboratory Reproducibility Studies 65

3. 95% CIs for HCV VL Results Presented in Survey 65

4. Predicted Decisions Based on HCV VL Test Results and Clinical Guidelines 66

5. Variables in the Survey 68

6. Responses to the Patient Management Questions from Draft HCV VL Survey 69

7. Decision Analytic Model Parameters 75

8. Comparison of Published and Inflation Adjusted Costs for HCV Health States 76

9. Respondent Preferences and Patient Management Practices 83

10. Sample Population Demographics 85

11. Multinomial Regression Statistics with Variable CI. All Responses 88

12. Multinomial Regression Statistics with Viral Load and the Presence or Absence of CI Variables. All Responses 89

13. All Data: Any Order Seen. Chi-square Analysis at Each VL level 90

14. Results of Pearson's Chi-square Test. All Three Decisions 91

15. Multinomial Regression Statistics: Two Populations Based on Responses to the First Presentation of HCV VL Only 92

16. Respondents' Decisions Split by Presentation of Results for First Question 93

17. Chi-square Test Result Based on Order Seen 95

18. Multinomial Regression of Decision Choices at 560 IU/ml 96

19. Logistic Regression Statistics for Continue Therapy Decision at HCV VL 560 IU/ml with CI 97

20. Correlation Matrix for the Decision to Continue Therapy at 560 IU/ml 98

21. Respondent Demographic, Practice and Preference Variables Demonstrating Significance in Respondent Decision Choices 103

22. Decisions Made at Each VL Level 105

23. Total Costs / QoL Comparisons from Guideline Directed Decisions and Survey Decisions Made With and Without CI 110

24. Value of Test Information: Difference in Total costs Realized as Cost Savings per Patient 112

25. Average Cost Savings per Patient in Removing Non-responders for Telaprevir Clinical Trial Population 115

26. Impact of Repeat Testing on Total Costs 116

27. Model Output Sensitivity to Costs and QoL Parameter Values 117

List of Figures

1. Example of a simple decision analytic tree model 33

2. Example of a simple Markov model 36

3. The natural course of HCV infection 47

4. Treatment strategies for chronic hepatitis C infection 51

5. Clinical path for HCV genotype 1 patients after 12 weeks of triple combination therapy 63

6. Percent change in cost as a function of percent retest rate 74

7. Groupings of data for analysis of survey responses 86

8. Decision path showing health states, costs, transition probabilities and QoL 107

9. Total costs resulting from guideline-dictated decisions and survey-guided decisions 110

10. Percent change in total costs from survey decisions made with and with CI compared to guideline-dictated decisions 111

11. Cost savings per patient at each VL due to the contribution of the HCV VL test in removing non-responders 113

12. Model sensitivity to costs and QoL parameter values 118

List of Abbreviations

AASLD - American association for the Study of Liver Diseases
AHRQ - Agency for Health Research and Quality
BCR - Benefit-Cost Ratio
CB - Cost Benefit
CE - Cost Effectiveness
CIs- Confidence Intervals
CEA - Cost Effectiveness Analysis
CED - Coverage with Evidence Development
CER - Comparative Effectiveness Research
CLSI - Clinical and Laboratory Standards Institute
CMS - Centers for Medicare and Medicaid Services
CU- Clinical Utility
DALY - Disability Adjusted Life-Year
EBM - Evidence Based Medicine
eGFR - Estimated Glomerular Filtration Rate
EGFR - Epidermal Growth Factor Receptor
ETR- End of Treatment Response
EVPI – Expected Value of Perfect Information
EVR - Early Virological Response
HBV- Hepatitis B Virus
HCV- Hepatitis C Virus
HE - Health Economics
HH - Hereditary Hemochromatosis
HIV - Human Immunodeficiency Virus
HPV - Human Papillomavirus
HTA - Health Technology Assessment
IDU - Injecting Drug User
ISPOR - International Society for Pharmacoeconomic Outcomes Research
ISO - International Standards Organization
MOA - Major Event Averted
NAT - Nucleic Acid Test
NHS - National Health Service
NICE - National Institute for Clinical Excellence
NIH - National Institutes of Health
NSCLC - Non Small Cell Lung Cancer
Pap - Papanicolaou
PPV - Positive Predictive Value
PT - Proficiency Testing

QALY - Quality Adjusted Life Year
QoL - Quality of Life
RNA - Ribonucleic Acid
RVR - Rapid Viral Response
SACGHS - Secretary's Advisory Committee on Genetics Health and Society
SVR - Sustained Virological Response
TKI - Tyrosine Kinase Inhibitor
VL - Viral Load
VOI - Value of Information

Abstract

THE IMPACT OF IMPRECISION IN HCV VIRAL LOAD TEST RESULTS ON CLINICIANS' THERAPEUTIC MANAGEMENT DECISIONS AND ON THE ECONOMIC VALUE OF THE TEST

By Roberta Marie Madej, MS, MBA

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2013

Major Director: William J. Korzun, Ph.D., Associate Professor
Department of Clinical Laboratory Sciences

Clinical laboratory test results are integral to patient management. Important aspects of laboratory tests' contributions are the use of the test information and the role they have in facilitating efficient and effective use of healthcare resources. Methods of measuring those contributions were examined using quantitative HCV RNA test results (HCV VL) in therapeutic management decisions as a model. Test precision is important in those decisions; therefore, the clinical use was evaluated by studying the impact that knowledge of inherent assay imprecision had on clinicians' decisions.

A survey describing a simulated patient at a decision point for HCV triple-combination therapy management was sent to 1491 hepatology clinicians. Participants saw HCV RNA results at five different levels and were asked to choose to: continue therapy, discontinue therapy, or repeat the test. Test results were presented both with and without the 95% confidence intervals

(CIs). Three of the VLs had CIs that overlapped the therapeutic decision level. Participants saw both sets of results in random order. Demographics and practice preferences were also surveyed.

One-hundred-thirty-eight responses were received. Adherence to clinical guidelines was demonstrated in self-reported behaviors and in most decisions. However, participants chose to repeat the test up to 37% of the time. The impact of the knowledge of assay imprecision did not have a statistically significant effect on clinicians' decisions.

To determine economic value, an analytic decision-tree model was developed. Transition probabilities, costs, and Quality of Life values were derived from published literature. Survey respondents' decisions were used as model inputs. Across all HCV VL levels, the calculated test value was approximately $2600, with up to $17,000 in treatment-related cost savings per patient at higher HCV VLs. The test value prevailed regardless of the presence or absence of CIs, and despite repeat testing.

The calculated value in cost savings/patient was up to 100 times the investment for HCV VL testing. Laboratory tests are investments in efficient uses of healthcare resources. Proper interpretation and use of their information is integral to that value. This type of analysis can inform institutional decisions and higher level policy discussions.

Chapter One: Introduction

Clinical laboratory test results are integral to patient care and management. At times the information provided by test results is pivotal in the choice or evaluation of a clinical strategy. Laboratory results are also used in developing, implementing and measuring the outcomes of clinical practice and evidence based medicine. Whether a clinician is screening individuals to determine disease risk, testing to inform diagnosis or prognosis, monitoring disease states, managing chronic health conditions, or individualizing therapeutic interventions, laboratory test information contributes to shaping the utilization of other medical services. Laboratory test results are used in quality measures of healthcare, objective clinical outcome measures, pay-for-performance measures, and as factual evidence in therapeutic clinical trials (The Lewin Group, 2005, 2008). While those who use laboratory test information may consider test results unconditionally robust, laboratory professionals understand that they are produced by assays, instruments, processes and people, all of which contribute to imprecision in that information. Clinicians and policy makers may not recognize the limitations of test results that can arise from the variability in the three phases of laboratory testing (pre-examination, examination and post-examination phases). These limitations should be considered when interpreting the results. Furthermore, they may not appreciate the strategies laboratory professionals develop to mitigate test result variability. Even for tests where little imprecision exists, research shows that laboratory test information is not always interpreted or used correctly. The concepts of disease probability are difficult, and management plans derived from test results may not be followed

optimally (Bramwell, West, & Salmon, 2006; Cahan, Gilon, Manor, & Paltiel, 2003; Gosh, Gosh, & Erwin, 2003; Steurer, Fischer, Bachmann. Koller & der Reit, 2002). This adds another facet to the quality and consistency of laboratory tests; the quality of the outcome of test information is dependent on the correct and optimal use of the results.

Conversely, laboratory professionals, whose customary focus are implementing the strategies to produce the most precise and accurate result, may not be aware of the broadest use of the test information by clinicians and policy makers. Evaluations of clinical laboratory tests have traditionally centered on the analytical characteristics of assay methods. The conventional laboratory test implementation plan includes verification or validation of sensitivity, specificity, precision, accuracy, reproducibility, and reportable range of the examination phase by best practice (Clinical and laboratory Standards Institute [CLSI] EP5A2, 2004; CLSI EP15A2, 2005) and in the United States, by law (Clinical Laboratory Improvement Act, 2003; Medical Device Quality Systems Regulation, 2008). The adoption of broader quality systems practices now necessitate consideration of the pre and post examination phases as part of the total testing process (CLSI GP26A3, 2004). In the fullest scope of practice, laboratory professionals are moving beyond the focused analytical evaluations to consider the applicability of the measurand to the disease process. However, there is little formal work assessing how the information generated from laboratory testing is interpreted and incorporated into the medical decision-making process. This is becoming more necessary as the field of companion diagnostics grows, and constraints in healthcare resources demand that the most effective, informative and efficient services and interventions be used.

Medical providers have the responsibility to evaluate and employ those interventions and services that will provide the best outcome for their patients and for the healthcare system

(Huninck et al., 2006; Price & Christenson, 2007). Clinical laboratory tests can be valuable tools in this effort, and are also subject to the same standard. The ultimate value of a laboratory test is more completely determined by its informative impact in medical decisions. The proper test must be ordered, and the information interpreted in context of the limitations of the tests and of the specific characteristics of the patient (Lundberg, 1998; Price & Christenson, 2007; Van Walraven & Naylor, 1998). The analytical characteristics of the test, the implementation of the test, and the actual performance of the test, added to its interpretation and application by the medical decision maker, define the utility and value of the results. It is important then as part of an assessment of a laboratory test's clinical utility, to understand the uses of that test's information, particularly when those uses are integrated in clinical practice guidelines.

With the increasing scrutiny on the most effective use of healthcare resources, it is also important to understand the economic value that laboratory tests have in the healthcare system. In most cases, they are used as information to reduce the uncertainty in medical decisions. Growing demands on expanding technological choices and resources require prudent utilization of healthcare dollars, and provoke increasing demands for evidence-based information as well as for the evaluation of how that information is used (Drummond & Sculpher, 2006; Tunis, 2004). Health economics has been employed for nearly two decades in some countries for the selection, pricing and reimbursement of therapeutics, and it is now being applied to a broader scope of health care services and programs (Drummond & Sculpher, 2006). In the United Kingdom, the National Health Service's (NHS) National Institute for Clinical Excellence (NICE) describes the process for evaluating healthcare technologies in three guidance documents: *Updated guide to the Methods of Technology Appraisals* (NICE, 2008), *Guide to Single Technology Appraisal Process* (NICE, 2010) and *Guide to the Multiple Technology Appraisal Process* (NICE, 2010).

Clinical effectiveness and cost effectiveness are aspects of the appraisals for those technologies approved by the NHS. An important aspect of technology assessment in the United Kingdom now includes the projected financial impact on healthcare resources. It is incumbent on the sponsor of a drug, intervention or device to demonstrate its cost effectiveness. There are now various types of Health Technology Assessment (HTA) requirements, which include clinical and economic evaluations in over 30 countries. A European collaborative project, the European Network for HTA (EUnetHTA) has been established in response to the need for developing a sustainable network to define and refine practices in this area (Minshall, 2008; Kristensen, 2008). The United States is also playing a role in these developments. The Agency for Healthcare Research and Quality (AHRQ) is a partner organization to EUnetHTA and guidelines for submission of economic assessments of therapeutics are already established (Fry, Avery, & Sullivan, 2003). As a part of the American Recovery and Reinvestment Act, the United States committed billions of dollars for comparative effectiveness research (CER), measuring costs and clinical utility of interventions, technologies, and services (Gaba, 2009). As one of the 12 agencies within the U.S. Department of Health and Human Services, the Agency for Health Research and Quality funds and disseminates comparative effectiveness and technology assessment research (http://www.ahrq.gov/about/).

Clinical laboratory testing, which has provided benchmark and experimental endpoint data to many health economic studies of therapeutics, devices and services, is now becoming the subject of these types of analyses; and its importance is growing. For example, while concerned with providing broader access to the appropriate use of genetic test results, the Secretary's Advisory Committee on Genetics Health and Society (SACGHS) made its first recommendation that evidence based information to drive coverage policies for that access (SACGHS, 2006).

Pharmacoeconomic experts criticize the lack of utility and value evidence for diagnostics (Ramsey et al., 2006). Because non-laboratory clinical interventions are different from laboratory information services, it is essential that methods to evaluate the clinical and economic utility of laboratory tests be appropriate to in vitro diagnostics.

HCV Viral Load Testing: A Model

The laboratory tests that drive the decisions for Hepatitis C disease management provide one model to examine the value of those tests in the decision-making process. Hepatitis C viral load assays (HCV VL) quantitatively measure the amount of HCV RNA in patients' plasma by nucleic acid testing (NAT), and the test results are a surrogate marker for the disease burden. HCV VL was integral to research on the natural history of the disease and is used in the clinical trials of therapeutics to treat it. The test now has a role in determining the management path of HCV patients. The American Association for Liver Diseases (AASLD) Clinical Practice guidelines and the American Gastroenterological Association recommend quantitative HCV RNA testing for viral load at several points during the diagnosis and management of the disease (Dienstag & McHutchinson, 2006; Ghany, Strader, Thomas, & Seeff , 2009; Ghany, Nelson, Strader, Thomas, & Seeff, 2011). HCV Viral load is considered one of the main sources of information for tailoring therapy and monitoring compliance (Chevaliez & Pawlotsky, 2009; Mangia et al., 2008; Paxton, 2011; Terrault et al., 2005; Zeuzem et al., 2003). The latest triple combination therapies for HCV genotype 1 patients include specific treatment recommendations based on absolute values of HCV VL (Ghany et al., 2011; Merck Sharp & Dohme Corporation, 2011; Vertex Corporation, 2011). Though HCV VL results are used as absolute values or relative value changes, there is little discussion of the inherent variability in viral load results except the published recommendations that patients be monitored with the same test throughout

their course. Global proficiency test assessments and shared-sample studies have demonstrated the variability in NAT (Best, Gust, Johnson, McGavin, & Dax, 2000; Schirm, van Loon, Valentine-Thon, Klapper, Reid, & Cleator, 2002; Yen-Lieberman et al., 1996). Practicing clinicians and laboratory professionals have published their observations regarding the variability of the assays and their possible impact on the decisions being made (Laperche et al., 2007, Pawlotsky, 1997). There are few professionals who serve in both capacities, and as VL testing becomes more mainstream, clinicians' understanding of this variability and its resulting impact on their decisions have not been studied.

Statement of Problem

Conventional assessments of diagnostic tests are technically oriented and focus on the performance characteristics of the analytical process and technologies. Diagnostic tests, however, are tools that produce information which is used in medical decisions. As it informs a decision, there are two significant variables to the contribution of a diagnostic test: 1) the overall test performance and 2) the use of the information. These are interdependent and contribute to the assessment of the utility and resulting value of a clinical laboratory test. An example of this is found in HCV VL testing to determine and direct treatment strategies. The test and methods have been technically validated and clinical trials support the basic utility of the results. Understanding the limitations, the use of the information resulting from these tests by clinicians, and the influence that knowledge of the limitations may have on their decisions have not been evaluated. Additionally, the value that the test information brings to specific healthcare situations and, if this value changes with the quality of the test, have not been assessed.

Research Questions

Will knowledge of the imprecision in quantitative laboratory test results affect clinicians' use of the information in managing patients? Can this be measured?

Can health economic tools (decision analytic models) be used to determine the value of laboratory test information?

Will changes in the use of test information result in changes in the value of the laboratory test?

Research Hypotheses and Objectives

This research studied the decisions made by clinicians using the viral load results from quantitative HCV RNA tests, with and without knowledge of the test imprecision (expressed as the 95% confidence intervals). In addition to determining if there were changes in decisions associated with the knowledge, the value of the test information was calculated using a decision analytic model. The overall objective was to determine if the clinical utility (CU), measured as patient management decisions of clinicians using the results and the value of the laboratory test information measured by cost savings, will change if clinicians are aware of the imprecision of the results.

Null Hypotheses

The hypotheses for this research were:

H1 null hypothesis: The presentation of HCV VL results with 95% confidence intervals will have no impact on the treatment decision for HCV genotype 1 patients.

H2 null hypothesis. There is no difference in the impact of awareness of CIs on the treatment decisions between different levels of HCV VL.

H3 null hypothesis. Clinicians' experience and demographics will have no impact on treatment decisions using HCV VL results presented with and without confidence intervals.

The Value of Clinical Laboratory Test Information in Healthcare

In addition to testing the hypotheses, the impact of these decisions on the economic value of the test information was evaluated through the use of decision analytic model outcomes.

Study Design

A non-experimental design using a survey targeted to a specific cohort of decision makers was employed. The survey was developed for clinicians who use HCV VL test results in the management of their chronic HCV patients. A defined clinical scenario was used describing a simulated HCV genotype 1 patient who had just completed 12 weeks of triple combination therapy with telaprevir, pegylated interferon-alpha and ribavirin. Different levels of quantitative HCV RNA results for the patient were presented to survey participants in the traditional manner, without confidence intervals, and also with the range of results that represent the 95% confidence intervals, which is an expression of the imprecision of the test. The order of presentation of these two sets of results was randomized. Participants were asked to decide whether to continue the patient onto double combination therapy, to discontinue therapy, or to repeat the test. The distribution of responses was analyzed for differences in the likelihood of using the test information (not repeating the test) as well as differences in the ultimate therapy decision made by those who used the test result. Further, the proportions of clinicians selecting each of the possible options were used to populate a decision analytic model of the case. Through the model, the expected value of the test results at each VL level and in a population of mixed VL levels, with and without corresponding confidence intervals was calculated.

Rationale for this Study

This research evaluated some of the interrelated factors in the provision and use of quality laboratory test results. In order for laboratory test information to be effective, the result must be robust and it must be interpreted and used appropriately in the correct context. As information providers, laboratory professionals make assumptions about the understanding and application of the information by the receivers of it without really knowing if those assumptions are accurate. This research examined whether the knowledge of the inherent imprecision of VL assays changes the decisions clinicians will make from those test results. There are recommendations to routinely report or make available on request the imprecision budgets with laboratory test results (CLSI C51A, 2011); yet there is no research to show that this addition would provide quality information to the clinician or to the patient. In addition, because the survey in this study was based on a well-defined clinical path where the use of the test is recommended to drive a specific treatment management decision, intended adherence to these guidelines was also evaluated.

The quality of a laboratory test result and its use in practice influences its value. Current health economic publications use clinical pathways and decision analytic models, assuming that clinicians invariably make the most suitable patient management choice based on the information received. Research has shown that this presumption is simplistic (Bramwell et al., 2006; Steurer et al., 2002), and that the consequences of misinterpretation and misuse of test results are not well understood or appreciated (Gandhi et al., 2006). Additionally, the actual use of information by clinicians, is rarely considered in health economic analyses.

In addition to patient diagnosis, prognosis, and medical management, the results of laboratory tests are inputs and endpoints in clinical trials and shape health economic assessment models; however, the value of the test information itself is not routinely measured or discussed. When

economic analyses of laboratory tests are undertaken, tools commonly used such as cost-benefit, cost-effectiveness and cost-utility analyses may not fully define the true impact of the information. Therefore, exploring the use of value calculations and clinical utility measures to evaluate the contributions of laboratory tests is important to the field of clinical laboratory science, health economics research, and healthcare resource assessment.

This study examined elements of the assumptions in current clinical research: the development of clinical decisions based on laboratory test information, compliance with guidelines in the face of test limitations, and the appropriateness of current methodologies to study laboratory test contributions to healthcare. The objective is that the outcome of this work provides insight, strategies, and tools specific to the assessment of laboratory test information's utility and value. If successful, these tools could be applied to other value analyses of laboratory tests in diverse healthcare situations.

Chapter Two: Literature Review

Clinical Laboratory Testing in Healthcare

Clinical laboratory testing is pervasive in most aspects of healthcare. Approximately 6.8 billion laboratory tests are performed annually in the United States. Over 4,000 different tests are available for clinical use, and approximately 500 of the over 1,162 tests reimbursed by Medicare are performed on a regular basis. In addition to being used as indicators of wellness, disease and patient treatment management, the results from testing are used in public health surveillance, in clinical trials, and in research. (The Lewin Group, 2008.)

Laboratory testing contributes to health management decisions from birth. Approximately four million newborns are screened by tests for congenital and metabolic disorders annually (Little & Lewis, 2008). For the 5,000 infants with severe disorders, more specific testing, and for some, life-long monitoring by laboratory tests will follow. Adults are accustomed to periodic screening tests such as cholesterol, fasting blood glucose, fecal occult blood, urinalysis, Papanicolaou (Pap) testing and prostate specific antigen (PSA) tests. Not only are test results incorporated into specific guidelines (America Diabetes Association [ADA], 2010; Nelson, Moser, Gaffey, & Waldron, 2009), but a majority of adults expect and desire these as part of their routine physical examinations (Oboler, Prochazka, Gonzales, Xu, & Anderson, 2002).

Advancements in treatment modalities, technological progress, and easy access to rapidly disseminated research information have facilitated the implementation of a growing number of new applications for medical laboratory testing. A relevant example of this is the establishment

of molecular tests to direct, monitor and manage patients with serious viral infections such as human immunodeficiency virus (HIV), hepatitis B virus (HBV) and hepatitis C virus (HCV). The role of quantitative molecular assays for viral load testing has been the subject of investigation and debate in several viral diseases since technological improvements made this type of measurement feasible and more robust. The precision and reproducibility of these tests continues to be discussed, particularly as they are integral to therapeutic management decisions (Chevaliez & Pawlotsky, 2009; Pisani et al., 2009).

Nucleic acid quantification for viral load was first applied in AIDS research to estimate the extent of HIV infection by measuring the number of RNA genome copies in the plasma (Ho, 1996). As molecular techniques were further refined, greater understanding of the pathogenicity of the virus, disease prognosis, and treatment success indicators developed (Mellors et al., 1996). Viral load measurements became surrogate markers of disease clearance in therapeutic trials; and the data gathered from those trials were applied to patient management guidelines when the agents were approved for therapy. Quantitative HIV RNA testing used with CD4+ counts for therapy initiation and serial quantitative HIV RNA tests for monitoring viral load during therapy are now part of standard practice in the management of AIDS (Hammer et al., 2008). Similar strategies have been applied to other viral diseases and several quantitative molecular assays have become integrated into patient management protocols. Examples include quantitative HBV DNA with genotyping assays which are integral to the selection of appropriate therapeutic regimens for hepatitis patients (Valsamakis, 2007). Quantitative HCV RNA tests are used to monitor patients on therapy, detect drug resistance, and follow patient compliance (Ghany et al., 2009; Paxton, 2011). Further, human cytomegalovirus (HCMV) viral load results distinguish between viral disease and organ rejection in transplant patients, facilitating the appropriate

choice between conflicting management strategies (Griffiths, Whitley, Snydman, Singh, & Boeckh, 2008).

The adoption of molecular viral load markers is one example of similar scenarios utilizing test results as evidence in clinical practice guideline development and in patient management decisions. The expanding recognition of companion diagnostics and the utilization of laboratory test results in all aspects of healthcare present the medical community with the challenge of determining their most effective and efficient applications. Consequently, when evaluating new laboratory tests, investigations of clinical utility and economic value will become as necessary as assessing analytical performance and clinical outcomes (Ramsey et al., 2006).

Clinical Utility: How Clinicians Use and Value Laboratory Test Information

Today the comprehensive assessment of necessary medical services includes an evaluation of clinical utility. The term clinical utility is used throughout healthcare but remains subjectively interpreted. First appearing in medical literature in 1961, its use has increased steadily, though a single consensus definition has not been developed (Smart, 2006). The characterization of clinical utility often reflects the perspective of the stakeholder; and the outcomes measured delineate the scope. For a laboratory test, clinical utility encompasses more than just the ability of the result to define a clinical state or disease. It is a multidimensional parameter in which the stakeholders are patients, clinicians, and others in healthcare. Evidence that a test result facilitates improved patient outcomes or that it is useful to those making medical decisions have been offered as definitions (The Lewin Group, 2009). Broader descriptions include economic and social outcomes (Grosse & Khoury, 2006).

A search of the term clinical utility in the medical literature identified several common attributes synonymous with clinical effectiveness. From the clinician's perspective, four main

characteristics are considered: appropriateness, acceptability, practicability and accessibility. For a service or an intervention to be appropriate, clinical evidence should demonstrate its impact on treatment and its importance to clinical decision making. To be acceptable, the effect on a patient's course, quality of life, and on the clinician's practice, are considered. The training and knowledge needed to implement the service and to maintain the functionality of using it are characteristics of practicability. Accessibility incorporates availability, costs, cost effectiveness and resources (Smart, 2006).

For laboratory tests, characteristics of appropriateness and acceptability, as defined here, are normally assessed when new tests are established. By regulation, manufacturers who sell *in vitro* diagnostic tests in the United States must submit data supporting the safety and effectiveness of their tests, either by showing substantial equivalence to pre-existing tests or demonstrating it through studies. The intended use of the test in question, the relative risk of the test and its placement in the healthcare system all impact the type of submission and the data needed to support it (USFDA 21CFR807, 21CRF814, 21CFR860). This is accomplished through clinical trials, where the new tests are validated analytically and evaluated according to clinical outcomes, or compared to other tests for which clinical outcomes have been established. Depending on the test and the scientific environment, intended use statements, clinical and professional guidelines, and marketing or post-market studies may incorporate the medical decisions to be made and the impact the test may have on patient care. Less well studied are the practicability and accessibility characteristics of clinical utility, those that incorporate clinician behaviors and economic effectiveness.

Clinicians are the gatekeepers to medical care and are the predominant prescribers and users of laboratory tests. It is appropriate then, to evaluate clinical utility from their perspective.

Clinicians interpret and utilize laboratory test results as one of their criteria in tailoring available treatment strategies to the medical and quality-of-life needs of their patients. The clinical utility of a laboratory test depends on the clinician's ability to prescribe it appropriately and to use the information derived from it effectively and correctly. This in turn depends on their understanding of the clinical information provided by the test and the limitations of the method. These are elements of the practicability aspect of clinical utility.

There is some published research regarding clinicians' understanding and use of diagnostic information, but less on the specific subject of laboratory test information interpretation and use. Because laboratory test results inform varied clinical states and stages, it is difficult to generalize the depth and breadth of understanding needed to apply them with only a few studies. Some test applications are a one-to-one correlation of result and disease state, while others demand a working knowledge of probability and risk, and still others require interpretation along a quantitative continuum. What has been demonstrated, though, is that many clinicians have difficulty with the practical applications of the concepts used to fully apply diagnostic information. Some examples follow.

Fifty practitioners in Australia completed self-assessment questionnaires where they rated their ability to understand and describe evidence-based medicine terms such as relative risk, absolute risk, sensitivity, specificity and positive predictive value (PPV). For each of the terms, there were only 14-30% positive responses to the question of whether they could understand and explain the term to others. In post-survey assessment interviews where the same physicians were asked to explain the terms, only three of the practitioners met some of description criteria: two partially explained the term, numbers needed to treat (NNT), and one satisfactorily explained

sensitivity. This study illustrated the practitioners' difficulty with the terms and also their inability to realistically assess their shortcomings (Young, Glasziou, & Ward, 2002).

In another study, researchers in Switzerland surveyed 263 general practice physicians to evaluate which types of additional test information would most effectively aid physicians' in correctly determining the probability of a patient having a disease. Their study also gave insight to physicians' ability to practically apply sensitivity and predictive values (Streurer et al., 2002). They designed a two-part survey, the first part consisting of multiple choice selections of definitions for sensitivity and PPV. In addition to answering the multiple choice questions, subjects were asked to calculate the probability of disease in a patient with a positive test. They were given the test characteristics of 95% sensitivity, 95% specificity, a population prevalence of 1%, and a range of choices for answers. Though 76% and 61% of the physicians chose the correct definitions for sensitivity and PPV, respectively, only 22% made the correct selection of less than or equal to 25% for the probability of disease in the presence of a positive test result (56% selected the choice of "nearly 100 %"). On first examination, the results seem more encouraging than the Australian group's findings, even with a large proportion of subjects selecting the grossly wrong PPV of near 100%. However, there are key differences. In this study, the participants were given a choice of definitions, as opposed to having their own definition scored against expected terms. Also, answers for the probability calculation were split into ranges and presented as selections. The large difference in percent correct for the definitions versus the percent correct for the PPV calculation illustrates a deficit in the ability to apply the information, even though the definitions could be memorized or reasoned.

In the second part of the survey, the participants were asked to estimate the probability of endometrial cancer in a 65 year old woman with abnormal uterine bleeding and were given 10%

as the prevalence of endometrial cancer in women with abnormal bleeding. The original group was split into three sub groups and each third received the results of a positive transvaginal ultrasound test in a different manner. One sub group received only the result ("pathological result compatible with cancer"). The second group received the result plus the sensitivity and specificity of the test; and the third received the result plus simple wording of the likelihood ratio of a positive test: "*A positive test is obtained twice as frequently in women with an endometrial cancer than in those without the disease*." In this phase of the study, the group that had only the positive test result grossly overestimated the probability of disease; those given numerical values for sensitivity and specificity gave a significantly lower estimate; and those for whom the test characteristics were described in simple words, gave an estimate of likelihood closest to that found in literature based groups. However, in a later analysis of the data, it was determined that the "simplified language" approach may have provided an opportunity to incorrectly apply a mathematical calculation that coincidently resulted in a number close to the correct one. In the amended analysis, it was still concluded that the added quantitative information regarding the test characteristics moderated the physicians' tendencies to overestimate the probability of disease (group one vs. group two plus group three). However, the comparison between the type of information provided, or comparing group three alone with group one in a stricter analysis were no longer statistically significant (Bachmann, Steurer, & ter Reit, 2003). The impact to the patient of overestimating or underestimating the probability of disease is specific to each clinical case and was not addressed by this study. It does raise the question of clinicians' expectations of the diagnostic utility of tests and the possibility that prejudicial expectations could lead to the incorrect application of the information. Inaccurate high expectations of utility can lead to inappropriate overuse of tests.

Of concern is that physicians' over-estimation of the pre-test probabilities of medical conditions in their patients could lead to improper test usage and bias of interpretations test results in support of the original assessments. To understand clinicians' ability to ration the probabilities of possible clinical conditions for a single patient, physicians and residents in another study were presented with a patient scenario and five possible clinical conditions for the symptomology and evidence. They were asked to estimate the probabilities of each condition for the patient in the scenario. Of the 84 physicians tested, 65% of them determined probabilities for each condition such that the total exceeded 100% (Cahan et al., 2003). This finding is not uncommon. Misapplying pre-test probabilities and baseline risks, as well as incorrect PPV calculations were recurrent findings in several other studies on this topic (Bramwell et al., 2006; Ghosh & Ghosh, 2004; Heller, Sandars, Patterson & McElduff, 2004).

The research discussed thus far focuses on qualitative tests, pre-and post-test probabilities and analysis of patient risk. There are few studies focused on the topic of the impact that clinicians' awareness of quantitative test imprecision will have on their decisions. A rare number of clinical guidelines exist that specify the use of tests with defined criteria for precision, such as those for cholesterol in lipid management and troponin in the assessment of cardiac arrest (National Cholesterol Education Program Laboratory Standardization Panel, 1988; Thygessen et al., 2012). Clinical specialists and physicians associated with laboratory practice have published their observations of test variability, examining analytical and biological sources; discussed the possible impacts, and have made recommendations for mitigation in laboratory practice, but not in medical practice (Fraser, 2003; Plebani, 2004). One example is an examination of the variability of creatinine measurements on the calculations of estimated glomerular filtration rate (eGFR) and the impact it has on the classification of renal disease (Klee, Schryver, Saenger, &

Larsen, 2007). Consecutive blood samples from patients were used to analyze the within-subject variance of creatinine testing, and mathematical simulation was used to create scenarios of different creatinine levels with varied levels of imprecision. It was determined that a 20 µmol/L negative shift in creatinine nearly doubled the percentage of patients classified as having decreased renal function, and the same positive shift nearly halved the percent. Recommendations for stringent control of the laboratory methods were made, but there was no mention of the need to understand or mitigate the influence of the variation on physicians' interpretations. In addition there was no recommendation to assess what physicians' actual interpretation would be for results very near the decision breakpoints.

Observations of imprecision in molecular viral load assays have led some practicing specialists to question patient monitoring results and recommend further comparative studies (Pawlotsky, 1997). However, no studies such as those described previously, addressing practicing clinicians' understanding of pre and post-test probability, PPV and likelihood ratios, or their understanding of the variation in the quantitative test results could be found in the literature, nor were they recommended by any of the researchers.

The clinical laboratory and regulatory communities have recognized the importance of the variability surrounding quantitative results and the possible implication this could have on the clinical utility of laboratory tests. This is most evident in the International Standards Organization's (ISO) requirements for determining and reporting the uncertainty for test methods in the 2012 standards for quality and competence in medical laboratories, and in guidelines developed to help laboratories provide this information (ISO, 2012 ; CLSI C51A, 2011). Studies of clinicians' understanding and application of these potentially confusing issues are necessary.

The provision of such information, along with quantitative results and reference ranges, could impact the clinical utility of current and future laboratory tests.

Optimal use of laboratory tests depends intrinsically on physicians' abilities to correctly apply the information derived from them in practice. This requires that they understand more than the results, the reference ranges, and the diagnostic conditions supported, but also the limitations of the test producing the results, and the robustness of that result in the context of clinical practice. When developing tests, calibrating the result with clinical outcomes may provide correct answers and have analytic value, making the test appropriate and acceptable; but it may not have complete clinical utility as used in routine practice. Studies measuring clinicians' interpretation, understanding, and use of quantitative information (i.e., the practicability aspect of clinical utility) are necessary.

Health Economics and the Assessment of Value

Growing demands for healthcare resources, escalating utilization and expanding technology options, require judicious allocation of those resources. Choosing appropriate services and interventions amid technology abundance and resource constraints is difficult. Yet physicians in their role as gatekeepers to healthcare services, must consider the most efficient and effective use of interventions and tools. Policy leaders also need clear, concise, and commutable data to inform their debates and decisions as they determine standard practice protocols and reimbursement strategies. To evaluate laboratory tests in this environment, the accessibility aspect of clinical utility must be considered in addition to analytical and clinical performance characteristics. This includes how the test augments or replaces conventional practices, the costs to the system, the benefits generated, and the resources needed to maintain the testing services. Laboratory tests are vehicles that provide information to reduce uncertainty in medical decisions.

New tests can be categorized as those that provide completely new information (as in a new analyte), refine existing diagnostic information, or provide similar information to what is available through the use of new technology. When evaluating the accessibility of a test, the information provided by the test as well as the resources needed to establish, maintain and effectively utilize the results from the test should be considered.

Evaluating the accessibility of a health care intervention or service includes demonstrating its value to the system in communicable terms. Led by the Australians in the 1990s, several countries have adopted a range of methods in the field of health economics (HE) to substantiate the utilization and reimbursement of therapeutics (Drummond & Sculpher, 2006). These methods are now incorporated into more expansive Health Technology Assessments (HTA) and are being applied to a broader scope of health care services and programs. HTA covers an expanded analysis of health technologies using the information from technical and clinical cost-effectiveness research and evidence based medicine (EBM) to evaluate where to place the technology relative to current practice and resources. In a global survey by the HTA principles working group of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), the term is defined as "the careful evaluation of a medical or health technology for evidence of its safety, efficacy, cost, cost-effectiveness and ethical and legal implications, both in absolute terms and in comparison with other competing technologies." (Stephens, Handke, & Doshi, 2010).

In the United Kingdom, the National Health Service's (NHS) National Institute for Clinical Excellence (NICE) lays out the process for evaluating healthcare technologies in three guidance documents: *Updated Guide to the Methods of Technology Appraisals* (NICE, 2008), *Guide to Single Technology Appraisal Process* (NICE, 2010) and *Guide to the Multiple Technology*

Appraisal Process (NICE, 2010). Clinical effectiveness and cost effectiveness are key parts of the submissions for approval by the NHS. The assessment includes the projected financial impact of the new technology on national healthcare resources. It is incumbent on the sponsor of a drug, intervention or device to demonstrate its cost effectiveness. Over 30 countries now employ some type of HTA. The majority of stated purposes for these HTAs are for coverage decisions and clinical guidance. Though the evaluation of costs is unpopular in many jurisdictions, a little over half of HTAs formally include cost and cost effectiveness (Stephens et al., 2010).

In response to the rapid adoption of HTA, a European collaborative project, the European Network for HTA (EUnetHTA) was established to develop a sustainable network defining and refining practices in this area (Kristensen, 2008; Minshall, 2008). Though HTAs vary among countries and even within countries, guiding principles to improve their conduct have been published by key experts in the field. The principles cover the structure of HTAs, the methods used, the processes for conduct, and the use of the HTA in healthcare decisions. The common threads in all the principles are: defining specificity in purpose, equal application to all health care goods and services and establishing processes for maintaining objectivity and transparency (Drummond et al., 2008).

The United States is active in the global debate and in forming its own policies. The Agency for Healthcare Research and Quality (AHRQ) is a partner organization in EUnetHTA. As a part of the American Recovery and Reinvestment Act, the United States committed billions of dollars for comparative effectiveness research (CER), measuring costs and clinical utility of interventions, technologies, and services (Gaba, 2009). As one of the 12 agencies within the U.S. Department of Health and Human Services, the Agency for Health Research and Quality (AHRQ) funds and disseminates comparative effectiveness and technology assessment research

(http://www.ahrq.gov/about/). The importance of these evaluations is growing and discussions continue to develop the structure, methods, and metrics of these programs (Avron, 2009; Garber & Tunis, 2009; Naik & Petersen, 2009). CER and HTAs are included in coverage with evidence development (CED) programs that are used by the Centers for Medicare and Medicaid Services (CMS) and large insurers for coverage and reimbursement decisions (Drummond et al., 2008). With respect to testing, the Secretary's Advisory Committee on Genetics Health and Society (SACGHS), while concerned with providing broader access to the appropriate use of genetic test results, listed as its first recommendation that evidence-based information drive coverage policies to ensure access (SACGHS, 2006). Since guidelines for the submission of economic assessments for therapeutics have already been established (Fry et al., 2003), health economists criticize the lack of utility and value evidence for diagnostics and recommend the same standards be applied to this field (Drummond et al., 2008; Ramsey et al., 2006).

New discoveries and rapid dispersion of information provide more options in a field of limited resources. It is important then to assess the impact of applying economic methods developed for pharmaceutical interventions on valuing diagnostic tests. These tools have not been used to appraise laboratory tests in the past, so there are few specific guidelines with which to compare the clinical and economic value of new tests and technologies. Pharmaceuticals are interventions; their delivery, function and outcomes are essentially different from laboratory tests, which supply information. From a laboratory test, there is usually an indirect and interdependent path leading to the final patient outcome. The laboratory test result must be recognized, understood, and used correctly in an intervention decision. The medical intervention must then be applied (or not) to the patient, and must work as predicted to produce the intended

outcome. It is essential to understand how current health economic methods can be suitably applied to isolate the value of the laboratory test in this complex relationship.

Economic Assessments in Healthcare

Three elemental types of economic analyses permeate the literature for the appraisal of therapeutics and interventions and are now being applied to evaluate laboratory tests: Cost-benefit analyses (CBA), Cost-effectiveness analyses (CEA), and a specific type of CEA: Cost-utility analyses (CUA) (Bloom, 2004; Neumann, Greenberg, Olchanski, Stone, & Rosen, 2005; Tunis, 2004). These assessments can provide information for decision making, but like diagnostic tests in patient management, they should not be the sole source of the decision. The elements of CBA, CEA and CUA are categorized in Table 1. Further detail is described in the following sections.

Table 1

Types of Health Economic Analyses

Type of Analysis	What is it asking?	Identification of Outcomes	Costs	Benefits	Report	Considerations
Cost Benefit Analysis (CBA)	What is the added cost for an added benefit?	Single or multiple outcomes.	Monetary	Monetary	Benefit /Cost ratio or NPV	Determine monetary value of intangibles
Cost Effectiveness Analysis (CEA)	What is the cost per unit of outcome?	Single and common outcome. Varying degrees.	Monetary	Natural units	Cost / natural unit selected	Determine the threshold value of the natural units
Cost Utility Analysis (CUA)	What is the cost per desired unit of outcome?	Single or multiple outcomes. Varying degrees.	Monetary	Value: combination of quant and qual of life	Cost/ QALY or DALY	Threshold values for cost per QALY/ DALY

Cost-benefit analysis. In CBA all the costs and the benefits from the programs being evaluated are converted to monetary terms. This method is employed in situations where market indicators are not useful or they are non-existent (such as health care policy); and where decisions must be made on projected future benefits. An investment can be considered favorable if the money derived from benefits exceeds that of the costs. Many times this is expressed as a ratio between the discounted net benefits (B) over the discounted net costs (C); if the benefit to cost ratio (BCR) is greater than one, a test or medical intervention has value (Folland, Goodman, & Stano, 1997). In theory, this statement is simple. However in reality, determining all costs, sourcing appropriate data for them, and quantifying the benefits in monetary terms from the proper perspective are challenges.

In CBA for clinical laboratory diagnostics, the costs include more than just the reagents and personnel, or as often presented in pharmacoeconomic studies, the list price or reimbursement price of the tests. The perspective of the analysis is an important issue in all health economic studies; in CBA this will determine the inputs used. If the study perspective for a CBA is at the institutional level, the costs to the institution beyond the clinical laboratory service, as well as the benefits gained by the whole institution are quantified. For example, if analyzing the introduction of a new test from the sole perspective of the laboratory, reagents, personnel, training, and comparisons to the charges for sending the test to reference labs might be made. The opportunity costs of using the space and resources for other laboratory tests might also be considered. But when assessing the impact at the institutional level, inputs to the costs and the benefits extend beyond the horizon of the laboratory and parameters such as patient length-of-stay, patient level of service, or physician utilization would be considered. CBA can be useful in different scenarios such as decisions whether or not to order a test, methodology comparisons,

and testing site comparisons (main laboratory vs. satellite or point-of-care service). The direct and indirect costs and benefits, and opportunity costs are relatively tangible and straightforward to measure in these narrowly focused CBAs.

Challenges emerge when clinical laboratory tests are discussed in the realm of total health care services and where competing, non-test intervention options might exist. For example, policy makers may compare the option of implementing a laboratory test or a different method of diagnosis to direct patients to specific types of care; or compare the implementation of a test-and-treat strategy to population vaccination for an infectious disease. In these situations, the societal perspective is used; the scope is broadened and other types of costs, unfamiliar to clinical laboratory professionals must be considered. In addition to capturing a broader scope of direct and indirect costs and benefits, some may be intangible, requiring a transformation to monetary value. This can be particularly challenging when intangibles, such as the value of a service or the value of a life, must be determined. Methods to derive the willingness to pay can be used for services, and several methods to estimate the value of an individual life are continually assessed and calibrated (Johansson, 2003; Messonnier & Meltzer, 2003).

CBA for diagnostic testing should carry further considerations based on the performance characteristics of the tests. Many past economic analyses of clinical programs assume that testing is 100% correct. This may be appropriate when the test information is a small part of a very large and complex study and its role is minimal. In more recent studies, where diagnostics are the central focus of the analysis, or the test information is pivotal in determining or assessing patient outcomes, the test performance characteristics play a key role in assessing the BCR. Costs associated with false negative and with false positive results must be added to the equations.

CBA facilitates the comparison of different types of programs because it strips them down to monetary measures, and when the net present value (NPV) of the BCR is calculated, CBA can provide data to compare options with different time horizons. When deciding to implement laboratory tests or programs surrounding the implementation of tests, when confronted with a choice between different options, or when ranking priorities within a set budget, CBA can be useful. Distilling options down to identical monetary units is in contrast to the other two common tools used in health economics, Cost-effective analysis and Cost-utility analysis, where the options compared are presented in terms of costs per unit of health outcome (Messonnier & Meltzer, 2003).

Cost-effectiveness analysis. Cost-effectiveness analyses systematically consider alternative approaches to an issue by comparing the costs and the consequences of each approach as they impact an appropriately selected health-related outcome unit. What distinguishes CEA is that the outcomes of all the alternatives are measured in the same unit and must be relevant to the clinical objective (Berger, Bingefors, Hedblom, Pashos, & Torrance, 2003). Examples are life-years, symptom free-days, and cases prevented. These are described as "natural units" (Drummond & Sculpher, 2006). Compared to CBA, which assesses programs with different objectives (and perhaps different outcomes); CEA evaluates programs with different approaches to the same objective. The costs to obtain that objective are compared. Therefore, a single, measurable, natural-unit outcome must be selected (Fos & Fine 2003).

CEAs are growing in popularity and are becoming the predominant method for evaluating the costs and effects of alternative strategies in health care (Messonnier & Meltzer, 2003). The programs and interventions being compared can be ranked, or several interventions can be studied, each for their incremental cost effectiveness. The perspective in CEA is usually from a

patient or public health view; and in CEA this perspective determines the outcomes measured. When a laboratory test is being evaluated, the impact that the test has on extending patients' lives, averting disease, or changing a specific state of health of patients is considered. The challenge for laboratory testing is to demonstrate the direct impact of the test on the patient state. An example of this is a new or more sensitive marker that informs the definitive treatment of patients, such as a marker for stroke or acute myocardial infarction. By using the new marker, the number of appropriately treated cases may increase. For this example, calculating the positive or negative balance of all the costs and consequences, as in CBA, may not be useful. A cost-effectiveness ratio (CER), the cost per life saved or the cost per complication averted, more appropriately conveys the value of the outcome. When compared against the standard test, or no test at all, an Incremental Cost Effectiveness Ratio (ICER) can be calculated illustrating the incremental cost per life saved (or other natural outcome) of the program under consideration. An issue for interpretation is assessing whether the value of the CER or the ICER is acceptable. Few countries and jurisdictions have established thresholds for cost effectiveness.

Cost-utility analysis. Because CEA expresses costs per a specific biological unit, interventions with different natural outcomes cannot be compared. In addition, there can be a range of interpretations that decision makers may have regarding the comparative value of the outcomes (e.g. determining the relative priority of blood pressure control compared to cholesterol management or to a year of life). A further refinement in CEA is the Cost-utility analysis. In CUA, the benefits or outcomes are expressed in parameters that account for quality as well as the quantity of the benefit in Quality Adjusted Life Years (QALY) or Disability Adjusted Life Years (DALY). QALYs are measures of mortality and the quality of life. The quality of life is measured on a scale from 0 (death) to 1 (full health) and the time gained in the

CEA is adjusted by the quality weight depending on the state of health (Gold et al., 1996 p. 88). The QALY has become a standard currency in health economic evaluations. QALYs are most useful when evaluating individual benefits in longer term outcomes. The values for QALYs encompass the preferences of individuals for specific states of health, and they can be determined in several ways. The important characteristic of these methods is that preference weights are not subjectively assigned but are determined through individual surveys and are published for use by researchers.

Less often used is the DALY. A DALY is defined as one year of healthy life lost due to death or disability from the disease or condition in question. It is the sum of the years of life lost (YLL) and the years of life with a disability (YLD) and thus, incorporates the time lost to death at each age, disability weights, age weights and time preference (Fos & Fine, 2005). DALYs provide a relative scale to compare different diseases and conditions, both physical and mental. Since DALYs are calculated from a theoretical maximum life span, that maximum must be carefully ascertained. For every subject studied, an individual life span should be calculated based on health and risk factors for each (minus the condition in question). Because this is impractical, actuarial data for specific populations are utilized. This makes the DALY a gross measurement at best and comparable among studies only when all the conditions used in the DALY calculation are transparent and pertinent. In addition to actuarial-determined life spans, the quality of life ratings are assigned by health professionals, not the individuals suffering with the disability. For these reasons, the DALY is considered less relevant.

CUA facilitates the comparison of programs producing different outcomes, as each is translated into directly comparable QALYs or DALYs. For example, if policy makers are deciding whether to institute a community smoking cessation or weight loss program, outcomes

from a CUA would be presented as the cost per QALY for each option. For laboratory services, a clinic might utilize CUA to decide between establishing a point-of-care cholesterol or a prothrombin-time testing service for their outpatient community. The question being addressed by CUA does not necessarily have to be a direct comparison of one or more options. Some jurisdictions have explicit or implicit values for the reasonable cost of a QALY or DALY. In those cases, if the analysis of a single option results in the cost/QALY less than the threshold, the program can be considered economically viable.

An example of interpreting the relevance of a CER derived from a CUA is in an analysis of *Chlamydia trachomatis* testing performed in The Netherlands. The impact of a systematic screening-treatment program that included partner treatment for young Dutch adults was evaluated for its impact on preventing the serious sequelae associated with the infection, pelvic inflammatory disease (PID) and sterility (de Vries, van Bergen, de Jong-van den Berg, & Postma, 2006). The calculated CER was €373 per major outcome averted (MOA). The only benchmark available in The Netherlands at the time was a threshold of €20,000 /life-year gained or QALY. In order to discuss the comparative relevance of their numbers, the researchers applied recently published quality adjusted weights from the United States to each of the disease states into their mathematical model to convert the events averted into QALYs. By this method, the calculated overall cost was €1000 per QALY. Since this was far below the policy threshold, they concluded that the program was cost effective.

Decision analysis and mathematical models. Health economic evaluations are commonly performed by constructing and populating mathematical models. The models may extrapolate data from clinical trials, or extract data from several sources including literature and public records. As models, they cannot duplicate life situations, so this limits applicability to the

assumptions made in building them. However, it can also be argued, that randomized clinical trials, considered the gold standard scientific in evidence gathering, also can have limitations in scope and applicability. In clinical trials, well-controlled patient selection criteria, strict operational management, and finite timelines, result in a representative sample and conditions that may not necessarily be generalize able to an entire population for a longer horizon. Mathematical models can incorporate the results from several trials, meta-analyses of publications, and numerous databases and can represent many scenarios. The assumptions in models, though limited, are transparent, and the impact of key assumptions can be tested in sensitivity analyses. Neither represents the total solution in healthcare policy analysis. Both should be used alone or in combination, with knowledge of the limitations and as appropriate for the situation analyzed (Cohen & Reynolds, 2008; Lang, Lopert, & Hill, 2003; Weinstein et al., 2001).

A framework for these analyses that is useful to health economics is decision analysis. Prior to use in healthcare, analytic modeling had demonstrated its utility in situations when investment choices had to be made in uncertain situations and the major downstream decisions and outcomes were considerations in the decision (Gallant, Kieffel, & Chatwin, 1997; Weinstein et al., 2001). Designing models for decision analysis gives structure to the question and the options, demands evidence, provides a format to evaluate uncertainty and variability within the assumptions, and can reveal a path to future research (Drummond, Sculpher, Torrance, O'Brien, & Stoddart, 2005). The nature of this discipline is a systematic approach to outline and quantify the options in decisions to be made under conditions of uncertainty (Goldie & Corso, 2003; Scott & Rafferty, 2007). A primary purpose of diagnostic testing is to provide information that reduces uncertainty in healthcare decisions; in some situations, a single diagnostic test can be

placed directly in the clinical decision path. In these cases, decision analytic modeling could be used to quantify the role of that test. Decision analysis can be particularly useful in healthcare because the expected value of different strategies can be compared after recording the relevant events with their probabilities and associated costs. When enough information exists, the modeling can be performed before resources are invested in long term trials. In pharmacoeconomics, decision modeling, first described for use in healthcare in 1980 (Fineberg, 1980; Weinstein & Fineberg, 1980) has become useful for analyzing data from many sources to better inform decisions in actual practice (Sanchez & Lee, 2000). The elements of a simple analytic decision model are shown in Figure 1. The first point in the tree is the decision node, represented by the square. A decision informed by a laboratory test will stratify the patients to different management schemes based on the results of the test. The next point of dispersion is the chance node (filled circle) that illustrates which of the several events beyond the control of the decision maker may occur. There may be several different branches; but, symmetry around each node is not required. The branches represent the natural sequence of consequent events, and each must be mutually exclusive. Collectively they must exhaust all possibilities. Two are shown in the figure for simplicity. The probabilities of each of these chance events are recorded along the branches. The final nodes are the terminal nodes which represent the end of each sequence along with the outcome of the events (life years/ QALYS/ length of stay). Value is assigned to the outcomes. The expected value of each branch can be compared by calculating from right to left, multiplying the probabilities by the cost of the outcomes for each decision alternative.

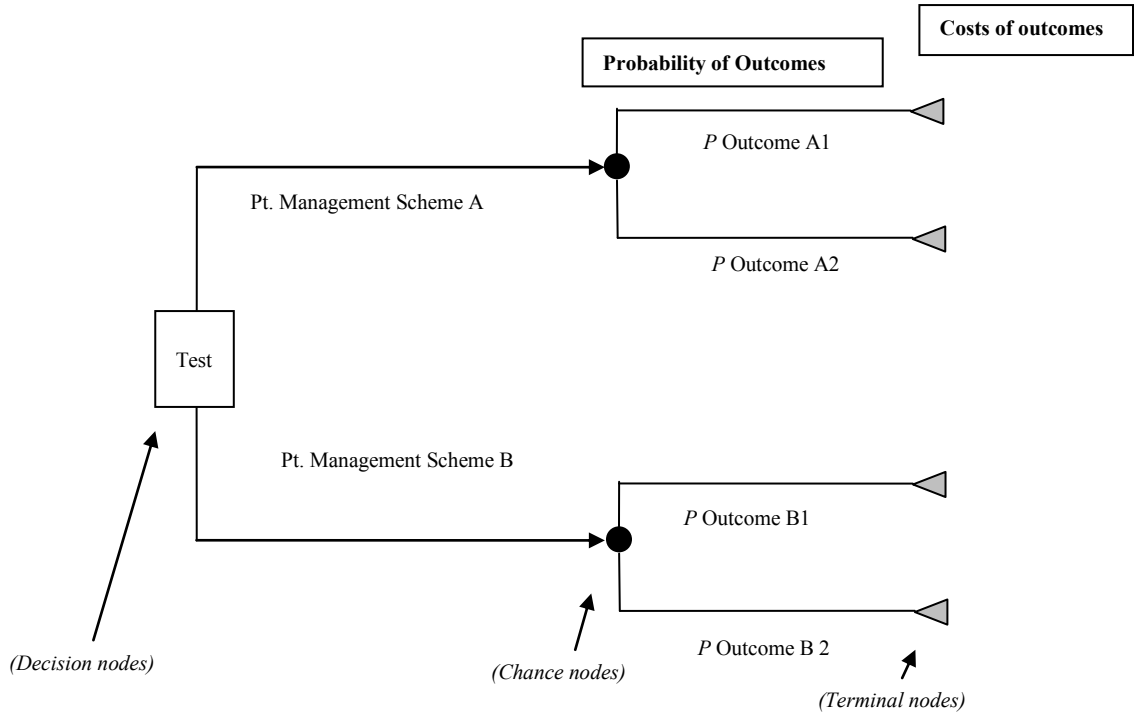

Figure 1. Example of a simple decision tree analytic model. This example describes a laboratory test informing patient management strategy (adapted from Weinstein & Fineberg, 1980).

The purpose of decision analytic modeling, according to an International Society for Pharmacoeconomics Outcomes Research (ISPOR) guideline, is to "structure evidence on clinical and economic outcomes in a form that can help to inform decisions about clinical practices and health-care resource allocations" (Weinstein et al., 2003). It is important that models are not mistaken for truth but instead used as tools to facilitate information gathering for disciplined analyses. Because a model has the ability to reveal logical connections between inputs and outputs in the form of valued consequences and costs, the model should be transparent and follow commonly accepted practice guidelines (Weinstein et al. 2001; Weinstein et al., 2003).

Technological advancements expand the options in healthcare services, and increase the utility of analytic modeling to inform research, clinical practice and policy decisions. Placing a

laboratory test in the clinical decision path and illustrating its impact on succeeding events, may be particularly suitable for the evaluation of pharmacogenomics and companion diagnostics, where test information is the gating item that guides therapy choices. The value of making the therapeutic decision with and without the test information can be calculated. An example is in the management of warfarin therapy, where genetic variants in metabolic enzymes have been associated with variations in the efficacy of the drug.

Researchers evaluating the use of genetic test information to modify doses of warfarin therapy have used decision modeling and cost effectiveness in their assessments. In one such study, You, Chan, Wong and Cheng (2004) employed decision analytic modeling to assess the cost-benefit of pharmacogenomics-oriented management of warfarin dosing. Their goal was to evaluate the potential clinical and economic outcomes of using CYP2C9 genotype data for patients receiving anticoagulant therapy. Patients with CYP2C9 polymorphisms are associated with reduced warfarin dosing requirements and higher risk of a major bleeding event from warfarin overdosing during the first three months of therapy. Using CYP2C9 polymorphism prevalence, event rates, and cost data from clinical trials and provider data, they constructed a decision tree to simulate the events of two theoretical cohorts initiating warfarin therapy: one with genotyping data prior to the institution of therapy and one without genotyping. Patients with at least one CYP2C9 variant (in the genotyped group) received intensified anticoagulation service while those without the significant polymorphisms and those not genotyped received standard care. Through the model outputs, it was determined that major bleeding events would be averted in the genotyped group; however, the total costs of care for this group would also be greater. The incremental cost of averting a bleeding event was $5778 per event. Using the CER benchmark of less than $50,000 per QALY as a goal (from Panel on Cost Effectiveness in Health

in Medicine), they converted their outcomes to this measure by estimating one month of illness at a utility score of 0.39 for events averted, resulting in 0.05 QALY. For those patients who hypothetically averted death from a major bleed, they conservatively estimated a life expectancy of one year and calculated one QALY. To be cost effective by the benchmark, they calculated that the genotyping scheme would have to show a gain of greater than 0.1 QALY per event averted ($5778/ $50,000); their results were between 0.05 and 1 QALY. One of the difficulties in the interpretation of this study is the projected infrequency of bleeding events. Though potentially devastating, the number of events in these patients was small, approximately eight per 100 patient years. This is weighed against the risk of delaying therapy (to wait for genotyping) and risking stroke; approximately two events per 100 patient years. Another issue is the 12-month time horizon of the model, appraising an intervention that is most useful in the initial three months. Specific to testing, a third challenge is measuring the impact of the laboratory test using conventional health economic tools. It is hard to prove that the test alone has a direct effect on the patient outcome. As in this study, the test result can lead to a more expensive health management path. However, the CUA and the information from this model are useful. Since the infrequency of the bleeding events is prohibitive to the trials needed to prospectively demonstrate the utility of the intervention, recommendations for more focused clinical trials can be made. In this case, limiting prospective clinical trials to the induction period of three months and examining the costs of the intensified anticoagulation service, perhaps as standard of care, could be pursued.

Analytic models can also be used to understand how best to focus technology investments to optimize patient outcome. Meyers et al. (2000) utilized a Markov model for the analysis of cost effectiveness of specific improvements in tests for cervical cancer screening. Figure 2 illustrates

a simple Markov model; an analytic tool used to represent disease progression states and events that recur over a specified period of time (Lang et al., 2003). In Figure 2, each mutually exclusive health state is in a circle, and the transitions between each of the states are represented by arrows (state transitions). The probabilities of moving from one state to another or staying in the same health state are assigned.

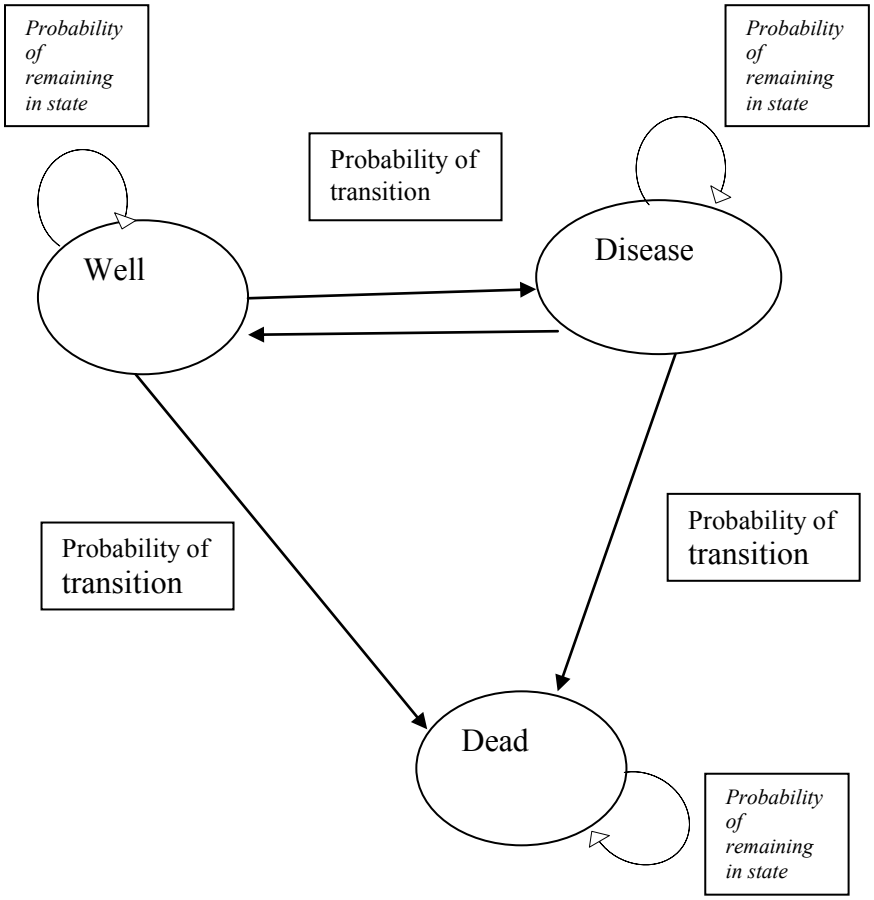

Figure 2. Example of a simple Markov model (adapted from Goldie & Corso, 2003).

In the Meyers study, researchers were able to provide insight into the marginal cost of performance specification improvements for new cervical cancer screening tests. Using the Papanicoloau (PAP) test as the base case, they modeled several levels of sensitivity and specificity of hypothetical screening tests with the downstream costs of the future events. While

they increased levels of sensitivity in the hypothetical test, they only altered the specificity by making it equal or poorer than the base case, assuming that a new technology would increase sensitivity at the expense of maintaining or decreasing specificity. They also examined different screening frequencies of the new versus hypothetical test. Improved sensitivity led to substantially greater costs than the health benefits obtained, mainly due to the magnified number of low grade lesions with the consequent increased resources needed to evaluate them. Findings such as these can help policy makers focus on the attributes that new technologies should present as they assess them. They can also influence the direction of future development of new tests. New screening tests for cervical cancer might only be viable if, in addition to sensitivity improvements, specificity is also improved or if the downstream procedures for evaluating all patients with positive test results became less expensive.

The optimal placement of a new test among conventional tests for hereditary hemochromatosis (HH) has been studied through analytic modeling in Canada (Gagne, Reihnarz, Laflamme, Adams, & Rousseau, 2007). Researchers built a modular simulation program to generate a virtual population of one million people with demographic, genetic and phenotypic characteristics specific for HH. They used this "population" to evaluate all screening algorithms that were practically possible using the available biochemical and genetic tests. In addition, they altered the population by varying HFE gene frequencies and penetrance. Using their model they were able to analyze the outcomes of 165 different screening algorithms in 91 different virtual populations. Among the top cost-effective strategies for most permutations of prevalence and penetrance was a combination of conventional biochemical tests: unbound iron-binding capacity plus transferrin saturation. The genetic test was optimally cost-effective as a confirmatory test. Additionally, analysis of population screening was also performed and was shown not to be cost-

effective when only hepatic complications were considered. This analytic model can be continually refined and used to provide insight on several research and policy questions regarding effective HH management.

For viral molecular tests, economic studies have focused on screening strategies for blood borne viruses such as HIV, HCV and HBV. In blood-donor screening, HE analyses have forecasted costs of up to $11.2 million per QALY using the combination of HCV, HIV and HBV NAT tests to avert infection and subsequent disease in blood-product recipients. Despite the high cost /QALY, the outcomes have not influenced policy makers to recommend against their adoption. However, the information has helped support the discontinuance of other non-essential or repetitive tests, such as anti-hepatitis B core and p24 antigen testing (Jackson, Busch, Stramer, & AuBuchon, 2003).

For individual diagnosis and patient management of these viral diseases, there is a paucity of health economic literature. Most work has focused on assessing the cost effectiveness of screening populations for the early detection and early intervention of HCV disease. The United Kingdom's National Health Service (NHS) published an HTA report on the cost effectiveness of actively screening former injecting drug users (IDUs) for HCV. They built a decision analytic model to evaluate the impact of case-finding (a proactive search of a targeted population) and treatment on a hypothetical cohort of 1000 IDUs. They compared the targeted, screened and treated cohort to one that spontaneously presented for testing. Clinical and epidemiological experts were used to inform the structure of the model and parameter estimates were populated through literature and electronic database searches. Their model consisted of a simple decision tree for the testing schemes, with a Markov model for the treatment and outcomes. The model ran for the life of the cohort cases; all entered at 37 years of age. For each 1,000 subjects in the

case-finding cohort, three cases of decompensated cirrhosis, three deaths (from HCV disease), and one case of hepatocellular carcinoma were averted. One liver transplant for each 10,000 cases would also be averted. Twenty-five more subjects in the case-finding cohort than in the spontaneous testing cohort would be prescribed combination therapy. In terms of cost-utility, £16,514 per QALY would be invested, below the £30,000 per QALY benchmark used by NHS policy makers. Though it resulted in higher absolute costs, £760,000 due to targeting, testing and treatment resources, the case-finding strategy in this scenario was within the acceptable NHS criteria (Castelnuovo et al., 2006). This type of study is used by NHS commissioners to adopt strategies, set policy or design further research.

An Italian group evaluated the cost effectiveness of screening in IDUs and patients who received major and minor surgery (potential exposure risk) for the Veneto region, using the demographic and public health records from this population of 4,750,000. They hypothesized that the socioeconomic burden of disease could be reduced by screening and early intervention in asymptomatic, at-risk groups. They further stratified the analysis by HCV genotype, which impacts treatment success. Both IDU and surgery patients' analyses were very sensitive to the prevalence of the more difficult to treat genotypes 1 and 4. However, only the IDU population screening was shown to be below the cost per QALY threshold of €30,000. Screening IDUs actually had a negative cost per QALY in those with genotypes 1 and 4 (€ -5.14). For cases with genotypes 2 and 3, the cost per QALY was €9,659. In the surgery patients, all costs per QALY were far above the threshold: €699,991 for genotypes 1and 4 and €2,324,471 for genotypes 2 and 3 (Tramarin et al, 2008). These outcomes could reflect the relative risks of these two populations. The surgery patients were not limited to only those that would receive blood

products. Even so, the adoption of HCV screening of blood products has significantly reduced the risk of infection for recipients.

The type of population, the risk of developing disease, and the prevalence of disease have been important modifiers in other studies. The results of an analysis in France showed that screening for HCV was the dominant (more-cost effective) strategy only in IDUs, when that group, transfusion patients, and the general population were compared (Loubiere, Rotily, & Moatti, 2003). An earlier US study using the prevalence of the 3.8% for the general population, also found that screening the average risk group in a relatively low prevalence setting was not cost effective as a general health policy (Singer & Younossi, 2001).

These examples illustrate some of the opportunities and difficulties in applying traditional economic analyses to studying the cost effectiveness of laboratory tests. Diagnostics are usually a new cost, unless one is replacing another. Understanding how to structure an analysis in specific populations with differing prevalence rates, or states of health and with different test performance characteristics is challenging.

Another way to consider the value of diagnostic tests is to regard the costs of laboratory tests as investments in information which will reduce uncertainty in specific medical decisions. Directing resources and investments towards developing and refining tests that provide the most clinically valuable information could result from these analyses. Value of information (VOI) analytical tools are becoming useful in clinical trial design to identify and evaluate the benefit of proposed additional research. Reducing uncertainty and the calculated clinical and economic value of that reduction are instructive in determining clinical trial strategy (Yokota & Thompson, 2004). Laboratory tests provide information which reduces uncertainty in health and patient management. In this regard, VOI might be applied to ascertain the value of diagnostic tests.

This idea is not new. In 1988, Phelps and Mushlin suggested methods to calculate the Expected Value of Diagnostic Information (EVDI) and the Expected Value of Clinical Information (EVCI) using CUA and decision analytic models as the analytical foundation. The probabilities of a patient being sick or well, and being treated or not in each of those states, with and without the diagnostic information to guide the treatment decisions, were illustrated in models that look very much like current analytical decision trees. While the publication was a theoretical description, it presented several principles. The most important was the view that diagnostic tests are information and their value should be measured in that regard. The base case analysis is then a specific clinical path without the diagnostic information, and it is compared to the clinical path utilizing the test results. The importance of understanding that diagnostic tests do not deliver perfect information and the costs and consequences of choosing the "incorrect path" based on that imperfect information were also described. Several applications for this type of analysis were proposed, including determining the range of disease probabilities over which devices may be used, and comparing one device against another without the expense of randomized controlled trials.

Methods calculating VOI have rarely been applied to analyses of laboratory tests. Recent work to determine the value of epidermal growth factor receptor (EGFR) in non-small-cell lung cancer (NSCLC) second-line treatment with (EGFR) tyrosine kinase inhibitors (TKI's) illustrates some of the challenges in measuring the value of laboratory testing and provides one use of VOI (Carlson, Garrison, Ramsey, & Veenstra, 2009). Two different test methods, a protein expression and a gene copy number test were evaluated. Sensitivity analyses of the model revealed that effectiveness was most sensitive to changes in disease-progression free and overall survival states. The CUA of pathways with and without the test results demonstrated that the

pathway without the testing was more "cost-effective" up to a threshold of $150,000 cost per QALY, after which, the gene copy number test was more cost effective. Testing for EGFR by gene copy number produced an added three weeks quality-adjusted life expectancy and a 1.4 month overall survival difference at the cost of $9200. (Test costs were $320; most of the additional costs were treatment costs associated with extended survival). The ICER of adding the test, and the outcomes associated with the testing was $162,000. It is not surprising that a test providing information that extends life by only a few weeks would produce a costly outcome. In addition to the CE analyses, the Expected Value of Perfect Information (EVPI) at a threshold of $US 100,000/ QALY was calculated. Over a five-year period, the EVPI was $31.4 million, representing the upper bound of opportunity costs that the added test information would contribute in appropriately triaging patients during that time. Currently, there is no agreed upon benchmark for information investment; therefore, the EVPI merely supports recommendations for further study into the role of this test, particularly if the efficacy of the treatment is improved.

The expense of NSCL cancer treatment, the increased response of those testing positive for EGFR, and the necessity of selecting only those patients that could respond provided a rationale to demonstrate the clinical and economic utility for a test to select these patients. However, the benefit of the intervention in the population was marginal, and the benefit added further health care costs to the economic outcome. Calculating the EVPI, however, demonstrated the value that the information itself could impart in the clinical scenario. While investing in healthcare costs and services, is it worth investing up to $31.4 million over five years to decrease the uncertainty with which patients are chosen for this therapy? Though there are no policy thresholds at this time, it is conceivable that clinicians and patients might be able to use this type

of information at the individual level or a cost per test basis if there was enough research around the VOI parameter.

The EGFR study calculated the VOI assuming the results derived from the tests were always correct; therefore, the value calculated was the upper limit of any contribution by the test. Combining more realistic technical characteristics of tests with the VOI derived from them, is the basis of a synthesis of analytical tools proposed in "ROTS analysis" (Laking, Lord, & Fischer, 2006). ROTS analysis combines two linear schools of thought: the ROC or Receiver – Operator Curve analysts and the VOI analysts. This multidimensional model links changing test thresholds and the cost-effectiveness of different interventions. Within the stated changing parameters, ROTS analysis provides a method to answer questions regarding the justification for doing a test, a test's optimum operating point, and comparison to competing tests or methodologies. However, the graphical representation and interpretation in ROTS may be beyond the current scope of CE discussions, particularly since the current HE methods were not specifically optimized to analyze laboratory tests. A subset of the conclusions in ROTS could be obtained by appropriate sensitivity analyses of more standard CE evaluations, though ROTS provides more interpolation and dimension. ROTS seems to be optimized for qualitative results, utilizing sensitivity and specificity as principle parameters. The results of quantitative tests would have to be translated into clinical sensitivity and specificity for the intended outcome, if the method were to be employed to assess quantitative results.

Among the methods surveyed, cost-benefit and cost-effective analyses are being applied to diagnostic testing, and used by policy makers and academics. Laboratory professionals utilize CB to justify new tests, methods, or instruments for their particular organization, but usually do not initiate more strategic CEA and CUA analyses. Conversely, health economists rarely

represent the performance characteristics of diagnostic tests in their analytical models. For laboratory professionals, the reasons may be that the analyses are too complex, require too much input, or have received too little exposure. Because of discreet institutional budgets, the relevance of developing CE models for health policy may have little perceived relevance to the operations of clinical laboratories. For academics and health economists, including test performance characteristics and their subsequent impact on the clinical pathways modeled may be too complex, or may be considered trivial in the scheme of the larger analyses. Most of the examples of HE analyses for laboratory tests in the research discussed here have been focused on diagnostics with a yes/ no result of disease states carrying different probabilities, which is only one use of laboratory test results. There is no work examining the value impact of the variation in quantitative tests which are used to monitor and adjust clinical therapies. Viral disease management depends on quantitative viral load test information for triaging patients to the proper therapy, for monitoring patients on therapy, and for determining clearance of the virus. For HCV, where a change in quantitative test results can catalyze an alteration in therapy, it is relevant to examine the clinical and economic value of those results and the impact that imprecision has on the decisions made by them.

Hepatitis C and Viral Load Testing

Hepatitis C imposes a major disease burden on the world (Sy & Jamal, 2006). The global prevalence of HCV has been estimated by WHO to be approximately 3%, affecting 170 million people (http://www.who.int/csr/disease/hepatitis/whocdscsrlyo2003/en/index3.html). This is considered an underestimate due to reporting inefficiencies, underreporting, and the fact that many of the rates are derived from disease prevalence in the blood donor population. Depending on the country and area surveyed, the prevalence of HCV can range from below 1% to above

20%. In the United States, underreporting and asymptomatic infections add to the recording challenges. Approximately 3.2 million people in the US have chronic HCV (http://www.cdc.gov/hepatitis/HCV/HCVfaq.htm#section1). The National Institutes of Health project 35,000 new cases per year and cautions that since most chronic infections have not yet been diagnosed, a fourfold increase in the number of infected adults is predicted from 1990 to 2015 (NIH, 2002).

HCV is the leading cause of liver transplantation in the United States, one of the principal causes of cirrhosis, and a common cause of hepatocellular carcinoma. It is the primary reason for death from liver disease in the United States. Between 1999 and 2007, HCV related deaths outpaced HIV and HBV related deaths and accounted for approximately 4.58 deaths per 100,000 persons per year (Ly et al., 2012). With an estimated prevalence of at least 1.8%, HCV is the most common chronic blood-borne infection in the United States (NIH, 2002).

The virus, transmission and course of disease. The hepatitis C virus is a small enveloped RNA virus constituting its own genus within the Flaviviridae family (Op De Beeck & Dubuisson, 2003). There are six genotypes of HCV, each having implications on viral persistence and patient management (Simmonds et al., 2005). Genotype 1 is predominant in the US, accounting for 70– 75% of all HCV infections. It is also associated with a lower response to current treatment regimens (NIH, 2002).

HCV is transmitted primarily through contact with infected blood or blood products. This is the main source of infection in several countries. In the US, the blood supply has been routinely tested for HCV antibody since 1992. For those who received transfusions, blood products or organ transplants prior to 1992 the risk is high. Since then, the risk from this source has been minimized and injection drug users currently have the highest risk. Other possible sources of

infection include exposure to an infected sexual partner, perinatal exposure, and occupational exposure to infected blood by healthcare workers (Ghany et al., 2009; & NIH, 2002; Sy et al., 2006).

Many individuals infected with HCV remain symptomless. For those that do develop clinical symptoms, the incubation period ranges from 15-150 days. Approximately 15-25% of infected individuals will spontaneously clear the virus. The remaining 75-85% will maintain chronic infections. These individuals are at risk for further, serious sequelae. Approximately 60-70% of those chronically infected will develop chronic liver disease. Of these, 5-10% are expected to develop cirrhosis and 1-5% will die from the consequences of the cirrhosis which includes hepatocellular carcinoma (Ghany et al., 2009; Seeff, 1999). A review of several natural history studies in the 1990s concluded that approximately 20% of infected individuals progress to cirrhosis, fibrosis and serious end-stage liver disease. The serious consequences of HCV infection usually manifest after approximately 20 years (Seeff, 1999). Figure 3 provides an illustration of HCV disease pathways.

Treatment and patient management. Because the symptoms of HCV infection can be inconspicuous while the virus remains transmissible, diagnosis is important. In addition, because the disease manifests slowly, early diagnosis provides a window of opportunity to intervene and prevent the more serious consequences associated with progression. For genotypes 2-6, the most common treatment strategies for chronic HCV infection employ combinations of long-acting pegylated interferon-alpha, and ribavirin (Ghany et al., 2009; Sulkowski, 2008). While there are several treatment protocols approved by the US Food and Drug Administration (FDA), the strategy that has demonstrated the highest rates of sustained viral response (SVR) and is the

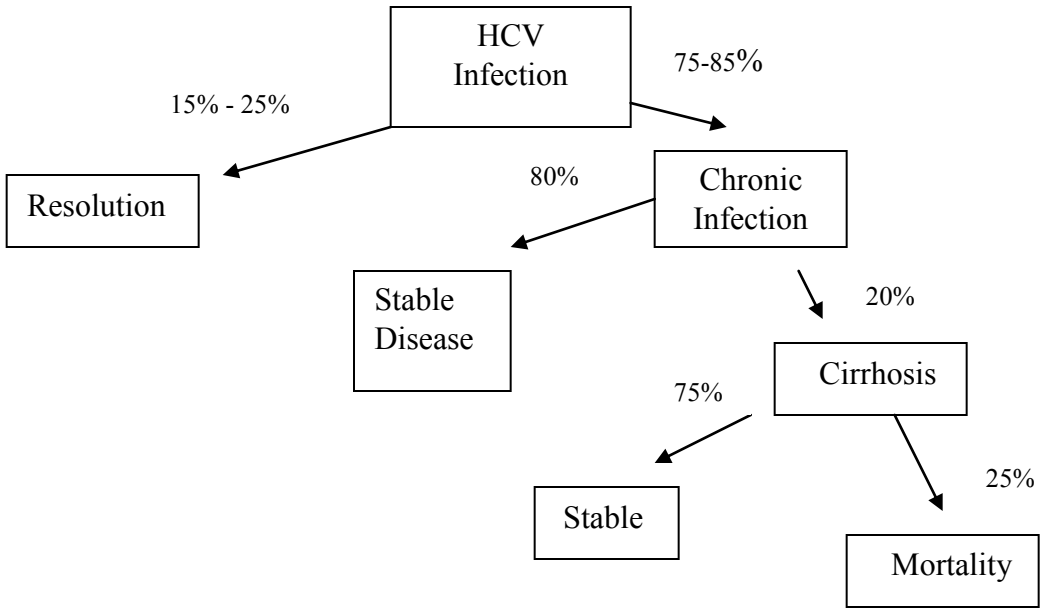

Figure 3. The natural course of HCV infection. Adapted from Seeff, 1999; Ghany et al., 2009; and CDC.gov 2012.

current standard of care, is a combination of peginterferon-alpha administered in weekly subcutaneous injections with daily oral ribavirin (Ghany et al., 2009). Besides the treatment inconvenience, the side effects of this combination therapy are not trivial. The management of the side effects is particularly frustrating for non-symptomatic individuals diagnosed with HCV. Adverse effects include profound fatigue, headache and other pain, neutropenia, thrombocytopenia, hypothyroidism and hyperthyroidism, visual disturbances, nausea, fever, depression, irritability, suicidal tendencies, and cognitive impairment. Other complaints include insomnia, inability to exercise, and the loss of motivation. The addition of concomitant medications and interventions to manage these adverse events impact patients' quality of life and productivity (Ghany et al., 2009; Hopwood, Treloar, & Redsull, 2006).

In 2011, a new class of therapeutics was added to the dual therapy for the treatment of HCV genotype 1, the most common and difficult to treat of the HCV viruses in the US. The use of one of the protease inhibitors, telaprevir (Incivek ™) or boceprevir (Victrelis™), with the dual combination therapy, increases the response rate, but rates of anemia are also increased and hemoglobin levels must be monitored closely (Ghany et al., 2011; Merck Sharp & Dohme Corporation, 2011; Vertex Corporation, 2011). There are several other therapeutics for HCV in development and for which HCV VL measurements will be critical in the assessment of their success (Cox, 2013).

Adverse events and inconvenience can also impact patients' compliance to therapy. Because of the associated debilitating events and the observation that not all infections are eliminated by therapeutics, a goal of many clinical research studies has been to find the most effective and shortest regimen. HCV viral load (VL) testing has played a key role in these studies; and as therapeutic strategies are individualized to patients, VL testing utilization will increase. Current treatment protocols require administration of therapy for a minimum of 12 weeks up to 72 weeks, depending on viral genotype, and the HCV VL test results at specified intervals during the treatment. The goal is to treat the patient aggressively, but humanely, to eradicate the virus. Through research, opportunities to remove patients from further therapeutic discomfort when VL results indicate therapy is not effective have been identified (Chevaliez & Pawlotsky, 2009; Ghany et al., 2009; Ghany et al., 2011).

Chronic HCV disease management and laboratory testing. Laboratory testing is integrated into all phases of hepatitis C patient care: diagnosis, therapy selection and management, and the determination of failure or sustained cure. Patients are usually first diagnosed by means of positive HCV antibody tests that are confirmed by qualitative or

quantitative (viral load) HCV RNA tests. Serum aminotranferase levels and liver biopsy histology help establish the severity of the disease and any sequelae present. Quantitative HCV RNA and HCV genotype tests along with biopsy results and patient data are used to plan the most appropriate and potentially effective treatment strategy (Chevaliez & Pawlotsky 2009; Ghany et al., 2009; Ghany et al., 2011; Sulkowski, 2008). A newer biomarker, IL28B has been added to the testing menu. Knowledge of mutations in this gene provides a probability of successful response for therapies using interferon in HCV genotype 1 infected patients. However, HCV VL reduction and clearance are still the definitive targets. (Ge et al., 2009; Sharifi & Alavian, 2011).

The goal of therapy is to avoid the future complications of HCV, which can only be achieved with certainty if the virus is eradicated. The most powerful predictor of treatment response and the definitive indicator of treatment outcome is HCV VL. When treatment begins, a viral load determination is made. Depending on the genotype, the severity of liver disease, the likelihood of a sustainable response to treatment, the probability of adverse events, and the presence of comorbid conditions, the length of treatment is individualized. The key information for adjusting the length of treatment is the succeeding HCV VL results. The best evidence of the eradication of the virus is provided by quantitative HCV RNA nucleic acid testing (NAT) of patients' blood throughout the course of therapy and continuing after therapy (Ghany et al., 2009; Sulkowski, 2008).

The American Association for the Study of Liver Diseases (AASLD) publishes data-supported recommendations for the diagnosis, treatment, and management of patients with hepatitis C disease. In the newly revised guidelines for HCV genotype 1, the predominant genotype in the United States and the most refractive to therapy, clinicians are recommended to

initiate their treatment-naïve HCV patients with one of the new protease inhibitors plus the standard dual combination therapy. After the initial treatment period using the three therapeutics, quantitative HCV RNA testing is performed to determine the viral load. If the HCV RNA level is not below a specified number (100 IU/ml for boceprevir or 1000 IU/ml for telaprevir) therapy can be discontinued (Ghany et.al, 2011). Clinical studies have demonstrated that there is virtually no probability of a sustained viral response (SVR) for patients that do not achieve this milestone (Chevaliez & Pawlotsky 2009; Ghany et al., 2011, Terrault et al., 2005; Zeuzem et al., 2003). For patients where the initial phase of therapy has been successful, treatment should continue with interferon-alpha and ribavirin for 12 to 36 additional weeks, depending on HCV VL results within that period. At the end of therapy, another viral load measurement is taken with the anticipation that the patient has achieved a "less than detectable" value. Twenty four weeks after therapy completion, an HCV VL test is performed to confirm the patients SVR status. AASLD guidelines also differentiate the treatment management for patients with genotype 2 or 3. These patients receive an HCV VL test at the beginning of their treatment and are scheduled for 24 weeks of interferon-alpha and ribavirin therapy. At the end of 24 weeks another viral load determination is made and then again at 48 weeks to determine the sustainability of the response (Ghany et al., 2009). Figure 4 illustrates these clinical pathways.

Clinical management guidelines illustrate the integral and essential role of quantitative HCV testing in HCV patient care. Considering the length of treatment, potential morbidity, and costs associated with current therapy, studies continue to focus on altering regimens specific to patient and disease characteristics. Newer clinical studies are exploring even more individualized

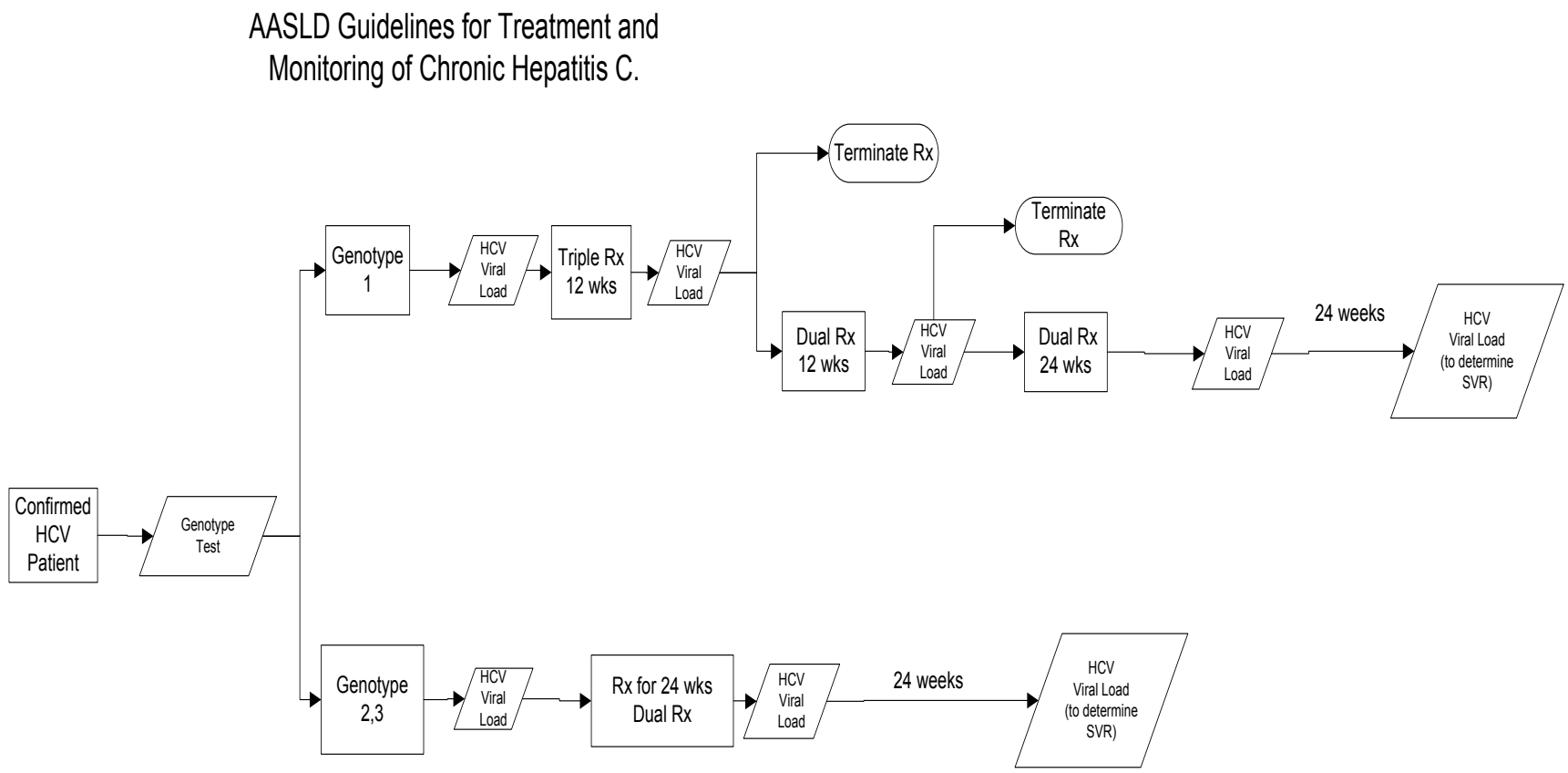

Figure 4. Treatment strategies for chronic hepatitis C infection. Derived from Ghany et al, 2009; and Ghany et al., 2011.

treatment strategies, from shortening treatment based on the first time HCV is undetectable, to lengthening it in other situations. The results of these studies are proposals that recommend shortening therapy when a rapid viral response (RVR) is established, halting therapy when it has been demonstrated to be ineffective, or lengthening it when prolonged therapy increases the chance of SVR (Chevaliez & Pawlotsky, 2009; Mangia et al., 2007; Poordad, Reddy, & Martin, 2008; Zeuzem et al., 2005). In all cases, the decisions are predicated on the HCV VL measurement. Establishing SVR by means of HCV VL was the surrogate endpoint for the clinical trials that led to the approval of the two new protease inhibitors for HCV genotype 1 patients (Jacobson et al., 2011). While the dependence on this measurement is essential, there is little technical discussion in the clinical literature on the imprecision of the assays and the clinical consequence of acting on such information, and no studies demonstrating how differences in the precision of the results will affect clinicians' decisions.

Quantitative HCV RNA viral load testing. The hepatitis C Virus was discovered in 1989; one of the first organisms to be linked to disease by molecular techniques alone (Choo et al., 1989). Establishing hepatitis C virus as the etiologic agent of one of the forms of non A/ non B hepatitis was accomplished by identifying and sequencing the viral RNA; without actually isolating or culturing the virus. Despite the newness of the molecular techniques used to isolate and quantify the virus, by 1995 viral load testing for HCV was being used regularly for prognosis and disease therapy management (Gretch et al., 1995; Martinot-Peignoux et al., 1995). By 2001, it was considered one of the most common tests performed in many molecular laboratories (Podzorski, 2001).

Early researchers demonstrated the discrepancies within and between the different available commercial and laboratory developed assays. There were no formal proficiency testing

programs for HCV VL at the time, so the discrepancies were observed by comparing the results of multi-center studies distributing shared samples. The earliest of these compared qualitative results on undiluted samples and on dilution series of samples. One early European study found that only five of the 31 participating laboratories from Europe, the US and Japan performed flawlessly on a panel of four HCV positive, six HCV negative and a dilution series with two samples (Zaaijer et al., 1993). The earliest College of American Pathologist (CAP) proficiency test panels demonstrated 13 – 95% correct results on the HCV RNA qualitative samples distributed in the first year to approximately 50 laboratories in the United States (Gretch, 1997). The Quality Control for Molecular Diagnostics (QCMD), a European Union supported proficiency testing effort, was the first to provide globally available qualitative and quantitative HCV proficiency testing panels. Two consecutive panels in 1999 and 2000 showed an improvement in qualitative testing where over 80% of global laboratories reported correct results. For those laboratories reporting quantitative results on the first panel, 60% were within the acceptable range of the geometric mean +/- 0.5 $\log_{10}$ and for the second panel in 2000, 73% were within this range (Schirm et al., 2002).

In addition to the differences in technologies and asynchronous states of development for available assays, there were no globally accepted standard reference materials, primarily because the virus could not be propagated. As a consequence, quantitative assays were developed with different reporting units: copies, genome equivalent units, and relative light units (Schirm et al., 2002). In 1996, the World Health Organization (WHO) in a global collaborative study with 22 laboratories established the first WHO International Standard for HCV RNA, expressing the quantity in International Units (IU) (Saldanha, Heath, Lelie, & the WHO collaborative study group, 1999). International units are assigned to biologic material WHO International Standards

and have no reference other than in relation to that standard material and the testing that quantified it (WHO, 2002). The difference in quantification among the global laboratories that tested the first standard, using either quantitative tests or limiting dilutions with qualitative testing, was greater than 2 $\log_{10}$. Although the inter laboratory imprecision was large, this was viewed as providing a step towards reducing the variability in the results of future tests.

Despite the establishment of the WHO International standard, assessment of proficiency testing for quantitative tests continued to be difficult. Participating laboratories still reported historical units, so conversion factors were used for multi-assay comparisons (Pembrey et al., 2003). Adding to the complexity was a technology shift to real time polymerase chain reaction (RT-PCR) which expanded the reportable range span from barely 5 $\log_{10}$ IU/ml to greater than 6 $\log_{10}$ IU/ ml (Barbeau, Goforth, Caliendo, & Nolte, 2004). Bias continues to exist between methods and among the genotypes in HCV VL, to the extent that it could influence treatment decisions. The recommendation to follow serial patient samples with one test method, and preferably through one laboratory, throughout the treatment experience of the patient, is still relevant (Caliendo et al., 2006; Laperche et al., 2007). In practice, however, this recommendation may not be easy to follow, and there is no information regarding actual compliance.

Even if clinicians had the ability to direct patients' serial samples to one laboratory using one method, there would still be reason for concern. Depending on the technology and the laboratory used, the inherent imprecision may make it difficult to determine the changes in viral load necessary to guide treatment decisions. At the time when therapy management decisions were made based on the detection of a 2 $\log_{10}$ reduction in viral load, the Instituto Superiore di Sanità in Italy conducted a multi-center study with 58 laboratories and two manufacturers using a panel

of samples designed to reflect the performance of this intended use. In addition to one panel sample at the sensitivity limit, six samples of the eight member panel were designed to test the ability of the tests to accurately detect a 2 $\log_{10}$ change. Three replicates each of two samples that were at 5.68 $\log_{10}$ and 3.61 $\log_{10}$ were distributed to the participants in each of two panel sets. In the data analysis, all combinations of the results between these two samples were calculated and the percent of unacceptable result-combinations was reported. The range of unacceptable results (less than 2 $\log_{10}$ IU/ml difference between the two samples) among the four methods tested was 30% to 66% with an overall rate of 40% of all the results. It was calculated that this could have led to the unnecessary discontinuation of therapy in up to 40% of patients. The researchers noted that if an uncertainty of 2 SD was considered in the evaluation of the results, the hypothetical incorrect decisions would have been reduced to 2% (Pisani et al., 2009). With the exception of the HCV VL results from the laboratories, the work that determined these conclusions was accomplished by mathematical modeling. The new therapeutic regimens for HCV genotype 1 incorporate the achievement of absolute reduction of viral load for the therapy management decision; the 2 $\log_{10}$ IU/ml change is no longer used. Though the limit of detection for the acceptable assays to determine the VL are specified in pharmaceutical labeling and clinical guidelines, the precision specifications are not, and there is variance around those absolute values in even the most reproducible tests. Whether or not physicians consider the imprecision in HCV VL tests as they are making therapeutic management decisions with these quantitative results is untested.

Justification for the Research

HCV Viral load testing is essential to the diagnosis and management of patients infected with the virus. The previous section discussed the recommendations for using HCV VL in clinical

decisions and some of the clinical research that demonstrated its utility. Based on research, the collective opinions of experts, and FDA submissions for therapeutics, HCV VL testing is standard practice to inform decisions in treatment of the disease. The efforts to understand and decrease the imprecision in HCV quantitative testing has also been discussed, as well as opinion leaders' views of the residual variation and the impact that could have on patient care. The research and discussion of these complex issues have been conducted among experts in laboratory testing, virology and hepatology. It is unknown if all treating physicians have integrated this knowledge into their practice. As new treatment regimens are introduced, it is important to understand where indecision, confusion, or non-adherence to recommended practices exist.

The current research regarding physicians' integration of test results in medical decision making illustrates a broad, inconsistent, and sometimes incorrect understanding of the precise meaning and application of the intended information supplied by laboratory testing. The types of laboratory tests that have been addressed in the studies discussed are primarily those informing the physician of the probable presence or absence of disease. There are no studies measuring physicians' understanding of continuous quantitative information, nor are there studies addressing the impact of integrating the knowledge of the imprecision in quantitative testing in their decisions.

It is appropriate to use HCV disease management as a model to study the implications of quantitative test variation on clinicians' decisions. The global impact of the morbidity and mortality on society has been discussed; but added to that is a growing body of publications outlining the costs of chronic HCV disease, cost-benefit impacts of intervention, and cost effectiveness of specific therapies. In the United States, the direct costs of HCV infection in

1998 were calculated to be one billion dollars (Wong, 2006). The extended incubation time before complications manifest will result in a calculated economic burden of $US 86.2 billion of direct and indirect medical expenses during 2010-2019. Adding quality of life, productivity and societal costs increases the estimate to $US 184 billion (Shah & Wong, 2006). The seriousness of the economic impact and the possibility of reducing that impact with suboptimal interventions drive studies to demonstrate the cost effectiveness of specific and comparative therapeutic regimens (Hornberger et al., 2006; Rajenrda & Wong, 2007). Paired with those studies and subsequent disease management strategies are the HCV VL tests that direct patients into the most effective therapeutic plan, but the monetary value of those tests is rarely studied. For example, the ability to cease treatment at 12 weeks for non-responding patients was calculated to save $US 15,000 per patient (Wong, 2006). Though HCV VL results drive the decision, there is no analysis of the contribution to savings the test may have. Further, there is no discussion of how the variation in the results might impact the decision and alter the utility and value of the pivotal determinant.

In the research presented here, these two important questions were studied. Referring to Smart's proposed model of clinical utility (2006), this study first examined the practicability, the knowledge and training needed to support HCV VL testing, by evaluating the understanding and impact the test results presented with imprecision information have on clinicians' use of the information. A survey presenting a specific scenario of a simulated patient completing 12 weeks of therapy and that patient's initial and 12 week HCV VL results at different levels was offered to clinicians (survey participants) who treat patients with chronic hepatitis C. In random order, each participant saw the test results with and without disclosing the imprecision, expressed as the 95% confidence intervals. The participants were asked to choose their course of action based on

the two different types of reports. The accessibility, or the economic and cost effectiveness consequences of using HCV VL results as "perfect" information, and as information with imprecision were then evaluated using decision analytic modeling and value of information calculations.

The purpose of this study was to test the hypothesis that clinician awareness of the expected imprecision in pivotal test results will impact their medical management decisions. The data from their responses were also used in decision analytic models to measure the changes in the value of the information provided by laboratory tests as clinicians alter their use of them. The work consisted of an observational study utilizing a clinician survey that presented a scenario where laboratory results provide the determining factor for the subsequent step in a patient's medical care.

The treatment of chronic hepatitis C using quantitative HCV RNA results was chosen for this study for several reasons: AASLD practice guidelines recommend the consideration of an action based on the results of the quantitative HCV RNA molecular tests. The guidelines and several publications recommend the use of a single test methodology by a single manufacturer throughout the care of the patient. It is unknown if clinicians comply with this recommendation and purposefully choose the same laboratory and method for their patient testing. Finally, there is an increasing reliance on both absolute and relative values derived from molecular viral load testing in the personalization of therapy regimes for several viral diseases. It is important to investigate the clinician mitigation and responses to the variation in this type of testing.

Contribution of the research. The hypothetical impact that the variation in VL testing could have on the ability to assess a 2 $\log_{10}$ reduction of HCV for patients on therapy was mathematically demonstrated (Pisani et al., 2009). However, whether or not clinicians would

have acted on the information as predicted in the European study was not tested. The new therapeutics and new guidelines for HCV genotype 1 patient management present an opportunity to analyze what physicians will do given the imprecision of the test results when making a decision around an absolute threshold. Physicians' abilities to appropriately utilize test results impact the value of those results, and consequently the tests from which the results are produced. Appraising the value of tests as information with realistic performance characteristics, as well as with data on the use of the information, is an approach not commonly employed in clinical utility discussions of laboratory tests. Additionally, the value of information as a measure of laboratory test value has not been published. It may provide additional information for the assessment, development and direction of future investment in clinical laboratory tests.

Chapter Three: Methods

The purpose of this research was to study if presenting imprecision information with laboratory test results impacts patient management decisions and the use of healthcare resources. The study utilized the quantitative test for HCV viral load (HCV VL), a pivotal test in clinical practice, as a model. The impact that knowing the imprecision, an inherent characteristic of the test, has on the decisions of those who use HCV VL for therapeutic management was measured in this study. Further, the test's economic value and the changes to that value based on the recorded decisions were calculated. For the first part of this research, a survey was developed to evaluate decisions made when the HCV RNA (HCV VL) results were provided with and without the 95% confidence intervals (CIs) that describe the imprecision. Using HCV VL results for a simulated HCV genotype 1 infected patient on triple-combination therapy, the survey was designed to compare the responses to VL results reported in the customary manner, without imprecision information, to responses to the same results reported with their 95% confidence intervals. Each survey participant was presented with both types of HCV VL presentations (with and without CI), but in random order. The characteristics of the simulated patient and the expected confidence intervals for the HCV RNA levels were derived from literature. It was expected that responses made viewing results with imprecision information would be different from the responses made viewing results in the conventional presentation.

Survey data were analyzed for differences in responses to the HCV VL results with CI and those without CI. The respondents' patient management decisions were then used as inputs to a

decision analytic model developed to calculate the value in cost savings and changes to that value based on changes in responses from the different presentations.

Survey Design

The survey was designed to focus on the decision made after 12 weeks of triple combination therapy for patients infected with HCV genotype 1 virus: telaprevir (Incivek™), alpha interferon and ribavirin. The selection of the decision point for the survey was based on information in the labeling of Incivek (Vertex Pharmaceuticals, 2011) and the updated American Association for the Study of Liver Diseases (AASLD) guidelines that incorporate treatment and evaluation guidelines for these therapeutic regimens (Ghany et al., 2011).

For treatment naïve patients without cirrhosis (the simulated patient in the survey), the following treatment recommendations are stated in the AASLD guidelines. (Class and levels refer to the strength of evidence supporting the recommendation):

> " The recommended dose of telaprevir is 750 mg administered with food (not low-fat) three times per day (every 7-9 hours) together with peginterferon alpha and weight-based ribavirin for 12 weeks followed by an additional 12-36 weeks of peginterferon alpha and ribavirin (Class 1, Level A).
>
> Patients without cirrhosis treated with telaprevir, peginterferon, and ribavirin, whose HCV RNA level at weeks 4 and 12 is undetectable should be considered for a shortened duration of therapy of 24 weeks (Class 2a, Level A).
>
> Treatment with all three drugs (telaprevir, peginterferon alpha, and ribavirin) should be stopped if the HCV RNA level is >1,000 IU/ml at treatment weeks 4 or 12 and/or detectable at treatment week 24 (Class 2a, Level B)."

The survey was designed according to the clinical path in Figure 5. The decisions that participants made based upon the HCV VL results after 12 weeks of therapy, with and without confidence intervals, were recorded and analyzed.

Participants were presented with a description of a simulated patient with HCV genotype 1 disease who had been on triple combination therapy for 12 weeks. The initial HCV VL for this patient, as well the description of that patient's tolerance to the therapy over the 12 week course were also described. Five different levels of HCV VL results were presented to the survey participants with and without the 95% confidence interval for each result. The order in which these were presented was randomized. The HCV VL levels in the survey included results for which the patient management choice made according to practice guidelines would be relatively clear, as well as results for which the choice could be equivocal. Participants were asked to consider each result individually and choose one of three patient management decisions: to continue therapy, to discontinue therapy, or to repeat the HCV VL test. The survey was hosted on Survey Monkey®, an internet based, secure survey service. (Appendix A contains the complete survey text.)

Although the HCV VL result can be pivotal in the 12-week treatment management decision for HCV patients on triple combination therapy, other patient variables are considered by clinicians, such as age, weight, sex, degree of cirrhosis, and tolerance to therapy (Ghany et al., 2009; Sulkowski, 2008). The survey was designed to reduce as many of these variables as possible by presenting one patient with average demographics and a moderate set of adverse events. Since professional practice, judgment, and experience can also influence the decisions, questions relating to the participants' demographics and practice were incorporated into the

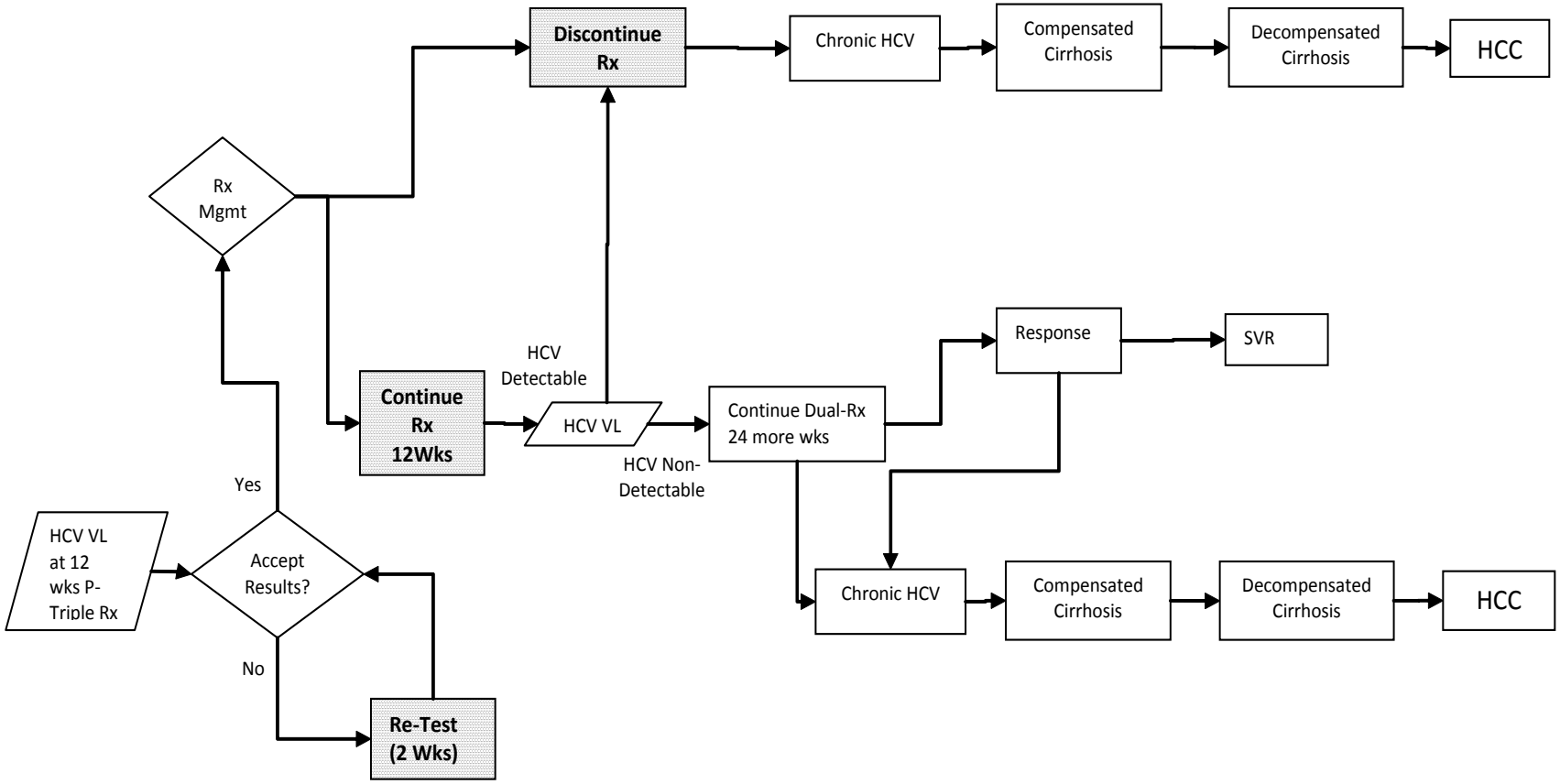

Figure 5. Clinical path for HCV genotype 1 patients after 12 weeks of triple combination therapy. The three shaded boxes represent the choices following the presentation of different HCV VL results with and without confidence intervals in the survey.

survey. These included professional credentials, the number of years the participant has been treating HCV, the average number of HCV patients seen per month, the area of practice specialty and the geographic region in which the participant practiced. Other items in the survey queried preferences regarding the prescription and use of laboratory tests, such as the ability to direct patient testing to a single testing laboratory or method, the desire to have confidence intervals reported with results, and the preferred guidelines used for managing patients with HCV disease.

Selection of HCV VL test results and imprecision ranges. One set of quantitative HCV RNA test results presented to the participants represented typical HCV VL results. From most laboratories these results are reported without information regarding the imprecision of the assays. The other set of results reported the identical test values, but with the 95% confidence intervals for those values, as well.

The 95% confidence intervals for the HCV VL results in the survey were based on published literature. Imprecision statistics are published from multi-site trials, assay studies and proficiency test summaries, and they provide insight to the expected confidence intervals within a laboratory. However, these are usually focused experiments, sometimes with artificial or pooled samples (Caliendo et al., 2006; La Perche et al., 2007; Pisani et al., 2009). The confidence intervals used in the survey for this research were calculated according to a recommendation by the Virology Quality Assurance Laboratory Program (VQA). The VQA program, formed by the virology committee of the AIDS Clinical Trials Group uses an intra-laboratory SD criterion of 0.15 $\log_{10}$ SD for viral load assays (Yen-Lieberman et al., 1996). It was first established for HIV assays, however, laboratory directors producing results for other VL assays recognize it as a goal (Caliendo et al., 2006). Though the recommendation was made

in 1996, later published studies either refer to this recommendation or their data demonstrates that 0.15 $\log_{10}$ = 1SD is still a reasonable estimate (Table 2).

Table 2

Three Studies with Published Intra-laboratory Reproducibility Statistics

Study	Year	Replicates	VL ranges: 0-2.9 $\log_{10}$ SD $\log_{10}$	3-4.9 $\log_{10}$ SD $\log_{10}$	> 5 $\log_{10}$ SD $\log_{10}$
Foreman &Valsamakis	2004	20	0.2	0.2	0.1
Caliendo et al.	2006	7	Approx. 0.15	Approx. 0.15	Approx. 0.15
Chevaliez et al.	2009	6	0.8	.04	0.21

For the imprecision budget of 1SD = 0.15 $\log_{10}$ used for the HCV VL results in the survey, the calculated 95% CI ranges were determined as shown in Table 3. They were rounded for the survey presentation.

Table 3

95% CIs for HCV VL Results in Survey. Based on 0.15 $\log_{10}$ =1SD

	HCV VL IU/ml	
Test Result	Low 95%	High 95%
562,000	286,000	1,107,000
31,620	16,060	62,230
1,950	991	3,837
1,000	508	1,968
562	286	1,107
100	51	197

Using the information in Table 3, if patients are managed according to absolute test values and clinical guidelines alone, without consideration of test imprecision and other factors, the responses in Table 4 could be predicted at each HCV VL level for the 12-week result.

Table 4

Predicted Decisions Based on HCV VL Test Result and Clinical Guidelines

Baseline (IU/ml)	Post 12 wk Rx HCV VL (IU/ml)	Decision according to guidelines
562,000		
	31,620	Discontinue therapy
	1,950	Discontinue therapy
	1,000	Continue/Discontinue therapy
	562	Continue therapy
	100	Continue therapy

Selection of the Study Sample

Clinicians who treat patients with chronic HCV disease were the intended population for this survey. Access to these clinicians was through the member directory of the AASLD-American Association for the Study of Liver Diseases. Members with physician (MD or DO), physician's assistant (PA), nurse practitioner (NP) or nurse (RN) credentials who did not indicate a pediatric specialty were invited to participate through a personally addressed e-mail containing a link to the survey.

Sample size. To compare the decisions made with and without the presentation of confidence intervals, it was determined that a sample size of 99 would be needed for an effect size of .1 (showing a 10% difference in responses across the three decisions) with 2 predictors (the level of the viral load and the presence/ absence of confidence intervals) an alpha of .05 and power of .8. (On-line sample size calculator (http://www.danielsoper.com/statcalc/calc01.aspx). Adding demographic and clinician practice variables, with up to 12 predictors, a sample size of 184 would be required for the same power and alpha. Invitations to join the survey with an enclosed link were sent by the primary researcher to 1491 selected AASLD members. The first page of

the survey contained an explanation and consent form. The study received approval by Virginia Commonwealth University IRB #HM14538.

Variables in the Survey

The decision to continue or discontinue treatment (therapy decision) is a function of the information presented to the clinician (patient and test result information), clinicians' judgment and experience (clinician variables), and standard or accepted clinical practices. The survey was designed to test many of these variables. The variables are listed in Table 5.

Testing a Draft Version of the Survey

Prior to the release of the final survey, a draft version of survey (HCVVL Survey- Draft 3) was sent via e-mail link to three physicians, one laboratory director and a clinical trials manager. Instructions to review the survey and send comments were provided in the e-mail with the internet link and PDF attachment of the survey. Some recipients of the e-mail chose to complete the survey and in addition to their comments, their survey responses were recorded.

For those that actually completed the survey, the responses were "analyzed" as a test of the survey and the internet hosted service. There were four recorded responses: two from physicians and two from clinical research managers (one was also a medical student). Three of the four respondents stated they managed HCV patients. The time to complete the survey was less than 11 minutes for all respondents.

All respondents chose integer values for the test results but one of three that answered the question stated that their laboratory reports are provided results both in integer and log values.

For the four that answered the main question, there was a difference in the decision choices made using the test values without the confidence intervals (No CI) and those results with the confidence intervals (with CI). The pairs where the decision was different between the two

Table 5

Variables in the Survey

Variable	Variable Type	Source	Analytical type
All participants			
Retest	Dependent	Survey	Categorical
Discontinue Therapy	Dependent	Survey	Categorical
Continue Therapy	Dependent	Survey	Categorical
Result presentation with or without 95% CI (intervention)	Independent	Provided	Categorical
HCV VL	Independent	Provided	Categorical
Contribution of CI results	Independent	Survey	Categorical
Desire to have CI results	Independent	Survey	Categorical
HCV Pt management professionals			
Years of experience	Independent	Survey	Continuous
No. of patients per month	Independent	Survey	Continuous
Guidelines used for managing HCV patients	Independent	Survey	Categorical
Use of HCV VL for Rx management	Independent	Survey	Categorical
HCV Pt management professionals utilizing HCV VL			
Send tests to same testing lab	Independent	Survey	Categorical
HCV VL test method specified	Independent	Survey	Categorical
Type of VL test results received from testing lab	Independent	Survey	Categorical
Guidelines used in treating HCV patients	Independent	Survey	Categorical
Physicians			
Field of practice	Independent	Survey	Categorical

presentations are highlighted in yellow in Table 6. The survey instrument appeared to be able to elicit and record responses and the data was retrievable in a manner that could be used for the planned analyses.

Final Survey Version

The final version of the survey distributed for the study was changed based on the reviews by this pilot group, and research committee members. Changes included:

Table 6

Responses to Patient Management Questions from Draft HCV VL Survey

Test Result (IU/ml)	% Response					
	Continue Rx with Dual Therapy		DC Rx		Retest	
	No CI	With CI	No CI	With CI	No CI	With CI
100,000	0	0	100	100	0	0
5,623	0	0	100	100	0	0
1,950	0	0	100	75	0	25
1000	25	0	0	0	75	100
562	100	75	0	0	0	25
100	100	100	0	0	0	0

1) The order of presentation of the HCV VL results with and without CIs was randomized. (In the pilot version, the order was constant: without CI presented first, with CI presented second.)

2) HCV RNA test result numbers were only presented as integers in the final survey. In the pilot survey, participants were given a choice between integer and $\log_{10}$ values. This choice was sacrificed on the final survey for the ability to randomize the question order. All those participants completing the pilot survey chose integers, and in discussions with laboratory directors on the subject, integer presentation was the predominant form for laboratories not providing both.

3) The pilot version included a question about the importance of other non- HCV VL test considerations for therapy management. All participants rated each of items at least moderately important. In view of the length of the survey and consideration that this did not contribute to the aim of the survey, this question was eliminated.

Other feedback from participants included questions regarding the drug of choice, suggestions to randomize the presentation of results with and without CIs and a request to simplify the survey. The three adjustments described above were made in response to their feedback.

The Decision Analytic Model

The treatment decision responses from the survey were used in an economic model to assess if changes in the treatment decisions influenced the value of the test. A development decision analytical model was constructed in Microsoft Excel 2007 to calculate the changes in test value as the use of the information changed.

The key assumptions in the development model were:

- The entry point in the model was at the HCV VL test following 12 weeks of triple combination therapy (Incivek, interferon –alpha and ribavirin); the point where AASLD guidelines recommend patients with HCV VL above 1000 IU/ml be removed from further treatment.
- The initial choice of decisions at the entry point was: Discontinue therapy, Continue therapy or to Repeat the test. (The same decisions as in the survey).
- Parameters and disease states were for HCV genotype 1 patients. For the development model, the health states were: Response, SVR, Chronic HCV, Cirrhosis, Hepatocellular Carcinoma, and Death.
- The time horizon of the development model was 72 weeks. This encompasses the length of influence of the HCV VL test in this clinical pathway, allowing the possibility of two weeks delay for re-testing, an additional 36 weeks of continuing therapy plus 24 weeks post therapy and testing to evaluate an SVR (SVR is defined as undetectable HCV RNA 24 weeks post the completion of therapy.) Sixty two weeks were needed to

accommodate the two week repeat-test decision, but since the model was in 12-week cycles, 72 weeks were used.

- The development model employed Markov chains to map the probabilities of moving through health states after the treatment management decisions had been made and entered.
- With the exception of the first two cycles, each time cycle was 12 weeks.
- In the first cycle, the population of patients entered in the "repeat test" decision remained in a chronic HCV state on dual combination therapy for two weeks and then were "re-entered" back into the population for 10 weeks based on the percent decision to discontinue or continue therapy from those that did not choose to re-test. This was designed to simulate the time for a new sample to be drawn, tested, and reported with subsequent clinician interpretation and decision.
- There were no Quality of Life (QoL) values calculated in the development model.

Parameters for the development model. The purpose of the development model was to test functionality and sensitivity to inputs. For the development model, transition probabilities and cost parameter values were derived from literature and bureau of statistics tables where accessible, however, estimations were made for several of the parameters. After testing the development model for functionality, all parameter values were sourced from literature and documented for the final model.

Testing the development model. The model was checked for mathematical accuracy by visually checking each equation, assuring all populations were 100% at each state, and after running the model, by analyzing the outputs for consistency with expected results.

A minimum of 235 scenarios were tested in the development model. Included were:

- All patients continuing on therapy (as if there were no test)
- All patients removed from therapy (therapy failure with no test)
- Using the response/ no-response population percentages cited in the clinical trials for Incivek (telaprevir)
- All combinations of the three decisions: Repeat the test/ Discontinue therapy and Continue therapy in 5% increments. (Simulating increased indecision based on test imprecision.)

The model output for a patient without therapy costs and without testing was $4,917 per person during this period. These patients continued on a path of chronic HCV and subsequent disease management cost consequences for the entire 72 weeks.

The total costs for remaining on therapy in the development model were $14,474 per patient. This represented the costs if there was no HCV VL testing at week 12 and therapy was continued for all patients, including those patients with VL greater than 1000 IU/ml (who should be removed from therapy). HCV VL testing saves medical costs and quality of life loss due to medication adverse events. The value of perfect test information was calculated to be the value of the savings from discontinuing a therapy regimen that was not appropriate (up to $9557) for the individual patient with a true HCV VL above 1000 IU/ml. However, the development model had no consideration of penalties for incorrectly removing a patient from therapy. There were no quality of life values for health states, and the costs of disease progression were not fully developed.

To test a scenario duplicating the expected population response rates in the clinical trials for telaprevir, inferences from the clinical trial literature were made. Though a cutoff of <1000 IU/ml for continuing therapy is given in literature and guidelines, there is no data in the studies

showing the percent of patients with HCV VLs greater than 1000 IU/ml at week 12. The telaprevir clinical trial reported that 58% of patients had an extended Rapid Virological Response (eRVR), defined as an undetectable HCV VL at week 4 and 12. Among the 42% of patients who did not have an eRVR, a Sustained Virological Response (SVR), which is the treatment goal, was achieved in 60%. Therefore, it was assumed that approximately 17% (42% x .40) of total patients starting therapy could be expected to have an HCV VL at week 12 that would indicate that they are non-responsive to the therapy.

In a population where approximately 17% would have an HCV VL that would dictate removal from therapy at week 12, the average per-patient cost would be $6542. Therefore, testing on a population basis, and removing patients with HCV VL's > 1000 IU/ml from unproductive therapy, should save approximately $7932 (14,474-6542) per patient, provided the test is reliable, and not repeated.

To model the consequences of imprecision, increasing retest rates were used as a surrogate. The percent retest population was started at 0% (completely accurate, precise and reliable information) and was increased up to 100% in 5% increments. Each of these were tested with all combinations of decisions to continue or discontinue therapy; also varied in 5% increments. The relative changes in total costs as the percent of decisions to repeat the test increased, were plotted from the model outputs and are shown in Figure 6.

Conclusion of development model testing and final model. Through the development model studies, it was concluded that mechanistically, the outputs from the model produced changes in total costs as a result of different percentages of each decision entered. For the final model used in the research, values for transition probabilities, costs and QoL were sourced from

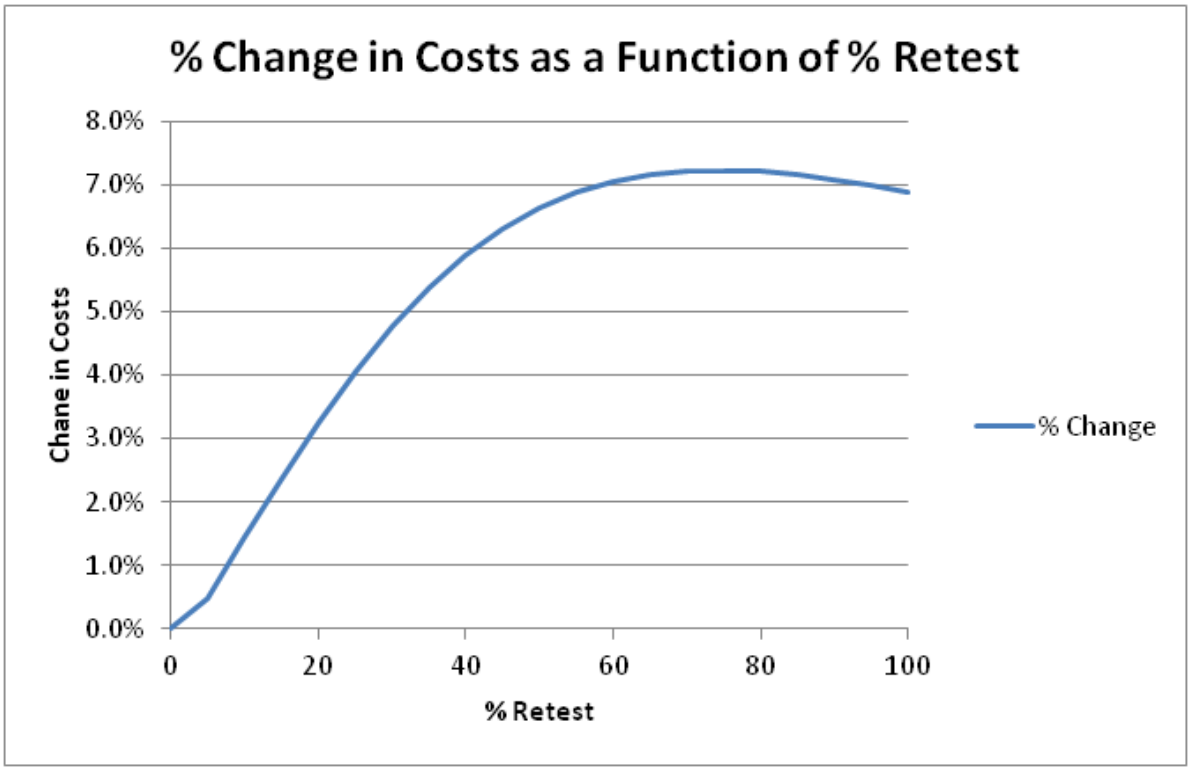

Figure 6. Percent change in costs as a function of percent retest rate.

literature. The dual therapy costs were adjusted for inflation. Table 7 lists the parameters used in the model and their source.

Costs for health states were derived from Pyenson, Fitch, and Iwasaki (2009). Commercial and Medicare costs (for those above and below 65years of age) were provided in their analysis. From that publication, the Medicare costs for those less than 65 years of age were used in the final version of the model. For several of the health states, they were more conservative than the commercial costs. The health state costs derived from Pyenson et al. were not inflation-adjusted. In the review of other literature, including an analysis of a large managed care organization claims database from July 2001 to March 2010, the variability of reporting costs of each health state for HCV disease was revealed (McAdam-Marx et al., 2011). Monthly costs at published

Table 7

Decision Analytic Model Parameters

Parameter (12 weeks)	Value	Source
Triple combination therapy responses		
Undetectable HCV VL p 12 wks dual Rx	0.58	Jacobson et al., 2011
Undetectable HCV RNA at Rx end	0.87	Jacobson et al., 2011
Continued Response	0.75	Jacobson et al.,2011
Health state transition probabilities		
Progression: Chronic HCV to cirrhosis	.006	Pyenson et al., 2009
Progression: Cirrhosis to Decompensated Cirrhosis	.010	Pyenson et al., 2009
Progression: Decompensated Cirrhosis to HepCA	.010	Pyenson et al., 2009
Costs		
Cost of HCV viral load test	$58.88	CMS, Laboratory Fee Schedule, April 2013
Cost of 12 wks dual therapy (interferon and ribavirin)	$5,721	Piwko et al, 2005 Adj to 2013 via http://www.bls.gov/cpi/cpicalc.htm
12 week period costs for patients with SVR[1]	$0	
12 week costs for patients with Chronic HCV[1]	$1545	Pyenson et al., 2009
12 week costs for patients with Cirrhosis[1]	$2345	Pyenson et al., 2009
12 week costs for patients with Decompensated Cirrhosis[1]	$21,300	Pyenson et al., 2009
12 week period costs for patients with Hepatocellular Ca[1]	$7200	Pyenson et al., 2009
QoL Values		
Chronic hepatitis	0.82	Australian Government, Dept of Health and Aging, 2002
Cirrhosis – diagnosed	0.74	Australian Government, Dept of Health and Aging, 2002
Decompensated cirrhosis	0.32	Australian Government, Dept of Health and Aging, 2002
Hepatocellular carcinoma	0.10	Australian Government, Dept of Health and Aging, 2002

[1]Excludes HCV dual therapy and HCV test costs. Includes other medical costs associated with health state

and inflation adjusted values were compared (Table 8). The values as published in Pyenson et al. were higher in some cases than McAdam-Marx et al. and lower in others. Therefore, the values as published in Pyenson et al. were used in the model as published (first column of costs in Table 8). A sensitivity analysis of the impact of cost variation was performed with the model and is reported in the results section.

Table 8

Comparison of Published and Inflation Adjusted Costs for HCV Health States

	Monthly Costs ($)			
HCV Health State	Pyenson et al., 2009	Pyenson et al. adjusted*	McAdam-Marx et al., 2011	McAdam-Marx et al. adjusted*
Chronic HCV	515	556	488	529
Compensated Cirrhosis	772	833	437	474
Decompensated Cirrhosis	4200	4536	2295	2488
HCC	2400	2592	3616	3919
* Adjusted for inflation to 2013 costs per http://www.bls.gov/cpi/cpicalc.htm				

Other changes made to the final version of the model used in this research were: removal of death as a health state, removal of SVR as a separate health state from Response, and reduction of the time from 72 to 60 weeks (more appropriately reflecting the time span over which the test has influence; that needed to evaluate SVR). In addition, the health state "Decompensated cirrhosis" was added and Quality of Life (QoL) values were accumulated at each state.

In addition, because the clinically prescribed decisions are not identical for all VL levels, different inputs and patient flow strategies were used for each VL level when testing the inputs from the survey at each distinct VL level.

For each viral load level, the model was run with at least three separate sets of inputs.

1. Inputs reflecting all patients on therapy for the duration of therapy, reflecting a situation where no test is available to guide removal of non-responders. (For this situation the model had to be adjusted to remove the 24-week HCV test cost and subsequent routing based on the results.)
2. Inputs according to AASLD guideline-directed decisions (e.g. no repeats and using the percent distribution of the population VLs at that level, based on the imprecision used in the survey 1SD =0.15 $\log_{10}$)
3. Inputs according to the proportion of decision choices by respondents in the survey. Only data from the first presentation of HCV results to the respondents were used. These data were used, as opposed to combining all responses, because clinicians normally are provided with a single presentation of results to review.

The decision proportions were entered into health states at time zero (review of HCV results post completion of 12 weeks of triple combination therapy) according to the following strategies:

1. For the Discontinue therapy arm, all decision proportions were entered in the Chronic HCV health state as no further therapy would be given.
2. For the Continue therapy arm of the model:
    a. The Repeat test decisions were combined with the Continue therapy decisions for entry. After a cycle of two weeks, this proportion of responses was re-distributed among the discontinue therapy and continue therapy arms according to the percent at in the original entry.
    b. For decisions recorded at a HCV VL=1950 IU/ml, proportions were entered into the health states based on the distribution of the VL, using the CIs in the survey. At an HCV VL =1950 IU/ml, the lower 95% confidence limit is 990 IU/ml. Approximately

2.7% of the HCV VL values would be at 1000 IU/ml or below. Therefore, in the Continue therapy arm, 2.7% of the decisions were entered in the Response health state and 97.3% were entered in the Chronic HCV health state. For guideline-directed decisions, the proportions used for each arm were: 0% Repeat test/97.3% Discontinue therapy/ 2.7% Continue therapy.

c. For decision choices recorded at HCV VL1000 IU/ml proportions in the Continue therapy arm were split equally between the Chronic HCV state and the Response state. For guideline-directed decisions, the proportions used for each arm were: 0% Repeat test/50% Discontinue therapy/ 50% Continue therapy.

d. For decision choices recorded at HCV VL 560 IU/ml proportions were entered into the health states based on the distribution of the VL using the CIs in the survey. At an HCV VL of 560 IU/ml the upper 95% confidence limit is 1110 IU/ml. If 97.5% is at 1110 IU/ml or below, then 95.2% is at 1000 IU/ml or below. Therefore, in the Continue therapy arm 95.2% of the decisions were entered in the Response state and 4.8% in the Chronic HCV health state. For guideline-directed decisions, the proportions used for each arm were: 0% Repeat test/4.8% Discontinue therapy/ 95.2% Continue therapy.

e. For decision choices recorded at HCV VL 100 IU/ml decisions for the Continue therapy arm were entered in the Response health state. For guideline-directed decisions, the proportions used for each arm were: 0% Repeat therapy/0% Discontinue therapy/100% Continue therapy.

As with the development model analyses, inputs representing a population of diverse HCV VL levels were also analyzed. The decision inputs were designed so that the therapy response

outcomes of the model reflected the telaprevir trial outcomes as closely as possible. The inputs are described in the results section.

Each decision set (proportions for Continue therapy, Discontinue therapy or Repeat test) was used for the initial inputs to a model run. The resulting total costs and the average QoL ratio from all branches of the model, each time the model was run, were recorded and compared. The total costs of the survey-recorded decision choices with and without CI were compared against the total costs of the guideline-directed decisions at each VL level. In addition, the total costs resulting from the removal of non-responders from therapy were subtracted from the total costs of keeping all patients on therapy for a determination of the value of the quantitative HCV RNA test information. Sensitivity analyses of the model to costs, QoL values and retest rates were also performed.

Hypothesis Testing

The recorded responses from the survey were analyzed using Chi-square, bivariate and multinomial logistic regression techniques to test the first three hypotheses defined for this research.

H1 null hypothesis: The presentation of HCV VL results with 95% confidence intervals will have no impact on the treatment decision for HCV Genotype 1 patients.

It was presumed that the presentation of results with CI would result in different decisions when compared to the decisions made with results presented in the conventional manner. For the HCV VL load levels immediately below the threshold, at the threshold, and immediately above the threshold it was expected that there would be increased ambiguity between the Discontinue and Continue therapy choices, and higher repeat testing decisions for the results presented with CI. Since the CI's for the HCV VL immediately above the threshold crossed slightly below the

threshold, it was also thought that there might be an increased proportion of clinicians deciding to continue therapy when compared to decisions made viewing the same HCV VL result without CI. Anticipating these findings, the second hypothesis was formed to further characterize the differences in decisions that might be observed.

H2 null hypothesis. There is no difference in the impact of awareness of CIs on the treatment decisions between different HCV VL levels.

H3 null hypothesis. Clinicians' experience and demographics will have no impact on treatment decisions using HCV VL results presented with and without confidence intervals. If there were differences in the decisions made viewing results with and without the presentation of CI, it was thought that differences in experience and practice might correlate with the ability to apply the information correctly.

In addition to testing these hypotheses, the recorded responses from the survey were used as inputs to the decision analytic model to evaluate the economic impact of the clinicians' decisions. It was proposed that the value of the HCV VL test at the 12- week decision point in therapy could be determined by calculating the cost of treatment and disease progression, leaving all patients on therapy throughout the course and comparing them to the costs in the scenario where patients who would fail therapy, revealed by the HCV VL test results, did not continue the dual-combination treatment.

Chapter Four: Results

The purpose of this research was to examine methods of measuring the contribution of laboratory tests to healthcare. Important aspects of a laboratory test's contribution are the clinical utility of the test information, and the role of that information in facilitating efficient and effective use of patient management resources. The use of quantitative HCV RNA test results (HCV VL) to decide whether to continue or discontinue antiviral therapy after 12 weeks of treatment was used as a model for this research. For the clinical utility of the HCV VL test, precision is crucial; therefore, the clinical utility was evaluated by studying the impact that the knowledge of the inherent assay imprecision might have on patient management decisions. To measure the test's economic contribution, a Markov decision-tree model was developed to determine the cost differences in leaving all patients on therapy versus removing those who would not respond given their HCV VL result. Further, the cost differences resulting from removing patients from therapy as directed solely by clinical guidelines and those resulting from removing patients according to the clinicians' decisions from the survey were determined and compared.

Impact of Test Imprecision on Clinician Decisions

A survey was designed to elicit clinicians' decisions when HCV VL results were provided with the 95% confidence intervals around five different levels of HCV viral load results. An invitation e-mail with the survey link was sent by the primary researcher to 1491 members of the American Association for the Study of Liver Diseases. At the close of the survey, 138 responses

were recorded (9.3%). Of those responding, 14 accepted only the consent and progressed no further, six of the respondents answered only one of the two questions with the HCL VL results, and the remaining 118 answered both. Though respondents could exit the survey at any point, 114 answered demographic and clinician preferences questions as well. Respondents used an average of 8.3 minutes to complete the survey. Recorded times greater than 60 minutes were removed from the data. It was assumed that these respondents had been interrupted or left their computers with the survey open and that the elapsed time was not reflective of time spent answering the questions.

Sample. The sample of respondents consisted of health care professionals who would utilize HCV VL results in their clinical practice. Physicians comprised 89% of respondents and 92% of respondents answered that they managed the treatment of hepatitis C patients. Responders were in practice an average of 20.5 years and saw an average of 47 patients per month. Ninety-one percent claimed their field as hepatology, gastroenterology or both. All respondents indicated they use HCV RNA results in the management of their patients with hepatitis C. Of those, 84.2% indicated they send their samples to the same testing lab for the pre-treatment and the 12-week test points most of the time. The remaining 15.8% responded it was not within their control. Sixty-one percent specified that the identical test method be used for the pre-treatment and the 12 week test points; while 28.9% responded it was not within their control. Of the 114 respondents to the question of practice guidelines used, 96% indicated the use of the AASLD guidelines. EASL guidelines were utilized by 15.8%, and information from pharmaceutical companies was utilized by 13.2%. These responses indicate knowledge of the current clinical practice guidelines for the management of hepatitis C disease, and the intent to comply. Table 9 contains the responses to the questions regarding practice preferences and logistics.

Table 9

Respondent Preferences and Patient Management Practices

Preferences and Practices	Percent of Responses
Preferences regarding presentation of CI:	
The results with CI n=118	
Neither added to nor detracted from the test result	44.9
Provided relevant added information to the test result	38.1
Detracted from or confused the test result	16.9
Desire to receive the 95% CI with patient results? n=114	
No	46.5
Yes	43
Not decided	7.9
Already do receive them	2.6
Practices regarding HCV RNA testing:	
Send tests to the same lab for Pretreatment and 12-wk HCV RNA. n=114	
Yes	84.2
Not within respondents control	15.8
No	0
Specify that identical methods be used for pre-treatment and 12-week HCV RNA. n=114	
Most of the time	61.4
Not within respondents control	28.9
Not at all	8.8
Sometimes	0.1
Type of VL results received from laboratory. n= 114	
Integers and Log_{10}	71.9
Integers	20.2
Log_{10}	6.1
Unsure	1.6
Guidelines used for patient management. (Non-exclusive choices.) n=114	
AASLD	96.5
EASL	15.8
Pharmaceutical Company	13.2
Institutional	7
Others	Less than 7

Descriptive statistics of the sample demographics are in Table 10. The responses were analyzed in total and then were split, based on whether the respondents saw the results presented without CI first (No CI) or with CI first (With CI). For categorical data, the chi-square test was used to test the independence of the observed versus expected frequencies in each of the two population samples (Microsoft Excel 2007). For continuous data, the TTEST function (Microsoft Excel 2007) was used to test for significance. The two groups, separated by the order of the presentation of results with confidence intervals, were demographically similar; no statistical differences were determined. They could be used be used as two separate sample groups for future comparisons.

Hypothesis testing. Different groupings of the data were used in the following analyses for hypothesis testing. Analyses of multiple decisions and variables employed Pearson's chi-square, logistic regression, and multinomial logistic regression (IBM SPSS version 20). Analyses at each VL level, comparing each of the three decisions made without CI (NoCI) and with CI (With CI) were performed by chi-square tests (Microsoft Excel 2007). Figure 7 outlines the different data groupings.

***H1 null hypothesis*.** The presentation of HCV VL results with 95% confidence intervals will have no impact on the treatment decision for HCV Genotype 1 patients.

H2 null hypothesis. There is no difference in the impact of awareness of CIs on the treatment decisions between different levels of HCV VL.

These two hypotheses are interdependent. In a thorough analysis of $H_0 1$, $H_0 2$ is included as the data was analyzed for the impact of CI on the three decisions in total and separately at each VL level.

Table 10

Sample Population Demographics

				Seen First				
				No CI		With CI		
		Total	Percent	Number	Percent	Number	Percent	p
Participants*		124		65	52.4	59	47.6	
	Location of practice							
	US	101	81.5	52	80.0	49	83.1	0.98
	Europe	6	4.8	3	4.6	3	5.1	0.99
	Canada	5	4.0	4	6.2	1	1.7	0.66
	AUS/NZ	2	1.6	2	3.1	0	0.0	0.61
	LATAM	2	1.6	1	1.5	1	1.7	0.99
	No response	8	6.5	3	4.6	5	8.5	0.58
	Profession							
	Physician	110	88.7	59	90.8	51	86.4	0.97
	Nurse Practitioner	1	0.8	0	0.0	1	1.7	0.77
	Nurse	1	0.8	1	1.5	0	0.0	0.82
	Physician's Assistant	1	0.8	0	0.0	1	1.7	0.78
	Other	1	0.8	1	1.5	0	0.0	0.82
	No response	10	8.1	4	6.2	6	10.2	0.87
	Manage hepatitis C patients							
	Yes	114	91.9	61	93.8	53	89.8	0.88
	No	4	3.2	2	3.1	2	3.4	0.99
	No response	6	4.8	2	3.1	4	6.8	0.82
	Number of patients per month							
	Responses	114	91.9	61	93.8	53	89.8	
	Range	0-300		2-300		0-300		
	Median	30		35		30		
	Avg	47.0		54.1		38.8		
	SD	56.6		61.4		49.8		0.15
	Number of years in practice							
	Responses	114	91.9	61	93.8	53	89.8	
	Range	1-45		1-40		4-45		
	Median	20		20		20		
	Avg	20.5		20.7		20.3		
	SD	10.0		9.8		10.3		0.82
	Field of practice	(Non-exclusive multiple choice)						
	Responses	114	91.9	61	93.8	53	89.8	
	Hepatology	88	71.0	47	72.3	41	69.5	0.99
	Gastroenterology	34	27.4	18	27.7	16	27.1	0.99
	Infectious diseases	9	7.3	5	7.7	4	6.8	0.99
	Internal Medicine	4	3.2	2	3.1	2	3.4	0.99
	Pediatric Hepatology	1	0.8	0	0.0	1	1.7	0.77
	Virology	1	0.8	1	1.5	0	0.0	0.82
	Pathology	1	0.8	0	0.0	1	1.7	0.77
	Research	1	0.8	0	0.0	1	1.7	0.77
	Blanks	10	8.1	4	6.2	6	10.2	0.88
	*Answering at least one question							

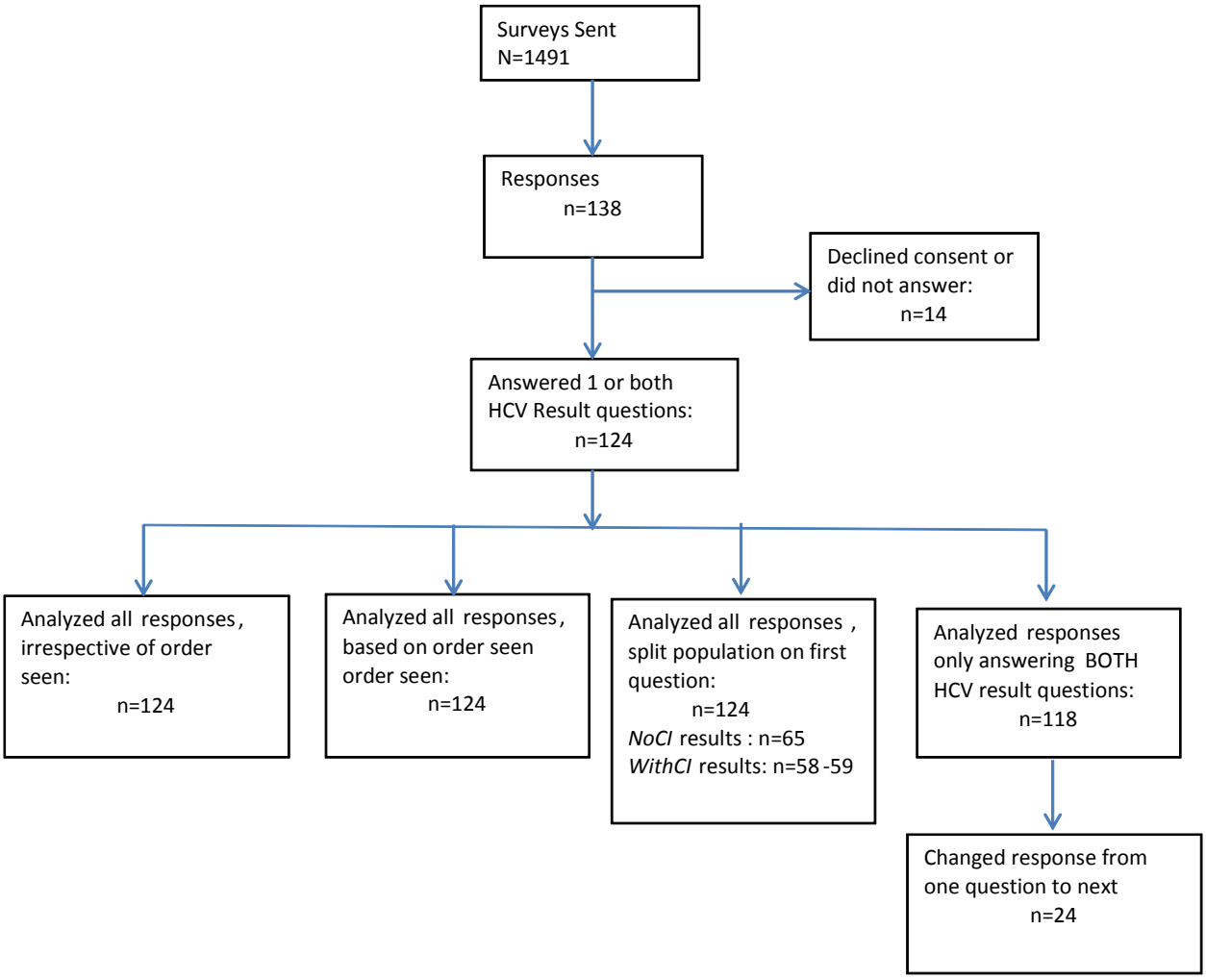

Figure 7. Groupings of data for analysis of survey responses.

The strategies used to analyze the responses in relation to the presentation of results with and without CI were (Figure 7):

- Analysis of the decisions from all respondents as one group.
- Analysis of the decisions between two groups, based upon whether the respondents saw HCV VL results with CI first or without CI first.
- Analysis of the impact of the presentation order of results with and without CI on the decisions made.

- Analysis of the impact of the presentation order of CI on individuals whose decisions were different between the first and second questions.

Analysis of decisions from all respondents independent of order seen. Data from 124 respondents were used, and decisions based upon the results with CI and the results without CI were compared.

Multinomial logistic regression was used to analyze the impact of CI on choosing one of the three possible decisions: Discontinue therapy, Continue therapy or Repeat the test. Through logistic regression methods (bivariate and multinomial) the probability of an outcome can be predicted based on the known values or frequencies of occurrence of the predictor variables. For the first analysis, the sole predictor variable was the presence or absence of CI with the quantitative HCV VL results. Table 11 lists the impact of the variable CI on the decision outcomes. Multinomial logistic regression uses one of the outcomes as a reference and returns one less than the number of outcomes as models. In this research, there were three outcomes so two models in relation to the reference are returned. The Repeat test outcome was the reference in these regressions. Separate analyses of the likelihood of repeating the test was performed by binomial logistic regression and is presented later. In logistic regression, the coefficients for each variable (B) are selected through maximum likelihood estimation, fitting the parameters to the observed data. They represent the change in the logit of the outcome variable (the natural logarithm of the odds of the outcome happening) associated with a one-unit change in the predictor variable. The Wald statistic denotes whether the coefficient is significantly different from zero, the strength of that predictor variable. It is a z-statistic with a chi-square distribution; therefore its significance can be reported. Exp B is the exponential of the predictor coefficient, B (natural logarithm) and as such indicates the change in odds based on a unit change of the

Table 11

Multinomial Regression Statistics with Variable CI. All Responses

Decision at all HCV VL levels[a]		B	Std. Error	Wald	df	p	95% Confidence Interval for Exp(B) Odds ratio		
							Lower Bound	Exp (B)	Upper Bound
Discontinue therapy									
	Intercept	1.213	.119	103.157	1	.000			
	Without CI	.141	.174	.654	1	.419	.818	1.151	1.620
	With CI	0[b]			0				
Continue therapy									
	Intercept	.841	.125	44.968	1	.000			
	Without CI	.196	.182	1.160	1	.281	.852	1.216	1.738
	With CI	0[b]			0				

a. The reference category is: Repeat test

b. This parameter is set to zero because it is redundant.

VL= Viral Load CI= Confidence Intervals

predictor; or the odds ratio. Numbers greater than one are associated with increasing odds and numbers less than one are associated with decreasing odds of the outcome. In Table 11, the analysis shows that the presentation of CI as a variable has no significant strength as a predictor (p=.419 to Discontinue therapy vs. Repeat test, and p= .281 to Continue therapy vs. Repeat test).

Since CI made no significant impact on the decision outcomes, the variable Viral load level was added. Predictor variables in this multinomial regression were, Viral load level (VL), whether the respondents saw results with CI (WCI) and the interaction of the two variables, (VL* WCI). The final model contained only the VL levels as a significant predictor (model chi-square 792.6, $p<0.05$, df=8). Addition of the variable CI, or the interaction of the two variables, CI and VL provided no additional model strength. In fact, the addition of CI as a variable detracted from model strength (p=.481). Therefore, the presence or absence of CI with the result

did not impact the decision outcomes in this data set. Table 12 lists the impact of the variable VL and WCI on the decision outcomes.

Table 12

Multinomial Regression Statistics with Viral Load and the Presence or Absence of CI Variables. All Responses

Decision at all HCV VLs[a]		B	Std. Error	Wald	df	p	95% Confidence Interval for Exp(B) Odds Ratio		
							Lower Bound	Exp (B)	Upper Bound
Discontinue therapy									
	Intercept	21.313	.527	1635.543	1	0.000			
	100 IU/ml	-21.378	.588	1320.312	1	.000	1.639E-10	5.193E-10	1.645E-09
	560 IU/ml	-20.760	.569	1329.446	1	.000	3.158E-10	9.641E-10	2.943E-09
	1000 IU/ml	-21.273	.539	1555.018	1	0.000	2.005E-10	5.770E-10	1.661E-09
	1950 IU/ml	-19.430	.488	1582.955	1	0.000	1.399E-09	3.644E-09	9.489E-09
	31620 IU/ml	0[b]			0				
	Without CI	.141	.196	.513	1	.474	.784	1.151	1.690
	With CI	0[b]			0				
Continue therapy									
	Intercept	17.590	.329	2865.757	1	0.000			
	100 IU/ml	-15.716	.377	1737.393	1	0.000	7.140E-08	1.495E-07	3.130E-07
	560 IU/ml	-15.950	.375	1813.406	1	0.000	5.677E-08	1.183E-07	2.465E-07
	1000 IU/ml	-18.134	.357	2581.260	1	0.000	6.620E-09	1.332E-08	2.682E-08
	1950 IU/ml	-18.269	0.000		1		1.163E-08	1.163E-08	1.163E-08
	31620 IU/ml	0[b]			0				
	Without CI	.241	.199	1.458	1	.227	.861	1.272	1.881
	With CI	0[b]			0				

a. The reference category is: "Repeat Test"

b. This parameter is set to zero because it is redundant.

VL= Viral Load level. WCI= With Confidence Intervals

Each VL level was then examined separately using all responses. Chi-square was used to test the independence of the observed versus expected frequencies of the decisions made at each VL, using HCV VL results with or without CI. A value of $p < 0.05$ for the CHISQ.TEST indicates a very low probability that differences are due to random error, therefore the data sets would be considered different. As shown in Table 13, no statistically significant probabilities were obtained for the any of the three possible decisions made from results with CI or without CI at

Table 13

All Data: Any Order Seen. Chi-square Analysis at Each VL Level

	Discontinue Therapy			Continue Therapy			Repeat Test		
	Question with			Question with			Question with		
	No CI	With CI		No CI	With CI		No CI	With CI	
Viral Load IU/ml	# (% resp)	# (% resp)	p	# (% resp)	# (% resp)	p	# (% resp)	# (% resp)	p
31620	117 (97.5)	119 (97.5)	1.00	3 (2.5)	3 (2.5)	1.00	0	0	
1950	99 (82.5)	98 (81.0)	0.99	8 (6.7)	8 (6.6)	1.00	13 (10.8)	15 (12.4)	0.99
1000	49 (40.8)	48 (39.7)	1.00	29 (24.2)	28 (23.1)	1.00	42 (35.0)	45 (37.2)	0.99
560	25 (20.8)	27 (22.1)	1.00	85 (70.8)	77 (63.1)	0.65	10 (8.3)	18 (14.8)	0.49
100	12 (10.0)	14 (11.5)	0.99	95 (79.2)	95 (77.9)	1.00	13 (10.8)	13 (10.7)	1.00

NoCI= Without Confidence Intervals. With CI= With Confidence Intervals

any of the viral load levels, indicating no difference in response at any level based on presentation of HCV VL results with or without CI.

To further verify the paired analyses performed for each single decision, all three decisions at each VL were further analyzed, utilizing Pearson's chi-square statistic to test for the independence of two variables. The variables were: HCV VL results presented with CI, and HCV VL results presented without CI. With a low significance value (p <0.05) the hypothesis that the variables are independent would be rejected and the confidence that they are related is increased. Table 14 contains the output of each of the five analyses. Pearson's chi-square statistic was significant at all VLs, indicating a strong relationship between decisions made using results with CI and those without CI, with little to no impact of the presentation of CI. Cramer's V, a measure of association that can be used for nominal variables was greater than 0.8 for each level tested, supporting the chi-square findings.

Analysis of responses between two groups based on the order of the presentation of results with and without CI. In practice, HCV VL results are presented to a clinician only once, currently without the CI. Therefore, respondents were split into two groups based on whether

Table 14

Results of Pearson's Chi-square Test. All Three Decisions

Viral Load IU/ml	Contingency Table	Pearson's X^2	df	p	Cramers V
31,620	NoCIxWCI*	118	1	0.00	1
1950	NoCIxWCI	174.8	4	0.00	0.86
1000	NoCIxWCI	162.9	4	0.00	0.83
560	NoCIxWCI	158.5	4	0.00	0.82
100	NoCIxWCI	199.9	4	0.00	0.921
*NoCI= results without CI. WCI= results with CI					

they saw results with CI first or without CI first. For each respondent, only the decisions recorded from the first question were analyzed (one answer-set of decisions from each respondent). Sixty-five respondents (53.4%) saw the results presented with no CI first, and 59 (47.6%) saw results with CI first. The data presented earlier in Table 10 demonstrate that the two groups were demographically similar.

To determine if there was a significant impact of confidence intervals or viral load level on the decisions based upon how the respondents saw the results first, multinomial logistic regression was used with the three decisions as outcomes for the dependent variable. As with the analysis using all responses, predictor variables were: VL (Viral load level), whether the respondents saw results with CI (Saw CI) and the interaction of the two variables, (VL* Saw CI). Again, the final model contained the two predictor variables independently; addition of the interaction provided no additional model strength (model chi-square 497.4, $p < .05$, df=8). Viral load level was the only significant predictor variable. The significance of the Wald statistic for the variable CI was p= .539 for Discontinue therapy and p= .128 for Continue therapy. The addition of CI detracted from overall model strength in this analysis (p= .051) but to a lesser

extent than in the analysis with all responses. Table 15 displays the regression statistics with the variables viral load level and the presentation of CI.

Table 15

Multinomial Regression Statistics: Two Populations Based on Responses to the First Presentation of HCV VL Only

Decision at all HCV VL's[a]		B	Std. Error	Wald	df	p	95% Confidence Interval for Exp(B) Odds Ratio		
							Lower Bound	Exp (B)	Upper Bound
Discontinue therapy									
	Intercept	21.317	.747	814.283	1	.000			
	100 IU/ml	-21.506	.837	659.641	1	.000	8.858E-11	4.572E-10	2.360E-09
	560 IU/ml	-20.769	.811	656.122	1	.000	1.949E-10	9.548E-10	4.679E-09
	1000 IU/ml	-21.393	.763	785.394	1	.000	1.146E-10	5.118E-10	2.285E-09
	1950 IU/ml	-19.432	.691	789.913	1	.000	9.381E-10	3.637E-09	1.410E-08
	31620 IU/ml	0[b]			0				
	Without CI	.173	.281	.378	1	.539	.685	1.189	2.063
	With CI	0[b]			0				
Continue therapy									
	Intercept	17.898	.463	1495.535	1	0.000			
	100 IU/ml	-15.626	.535	853.550	1	.000	5.735E-08	1.636E-07	4.667E-07
	560 IU/ml	-15.773	.536	866.571	1	.000	4.940E-08	1.412E-07	4.035E-07
	1000 IU/ml	-18.113	.502	1299.566	1	.000	5.083E-09	1.361E-08	3.643E-08
	1950 IU/ml	-18.261	0.000		1		1.173E-08	1.173E-08	1.173E-08
	31620 IU/ml	0[b]			0				
	Without CI	-.432	.284	2.315	1	.128	.372	.649	1.132
	With CI	0[b]			0				

a. The reference category is: Repeat Test

b. This parameter is set to zero because it is redundant.

RL= Viral load level. CI = Confidence Intervals

As with the analysis of all responses, chi-square tests were used for this grouping to test for a probable correlation of decision with the presentation of CI at each VL (Table 16). Though no statistically significant correlations were found, some interesting observations were made. When presented with HCV VL with CI just above the decision threshold [1950 (990-3800 IU/ml)], 7% more respondents chose the decision to continue therapy, and approximately 11% fewer chose to discontinue therapy than those presented with results without CI. A small percent more (2%)

Table 16

Respondents' Decisions Split by Presentation of Results for First Question

	Discontinue Therapy "Survey with"			Continue Therapy "Survey with"			Repeat Test "Survey with"		
	No CI	With CI		No CI	With CI		No CI	With CI	
Viral Load IU/ml	# (% resp)	# (% resp)	p	# (% resp)	# (% resp)	p	# (% resp)	# (% resp)	p
31620	64 (98.5)	57 (95.0)	0.93	1 (1.5)	2 (5.0)	0.93	0	0	
1950	56 (86.2)	45 (77.6)	0.67	2 (3.1)	6 (10.3)	0.45	7 (10.8)	7 (12.1)	1.00
1000	28 (43.1)	19 (32.8)	0.71	16 (24.6)	14 (24.1)	1.00	21 (32.3)	25 (43.1)	0.68
560	15 (23.1)	10 (16.9)	0.87	41 (63.1)	45 (76.3)	0.47	9 (13.8)	4 (6.8)	0.65
100	7 (10.8)	5 (8.5)	0.98	48 (73.8)	51 (86.4)	0.38	10 (15.4)	3 (5.1)	0.32

NoCI= Without Confidence Intervals. With CI= With Confidence Intervals

chose the Repeat test option. When presented with results just below the decision point and with CI [560 (290-1,110) IU/ml], 13% more respondents chose the decision to continue therapy, 5% less chose to discontinue therapy, and 7% fewer chose to repeat the test. With the exception of the repeat testing at this level, these decision differences were more pronounced with the responses split into two groups based on the first presentation seen, than those observed when all responses were analyzed as one group. At the decision threshold of 1000 IU/ml, 11% fewer who saw the results with CI (510- 1,970 IU/ml) chose to discontinue therapy than those who saw results without CI, and 7% more repeated the test. The percentage of respondents who chose to continue therapy was the same between these two groups at this level. Since the confidence intervals presented with the HCV VL results immediately above and below the threshold showed the VL crossing the decision threshold, these results reveal some impact on the decision outcomes. However the impact was not statistically significant by chi-square tests.

To determine if there were differences in response to CI between VLs, the number and percent responses for each decision were analyzed by two-tailed T tests. No statistical significance was found for any of the response differences (p= .66 - .91).

Analysis based on order of the presentation of results, with or without CI. The order of presentation of the HCV VL results, with or without CI, was randomized among the survey participants. Therefore, responses could be analyzed based on the order in which they were seen. As with prior analyses, chi-square tests were used separately for each VL level. There was only one decision outcome that was initially observed as statistically significant, the decision to continue therapy based on HCV VL results of 560 IU/ml with CI [560 (290-1,110) IU/ml]. For this case, the order in which the HCV VL results were seen, with or without CI, had an impact on the decision made. If the respondents saw the HCV VL without CI first, fewer decided to continue therapy than those who saw the result with CI first ($p = 0.04$). More decided to discontinue therapy than those who saw the CI first, and more decided to repeat the test than those who saw the CI first. The latter two comparisons were not statistically significant by the original criterion of $p<0.05$. However, even the one statistically significant effect by the original criterion may be a result of an inflated chance of a type I error due to the multiple tests performed (Field, 2013). Applying the Bonferroni correction, using 28 as the number of tests, a more rigorous threshold for significance would be 0.002 ($p_{crit} = \alpha/\kappa$). In this case, no finding is statistically significant. Table 17 shows all the pair wise chi-square tests for these analyses and the associated *p*-values with each pair. The first column labeled "Viral Load" is the item seen by the participant. The decision choices are in the succeeding columns, but split by the type of presentation the respondents saw first. For example, for the first viral load presented (31,620 IU/ml without CI), 117 respondents decided to discontinue therapy. Sixty-four of them answered this seeing this as the first presentation (NoCI), while 53 of the respondents had seen the results with CI before viewing this result.

Table 17

Chi-square Test Result Based on Order Seen

	Discontinue Therapy			Continue Therapy			Repeat Test		
	Seen First			Seen First			Seen First		
	No CI	With CI		No CI	With CI		No CI	With CI	
Viral Load	# (%resp)	# (%resp)	p	# (%resp)	# (%resp)	p	# (%resp)	# (%resp)	p
31620	64 (98.5)	53 (96.4)	0.91	1 (1.5)	2 (3.6)	0.91	0	0	
1950	56 (86.2)	43 (78.2)	0.73	2 (3.1)	6 (10.9)	0.40	7 (10.8)	6 (10.9)	1.00
1000	28 (43.1)	21 (38.2)	0.96	16 (24.6)	13 (23.6)	1.00	21 (32.3)	21 (38.2)	0.93
560	15 (23.1)	10 (18.2)	0.93	41 (63.1)	44 (80.0)	0.25	9 (13.8)	1 (1.8)	0.13
100	7 (10.8)	5 (9.1)	0.99	48 (73.8)	47 (85.5)	0.49	10 (15.4)	3 (5.5)	0.39
31620 (16,060 - 62,230)	62 (98.4)	57 (95.0)	0.77	1 (1.6)	3 (5.0)	0.77	0	0	
1950 (990- 3,800)	53 (84.1)	45 (77.6)	0.84	2 (3.2)	6 (10.3)	0.47	8 (12.7)	7 (12.1)	1.00
1000 (510 – 1,970)	29 (46.0)	19 (32.8)	0.53	14 (22.2)	14 (24.1)	1.00	20 (31.7)	25 (43.1)	0.64
560 (290 - 1,110)	17 (27.0)	10 (16.9)	0.62	32 (50.8)	45 (76.3)	0.04	14 (22.2)	4 (6.8)	0.12
100 (50 – 200)	9 (14.3)	5 (8.5)	0.80	44 (69.8)	51 (86.4)	0.18	10 (15.9)	3 (5.1)	0.29

NoCI= Without Confidence Intervals. With CI= With Confidence Intervals

Further analysis of the decision outcomes for a VL of 560 IU/ml with CI. Though there were no statistically significant findings at the VL of 560 IU/ ml with CI, this result presented some of the more interesting observations and was used for further analysis. In the survey, respondents were asked if the results with CI provided relevant information, if it detracted from or confused the result, or if neither of these applied. This variable (Relevant) was used in a multinomial logistic regression with all three decisions as outcomes at the VL level of 560 IU/ml presented with CI. The other predictor variable was which test results were seen first, those without CI or those with CI. The only outcome having a significant predictor was the decision to continue therapy. The significant predictor for Continue therapy was the order of the presentation of results: viewing the HCV VL results with no CI first. The regression model and individual chi-

square tests had the same outcome. However from the model it could be determined that if clinicians saw a VL of 560 IU/ml without CI first, they were 7.5 times less likely (odds ratio 0.134) to decide to continue their patients on therapy than to repeat the test. Table 18 displays the statistical data.

Table 18

Multinomial Regression of Decision Choices at 560 IU/ml

560 IU/ml with CI[a]		B	Std. Error	Wald	df	p	95% Confidence Interval for Exp(B)		
							Lower Bound	Exp(B)	Upper Bound
Discontinue Therapy									
	Intercept	1.441	.710	4.120	1	.042			
	No CI seen first	-.982	.757	1.682	1	.195	.085	.375	1.652
	With CI seen first	0[b]			0				
	CI added to information	-.696	.665	1.097	1	.295	.136	.499	1.834
	CI detracted from information	.112	1.305	.007	1	.932	.087	1.118	14.437
	CI neither added or detracted	0[b]			0				
Continue Therapy									
	Intercept	2.418	.658	13.523	1	.000			
	No CI seen first	-2.009	.692	8.436	1	.004	.035	.134	.520
	With CI seen first	0[b]			0				
	CI added to information	.169	.600	.079	1	.778	.366	1.184	3.833
	CI detracted from information	2.017	1.132	3.178	1	.075	.818	7.519	69.104
	CI neither added or detracted	0[b]			0				

a. The reference category is: Repeat Test

b. This parameter is set to zero because it is redundant.

Narrowing the analysis further just to the decision to continue therapy with the presentation of results at 560 IU/ml, binomial logistic regression analyses could be used and other factors could be examined. Predictor variables were: which test results were seen first, with or without CI (Seen first), and whether the respondent thought that the results with CI provided relevant information (Relevant). The predictor variables were added in hierarchical progression: Seen first, Relevant and the interaction Seen first *Relevant. The chi-square of each new model over

the previous was significant at $p<0.05$. The final model contained both predictors and the interaction of the two predictor variables. However, the model only correctly predicted 67.8% of the outcomes. Of the predictor variables, the two parameters most informative (Wald statistic $p<0.5$) were: viewing HCV VL *without* CI first (from Seen first) and, if the CI detracted from or confused the result (from the variable Relevant). At HCV VL of 560 IU/ml, clinicians who saw results without CI first were nearly four times less likely to choose to continue therapy (odds ratio 0.256) than those who saw the results with CI first. However, respondents who answered that the CIs detracted from the result were seven times more likely to choose the "continue with therapy" option at this level (odds ratio 7.12), than those who responded that the CI presentation did not add information or it detracted/ confused the result (the reference). Table 19 lists the variables with their corresponding statistics.

Table 19

Logistic Regression Statistics for Continue Therapy Decision at HCV VL 560IU/ml with CI

Continue Therapy at 560 IU/ml with CI		B	S.E.	Wald	df	p	95% C.I.for EXP(B) Odds ratio		
							Lower	Exp(B)	Upper
	No CI seen first	-1.363	.429	10.073	1	.002	.110	.256	.594
	Relevant			7.864	2	.020			
	CI added to information	.598	.442	1.834	1	.176	.765	1.819	4.322
	CI detracted from information	1.963	.715	7.538	1	.006	1.754	7.121	28.915
	Constant	.794	.364	4.760	1	.029		2.213	

Prior analyses suggested that there were no interdependencies between the variables in the survey; however, to specifically illustrate this, a correlation matrix utilizing Pearson's chi-square was developed. Table 20 shows that the only statistically significant correlation is between the decision to continue therapy when the HCV VL = 560 IU/ml with CI and whether or not the CI was seen first. This is consistent with the other findings.

Table 20

Correlation Matrix for the Decision to Continue Therapy at 560 IU/ml

Correlations		560 IU/ml With CI Continue	Seen first: With or Without CI	CI provided relevant information	Want to have CI with results
560 IU/ml With CI Continue Therapy	Pearson Correlation	1	.264**	-.116	.031
	Sig. (2-tailed)		.003	.211	.746
	N	122	122	118	114
Seen first: With or Without CI	Pearson Correlation	.264**	1	.042	-.034
	Sig. (2-tailed)	.003		.648	.718
	N	122	124	118	114
CI provided relevant information	Pearson Correlation	-.116	.042	1	-.143
	Sig. (2-tailed)	.211	.648		.130
	N	118	118	118	114
Want to have CI with results	Pearson Correlation	.031	-.034	-.143	1
	Sig. (2-tailed)	.746	.718	.130	
	N	114	114	114	114

**. Correlation is significant at the 0.01 level (2-tailed).

Impact of presentation order, with and without CI, on changes in decisions from the first to the second set of results. Data from the 118 respondents who answered both questions were included in the data set. Twenty-four (20.3%) gave different decisions for the second question than they did for the first question. Of those, 18 (75%) saw the test results without CIs first, while 6 (25%) saw the results with CIs first. By the chi-square test, this difference was not significant: $p = 0.13$).

Binomial logistic regression was performed on the dependent variable Changed decision, using each decision at each level (31 changed decisions out of 588 individual decisions at all levels). Predictor variables were VL and Seen first (whether the respondent saw the results without CI first or saw the results with CI first). The predictor variables VL and Seen First were added in hierarchical progression: VL, Seen first and the interaction of VL*Seen first. In a final

model, excluding the interaction of the two variables, Seen first was significant (Wald = 3.176, p=0.001, df=1). For those who viewed results without CIs first, the odds of changing the response on the next result with CI were approximately five times more likely than the converse (95% CI of odds ratio: 1.11-13.39). The addition of the interaction of VL*Seen first was not informative and masked the significance of the variable Seen First.

H1 and H2 null hypothesis findings. The H1 null hypothesis, that the presentation of CIs will have no impact on the clinicians' therapy management decisions, cannot be rejected. The presentation of HCV VL results with 95% confidence intervals had no statistically significant impact on the treatment decision for HCV genotype 1 patients. There were no statistically significant findings at a significance threshold that corrected for the multiple analyses performed. The one finding with a $p<.05$ did continue to be seen when analyzed further, however, it describes a scenario that does not reflect clinical practice. At 560 IU/ml, when respondents first saw results without CI and then were presented with results displaying the CI, they were less likely to choose to continue therapy than their counterparts who saw the results with CI first. Conversely, clinicians who stated the presentation of CIs detracted from or confused the results were more likely to continue therapy. In addition, clinicians seeing results without CI first and then with CI were more likely to change their decision on the second question. Though these are interesting findings that could provide some insight to further research, they do not support the rejection of the null hypothesis.

The H2 null hypothesis is also not rejected. $H_0$2, that there is no impact of CIs on the decisions at different VLs, depended on at least one finding in $H_0$1 at one of the VL levels. Viral load levels themselves, had the greatest impact on decision outcomes irrespective of whether the results were presented with or without CIs.

H3 null hypothesis. Clinicians' experience and demographics will have no impact on treatment decisions using HCV VL results presented with and without confidence intervals.

Multinomial logistic regression was used to evaluate possible correlations between physician demographics, clinical experience, practice preferences and the three outcome decisions. Predictors variables in the analysis were the five viral load levels (VL), presentation of HCV VL results with or without CI (With CI / NoCI), where the respondents practiced (Where practice), the respondents' professions, field of medicine, whether they manage hepatitis C patients, use AASLD guidelines, the number of years in practice (Years practice), and the number of hepatitis C patients seen per month (Patients month). Besides VL, only where the clinicians practiced, the number of years in practice and the number of patients seen per month demonstrated any significance to the outcome (Wald statistic at $p < .05$). The first two were significant in the decision to discontinue therapy and the third was significant in the decision to continue therapy.

To examine the role of clinicians' practice preferences, the variables Relevant (whether the respondents thought the presentation of CIs added information to the result) and whether or not the respondent wanted to see CI with their results (Want CI), were added to the model with the variables HCV VL levels, With CI, Where practice, Years practice and Patients month. With addition of these variables, where clinicians practiced was no longer statistically significant in the overall model. The Wald statistic for practicing in Europe went from p=0.031 to p=0.325 and for practicing in Canada, it went from p=0.043 to p=0.43. The number of respondents in these categories was small compared to those that practice in the United States (less than 10% combined).

Across all HCV VL levels, the number of years in practice had a negative influence on the decision to discontinue therapy (Exp B/ odds ratio 0.974). However, the impact was slight. For

each increasing year in practice the choice of discontinuing therapy at all VLs is 1.02 times less likely than repeating the test. Responses indicating a desire to have CIs reported with results and those indicating the converse both had a negative association with this decision. The odds ratios for both these responses were less than one and the frequency of these two choices was very close (42.0% wanted CI and 47.3% did not). The other choices for this variable were "Don't know" and "I already receive CI", and the responses for these two variables together were approximately 10%. Because there was no distinguishable difference in the responses of those that did and did not want CI, this variable was not considered useful in predicting the decision to discontinue therapy.

For the decision to continue therapy, the number of patients seen per month had a small negative impact. However clinicians who thought the presentation of CI's detracted from or confused the result, were 3.2 times more likely to choose this response at all VLs (odds ratio 3.197) than to repeat the test. This was a statistically significant finding and is similar to the findings at 560 IU/ml in the presentation of results with CI.

The decisions at each VL were further analyzed by multinomial logistic regression. Only the number of years in practice and the number of hepatitis C patients seen per month were significantly associated with decisions at specific HCV VL levels (Wald statistics $p<0.05$), and those were only at HCV VLs of 1950 IU/ml and 100 IU/ml respectively. The effects were small. For example, at 1950 IU/ml, with each increasing year of practice, the respondents were 1.07 times less likely to choose the decision to discontinue therapy versus repeating the test. The significant effect of respondents who stated that the results with CI detracted/ confused the result, which was seen in the model with all decisions from all VLs, was not significant at any

single specific VL with or without CI. All variables with statistical significance are listed in Table 21.

The H3 null hypothesis, that experience and demographics will have no impact on the therapy management decisions, is rejected. Some respondent demographics and practice preferences did impact the HCV therapy decision outcomes. Over all decision outcomes at any VL, the significant demographics and characteristics were: the years in practice, the number of patients seen per month, and whether or not the presentation of results detracted or confused the results. At 1950 IU/ml, the number of years in practice was significant in the Discontinue therapy decision; and at 100 IU/ml, the number of patients seen per month was significant to the Continue therapy decision. The presentation of results with or without CI made no impact.

Other analyses for clinical utility. The amount of repeat testing was notable at some VL levels. Repeat testing delays the implementation of results and can add costs to a patient's care. Analysis of the data to determine if any of the variables contributed to the likelihood of repeating the test was performed.

Decision to repeat the test. Respondents chose to repeat the test at all VLs except the highest one: 31,620 IU/ml. The choice Repeat test was 17.5% of all decisions made at all levels. For those respondents answering both HCV VL questions, the decision to repeat did not decrease with the second viewing of the HCV VLs: 8.2% of all first-question response decisions and 8.5% of all second-question response decisions were Repeat test decisions. The decision to repeat tests at the therapy decision threshold and the two immediately flanking levels (560, 1000, and 1950 IU/ml) was 19.8% of all decisions made. At the decision threshold (1000 IU/ml) Repeat test was 36.1% of all decisions made.

.

Table 21

Respondent Demographic, Practice and Preference Variables Demonstrating Significance in Respondent Decision Choices

					95% Confidence Interval for Exp(B)		
Dependant variable		Variable parameter	Variable Type	p	Lower bound	Exp (B)	Upper bound
Decisions at all VL	Decicion to Discontinue Therapy	*Where practice: Europe	Categorical	0.031 0.325	0.025 .075	0.144 .421	0.838 2.363
		*Where practice :Canada	Categorical	0.043 0.433	0.013 0.081	0.112 0.488	0.937 2.932
		Number of years in practice	Continuous	0.025	0.952	0.974	0.997
		Want CI with HCV VL results	Categorical	0.011	0.026	0.127	0.625
		Do NOT want CI with HCV VL results	Categorical	0.001	0.014	0.067	0.332
	Decision to Continue Therapy	Number of patients per month		0.014	0.991	0.995	0.999
		CI's detracted from or confused result	Categorical	0.001	1.603	3.197	6.375
Decisions at VL 1950 IU/ml	Decicion to Discontinue Therapy	Number of years in practice	Continuous	0.026	0.868	0.928	0.991
Decisions at VL 100 IU/ml	Decision to Continue Therapy	Number of patients per month	Continuous	0.01	0.974	0.985	0.996

*Statistical significance of parameters in Where_practice were only observed in model without Relevant and Want_CI: top p value for each.

Table 22 lists the percent of decisions at all VLs presented with and without CI. In the previous chi-square analyses of the decisions made at each VL the decision to repeat the test was not significantly influenced by the presentation of CI. In the multinomial regression analyses, the repeat decision outcome was the reference and only the variables examined in H3 null hypothesis analyses showed any significance. Using the Repeat test as the only dependant variable with a yes/no outcome, binomial logistic regression could be utilized for a specific analysis of this decision.

Binomial logistic regression analyses were performed on the decision of whether or not to repeat the test to further examine if there was any effect from viral load levels or the presentation of results with CI. Data from all responses were used first. Predictor variables were: VL and the presentation of CI. The model chi-square was significant ($p<0.05$) and 86.4% of cases could be correctly classified. The only significant variable however was the HCV VL level (Wald=33.12, $p<0.5$, $df=4$). Whether or not the respondents thought the presentation of CI provided relevant information (predictor variable-Relevant) was added to the model. There was no significant correlation (Wald = 3.18, $p= 0.204$, $df=2$).

Similarly, logistic regression analysis was performed using only the responses from the first question. VL and whether the respondent viewed results with CI (Saw CI) were predictor variables. The presence or absence of CI was not significant in the decision to repeat tests either alone ($p=0.663$) or interacting with VL ($p=0.196$). Viral Load level is the strongest predictor of the decision to repeat the test. This is consistent with the visual examination of the data.

Summary of survey analyses. From the survey, 138 responses were received. Strong clinical guideline adherence was demonstrated in self-reported behaviors and in decision choices recorded. However, up to 37% of decisions were to repeat the test. The impact of knowing the

Table 22

Decisions Made at Each VL Level

HCV VL Level IU/ml		Discontinue Therapy (%)	Decisions Continue Therapy (%)	Repeat test (%)
31620	Combined	97.5	2.5	0.0
	NoCI	97.5	2.5	0.0
	With CI	97.5	2.5	0.0
1950	Combined	81.7	6.6	11.6
	NoCI	82.5	6.8	10.8
	With CI	81.0	6.6	12.4
1000	Combined	40.2	23.7	36.1
	NoCI	40.8	24.2	35.0
	With CI	39.7	23.1	37.2
500	Combined	21.5	67.0	11.6
	NoCI	20.8	70.8	8.3
	With CI	22.1	63.1	14.8
100	Combined	10.7	78.5	10.7
	NoCI	10.0	79.2	10.8
	With CI	11.5	77.9	10.7

Combined: All responses from Q1 and Q2 combined

NoCI: Responses from all results reported without confidence intervals

With CI: Responses from all results reported with confidence intervals

assay imprecision was not statistically significant to clinicians' decision regarding the therapy management of the simulated patient. There was no significant impact of clinician demographics, experience or practice preferences for the decisions made using the results with or without CI, however those who thought the CIs detracted from or confused the result were more likely to decide to continue therapy than repeat the test. Those who thought the CIs confused or

detracted from the result were only 16.9% of responses however. Some associations were observed between demographics, practice preferences, and decisions regardless of the presentation of CI at each viral load level. The most impactful variable on the decisions recorded for repeating the test was the level of viral load being reviewed.

Determining Value and the Impact of Clinicians' Use of Test Information on that Value

A Markov decision-tree model was developed as described in the Methods section. Changes from the pilot model were: removal of death as a health state, removal of SVR as a separate health state from the Response health state, and reduction of the time horizon from 72 to 60 weeks (more appropriately reflecting the time span over which the test has influence; that needed to evaluate SVR). In addition, the health state "Decompensated cirrhosis" was added and Quality of Life (QoL) values were accumulated at each state. Transition probabilities, costs and QoL parameters for health states were populated from literature and older costs were inflation adjusted where appropriate (Table 7). Decisions made by clinicians' test use were represented by the survey data for initial model inputs. The circled boxes in Figure 8 illustrate where in the decision tree initial inputs were entered.

Analyses to evaluate the economic impact of the clinicians' decisions. The impact of decisions made from HCV VL results with and without CI on the value of the test was evaluated using the model described in the methods and the preceding paragraphs.

Because the clinically prescribed decisions are not identical for all VL levels, data for each VL in the survey was first evaluated separately, and different inputs and patient flow strategies were used for each level accordingly. At each VL level, the analysis was run with all patients on therapy to determine treatment costs without the therapy management information provided by HCV VL, then with decision inputs according to established guidelines (AASLD, 2011; Vertex

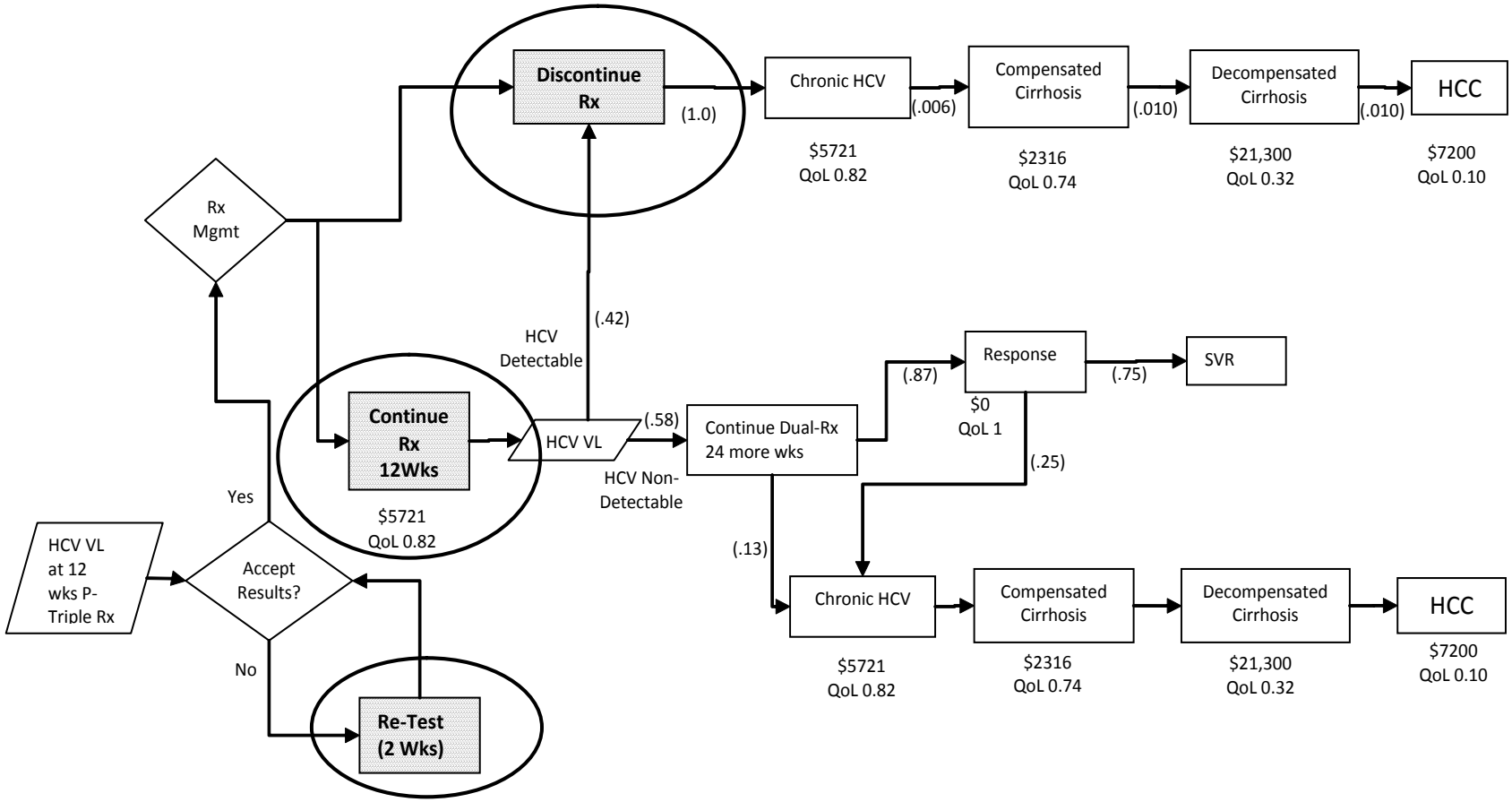

Figure 8. Decision path showing health states, costs, transition probabilities and QoL. The circled boxes show the entry point for the three decisions: Continue therapy, Discontinue therapy or Repeat the test.

Pharmaceuticals, 2011), and finally, with decision inputs from the survey described (made with and without CI). For decision inputs according to AASLD Guidelines, no entries were made to the Repeat test input, and the Continue and Discontinue therapy inputs were made using the percent distribution of the population VLs at that level (based on the imprecision used in the survey of 1SD =0.15 $\log_{10}$). For example, at 560 IU/ml, based on the imprecision of the test, approximately 4.8% of results would be expected to fall at or above 1000 IU/ml. Therefore for the guideline- dictated decision, 4.8% were placed in the Discontinue therapy arm. For 31,620 IU/ml, all were placed in the Discontinue therapy arm. For decision inputs according to the proportions from the survey, only responses from the first presentation of HCV results to the respondents were used. The data were used in this manner (as opposed to combining all responses) because the presentation of HCV VL results twice (with and without CI) does not reflect common clinical practice. When using Repeat test decisions, those proportions were first combined with the Continue therapy decisions for entry into the model. After an initial step of "two weeks" the proportion of Repeat test responses was re-distributed among the Discontinue therapy and Continue therapy arms according to the original proportions of those two decisions, adjusted for no repeat testing. For example, if the original decision inputs were 40% Discontinue therapy, 40% Continue therapy and 20% Repeat test, the 20% Repeat test would be combined with the 40% Continue therapy decisions for initial input. After the first two week cycle, the 20% would be redistributed with half (10%) going to the Discontinue therapy arm and the other half (10%) remaining in the Continue therapy arm.

The resulting total costs and the average QoL ratio for each of the model inputs were recorded. The total costs of the survey-guided decisions with and without CI were compared against the total costs of the decisions according to guidelines at each viral load level. Since the

survey respondents were primarily specialists, their recorded decisions and the guideline-dictated decisions were close for most of the viral load levels; with the exception of Repeat test decisions. This was reflected in the comparisons of the total costs between the guideline decisions, those made viewing HCV VLs with CI, and those made viewing HCV VLs without CI. For most comparisons, there was less than a 10% difference in total costs. As expected, the greatest differences were seen at or near the decision point. For those VLs above the 1000 IU/ml threshold, deviations from guidelines were associated with higher costs as a result of keeping patients on therapy. Notably, at 1950 IU/ml the decisions made using results with CI were associated with 16% greater costs than the guideline-dictated decisions and 12% higher costs than those decisions made using results without CI. The impact to the QoL values between the three types of decision inputs was negligible.

Table 23 displays the decision model inputs and outputs for each VL using the 12-week decision point inputs defined by guidelines and from the survey data. From the survey data the input decisions are further split to those made with CI and those made without CI presentation. The first column in the table lists the five viral load levels in IU/ml. For each of the decision inputs (guideline-directed or survey-directed), the actual percentages of each decision are stated with the output total costs from the model and the output QoL value. Using the guideline directed decision percentages as a benchmark, the total cost outputs from decisions made by respondents in the survey using results without and with precision information are compared (labeled in the table as "% Change from Guide"). Additionally, total costs outputs from the survey-directed decision inputs with CI are compared to those made without CI (labeled as "%Change from NoCI").

Table 23

Total Costs / QoL Comparisons from Guideline-Directed Decisions and Survey Decisions Made with and without CI

	Guideline			Survey No CI				Survey with CI				
HCV VL (IU/ml)	Model Inputs %R/%DC RX/%ContRX	Total costs $ (60 Wks)	Av QOL	Model Inputs %R/%DC RX/%ContRX	Total costs $ (60 Wks)	Av QOL	%Change from Guide	Model Inputs %R/%DC RX/%ContRX	Total costs $ (60 Wks)	Av QOL	%Change from Guide	%Change from NoCI
31620	0/100/0	7813	0.82	0/98.5/1.5	8000	0.82	2	0/95/5	8436	0.82	8	5
1950	0/97.3/2.7	8147	0.82	10.8/86.2/3.1	8377	0.82	3	12.1/77.6/10.3	9421	0.82	16	12
1000	0/50/50	13424	0.86	32.3/43.1/24.6	12461	0.87	-7	43.1/32.8/24.1	13590	0.87	1	9
560	0/4.8/95.2	17414	0.97	13.8/23.1/63.1	15171	0.94	-13	6.8/16.9/76.3	16051	0.95	-8	6
100	0/0/100	17778	0.99	15.4/10.8/73.8	16494	0.97	-7	5.1/8.5/86.4	16877	0.97	-5	2

VL= Viral load levels. RX= therapy. NoCI= Without Confidence Intervals. With CI= With Confidence Intervals.

Using the data from Table 23, the total costs at each VL resulting from guideline-directed decisions versus decisions made by survey respondents using results with and without CI presentation are shown in Figure 9.

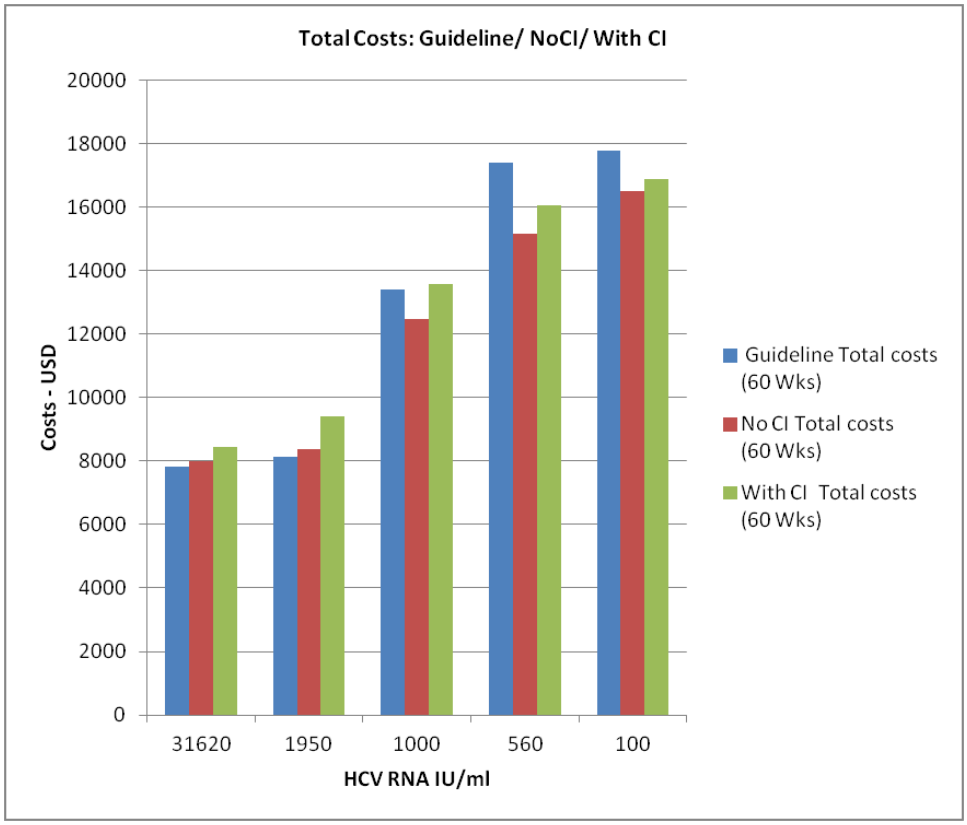

Figure 9. Total costs resulting from guideline-dictated decisions and survey-guided decisions.

Figure 10 displays a graphical representation of the percent change in total costs of the survey-directed decisions made with and without CI from the guideline directed decisions.

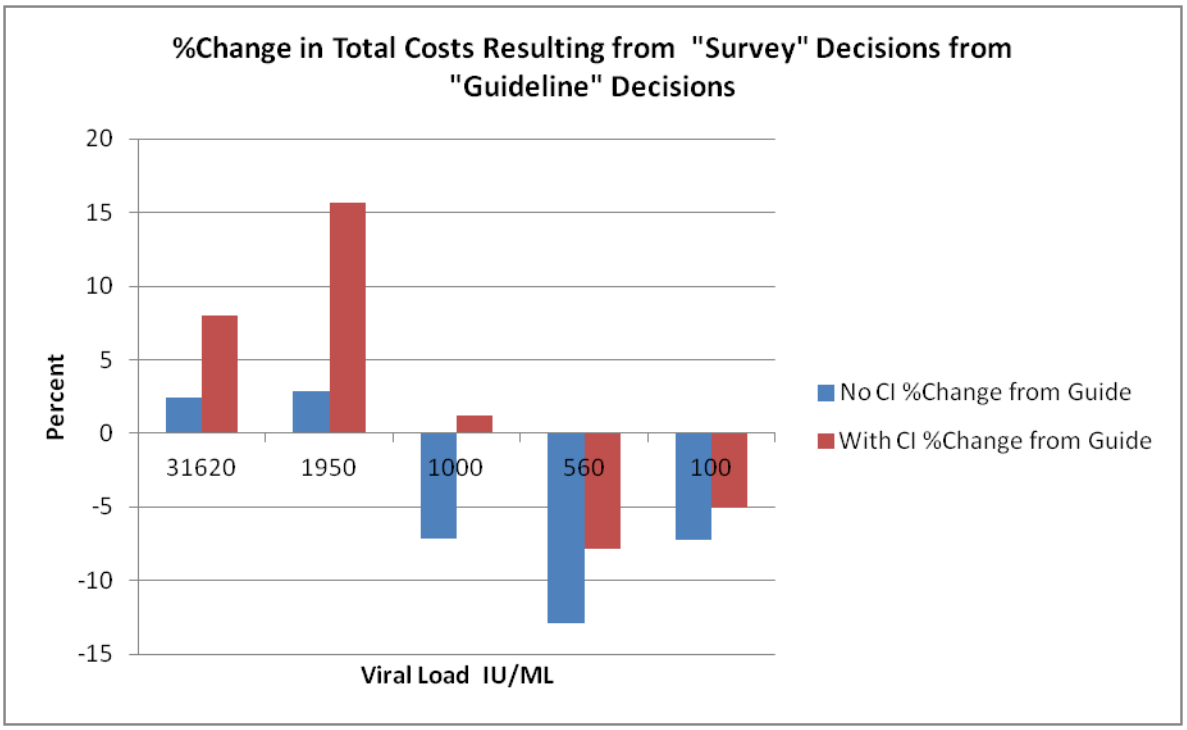

Figure 10. Percent change in total costs from survey decisions made with and with CI compared to guideline-dictated decisions.

The value of HCV VL test information. To determine the value in cost differences per patient that the HCV test information brings to the 12-week therapy decision, analyses were performed with 100% of the decisions in the Continue therapy arm for each viral load level. This provided a base case at each VL level reflecting all patients on therapy regimens without a test to help manage the decision to remove non-responders. The all-on-therapy analyses were followed by using inputs to the Discontinue and Continue therapy arms for each VL according to the AASLD guidelines (guideline-directed). Finally, inputs according to the decisions made in the survey were used. Since the value of the test at this specific point in the therapy management course is demonstrated by removing non-responders from inappropriate therapy, it was measured in the cost differences by subtracting the costs of the decisions made with the test

information from the costs of continuing all patients on therapy. Table 24 shows the costs of the base case (all patients continuing therapy) compared to costs using the test according to guidelines, and according to the combined responses on the survey as well as with those responses using HCV VL with CI and those using HCV VL without CI. As expected, the highest cost differences are among those that have high viral load levels. For those at the 100 IU/ml level, any differences seen as "cost savings" that are greater than the decision-directed savings are inappropriate, as removing a potential responder from therapy is incorrect. (The calculation of the guideline-directed value at 100 IU/ml is appropriately $95 due to the removal

Table 24

Value of Test Information: Average Differences in Total Costs Realized as Savings per Patient

		Guideline Directed		Survey - All Responses		Survey - NoCI only		Survey - With CI Only	
VL Result: HCV RNA IU/mL	Total Costs $: All Continue Therapy	Total Costs $	Value of test information Costs savings $[1]	Total Costs $	Value of test information Costs savings $[1]	Total Costs $	Value of test information Costs savings $[1]	Total Costs $	Value of test information Costs savings $[1]
31620	24976	7813	17163	8125	16851	8000	16976	8436	16540
1950	24900	8147	16753	8888	16012	8377	16523	9421	15479
1000	21424	13423	8001	12713	8711	12461	8963	13590	7834
560	18726	17414	1312	15431	3295	15171	3555	16051	2675
100	17873	17778	95	16569	1304	16494	1379	16877	996

[1]Based on the difference in costs of leaving all patients all on therapy and removing non-responders per guideline-dictated and survey response decisions.

of the proportion of patients that show the presence of HCV RNA at 12 weeks post dual-combination therapy. This proportion, derived from literature, is coded into the model. At 560 IU/ml, due to the distribution of the population of VLs with the imprecision stated, approximately 4.8% cross the 1000 IU/ml threshold. Figure 11 shows the cost differences, realized as savings at each VL level from Table 24.

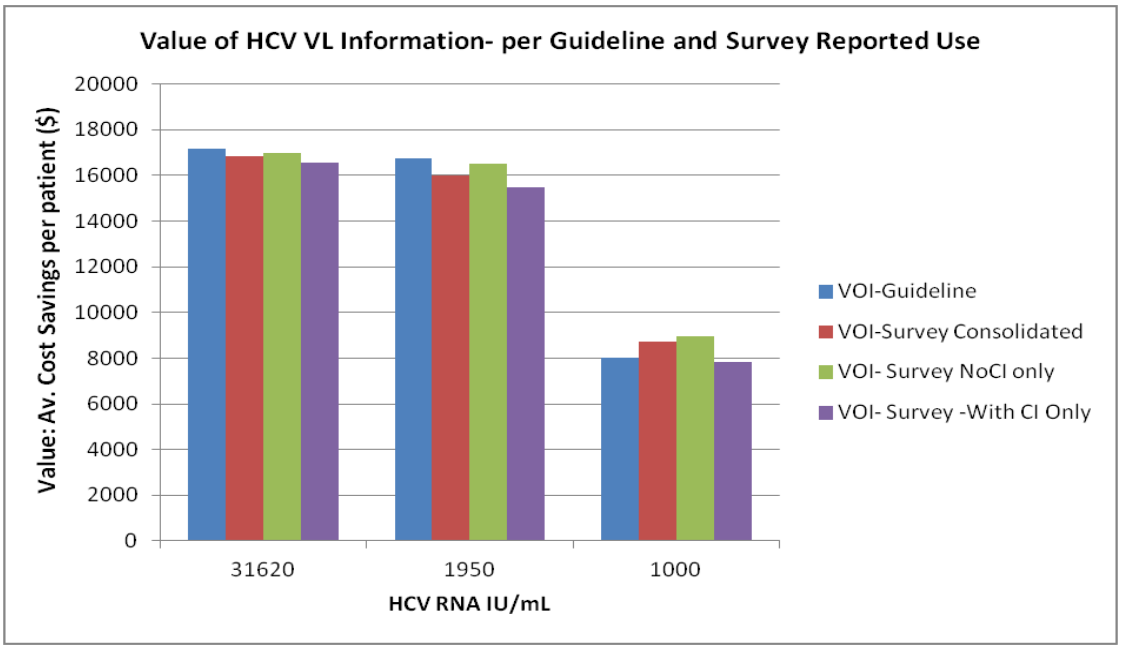

Figure 11. Cost differences realized as cost savings per patient at each VL due to the contribution of the HCV VL test in removing non-responders.

From this evaluation, it was demonstrated that there were some differences in total cost savings per patient based on clinicians use. When compared to decisions made on the basis of HCV VL according to the AASLD guidelines, the different decisions by clinicians produced up to 16% increase in average total costs per patient in VLs at 1000 IU/ml and above. Below 1000 IU/ml, cost differences seen as "savings" greater than those realized by the guidance directed decisions are not valid as these savings are associated with patients who were inappropriately removed from therapy. QoL values did decrease in these cases, but only by one to two percent. The time horizon of the model may be too short to show a more dramatic effect. Interestingly, decisions made with the presentation of CI's resulted in an increase of costs per patient at all viral load levels (from 1 to12%). Since the economic contribution of the HCV VL is to produce cost savings in the treatment management of non-responders, the presentation of CIs diminished

the value of the test information. QoL values at each VL were not affected by presentation of CI when compared within a VL level.

Analysis of a representative population with several VLs. To simulate decisions for a population of patients having a distribution of VLs at the 12-week test point, initial inputs designed to produce therapeutic outcomes relatively consistent with the population in the telaprevir clinical trial were used (Jacobson, et al., 2011). During the trial, all patients were on therapy throughout the course, unless there was a clinical reason for removal. Seventy-three percent of patients had an SVR, while 17% relapsed. At the start of triple combination therapy (12 weeks prior to the starting point of this model), 77% of patients had greater than 800,000 IU/ml HCV RNA. In addition, approximately 6% already had cirrhosis. To represent the population in the relatively simple model that was developed for this work, 86% were entered into the Continue therapy arm in the Response health state with 14% in the Discontinue therapy arm, splitting the Discontinue therapy proportion further with half in the Chronic HCV health state and half in the Cirrhosis health state. The final response percent at week 60 was 73%, consistent with the response outcome of the trial.

The output of the model simulating this population was used to determine the total costs of all patients continuing therapy, a scenario without a test available. Inputs were then adjusted to enter the proportion of non-responders stated in the trial into the discontinue therapy arm to assess the cost differences of removing them at week 12 based on their HCV VL. The results ($2633 in average cost savings per patient) are shown in Table 25.

Impact of repeat-testing on test value. The impact of repeat-testing on the costs was examined using this clinical study population. For this purpose, adjustments were made to the

Table 25

Average Cost Savings per Patient in Removing Non-responders for Telaprevir Clinical Trial Population

Telaprevir Study "Population"					Total Costs $	Av QoL Value
All patients Continue therapy					19472	0.95
Pts who would not have responded: off therapy at 12 and 24 wks					16839	0.96
				Cost Savings per Patient	2633	

model to maintain the ratios of responders and non-responders as close to the final outcome as possible. Table 26 shows the different combination of inputs with repeat testing occurring in eventual non-responders as well as responders. In this model, changes in repeat testing up to 50% had 6% or less impact on the total costs. This may be due to the fact that repeat-testing was handled by keeping the patient in the Continue therapy arm for the two weeks while "waiting" for the test result. Following the two-week step, those who would have been removed from therapy were sent back to the Discontinue therapy arm of the model. In this situation, where the response rate is high, at most only 14% of those retested had the extra two weeks of therapy. For these analyses, the HCV RNA test cost is only $58.88 and two weeks of therapy is $953. Per patient, this is about $1000 in costs. This small impact on costs, even in high repeat testing scenarios may give insight to the robustness of the test value.

Analyzing model-outputs' sensitivity to variation in parameters. The model developed for this research was not complex and had a short time horizon (60 weeks). Inputs and parameters were based on literature and on survey data. In the analyses described, several different levels of inputs were tested with variances based on the data received, and the outputs directly reflected them. To determine the sensitivity of the model outputs to variances in the cost and QoL

Table 26

Impact of Repeat Testing on Total Costs

Inputs			Outputs	
Repeat Rate	Discontinue Therapy	Continue Therapy	Costs ($)	% Change from Index
0*	0.14*	0.86*	16831	Index
0.05	0.09	0.86	17166	2
0.1	0.04	0.86	17501	4
0.15	0	0.85	17769	6
0.15	0.05	0.8	17434	4
0.2	0.05	0.75	17434	4
0.2	0.1	0.7	17099	2
0.25	0.1	0.65	17099	2
0.3	0.1	0.6	17099	2
0.35	0.1	0.55	17099	2
0.4	0.05	0.55	17434	4
0.4	0.1	0.5	17099	2
0.45	0.05	0.5	17434	4
0.45	0.1	0.45	17099	2
0.5	0.05	0.45	17434	4
0.5	0.1	0.4	17099	2
0.5	0.14	0.36	16781	0
* Index				

parameters, those values were varied and the model was run with each change using the telaprevir clinical trial population inputs. Total costs were changed in 10% increments from minus 50% of the parameter value used in the analyses to plus 50% of the value. Additionally, treatment costs and the HCV RNA test costs were tested separately for their contribution to model output sensitivity. Table 27 and Figure 12 are tabular and graphical representations of the outputs of this deterministic sensitivity analysis.

Table 27

Model Output Sensitivity to Costs and QoL Parameter Values

	Total Cost Changes		Therapy Cost Changes Only		HCV VL Test Cost Changes Only		QoL Ratio Changes	
%Change from Baseline	New Total Costs ($)	%Change From Baseline	New Total Costs ($)	%Change From Baseline	New Total Costs ($)	%Change From Baseline	New Average QoL	%Change From Baseline
-50	8420	-50.0	9525	-43.4	16814	-0.1	0.48	-50.0
-40	10104	-40.0	10988	-34.7			0.57	-40.6
-30	11788	-30.0	12450	-26.1	16824	-0.1	0.67	-30.2
-20	13471	-20.0	13914	-17.4			0.77	-19.8
-10	15155	-10.0	15372	-8.7			0.86	-10.4
0*	16839*	0*	16839*	0*	16839*	0*	0.96*	0*
+10	18523	10.0	18202	8.1			0.97	1.0
+20	20207	20.0	19765	17.4			0.99	3.1
+30	21891	30.0	21228	26.1	16854	0.1	1	4.2
+40	23575	40.0	22690	34.7			1	4.2
+50	25259	50.0	24153	43.4	16864	0.1	1	4.2
+100					16890	0.3		
+250					16965	0.7		
+500					17090	1.5		
+750					17215	2.2		
+1000					17341	3.0		
+2000					17843	6.0		
*Baseline								
** QoL values not adjusted beyond 1								

For total costs, the increases or decreases resulted in similar changes of magnitude in the output. Table 27 shows the variation of the parameter values and resulting change in the model outputs. Treatment costs were about 87% of the total costs, and the model was demonstrated to be sensitive to them. HCV RNA test costs had very little impact on the model outputs, even

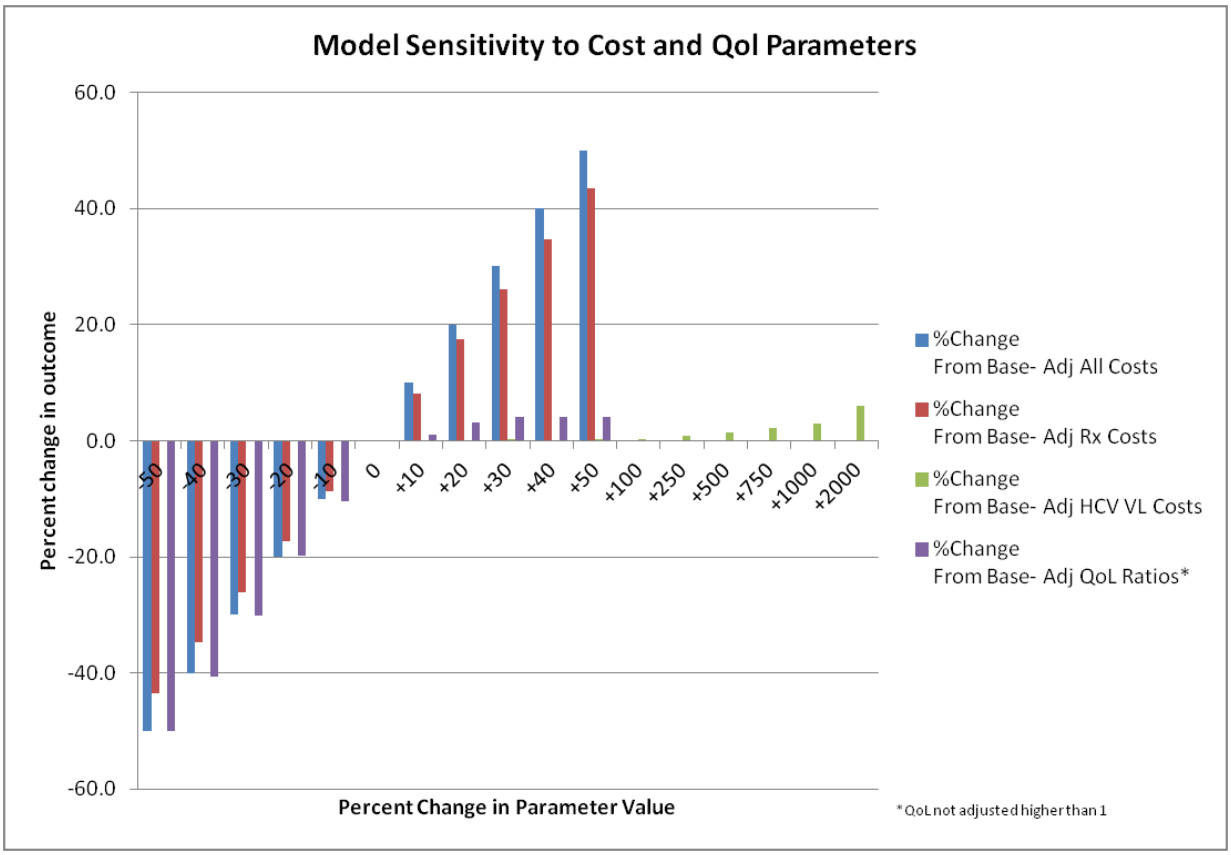

Figure 12. Model sensitivity to costs and QoL parameters values.

when increased by 2000% (resulting in a test cost of $1237). The base case "cost" used was the Medicare national limit of $58.88, however the cost of an HCV RNA test does vary widely and private-pay patients can pay above $1000. The analyses in this research demonstrated that even at that price, the test maintains its contributory economic value in the decisions being made regarding therapy management.

QoL ratios were also varied in the manner described for costs, except when each QoL ratio reached one, or full health, no further adjustments were made. For QoL ratios, the changes below the base case mirrored the percent change in QoL values.

Section summary: determining value. Through decision-analytic modeling, the economic contribution of the HCV VL test at week 12 was demonstrated. In this evaluation, changes in the

use of the information were shown to affect the value of the test. Changes in the use of the information from that directed by clinical guidelines could be measured in a simulated population with a representative distribution of viral load levels. Despite variations in use, the test held value.

Summary of Results.

The analyses and results in this section described two aspects of the clinical utility of a laboratory test, using HCV VL as a model, the interpretation and use of the test information and the economic contribution it has to the efficient use of clinical resources. It was found that among the hepatology specialists who responded to the survey, test information use was close to clinical guidelines, with the exception of repeat testing. Although differences were observed, there was no statistically significant impact on the use of the information when CIs were present.

Through the use of a specifically developed decision analytic model, economic contributions in average cost differences per patient could be demonstrated in simulated scenarios. The HCV VL test contributed an average of over $2600 in cost differences over all subjects in a hypothetical population of HCV patients, with savings up to $17,000 at higher VL levels. Varying the cost of the test up to $1200 had little impact on the value of the information produced from it.

Chapter Five: Discussion and Conclusions

Summary of the Study

The goal of this study was to examine the impact of physician awareness of a key performance characteristic for a pivotal laboratory test utilized in patient management decisions, and the economic contribution of that test to patient care. Quantitative HCV VL was used as the model since it is the principal indicator of the success or failure of current anti-HCV therapies. For HCV genotype 1 patients on triple-combination therapy, the HCV VL assessed at several time points during the course of therapy determines whether or not the patient should continue treatment. Since the treatment regimen is costly and can be debilitating, removing non-responders is appropriate when the HCV VL indicates they will not benefit from further therapy. Clinicians' uses of quantitative HCV RNA results to make this decision after 12 weeks of the triple-combination-therapy part of the HCV-treatment regimen were studied.

The key performance characteristic examined for the HCV VL test was its imprecision, which in this study was operationally defined as the 95% CI based on an SD of 0.15 $\log_{10}$ IU/ml. Through a survey, decisions made by clinicians after reviewing HCV VL results with and without those CIs were compared. Demographics, practice behaviors, and test presentation preferences were also analyzed to determine if they played a role in their decisions. In addition, a decision-analytic model was developed to calculate and evaluate how the value of the HCV VL test information was affected by the clinicians' utilization of the HCV VL results in their clinical decision making process. The hypothesis was that the clinical utility (CU) and the value of the

laboratory test information would change if clinicians were aware of the imprecision of the results.

Discussion of Survey Findings and Implications

Of the 1491 e-mail invitations to the survey sent, 124 responses were received with usable data (8.3%). The subset was small, however over 90% of those responding were representative of the target population, clinicians who manage HCV patients.

In an analysis of all responses, there was less than a 10% difference between decisions made when viewing quantitative HCV VL results with CIs, and those made without the CIs, at all HCV VL levels. Analysis of responses based solely on the presentation format of the first question seen by respondents showed greater differences between the two presentations, particularly for those HCV VL levels flanking either side of the decision threshold. For the HCV VL level immediately above the 1000 IU/ml threshold (1950 IU/ml) and immediately below (560 IU/ml), the presentation of CIs led to a 7-13% increase in the decision to continue therapy. At the 560 IU/ml HCV VL level, the presentation of CIs was also associated with a decrease in the decision to repeat the test compared to decisions made without CIs. These differences were not statistically significant, however. Small response numbers may have had a role: data from the first question seen provided only 65 responses to the HCV VL results without CI and 59 from the results with CI.

When the responses were analyzed to examine the impact that the order of presentation had on decisions, those clinicians who viewed results without CIs, and then saw results with CIs, were less likely to choose to continue therapy at all VL levels, except the actual threshold level of 1000 IU/ml (the decisions were the same). The VL level demonstrating the highest difference between the responses with the two presentations was at 560 IU/ml; just below the threshold. At

this level, while most of the assay variance was below the threshold (lower bound = 290 IU/ml) the upper bound of the CI crossed the threshold slightly (upper bound = 1,110 IU/ml). This may have influenced the decision not to continue treatment or to repeat the test. Repeating the test, where the greater number of the decisions shifted, delays the decision. These findings, though interesting, do not reflect current practice, as clinicians are not presented with the same results twice. However, they could provoke further discussion regarding the presentation of imprecision information to a community not accustomed to seeing it or instructed how to incorporate it into their decisions.

The tendency to continue the current course when the information is doubtful was demonstrated in the decisions made by those respondents who thought the imprecision information confused or detracted from the result. They were more likely to continue therapy than even to repeat the test. The decision to continue therapy is one of no action in this scenario. Patients are already on therapy at the decision point; and the action decision would be to remove the treatment. In the HCV triple combination therapy regimen, there is another HCV VL testing point 12 weeks after the one addressed in the survey (Ghany et al., 2011). At that time, the HCV VL should be undetectable for the patient to continue the last 24 weeks of therapy, giving the clinician and the patient time to revisit the earlier decision and compare the HCV VL results. Whether these considerations were part of the thought process for this group of experienced respondents was not tested.

Behavioral decision research is a field of study having roots in behavioral economics, and has grown to several applied areas (Medin & Bazerman, 1999; Thaler, 1991). Some of the tenets discussed in this field may provide insight into the observed tendency to choose no action (Continue therapy or Repeat the test) over action when faced with risk or uncertainty. It was not

the aim of this study to explain why clinicians made their choices, but to observe what the choices were and assess their impact on the value of the test information. Questions regarding the reason for choices, or the next course of action, might be added in future studies to focus on the question of continuing a path for patients when guidelines suggest an opposite action.

Repeat testing levels were high, and not surprisingly, most predominately high near the decision threshold. Repeat test rates were not influenced by the presentation of imprecision information or the order in which that information was seen. The Repeat test option was 8.2% of all first question responses and 8.5% of all second question responses, indicating that the second presentation of results did not fulfill the role of a repeat test for respondents. If the second viewing did fulfill this role, the repeat rate would be expected to be less for the second response.

The impact that the HCV VL level had on the actual decisions was anticipated. The therapeutic management of HCV patients hinges on HCV VL testing. The fact that the presentation of CIs had no statistically significant impact on the responses was unexpected, especially considering the studies discussed earlier that demonstrated clinicians' difficulty with more complex laboratory data (Bachmann et al., 2003; Bramwell et al., 2006; Cahan et al., 2003; Ghosh & Ghosh, 2004; Heller et al., 2004). The findings in this work could be attributed to several factors.

The targeted population was one of hepatology specialists; and the returned surveys were from the AASLD members, which may account for the strong knowledge of, and intended adherence to the treatment guidelines. It was appropriate to survey this group however, as it represents the community of specialists primarily managing HCV disease. Experienced clinicians using quantitative HCV RNA tests in the management of their patients have worked with quantitative HCV RNA tests as they have evolved, and the discussions surrounding the

relative imprecision in HCV VL testing have been active among them. An inherent knowledge of the test imprecision and having to make treatment decisions while considering the variance in test methods and in individual laboratories may already exist. The presentation of results with CIs in the survey may have served to substantiate this knowledge. It was notable, though, that the greatest difference in decisions made with and without CI was for those clinicians that had made decisions without CI first and then were presented with CI at the HCV VL level of 560 IU/ml, which is very near the decision threshold. If the presentation of CIs with a result that is near a threshold impacts this experienced and specialized group, surveying a sample of non-specialists with a relevant diagnostic/ therapeutic combination could possibly demonstrate greater differences in dealing with the presentation of imprecision.

The CIs defining the imprecision in this study were conservative, particularly at the lower levels where imprecision is greatest in quantitative viral load assays. Table 2 described published studies with these relative values. The CIs presented with the HCV VL results in the in the survey may have illustrated an optimal situation. A larger imprecision range might have produced different findings. In addition, the decision threshold for telaprevir is 1000 IU/ml, which is 40 times the limit of detection for quantitative HCV RNA assays used for these decisions and comfortably within a reproducible part of the reportable range. Another protease inhibitor (boceprevir), used in a similar manner with interferon and ribavirin, has a decision threshold of 100 IU/ml (Ghany et al., 2011). The imprecision at this point would be expected to be much larger, and in some cases cross the lower limits of detection for the tests. Surveying clinicians using this decision threshold and associated imprecision characteristics may also have produced greater differences between groups seeing the results with CI and those seeing results without them, particularly at the threshold.

An interesting finding was the frequent choice of the decision to repeat the test. Respondents chose the decision to repeat at all HCV VL levels except the highest (31,620 IU/ml). The Repeat test decision was 17.5% of all decisions. The decision to repeat tests at the threshold of 1000 IU/ml and the two immediately flanking levels (560 and 1950 IU/ml) was 19.8% of all decisions, and at the decision threshold (1000 IU/ml) it was 36.1% of all decisions. At these three HCV VL levels, the decision to repeat was 2-6% higher when the results were presented with CIs. The presentation of CIs had a consistent and small, but non-statistically significant impact.

Though there were no statistically significant associations with the decision to repeat the HCV RNA test, the frequency of this decision among respondents is an important consideration. The Repeat test decision had little impact on the economic value of the test; however, indications for repeat testing are not part of the labeling for quantitative HCV RNA tests, the therapeutics for which they are companions, nor is repeat testing considered in the guidance documents. The inherent imprecision of the quantitative HCV RNA test is already a factor in the establishment of decision thresholds by means of the therapeutic clinical trials; at least for the specific test used in those trials. As such, a repeat test may be considered medically unnecessary by reimbursement policy makers; and the patient or the institution may be required to bear the cost when the clinician makes this decision. Adding information about the imprecision of the test to the result, which, in this study, was demonstrated to increase repeat testing behavior, may have economically adverse consequences.

Repeat testing is not uncommon in clinical laboratory medicine, within the laboratory and at clinicians' requests. Several publications discuss the reasons for and the implications of this practice (Deetz, Nolan, & Scott, 2012; Hawkins, 2005; van Walraven et al., 2003). However, the tests studied in these publications were typically chemistry and hematology tests, which have

lower cost, turn-around time and resource impact than molecular tests for viral load. As tests with more expensive technology, time and resource impact are implemented, examining this tendency further or outlining in guidelines where repeat testing is appropriate may be necessary.

The manner in which the imprecision information was presented to the clinicians may also have had an impact on their decisions. In this survey, the imprecision information was presented numerically and consistent with the visual format of the test result. No additional explanation accompanied the 95% CI range. Presentation in words, graphically, or with interpretive language may have led to different decisions. An explanation of how to apply the imprecision information or a graphical representation of the odds of the value crossing the decision threshold might have had an impact on the Repeat test decisions or those decisions that were inconsistent with the clinical guideline recommendations. Bachmann and colleagues noted that among the different presentation of results they provided to their participants, the most effective way of presenting complex laboratory information was with an explanatory statement of the odds of the patient having the disease in question (2003).

Limitations of the survey research. Assessing the impact of a laboratory test in a healthcare setting necessitates the mitigation of as many variables as possible. One way to accomplish this is to choose a clinical path where a specific decision employs a prescribed laboratory test, and then use that in a setting with a single patient. The scenario chosen for this research was based on professional guidelines of treatment and assessment; and the description of the simulated patient was intended to present one with an average burden of therapeutic-associated adverse events. However, a single patient scenario in one specific clinical path limits the ability to generalize the findings, or to assess if the bias towards patient types exists among respondents. Since patient-treatment biases were not the focus of this research, the single simulated patient

scenario provided the opportunity to hold a greater number of variables constant to assess the effect of the HCV RNA test variables and interpretation. The situation presented in this study may have a seemingly narrow focus, but the approach used can serve as a basis for research questions in other disease management areas, or for other questions regarding viral disease management.

There are no publicly available, validated instruments measuring clinicians' experiences with test result imprecision. The survey, through which the hypotheses in this study were tested, was developed specifically for this work. The questions presented were designed to be consistent with standard practice guidelines and the HCV VL numbers and imprecision budget used for the test results were calculated using published studies and recommendations of minimum performance standards by expert groups. Patient management for HCV disease is evolving quickly. Even during the development phase of this research, the treatment protocols changed affecting the patient scenario and the survey design (Ghany, et al., 2009; Ghany et al., 2011). However the reliance on the clinicians' interpretations of quantitative HCV RNA results as a critical component of the therapeutic management of the patient remained constant and will continue to be used for this purpose for the foreseeable future.

Relying on a survey as a surrogate for clinicians' decision also has limitations. It is possible that the participants chose their answers according to how they believed they should answer. A survey cannot represent the full complexity of the authentic medical situation. Further, there are no consequences for a wrong decision. A more robust way of gathering data would be to review patient records. The HCV VL result and the continuance or removal of therapy would be apparent information in the patient record, and easily extractable. However, a multi-site or multi-system study would have to be done in order to access a significant number of clinicians.

Patients would have to be matched by disease severity, viral load, adverse events and demographics. A small study to verify or dispute the findings in this study may be justified but the larger one may not be a priority until the use of electronic health records across systems is more established and easily accessed by legitimate researchers.

Discussion of Economic Model Findings and Implications

A Markov decision-tree model was developed to determine the value of the HCV VL test information. Transition probabilities, costs, and QoL values for health states were populated from published literature. Clinicians' actual use of the HCV VL information was represented by the survey data and used as inputs to the model. The definition of value for this study was the potential average differences in costs per patient obtained by removing non-responders from inappropriate therapy. This would result in potential savings that could be attributed to having the HCV VL information. The HCV VL test information contributes to an average of over $2600 in cost savings in a model population of HCV patients, with savings up to $17,000 at higher HCV VL levels. Presentation of CI with the HCV VL results showed a slight decrease in the value, as decisions made after reviewing HCV VL with CIs resulted in greater costs than guideline-directed decisions made without CIs.

Though cost savings of $2,600 to $17,000 may be considered small compared to other health economic analyses, when the investment made to facilitate those savings ($58.88 Medicare reimbursement rate) is considered, the HCV VL test in the clinical scenario used for this research was determined to be economically effective. Even when the cost of the test was mathematically increased to nearly $1200, there was little impact to its value in the model. The optimal value does depend on realizing cost savings through the proper use of the test in patient management. It was shown that when decisions were made according to guidelines, the test was most cost

effective. This type of analysis might help demonstrate the importance of investing in programs to assure that clinicians interpret and use testing information properly, in addition to the efforts spent in the technical considerations of choosing an assay.

Limitations of the health economic research. Mathematical models have advantages as well as limitations, and are as robust as the assumptions and data built into them. The costs, disease state transitions, and QoL values in the model developed for this research were sourced from a rich literature base and conservative choices were made. The sensitivity of the model outputs to costs was demonstrated; higher or lower treatment or disease associated costs will have an impact on the outcome of test value. With the small investment needed for the current HCV laboratory test though, its relative value prevails. For HCV VL testing at the 12-week decision point, testing costs up to $2600 would still realize a net savings, using the assumptions made in this model.

A limitation of the decision analytic model developed for this study was seen at the lower VL levels. Adverse outcomes for inappropriately choosing to remove a patient from therapy exist due to the patient's disease progression. However, the full costs of these adverse outcomes could not be assessed in this study due to the short time horizon of the model. Discontinuing 24 weeks of therapy for an individual would "save" approximately $11,400 for approximately the first half-year; however, depending on the rate of disease progression due to therapy discontinuance, the next four-and-a-half years would incur six month costs of $3090 to $42,600. Liver transplant, which could happen after this time period, and which would add considerable cost was not included in the model. Since the proportion of patients that progress through the more expensive stages of hepatitis C is small, and those costs do not always outweigh the treatment savings in the short run, QoL ratios were also evaluated. To assess a quality of life penalty, QoL

values were calculated at each health state and for some inputs having a greater proportion of decisions to discontinue therapy than was appropriate, the QoL value did decrease. However the decrease in QoL was less than 2%. Again, the measurable change in health state in 60 months for the majority of patients is subtle. Attempts to build economic models to appropriately address the scope of influence of diagnostic tests would benefit from inter-disciplinary discussions among laboratory professionals, health economists, clinicians and policy makers.

The costs of producing and implementing laboratory reports with CIs on each result were not considered. It was assumed that they would be minimal in relation to the overall costs of the treatments and disease management. However they could be significant with respect to the current cost of generating a billable laboratory result. Costs to be considered are: the cost of performing studies to determine the imprecision of the results within the laboratory, changes to the laboratory information systems which generate the reports, education and training of clinicians, and of laboratory staff to answer clinicians' questions. The value of the information in the population of HCV patients was calculated to be an average of $2600 per patient. The cost of implementing CIs, as well as the true costs of performing the test (instead of the Medicare reimbursement used in this study), could be used to determine the cost/benefit of including CIs with laboratory test results for HCV VL.

Conclusions and Recommendations for Future Research

The results of this research led to the following conclusions:

- The presentation of test imprecision with HCV VL results had little impact on clinicians' treatment management choices at the 12-week decision point for HCV patients.
- There was a high percent of repeat testing as the HCV VL test result neared the decision threshold. Presentation of imprecision information did not influence this rate.

- When treatment management decisions changed from viewing results with and without CIs, they were most often associated with not having imprecision information first, followed by seeing results with imprecision information.
- Most clinicians surveyed preferred *not* to have imprecision information presented with HCV VL test results.
- The monetary value of HCV VL information at the 12 week decision point for triple combination therapy could be calculated through a decision analytic model. The monetary value of the test information prevailed despite knowledge of imprecision and high rates of repeat testing.

Recommendations for future research in the context of these conclusions are discussed below.

The responses to the survey showed that the presentation of test imprecision with the HCV VL results had little impact on the clinicians' treatment management choices. As noted, this was unexpected and inconsistent with predictions modeled by laboratory professionals (Pisani et al., 2009). The high percent of the Repeat test choice, though not entirely unexpected, was also notable. Since many clinical path recommendations are developed without reference to test specifications, and the conditions for repeat testing are not usually discussed, these findings underscore the importance of conducting research in the interpretation and employment of laboratory test results by the actual users of the information.

The clinicians surveyed in this study were an elite group of specialists. By the number of years in practice, it can be assumed that their professional experience had developed with the development of the quantitative molecular HCV RNA tests. Their responses provide a benchmark for appropriate interpretation and use of complex quantitative test information. Laboratory test results, however, are available to and used by many types of providers: general

practice physicians, nurses, pharmacists, other non-provider professionals and patients. Coincidently, with the advances in research and technology, laboratory test information is becoming more complex.

Future research in test interpretation and clinical use should study test results for complex specialty areas, which highly impact medical decisions and are used by non-specialists. A relevant example is cardiac marker testing to determine and predict myocardial events. Mathematical models demonstrating the impact of test imprecision on those decisions have been performed and the recommendations of useable assays include a tolerance for imprecision (Wu & Christensen, 2013). However, the determination of a cardiac event is not always made by a specialist who is familiar with the latest studies in this field regarding the biological variation and assay imprecision of standard and ultra sensitive tests. Research presenting emergency department physicians with results of these tests that include their imprecision could illustrate the impact of this additional information on high risk, high impact decisions. Another example is quantitative CMV testing for transplant patients. In this case, specialists from different fields (transplantation and infectious diseases) use quantitative CMV VL to discriminate between graft rejection, infection, clearance of infection or viral resistance (Bienek, Kirby, Cheng, Eichelberger, & Qian, 2011; Perrottet et al., 2010). CMV VL results inform very different management strategies for each of these indications. Research on the clarity of the result at the decision point considering the test imprecision could help refine the precision specifications needed for these tests. A third area is in disease prediction tests, which are being used to persuade patients to make behavioral and lifestyle changes to avert them. Research in the interpretation, use and the actual impact of that information on the intended outcome is critically

important in the current heath care environment where resource prioritization must be supported by data.

Clinicians' knowledge and use of imprecision information is also an important topic for developing clinical laboratory policy. Laboratory guidelines exist that articulate the determination and reporting of imprecision with laboratory test results (CLSI C51A, 2011& IS0, 2012). In this study, it was shown that most changed decisions occurred when respondents viewed results with imprecision after having viewed the same results without it. Further, the most frequent response to the question of whether or not respondents wanted imprecision information with laboratory test results was negative. It is currently mandated that test imprecision information be available on request, and it is anticipated that future guidance will require it. Standardization of how it is to be calculated and portrayed; and an understanding of how users of the information will respond have not been studied. Further research to help guide policy in this area is warranted.

Another area for future research is in methods development for the value of laboratory test information analyses. The value of the HCV VL test information at the 12-week decision point for triple combination therapy was determined through a decision analytic model evaluating the monetary value of the information provided for that specific decision. Typically, health economic analyses examine the extension of life or increased quality of life facilitated though service or therapeutic interventions where incremental cost effectiveness ratios are measured and compared. When applied to evaluate diagnostic tests, these methods presume that the primary intended use of the diagnostic test is to extend life and quality. While those are the ultimate outcomes, evaluating diagnostic information in this manner and comparing its contribution to interventions and services with direct impact on them, can obscure the primary contribution of

the diagnostic test. Examining the value of the test information that directs a decision along that ultimate path may more clearly represent the contribution of the laboratory test. The analysis presented in this study demonstrated this strategy and is one way the value of the test result could be calculated. Development of the decision analytic model revealed several issues to be considered such as: the required complexity of the model, the aspects of the outcome to be measured, the appropriate time horizon of the model, the representation of actual decisions made by users, and the impact of decisions made as a result of incorrect results.

The value of information is a concept already gaining use in clinical trial decisions for extending or truncating arms of trials, determining the value of additional research, and adding different indications (Meltzer, Hoomans, Chung, & Basu, 2011). Though the value of information in a decision path can be calculated, it has not been used as a tool to measure or compare laboratory tests as was presented here. The field of companion diagnostics is growing, where more laboratory tests will have direct roles in deciding treatment paths. Further development and consideration of this type of analysis will be useful to guide investment in the diagnostic tests that will be necessary to realize personalized medicine. Interdisciplinary work is critical to define the acceptable perspective and scope of economic analyses specifically relevant to clinical laboratory diagnostics. Information derived from such analyses may help support policy decisions needed to make available those diagnostics for which clinical and economic utility can be demonstrated.

The predicted economic burden of HCV disease has catalyzed a strong pipeline of HCV therapeutics currently in trials (Cox, 2013). Health economic literature on the cost effectiveness of new therapeutics is now available shortly after their approval for use (Camma et al., 2012). However, there are no economic analyses for the laboratory tests which are integral to the

clinical trials and which determine the efficacy of the drugs; the same tests that will be used to triage patients to therapies and manage their therapy for response, compliance and drug resistance. As new therapies in hepatitis C and hepatitis B are developed, the viral load assays will be critical in assessing outcomes and then managing patients. Knowing the technical limitations of the assays only answers part of the question of their clinical utility. Whether or not clinicians accurately account for those technical limitations when using the information is also important. To be able to communicate the value that those assays bring to patient management decisions will also help secure development support and future access to these tests for the patients who need them.

List of References

American Diabetes Association. (2010). Executive summary: standards of medical care in diabetes – 2010. *Diabetes Care, 33*, (supp 1), s4-s10. doi: 10.2337/dc10-S004

Australian Government. Department of Health and Aging. (2002) Methodology for estimating the QALYs gained from the prevention of HIV and HCV infections. Section 5.3 in *Return on investment in needle and syringe programs in Australia: report*. Retrieved from http://www.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-drugs-illicit-index.htm

Avorn, J. (2009). Debate about funding comparative-effectiveness research. *New England Journal of Medicine, 360*, 1927-1929. Retrieved from http://www.nejm.org/doi/full/10.1056/NEJMp0902427

Backus, L.I., Boothroyd, D.B., Phillips, B.R., Belpirio, P., Halloran, J., & Mole, L.A. (2011). A sustained virologic response reduces risk of all-cause mortality in patients with hepatitis C. *Clinical Gastroenterology and Hepatology, 9*, 509 – 516. doi:10.1016/j.cgh.2011.03.004

Bachman, L.M., Struerer, J., & ter Reit, G. (2003). Simple presentation of test accuracy may lead to inflated disease probabilities. *British Medical Journal, 326,* 393-394. doi:10.1136/bmj.326.7385.393/a

Barbeau, J.M., Goforth, J., Caliendo, A.M., & Nolte, F.S. (2004). Performance characteristics of a quantitative taqman hepatitis C virus RNA analyte specific reagent. *Journal of Clinical Microbiology, 42*, 3739 – 3746. doi: 10.1128/JCM.42.8.3739–3746

Berger, M.L., Bingefors, K., Hedblom, E.C., Pashos, C.L., & Torrance, G.W. (2003). *Health care cost, quality, and outcomes. ISPOR book of terms*. Princeton-Pike, NJ: International Society for Pharmacoeconomics and Outcomes Research Press.

Best, S.J., Gust, A.P., Johnson, E.I.M., McGavin, C., & Dax, E. M. (2000). Quality of human deficiency virus viral load testing in Australia. *Journal of Clinical Microbiology*, *38,* 4015-4020. Retrieved from http://jcm.asm.org/content/38/11/4015

Bieniek, R., Kirby, J., Cheng, A., Eichelberger, K., & Qian, Q. (2011). Effective use of PCR for the detection of cytomegalovirus viremia and monitoring therapy in immunocompromised patients. *Laboratory Medicine, 42*, 339-343. doi:10.1309/LMP4IL3XGU6MQJCE

Bloom, B.S. (2004). Use of formal benefit/cost evaluations in health system decision making. *The American Journal of Managed Care, 10*, 329-335. Retrieved from http://www.ajmc.com/publications/issue/2004/2004-05-vol10-n5/May04-1791p329-335

Bramwell, R., West, H., & Salmon, P. (2006). Health professionals' and service users' interpretation of screening test results: experimental study. *British Medical Journal*, *333*, 284-286. doi: 10.1128/JCM.42.8.3739–3746.2004

Cahan, A., Gilon, D., Manor, O., & Paltiel, O. (2003). Probabilistic reasoning and clinical decision making: do doctors overestimate diagnostic probabilities? *Quality Journal of Medicine, 96,* 763-769. doi:10.1093/qjmed/hcg122

Caliendo, A.M., Valsamakis, A., Zhou, Y., Yen-Lieberman, B., Andersen, J., Young, S., … Lurain, N.S. (2006). Multilaboratory comparison of hepatitis C virus viral load assays. *Journal of Clinical Microbiology, 44*, 1726-1732. doi:10.1128/JCM.44.5.1726–1732.2006

Camma, C., Petta, S., Enea, M., Bruno, R., Bronte, F., Capursi, V.,… Craxi, A., on behalf of the WEF study group. (2012). Cost effectiveness of boceprevir or telaprevir for untreated patients with genotype 1 chronic hepatitis C. *Hepatology, 56*, 850-860. doi:10.1002/hep.2574

Carlson, J.J., Garrison, L.P., Ramsey, S.D., & Veenstra, D.L. (2009). The potential and economic outcomes of pharmacogenomic approaches to EGFR-tyrosine kinase inhibitor therapy in non-small-cell lung cancer. *Value in Health, 12*, 20-27. doi:10.1111/j.1524-4733.2008.00415.x

Castelnuovo, E., Thompson-Coon, J., Pitt, M., Cramp, M., Siebert, U., Proce, A., & Stein, K. (2006). The cost-effectiveness of testing for hepatitis C in former injecting drug users. *Health Technology Assessment Journal, 10*, 1-122. doi:10.3310/hta10320

Centers for Disease Control and Prevention. Hepatitis C Information for Health Professionals Website. http://www.cdc.gov/hepatitis/HCV/HCVfaq.htm#section1 accessed 20 Feb 2012

Chevaliez, S., Bouvier-Alias, M., & Pawlotsky, J.-M. (2009). Performance of the Abbott real-time PCR assay using m2000sp and m2000 rt for hepatitis C virus RNA quantification. *Journal of Clinical Microbiology, 47,* 1726-1732. doi:10.1128/JCM.01300-08

Chevaliez, S., & Pawlotsky, J.-M. (2009). How to use virological tools for optimal management of chronic hepatitis C. *Liver International, 29*(s1), 9-14. doi:10.1111/j.1478-3231.2008.01926.x

Choo, Q-L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., & Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral genome. *Science, 244,* (4902) 359 – 362. Retrieved from http://www.jstor.org/stable/1703266

Clinical and Laboratory Standards Institute. (2004). *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline Second Edition, EP5-A2*. Wayne, PA.

Clinical and Laboratory Standards Institute. (2004). *Application of a Quality Management System Model for laboratory Services; Approved Guideline Third Edition, GP26-A3*. Wayne, PA.

Clinical and Laboratory Standards Institute. (2005). *User Verification of Performance for Precision and Trueness; Approved Guideline Second Edition, EP15-A2*. Wayne, PA.

Clinical and Laboratory Standards Institute. (2011). *Expression of Measurement Uncertainty in Laboratory Medicine; Approved Guideline, C51-A*. Wayne, PA.

Clinical Laboratory Improvement Amendments of 1988. Establishment and verification of performance specifications. Laboratory Requirements: 42 C.F.R.493 §1253 (2003).

Clinical Laboratory Fee schedule: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ClinicalLabFeeSched/clinlab.html accessed 6/12/13

Cohen, D.J., & Reynolds, M.R. (2008). Interpreting the results of cost-effectiveness studies. *Journal of the American College of Cardiology, 52,* 2119-2126. doi:10.1016/j.jacc.2008.09.018

Cox, R.K. (2013). Staying Informed: anticipating new regimens with direct acting antivirals for Hepatitis C. US Department of Veterans' Affairs Website. http://www.hepatitis.va.gov/provider/topics/DAA-update.asp accessed 4 July 2013

de Vries, R., van Bergen, J., deJong-van den Berg, L.T.W., & Postma, M.J., for the PILOT-CT Study Group. (2006). Systematic screening for chlamydia trachomatis: estimating cost-effectiveness using dynamic modeling and Dutch data. *Value in Health, 9,* 1-11. doi:10.1111/j.1524-4733.2006.00075.x.

Deetz, C.O., Nolan, D.K., & Scott, M.G. (2012). An examination of the usefulness of repeat testing practices in a large hospital clinical chemistry laboratory. *American Journal of Clinical Pathology, 137*, 20-15. doi:10/1309/AJCPWPBF62UGEFOR.

Dienstag, J.L., & McHutchison, J.G. (2006). American gastroenterological association medical position statement on the management of hepatitis C. *Gastroenterology*, *130*, 225-230. doi:10.1053/j.gastro.2005.11.011

Drummond, M., & Sculpher, M. (2006) Better analysis for better decisions: has pharmacoeconomics come of age? [Editorial]. *Pharmacoeconomics, 24,* 107-108. Retrieved from http://adisonline.com.proxy.library.vcu.edu/pharmacoeconomics/Fulltext/2006/24020/Better_Analysis_for_Better_Decisions__Has.1.aspx

Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J., & Stoddart, G.L. (2005) *Methods for the economic evaluation of health care programs*. New York: Oxford University Press.

Drummond, M.F., Schwartz, J.S., Jönsson, B., Luce, B.R., Neumann, P.J., Siebert, U., & Sullivan, S.D. (2008). Key principles for the improved conduct of health technology assessments for resource allocation decisions. *International Journal of technology Assessment in Health Care, 24* (3), 244-258. doi:10.1017/S0266462308080343

El-Kamary, S.S., Jhaveri, R., & Shardell, M.D. (2011). All-cause liver-related and non-liver related mortality among HCV-infected individuals in the general US population. *Clinical Infectious Diseases, 53,* 150-157. doi: 10.1093/cid/cir306

Field, A. (2013). Inflated error rates. In *Discovering statistics using IBM SPSS statistics* (pp. 68-69). London, England: Sage.

Fineberg, H.V. (1980). Decision trees: construction, uses and limits. *Bulletin du cancer*, 67, 395-404.

Folland, S., Goodman, A.C., & Stano, M. (1997). *The economics of health and health care.* Upper Saddle River, NJ: Prentice-Hall, Inc.

Fos, P.J., & Fine, D.J. (2005). Chapter 10- Economic analysis of health care for populations. In *Managerial epidemiology for health care organizations* (2nd ed.). San Francisco, CA: Jossey- Bass

Fraser, C. (2004). Test result variation and the quality of evidence-based clinical guidelines. *Clinica Chimica Acta, 346*, 19-24. Retrieved from http://dx.doi.org/10.1016/j.cccn.2003.12.032

Fry, R., Avery, S., & Sullivan, S. (2003). The academy of managed health care pharmacy format for formulary submissions; an evolving standard. A Foundation for Managed Care Pharmacy Task Force report. *Value in Health, 6*, 505-521. doi: 10.1046/j.1524-4733.2003.65327.x

Gaba, M. (2009). Comparative effectiveness. Will it foster FDA-CMS collaboration or drive the agencies apart? Update. Food and Drug Law. Regulation and Education, 2009, 14- 17. http://www.fdli.org/home

Gagné, G, Reinharz, D., Laflamme, N., Adams, P.C., & Rousseau, F. (2007). Hereditary hemochromatosis screening: effect of mutation penetrance and prevalence on cost-effectiveness of testing algorithms. *Clinical Genetics*, 71, 46-58. doi: 10.1111/j.1399-0004.2006.00727.x

Gallant, L., Kieffel, H., Chatwin, R., & Smith, J.E. (1999, March). Using learning models to capture dynamic complexity in petroleum exploration (SPE 5294). Proceedings of the Society of Petroleum Engineers Hydrocarbon and Evaluation Symposium, Dallas, Texas.

Garber, A.M., & Tunis, S.R. (2009). Does comparative effectiveness research threaten personalized medicine? *New England Journal of Medicine, 360*, 1925-1927. Retrieved from http://www.nejm.org.

Gandhi, T.J., Kachalia, A., Thomas, E., Puopolo, A.L., Yoon, C., Brennan T.A., & Studdard, D.M.. (2006). Missed and delayed diagnosis in the ambulatory setting: a study of closed malpractice claims. *Annals of Internal Medicine, 145,* 488- 496. Retrieved from http://annals.org/journal.aspx

Ge, D., Fellay, J., Thompson, A,J., Simon, J.S., Shianna K.V., Urban, T.J., …Goldstein, D.B. (2009). Genetic variation in IL2B with response to pegylated interferon-alpha and ribavirin therapy. Nature, *461*(7262), 399-401. doi:10.1038 nature/08309

Ghany, M.C., Strader, D.B., Thomas, D, L., & Seef, L.B. (2009). AASLD practice Guidelines: diagnosis management, treatment of hepatitis C: an update. *Hepatology*, 49, 1335 – 1374. doi: 10.1002/hep.22759

Ghany, M.C., Nelson, D.R., Strader, D.B., Thomas, D.L., & Seeff, L.B. (2011). An update on treatment of genotype 1 chronic hepatitis C virus infection: 2011 practice guideline by the American Association for the Study of Liver Diseases. *Hepatology*, 54, 1433-1444. doi: 10.1002/hep.24641

Ghosh, A.K., & Ghosh, K. (2004). Do medical students and physicians understand probability? *Quality Journal of Medicine, 97*, 53-56.

Gold, M.R., Siegel, J.E., Russell, L.B., & Weinstein, M.C. (Eds) (1996). *Cost-effectiveness in health and medicine*. (p 88). New York, NY: Oxford University Press.

Goldie, S.J., & Corso, P.S. (2003). Decision analysis. In Haddix, A.C., Teutsch, S.M., & Corso, P.S. (Eds.) *Preventive effectiveness* (pp 127 – 155). New York, NY: Oxford University Press.

Gretch, D.R. (1997). Diagnostic tests for hepatitis C. *Hepatology, 26,* [Supplement.1], 43s-47s. doi: 10.1002/hep.510260708

Gretch, D.R., dela Rosa, C., Carithers Jr., R.L., Willson, R.A., Williams, B., & Corey, L. (1995). Assessment of hepatitis C viremia using molecular amplification technologies: correlations and clinical implications. *Annals of Internal Medicine, 123* (5), 321-329. Retrieved from http://www.acponline.org/journals/annals/annaltoc.htm

Griffiths, P., Whitley, R., Snydman, D.R., Singh, N., & Boeckh, M.(2008). Contemporary management of cytomegalovirus infection in transplant recipient: guidelines from an IHMF® workshop, 2007. *Herpes*, 15 (1) 4- 12. Retrieved from http://ww35.ihmf.org/journal/journal.asp

Grosse, S. D., & Khoury, M.J. (2006) What is the clinical utility of genetic testing? *Genetics in Medicine*, *8*, 448- 450. doi: 10.1097/01.gim.0000227935.26763.c6

Hammer, S.M, Eron, J.J., Reis, P., Schooley, R., Thompson, M., Walmsley, S., …Volberding, P.A. (2008). Antiretroviral treatment of adult HIV infection. 2008 recommendations of the international AIDS society-USA panel. *JAMA*, *300*, 555 – 570. doi:10.1001/jama.2012.9885

Hawkins, R.C. (2006). Potentially inappropriate repeat laboratory testing in inpatients. *Clinical Chemistry, 52*, 784-785. doi: 10.1373/clinchem.2005.064139

Heller, R.F., Sandars, J.E., Patterson, L., & McElduff, P. (2004). GPs' and physicians' interpretation of risks, benefits and diagnostic tests results. *Family Practice*, *21*, 155 – 159. doi: 10.1093/fampra/cmh209

Ho, D.D. (1996). Viral counts in HIV infections. *Science, 272*, 1124-1125. Retrieved from http://www.jstor.org/stable/2889933

Hopwood, M., Treloar, C., & Redsull, L. (2006). Experiences of hepatitis C treatment and its management. What some patients and professionals say. Monograph 4/2006. National Centre in HIV Social Research. The University of New South Wales. Retrieved from http://nchsr.arts.unsw.edu.au/media/File/HepCtreatment.pdf

Hornberger, J., Farci, P., Prati, D., Zeuzam, S., Green, J., & Patel, K.K. (2006). The economics of treating chronic hepatitis C patients with peginterferon α-2a (40 kDa) plus ribavirin presenting with persistently normal aminotransferase. *Journal of Viral Hepatitis, 13,* 377-386. doi:10.1111/j.1365-2893.2005.00700.x

Huninck, M.G.M., Glasziou, P.P., Siegel, J., Weeks, J.C., Pliskin, J.S., Elstein, A.S., & Weinstein, M.C. (2006). *Decision making in health and medicine. Integrating evidence and values.* Cambridge University Press, Cambridge, UK.

International Standards Organization. (2012) Medical laboratories-Particular requirements for quality and competence: ISO 15189:2012. Geneva, Switzerland.

Jacobson, I. M., McHutchinson, J.G., Dusheiko, G, Di Bisceglie, A.M., Reddy, K.R., Bzwoej, N.H., … Zeuzem, S. for the ADVANCE study team. (2011). Telaprevir for previously untreated chronic hepatitis C infection. *The New England Journal of Medicine, 364,* 2405-2416. Retrieved from http://www.nejm.org

Jackson, B.R., Busch, M.P., Stramer, S.L., & AuBuchon, J.P. (2003). The cost effectiveness of NAT for HIV, HCV and HBV in whole blood donations. *Transfusion*, *43*, 721- 729. doi:10.1046/j.1537-2995.2003.00392.x

Johansson, P.-O. (2003). The empirical value of a statistical life: theoretical and empirical evidence. *British Medical Journal, Applied Health Economics and Health Policy 2003*, Special Issue, 25-33.

Jones, R., & Cuckle, H. (2007). Evidence-based laboratory medicine: impact of analytical performance on outcomes. In C. P. Price, & R. H. Christenson (Eds.), *Evidence-based laboratory medicine. Principles, practice and outcomes, second edition* (pp. 141- 164), Washington, DC: AACC Press.

Klee, G.G., Shryver, P.G., Saenger, A.K., & Larsen, T.S. (2007). Effects of analytic variations in creatinine measurements on the classification of renal disease using estimated glomerular filtration rate (eGFR). *Clinical Chemistry and Laboratory Medicine, 45,* 737-741. doi: 10.1515/CCLM.2007.168

Kristensen, F.B. (2008). What is Health Technology Assessment- and how is the European network for HTA (EUnetHTA) developing transnational collaboration? *ISPOR Connections, September/ October 2008*, 14-16. Retrieved from http://www.ispor.org/news/IConnections.aspx

Laking, G., Lord, J., & Fisher, A. (2006). The economics of diagnosis. *Health Economics, 15*, 1109-1120. doi:10.1002/hec.1114

Lang, D.L., Lopert, R., & Hill, S.R. (2003). Use of pharmacoeconomics in prescribing research. Part 5: Modeling, beyond clinical trials. *Journal of Clinical Pharmacy and Therapeutics, 28*, 433-439. Retrieved from http://www.wiley.com/WileyCDA/WileyTitle/productCd-JCPT.html

Laperche, S., Bourchardeu, F., Thibault, F., Pozzetto, B., Vallet, S., Rosenberg, A.R., … Lefrère, J.-J. (2007). Multicenter trials need to use the same assay for hepatitis C virus viral load determination. *Journal of Clinical Microbiology, 45(11)*. 3788-3790. doi: 10.1128/JCM.00825-07.

Little, C., & Lewis, J. (2008) Newborn screening. *Newborn and Infant Nursing Reviews*, 8 (1), 3-9. doi:10.1053/j.nainr.2007.12.004

Loubiére, S., Rotily, M., & Moatti, J-P. (2003). Prevention could be less cost effective than cure: the case of hepatitis C screening policies in France. *International Journal of Technology Assessment in Health Care, 19* (4), 632-645.

Lundberg, G. (1998). The need for an outcomes research agenda for clinical laboratory testing. Editorial. *JAMA, 280*, 565-566. Retrieved from http://jama.jamanetwork.com

Ly, K.N., Xing, J., Klevens, M., Jiles, R.B., Ward, J.W., & Holmberg, S.D. (2012). The increasing burden of mortality from viral hepatitis in the united states between 1999 and 2007. *Annals of Internal Medicine, 156 (4),* 271-278. Retrieved from http://www.acponline.org/journals/annals/annaltoc.htm

Mangia, A., Minerva, N., Bacca, D., Cozzolongo, R., Ricci, G., & Carretta, V. (2008). Individualized treatment duration for Hepatitis C genotype 1 patients: A randomized controlled trial. *Hepatology*, 47 (1), 43-50. doi: 10.1002/hep.22061

Martinot-Peignoux, M., Marcellin, P., Pouteau, M., Castelnau, C., Boyer, N., Poliquin, … Erlinger, S. (1995). Pretreatment serum hepatitis C RNA levels and hepatitis C virus genotype are the main and independent prognostics factors of sustained response to interferon alfa therapy in chronic hepatitis C. *Hepatology, 22*, 1050 – 1056. doi:10.1002/hep.1840220406

McAdam-Marx, C., McGarry, L.J., Hane, C.A., Biskupiak, J., Deniz, B., & Brixner, D.I. (2011). All-Cause and incremental per patient per year costs associated with chronic hepatitis C virus and associated liver complications in the United States: a managed care perspective. *Journal of Managed Care Pharmacy, 17*, 531-546.

Medical Device Quality Systems Regulation, 21 C.F.R. §820, (2012).

Medlin, D.L., & Bazerman, M.H. (1999). Broadening behavioral decision research: multiple levels of cognitive processing. *Psychonomic Bulletin & Review, 6*, 533-546. http://www.springer.com/psychology/cognitive+psychology/journal/13423

Mellors, J.W., Rinaldo, C.R., Gupta, P., White, R.M., Todd, J.A., & Kingsley, L.A. (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science, 272*, 1167-1170. doi: 10.1126/science.272.5265.1167

Merck Sharp & Dohme Corp., a subsidiary of Merck& Co., Inc. (2011). Victrelis. Highlights of prescribing information. Whitehouse Station, NJ. Retrieved from http://www.merck.com/product/usa/pi_circulars/v/victrelis/victrelis_pi.pdf

Messonnier, M., & Meltzer, M. (2003).Cost-benefit analysis. In Haddix, A.C, Teutsch, S.M, & Corso, P.S. (Eds.) *Preventive effectiveness* (pp 127 – 155). New York, NY: Oxford University Press, Inc.

Meltzer, D.O., Hoomans, T., Chung, J., & Basu, A. (2011). Minimal modeling approaches to value of information analysis for health research. Medical Decision Making, *31*, E1-E19. doi:10.1177/0272989X11412975

Meyers, E.R., McCrory, D.C., Subramanian, S., McCall, N., Nanda, K., Datta, S., & Matchar, D. (2000). Setting the target for a better cervical screening test: characteristics of a cost-effective test for cervical neoplasia screening. *Obstetrics and Gynecology*, *96*, 645-652. Retrieved from http://journals.lww.com/greenjournal/pages/default.aspx

Minshall, M. (2008). Raising the bar in the USA: the impact of heightened awareness of the need for health economic data in the absence of a regulatory mandate. *ISPOR Connections, January/February 2008*, 9-12. Retrieved from http://www.ispor.org/news/IConnections.aspx

Naik, A.D., & Petersen, L.A. (2009). The neglected purpose of comparative-effectiveness research. *New England Journal of Medicine, 360*, 1929-1931. Retrieved from http://www.nejm.org

National Cholesterol Education Program Laboratory Standardization Panel (1988). Current status of blood cholesterol measurement in clinical laboratories in the United States. *Clinical Chemistry, 34*, 193-220. Retrieved from http://www.clinchem.org

National Institute for Clinical Excellence – NICE. (2008). Guide to the methods of technology appraisal. National Institute for Clinical Excellence, London, UK. Retrieved, April 7, 2009, from http://www.nice.org.uk.

National Institute for Clinical Excellence – NICE. (2010). Guide to the multiple technology appraisal process National Institute for Clinical Excellence, London, UK. Retrieved, February 4, 2012, from http://www.nice.org.uk.

National Institute for Clinical Excellence – NICE. (2010). Guide to the single technology appraisal process. National Institute for Clinical Excellence, London, UK. Retrieved, February 4, 2012, from http://www.nice.org.uk.

National Institute of Health. (2002). NIH consensus statement on management of hepatitis C: 2002. *NIH Consensus and State -of-the-Science Statements*, 19 (3), 1-46. http://consensus.nih.gov/2002/2002HepatitisC2002116html.htm

National vital statistics report, *58* (10) 7Dec 2011 p1. Retrieved from http://www.cdc.gov/nchs

Nelson, W., Moser, R., Gaffey, A., & Waldron, W. (2009). Adherence to cervical cancer screening guidelines for U.S. women aged 25-64: Data from the 2005 health information national trends survey (HINTS). *Journal of Women's Health*, *18*, 1759-1768. doi: 10.1089=jwh.2009.1430

Neumann, P.J., Greenberg, D., Olchamski, N. Stone, P.W., & Rosen, A.B. (2005). Growth and Quality of the Cost-Utility Literature, 1976 – 2001. *Value in Health, 8,* 3-9. doi:10.1111/j.1524-4733.2005.04010.x

Oboler, S.K, Prochazka, A., Gonzales, R., Xu, S., & Anderson, R.J. (2002). Public expectations and attitudes for annual physical examinations and testing. *Annals of Internal Medicine*, 136 (9), 652-659. http://www.acponline.org/journals/annals/annaltoc.htm

Op De Beeck, A., & Dubuisson, J. (2003). Topology of hepatitis C virus envelope glycoproteins. *Reviews in Medical Virology*, *13*, 233-241. doi: 10.1002/rmv.391

Pawlotsky, J.-M. (1997). Measuring hepatitis C viremia in clinical samples: can we trust the assays? *Hepatology, 26*, 1-3. doi: 10.1002/hep.510260131

Paxton, A. (2011). HCV superstars spotlight viral load testing. *CAP Today.* Retrieved from http://www.cap.org. accessed 6/30/2013

Perrottet, N., Manuel, O., Lamoth, F., Venetz, J-P., Sahli, R., Decosterd, L.A.,…Meylan, P. (2010). Variable viral clearance despite adequate ganciclovir plasma levels during valgancyclovir treatment for cytomegalovirus disease in D+/R- transplant recipients. *BMC Infectious Diseases*, *10*, 2. doi:10.1186/1471-2334-10-2

Phelps, C.E., & Mushlin, A.I. (1988). Focusing technology using medical decision theory. *Medical Decision Making*, *8*, 279-289. doi: 101177/0272989X8800800409

Plebani, M. (2004). What information on quality specifications should be communicated to clinicians, and how? *Clinica Chimica Acta, 346*, 25-35. Retrieved from http://dx.doi.org.proxy.library.vcu.edu/10.1016/j.cccn.2004.03.019

Pembrey, L., Newell, M-L., Tovo, P-A., van Drimmelen, H., Quinti, I., Furlini, … Alfarano, A. (2003). Inter-laboratory comparison of HCV-RNA assay results: implications for multi-centre research. *Journal of Medical Virology, 69*, 195-201. doi: 10.1002/jmv.10283

Pisani, G., Cristiano, K., Marino, F., Luciani, F., Bisso, G.M., Mele, C., Adriani, D., Gentili, G., et al. (2009). Quantification of hepatitis C. Virus (HCV) RNA in a multicenter study: implications for management of HCV genotype 1-infected patients. *Journal of Clinical Microbiology, 47,* 2931-2936. doi:10.1128/JCM.00532-09

Podzorski, R.P. (2002). Molecular testing in the diagnosis and management of hepatitis C virus infection. *Archives of Pathology and Laboratory Medicine, 126*, 285 – 290. Retrieved from http://www.archivesofpathology.org

Poordad, F., Reddy, K.R., Martin, P. (2008). Rapid virological response: a new milestone in the management of chronic hepatitis C. *Clinical Infectious Diseases, 46,* 78-84. doi: 10.1086/523585

Price, C.P., & Christenson, R.H. (2007). Preface to the Second Edition. In C. P. Price, & R. H. Christenson (Eds.), *Evidence-based laboratory medicine. Principles, practice and outcomes, second edition* (pp. ix - xi). Washington, DC: AACC Press.

Pyenson, B., Fitch, K., & Iwasaki, K. (2009). Consequences of hepatitis C virus (HCV): Costs of a baby boomer epidemic of liver disease. A report commissioned by Vertex Pharmaceuticals Incorporated. Milliman, Inc, New York, NY. http://publications.milliman.com/research/health-rr/pdfs/consequences-hepatitis-c-virus-RR05-18-09.pdf. Accessed 31 March 2013.

Rajendra, A., & Wong, J.B. (2007). Economics of hepatitis B and hepatitis C. *Journal of Hepatology, 47,* 608-617. doi:10.1016/j.jhep.2007.07.016

Ramsey, S.D., Veenstra, D.L., Garrison, L.P., Carlson, R., Billings, P., Carlson, J. & Sullivan, S. (2006). Toward evidence-based assessment for coverage and reimbursement of laboratory-based diagnostic and genetic tests. *The American Journal of Managed Care*, *12*, 197-202. Retrieved from http://www.ajmc.com/publications

Sanchez, L. A., & Lee, J.L. (2000) Applied pharmacogenomics: modeling data from internal and external resources. *American Journal of Health-System Pharmacists, 57*, 146-155. doi:10.1053/j.gastro.2006.05.016

Saldanha, J., Heath, A., Lelie, N., and the WHO collaborative study group. (1999). Establishment of the first international standard for nucleic acid amplification technology assays for HCV RNA. *Vox Sanguinis, 76*,149-158. doi: 10.1046/j.1423-0410.1999.7630149.x

Schirm, J., van Loon, A.M., Valentine-Thon, E., Klapper, P.E., Reid, J., & Cleator, G.M. (2002). External quality assessment program for qualitative and quantitative detection of hepatitis C virus RNA in diagnostic virology. *Journal of Clinical Virology, 40*, 2973-2980. doi: 10.1128/JCM.40.8.2973–2980.2002

Scott, M.A., & Rafferty, J.P. (2007). Evidence –based laboratory medicine: health economics, reimbursement, and the value proposition. *In Evidence-Based Laboratory Medicine. Principles, Practices and Outcomes,* 2nd Edition. Price, C.P. & Christensen, R.H. Eds. AACC Press, Washington, D.C.

Secretary's Advisory Committee on Genetics, Health and Society (SACGHS) (2006). Coverage and Reimbursement of Genetic Tests and Services. Retrieved from http://oba.od.nih.gov/SACGHS/sacghs_documents.html

Seeff, L. (1999). Natural history of hepatitis C. *The American Journal of Medicine, 107* (6, supplement 2), 10S-15S. Retrieved from http://dx.doi.org.proxy.library.vcu.edu/10.1016/S0002-9343(99)00374-5

Shah, B.B., & Wong, J.B. (2006) The economics of hepatitis C. *Clinics in Liver Disease, 10*, 717-734. doi:10.1016/j.cld.2006.08.026

Sharafi, H., & Alavain, S.M. (2011). IL28B polymorphisms, explanation for the different responses to therapy in hepatitis C patients. *Hepatitis Monthly, 11*, 958-959. doi: 10.5812/kowsar.173514X.794

Simmonds, P., Bukh, J., Combet, C., Deléage, G., Enomoto, N., Feinstone, S., ... Widell, A. (2005). Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology*, *42*, 962-973. doi: 10.1002/hep.20819

Singer, M.E., & Younossi, Z.M. (2001). Cost effectiveness of screening for hepatitis C virus in asymptomatic, average-risk adults. *The American Journal of Medicine, 111*, 614-621. Retrieved from http://dx.doi.org.proxy.library.vcu.edu/10.1016/S0002-9343(01)00951-2

Smart, A. (2006). A multi-dimensional model of clinical utility. *International Journal for Quality in Healthcare, 18*(5) 377-382. doi: 10.1093/intqhc/mzl034

Stephens, J.M., Handke, B., & Doshi, J.A. on behalf of the HTA Principles Working group, part of the International Society for Pharmacoeconomics and outcomes Research (ISPOR) HTA Special Interest Group. (2010). A Survey of HTA Research Methods and Trends in Europe Poster Presentation. ISPOR 13th Annual European Congress Poster Presentation, Prague, Czech Republic. Retrieved from http://www.ispor.org/sigs/HTA_MD/2010_HTATrendsinEurope.pdf

Struerer, J., Fischer, J.E., Bachmann, L.M., Koller, M., & ter Reit, G. (2002). Communicating accuracy of tests to general practitioners: a controlled study. *British Medical Journal*, *324* (7341), 824- 826. doi: 10.1136/bmj.324.7341.824

Sulkowski, M. (2008). New data on currently available HCV therapies and their impact on initial treatment strategies. *Hepatitis Annual Update 2008*. Retrieved from http://www.clinicaloptions.com/ccohep2008

Sy, T.,& Jamal, M. M. (2006). Epidemiology of hepatitis C virus (HCV) infection. *International Journal of Medical Sciences*, 3(2), 41-46. Retrieved from http://www.medsci.org

Terrault, N.A., Pawlotsky, J-M., McHutchison, Anderson, F., Krajden, M., Gordon, S., … Friesenhahn, M. (2005). Clinical utility of viral load measurements in individuals with chronic hepatitis C infection on antiviral therapy. *Journal of Viral Hepatitis, 12*, 465-472. doi:10.1111/j.1365-2893.2005.00615.x

Thaler, R.H. (1991). The psychology of choice and the assumptions of economics. In A. Thaler (Ed.), *Quasi rational economics* (pp.137-167). New York, NY: Russell Sage Foundation. (Reprinted from *Laboratory experiments in economics: six points of view*. By Alvin Roth, Ed., 1987, New York, NY: Cambridge University Press)

The Lewin Group (2005). The value of diagnostics, innovation, adoption and diffusion into healthcare. A report prepared for AdvaMed. Retrieved from http://www.lewin.com

The Lewin Group (2008, May). *Laboratory medicine: A national status report.* Prepared for Division of Laboratory Systems, National Center for Preparedness, Detection and Control of Infectious Diseases, Center for Disease Control and Prevention. Wolcott, J., Schwartz, A., Goodman, C. Retrieved from http://www.lewin.com

The Lewin Group (2009, September). *The Value of Laboratory Screening and Diagnostic tests for prevention and health care Improvement.* Prepared for American Clinical laboratory Association and Advanced medical technology Association (Advamed). Wolcott, J., & Goodman, C. Retrieved from http://www.lewin.com

Thygessen, K., Alpert, J.S., Jaffe, A.S., Simoons, M.L., Chaitman, B.R., & White, H.D. the Writing Group on behalf of the Joint ESC/ACCF/AHA/WHF Task Force for the Universal Definition of Myocardial Infarction (2012). Third universal definition of mydocardial infarction. *European Heart Journal.* Advance online publication. doi:10.1093/eurheartj/ehs184

Tramarin, A., Gennaro, N., Compostella, F.A., Gallo, C., Wendelaar Bonga, L.J., & Postma, M.J. (2008). HCV screening to enable early treatment of hepatitis C: a mathematical model to analyze costs and outcomes in two populations. *Current Pharmaceutical Design, 14,* 1655 – 1600. doi: http://dx.doi.org.proxy.library.vcu.edu/10.2174/138161208784746833

Tunis, S.R. (2004). Economic analysis in healthcare decisions. *The American Journal of Managed Care*, 10(5), 301-304. Retrieved from http://www.ajmc.com

United States Food and Drug Administration (2009). 21CFR807. Establishment registration and device listing for manufacturers and initial importers of devices. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=807&showFR=1 , accessed 7/30/2010.

United States Food and Drug Administration (2009). 21CFR814 Premarket approval of medical devices. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=814&showFR=1. accessed 7/30/2010.

United States Food and Drug Administration (2009). 21CFR860 Medical device classification procedures. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=860. accessed 7/30/2010.

Valentine-Thon, E., van loon, A.M., Schirm, J., Reid, J., Klapper, P.E., & Cleator, G.M. (2001). European proficiency testing program for molecular detection and quantification of hepatitis B virus DNA. *Journal of Clinical Microbiology, 39,* 4407- 4412. doi: 10.1128/ JCM.39.12.4407-4412.2001

Valsamakis, A. (2007) Molecular testing in the diagnosis and management of chronic hepatitis B. *Clinical Microbiology Reviews*, *20*, 426 – 439. doi: 10.1128/CMR.00009-07

Van Walraven, C., & Naylor, C.D. (1998). Do we know what inappropriate laboratory utilization is? *JAMA, 280*, 550- 559. doi:10.1001/jama.280.6.550

Van Walraven, C., & Raymond, M. for the network of Eastern Ontario medical laboratories (NEO-MeL). (2003). Population –based study of repeat laboratory testing. *Clinical Chemistry, 49*, 1997-2005.

Vertex Pharmaceuticals Incorporated. (2011). Incivek. Highlights of prescribing information. Cambridge, MA. Retrieved from http://pi.vrtx.com/files/uspi_telaprevir.pdf

World Health Organization (2002). Annex II. Recommendations for the preparation, characterization and establishment of international and other biological reference standards (revised 2004). WHO Technical report series, No. 932, Geneva, World Health Organization. Retrieved from http://www.who.int/bloodproducts/publications/Final_Prep_Stds_Nov2004.pdf

World Health Organization Global Alert and Response Website: Hepatitis C. http://www.who.int/csr/disease/hepatitis/whocdscsrlyo2003/en/index3.html accessed 2/20/2012.

Weinstein, M.C., O' Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C., & Luce, B.R. (2003). Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR task force on good research practices-Modeling studies. *Value in Health, 6*, 9-17. doi:10.1046/j.1524-4733.2003.00234.x

Weinstein, M.C. (1980). *Clinical decision analysis*. Philadelphia: WB Saunders.

Weinstein, M.C., Toy, E.L., Sandberg, E.A., Neumann, P.J., Evans, J.S., Kuntz, K.M., … Hammitt, J.K. (2001). Modeling for health care and other policy decisions: Uses Roles, and validity. *Value in Health*, *4*, 348 – 361. doi: 10.1046/j.1524-4733.2001.45061.x

Wong, J.B. (2006). Hepatitis C cost of illness and considerations for the economic evaluation of antiviral therapies. *Pharmacoeconomics, 24*, 661-672. Retrieved from http://adisonline.com/pharmacoeconomics

Wu, A.H.B., & Christensen, R.H. (2013). Analytical and assay issues for use of cardiac troponin testing for risk stratification in primary care. *Clinical Biochemistry*. Retrieved from http://dx.doi.org/10.1016.j.clinbiochem.2013.04.013

Yokota, F., & Thompson, K.M. (2004). Value of information literature analysis: a review of applications in health risk management. *Medical Decision Making, 24,* 287-298. doi:10.1177/0272989X04263157

You, J.H.S., Chan, F.W.H., Wong, R.S.M., & Cheng, G. (2004). The potential clinical and economic outcomes of pharmacogenetics-oriented management of warfarin therapy- a decision analysis. *Thrombosis and Haemostasis, 92*, 590-597. doi: 10.1160/TH04-03-0161

Young, J.M., Glasziou, P., & Ward, J.E. (2002). General practitioners' self-rating of skills in evidence based medicine: validation study. *British Medical Journal, 324*, 950-951. doi: 10.1136/bmj.324.7343.950

Yen-Lieberman, B., Brambilla, D., Jackson, B., Bremer, J., Coombs, R., Cronin, M., … Reichelderfer, P. (1996). Evaluation of a quality assurance program for quantitation of human immunodeficiency virus type 1 RNA in plasma by the AIDS clinical trials group virology laboratories. *Journal of Clinical Microbiology, 34,* 2695 -2701. Retrieved from http//journals.asm.org

Zaaijer, H.L., Cuypers, H.T.M., Reesink, H.W., Winkel, I.N., Gerken, G., & Lelie, P.N. (1993). Reliability of polymerase chain reaction for detection of hepatitis C virus. *The Lancet, 341,* (8847) 722- 725.

Zeuzem, S., Heathcote, E.J., Shiffman, M.L., Wright, T.L., Bain, V.G., Sherman, M., … Sederati, F. (2003). Twelve weeks of follow-up is sufficient for the determination of sustained virologic response in patients treated with interferon α for chronic hepatitis C. *Journal of Hepatology, 39*, 106-111. Retrieved from http://www.elsevier.com/locate/jhep

Zeuzem, S, Pawlotsky, J.-M., Lukasiewicz, E., von Wagner, M., Goulis, I., Lurie, Y., … Neumann, A. U. (2005). International, multicenter, randomized, controlled study comparing dynamically individualized versus standard treatment in patients with chronic hepatitis C. *Journal of Hepatology*, *43*, 250-257. Retrieved from www.elsevier.com/locate/jhep

Appendix A: HCV VL Survey

Text of Survey: HCV-1 Triple Combination Therapy- Incivek.

Consent

We are conducting a study of laboratory tests and their impact on medical decisions. If you are willing to participate, we would like you to complete a brief, confidential survey. The principal investigator on this survey is William Korzun, PhD, Associate Professor, Department of Clinical Laboratory Sciences, Virginia Commonwealth University.

Time Required: We expect the survey will take less than 10 minutes.

Risks: This is a confidential, individual based survey. There are no reasonably foreseeable risks or discomforts involved in taking part in this survey.

Compensation: There is no compensation for completing this survey. However if you would like a summary of the results, please send a separate email with the title "HCV Survey Summary" to madejrm@mymail.vcu.edu. A summary will be sent to you in 2013.

Confidentiality: Your responses are anonymous and you cannot be identified. Individual responses to the survey questions will be kept confidential, and the output of the study will be based on aggregation of the data collected from participants.

Participation: In the event you choose not to participate, you will not be penalized.

Study contact for questions: If you have questions, concerns, or complaints, please feel free to email Dr. William Korzun:

WJKORZUN@VCU.EDU or Roberta Madej at madejrm@mymail.vcu.edu.

If you have any questions about your rights as a participant in this study, you may contact:
Office for Research
Virginia Commonwealth University
800 East Leigh Street, Suite 113
P.O. Box 980568
Richmond, VA 23298
To proceed with the survey please, choose "I consent and agree to participate". If you do not wish to take the survey, please choose "I do not wish to take this survey".
Thank you for your time.

Please choose one of the following:

○ I consent and agree to participate

○ I do not wish to take this survey

Introduction: HCV Triple Combination Therapy with Incivek (Telaprevir)

Welcome to the Survey. Thank you for your participation. There are no correct or incorrect answers. The best estimate of what your decision would be in the following scenario is requested.

The purpose of this survey is to evaluate responses to quantitative lab test results presented in different ways by the clinical laboratory, using quantitative HCV RNA (HCV viral load) as a model. You will be presented with a description of a patient followed by laboratory results and will be asked to choose the responses closest to what your decisions would be for each result presented. The results will be presented to you twice – each in a different way.

The decision for this "patient" is the treatment management decision made at Treatment Week (TW) 12 for patients with HCV genotype 1 disease who have been on triple combination therapy with Incivek (telaprevir), peginterferon alpha and ribavirin therapy. According to the AASLD guidelines published in October 2011, the recommended course is to treat with the three drugs for 12 weeks, then to continue with peginterferon alfa and ribavirin for an additional 12- 36 weeks, or discontinue all therapy. Important to this decision are the results of quantitative HCV Viral Load (HCV VL) testing at TW 12.

At the 12week decision point, the AASLD guidelines recommend the following:

"Patients without cirrhosis treated with telaprevir, peginterferon, and ribavirin, whose HCV RNA level at weeks 4 and 12 is undetectable should be considered for a shortened duration of therapy of 24 weeks (Class 2a, Level A).
Treatment with all three drugs (telaprevir, peginterferon alfa, and ribavirin) should be stopped if the HCV RNA level is >1,000 IU/mL {3Log10 IU/ml} at treatment weeks 4 or 12 and/or detectable at treatment week 24 (Class 2a, Level B)."

Ghany et al. (2011) Hepatology, 54, 1433-1444.

Description of case and results – Integers

Patient: middle aged with chronic HCV genotype 1a. No cirrhosis.
Your previously untreated (for HCV) patient has been on triple combination therapy for 12 weeks with: Incivek (telaprevir) 750mg three times a day, together with peg interferon alfa weekly and weight based ribavirin. Over the 12 weeks your patient has experienced the following side effects: nausea, moderate to severe fatigue, joint and muscle aches, irritability and headaches. Hemoglobin has gone from 12.3 to 9.1 gm/dl. The patient is employed fulltime and

has missed approximately 10 days of work in the last 12 weeks due to the fatigue, nausea and headaches.

You are reviewing the Treatment Week 12 HCV viral load results in order to make a decision regarding the continuing management of this patient. The quantitative HCV RNA test result for this patient performed before the start of therapy (week zero, baseline) was 560,000 IU/ml.

Consider each of the following 12 week HCV RNA results individually with regards to this patient. For each, please indicate what your most likely decision would be regarding this patient's therapy?

	Continue Rx with dual therapy	Discontinue Rx	Repeat the HCV viral load test
31,620 IU/ml	○	○	○
1,950 IU/ml	○	○	○
1,000 IU/ml	○	○	○
560 IU/ml	○	○	○
100 IU/ml	○	○	○

Description of case and results – Integers with CI

Patient: middle aged with chronic HCV genotype 1a. No cirrhosis.
Your previously untreated (for HCV) patient has been on triple combination therapy for 12 weeks with: Incivek (telaprevir) 750mg three times a day, together with peg interferon alfa weekly and weight based ribavirin. Over the 12 weeks your patient has experienced the following side effects: nausea, moderate to severe fatigue, joint and muscle aches, irritability and headaches. Hemoglobin has gone from 12.3 to 9.1 gm/dl. The patient is employed fulltime and has missed approximately 10 days of work in the last 12 weeks due to the fatigue, nausea and headaches.

You are reviewing the Treatment Week 12 HCV viral load results in order to make a decision regarding the continuing management of this patient. The quantitative HCV RNA test result for this patient performed before the start of therapy (week zero, baseline) was 560,000 IU/ml (95% CI: 286,000 - 1,107,000 IU/ml).

Consider each of the following 12 week HCV RNA results individually with regards to this patient. For each, please indicate what your most likely decision would be regarding this patient's therapy?

	Continue Rx with dual therapy	Discontinue Rx	Repeat the HCV viral load test
31,620 IU/ml (95% CI: 16,060 - 62,230 IU/ml)	○	○	○
1,950 IU/ml (95% CI: 990 - 3,800 IU/ml)	○	○	○
1,000 IU/ml (95% CI: 510 - 1,970 IU/ml)	○	○	○
560 IU/ml (95% CI: 290 - 1,110 IU/ml)	○	○	○
100 IU/ml (95% CI: 50 – 200 IU/ml)	○	○	○

Question regarding the utility of the test results with CI

One set of test results you reviewed contained the measurement uncertainty of the result expressed as the 95% confidence interval (CI). The test results with the 95% confidence intervals:

○ Provided relevant added information to the test result.

○ Detracted from or confused the test result.

○ Neither added to nor detracted from the test result.

Participant demographics

Please answer the following questions regarding your professional demographics and practice.

Where do you currently practice/ work?

○ Africa

○ Australia/ New Zealand

○ Canada

○ Europe

○ Latin America/ South America

○ United States

Do you manage the treatment of Hepatitis C patients?

○ Yes

○ No

HCV patient managing professionals (For those choosing "Yes" to the question "Do you manage the treatment of HCV patients?)

Which of the following best describes your profession?

○ Physician

○ Physician Assistant

○ Nurse Practitioner

○ Nurse

○ Other (please specify)

How many YEARS have you been practicing medicine? _____

Approximately how many Hepatitis C patients do you see per MONTH? _____

Which most appropriately describes your practice field?

○ General Medicine

○ Internal Medicine

○ Infectious Diseases

○ Hepatology

○ Gastroenterology

○ Urology

○ Pathology

○ Family Medicine

○ Pediatrics

○ Oncology

○ Other (please specify)

Please select the guidelines that you use when managing Hepatitis C patients (select all that apply):

○ American Association for the Study of Liver Diseases (AASLD)

○ American College of Gastroenterology (ACG)

○ American Gastroenterological Association (AGA)

○ CDC

○ EASL

○ Information from the pharmaceutical company

○ Institutional guidelines

○ Other (please specify)

Would you like to receive the 95% confidence intervals of laboratory values reported with patient results?

○ Yes

○ No

○ I already receive 95% confidence intervals from my testing laboratory

○ I don't know

Do you use quantitative HCV RNA test results (HCV viral load tests) in the treatment management of your patients with Hepatitis C?

○ Yes

○ No

(For "No") **Why not? ___**

(For "Yes" to the use of quantitative HCV RNA results) **Do you send patients / samples to the same testing laboratory for the pretreatment and 12 week quantitative HCV RNA ?**

○ Sometimes,

○ Most of the time,

○ Not at all

○ It is not within my control

(For "Yes" to the use of quantitative HCV RNA results) **Do you specify that the identical test method be used for the pretreatment and 12 week quantitative HCV RNA tests?**

○ Sometimes,

○ Most of the time,

○ Not at all

○ It is not within my control

(For "Yes" to the use of quantitative HCV RNA results) **Please choose the type of HCV viral load results that are closest to the test results you receive from your testing laboratory:**

○ Integers (i.e. 1,000,000 IU/ml)

○ Log_{10} (i.e. 6.0 Log_{10} IU/ml)

○ Both integers and Log_{10} values are provided

○ I am not sure

Non HCV Managing professionals (For those choosing "No" to the question, "Do you manage the treatment of HCV patients?)

Which of the following best describes your profession?

○ Physician

○ Pathologist

○ Non-Pathologist Laboratory Director

- ○ Laboratory Professional (Non Pathologist/ Non Director)
- ○ Pharmacist
- ○ Other (please specify)

Thank you for the time that you have spent in taking this survey. If you would like a copy of the results please send a separate email with the title "HCV Survey Summary" to madejrm@mymail.vcu.edu.

Vita

Roberta Marie Madej was born in San Francisco, CA on July 20, 1955 (Roberta Marie Bianco). She is an America Citizen. She completed a Bachelor of Science in Microbiology/ Medical Technology from San Jose State University in 1977. She completed a Master of Science in Clinical Laboratory Sciences from San Francisco State University in 1988 where she received the graduate student award for distinguished achievement. She completed a Masters of Business Administration from St. Mary's College University in 1992.

Roberta has worked in several biotech companies in the area of molecular diagnostics since 1980, including Cetus, Chiron, Roche, Tethys and Cepheid. Her roles have been in research, development, quality, scientific affairs and technical support. She has also maintained part time positions in clinical laboratories in the San Francisco East Bay area since 1975.