Application of Near-Infrared Spectroscopy and ChemoInformatic Analyses to Biological Fluids for the Assessment of Maternal and Fetal Health Status

Javier E. Sanchez-Galan

Doctor of Philosophy

Division of Experimental Medicine, Department of Medicine
Faculty of Medicine

McGill University

Montreal, Quebec

December 2014

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Doctor of Philosophy

DEDICATION

To my mother, Osmila.

"I learned this, at least, by my experiment: that if one advances confidently in the direction of his dreams, and endeavors to live the life which he has imagined, he will meet with a success unexpected in common hours. He will put some things behind, will pass an invisible boundary; new, universal, and more liberal laws will begin to establish themselves around and within him; or the old laws be expanded, and interpreted in his favor in a more liberal sense, and he will live with the license of a higher order of beings."

Henry David Thoreau, Walden (1854)

ACKNOWLEDGEMENTS

First, I would like to thanks my supervisors for their support, guidance and overall patience, Dr. David Burns for helping me develop my scientific intuition and ingenuity, and Dr. Kristine Koski for teaching me scientific rigor and my PhD. committee for helping me in my transition into this challenging interdisciplinary project.

I gratefully thank the government of Panama who gave me full financial support to pursue my master's and doctoral studies through the *Professional Excellence Scholarship Program* (2007-2012) organized by IFARHU and SENACYT. I also want to acknowledge the funds I received through the *James P Kenney Award* (2012-2013) and the NSERC funding I received during the 2013-2014 period.

My time McGill would have not been the same without my academic colleagues in the Blanchette's lab (Mathieu L-A, Mathieu R, Ethan, Glenn), Burns lab (Andrien, Andrew, Peter) and Koski-Scott's lab (Lisa, Zsofia, Hilary, Doris, Carli). My list of Montreal friends is way too long, but I still want to mention a few key people that helped me keep my (in)sanity: Jesus, Fannie, Fred, Eleazar, Mauser and Elisa.

My return home and the final writing of this document would not have been as easy without the help of my mentors: Prof. Victor Lopez, Dr. Rodriguez, Dr. Alvarez, and colleagues and co-workers at UTP-Tocumen: Grimaldo, Boris and Ubaldo; thank you all for the insightful discussions (good luck in your doctoral studies!).

Lastly, I would like to thank my parents and family for their full support and encouragement in all my personal, musical and professional endeavours over the years.

ABSTRACT

Spectroscopy methods are usually employed for the analysis biological fluids, to assess the presence of specific metabolites that are present and that might be associated with adverse health conditions. Among these vibrational methods Near-Infrared Spectroscopy (NIRS) is of great importance since it provides rapid and accurate way to measure metabolites present in biological fluids and tissues. Most NIRS research has been focused in determining changes in biological fluids and tissue in the "medical spectral window" or the 600 to 1100nm region of the electromagnetic spectrum. However, much less research has focused in the analysis of biological fluids in the longer wavelengths of the NIR region in the 1100 to 2500nm region and the analysis of metabolic alternations that could arise during pregnancy and lactation. The main objective of this thesis was to determine if NIRS analysis of two biological fluids, serum and amniotic fluid in the 1600 to 2400nm region could provide insight into maternal-fetal conditions including: maternal infections and diet deficiencies during pregnancy and lactation and infant size. In this thesis a chemoinformatics method for the analysis of long wave NIR spectral signals was developed using multivariate statistical analysis (ANOVA) and was applied to either serum or amniotic fluid samples from either animal or human sources. Our results taken together show the possibility that long wave NIR spectroscopy can be used to identify candidate spectral regions that could be related to infection, diet and perinatal growth in the context of maternal-fetal health. These findings may provide a platform for the elucidation of novel spectral markers in two biological fluids that could be used in the research and development of clinical instrumentation for the assessment of maternal and fetal health.

ABRÉGÉ

Habituellement, afin d'évaluer la présence de métabolites spécifiques qui pourraient être associés avec des mauvaises conditions de santé, des méthodes de spectroscopie sont utilisées pour analyser les fluides biologiques.Parmi celles-ci, la spectroscopie proche infrarouge (NIRS) est d'une grande importance, car elle permet de mesurer rapidement et avec précision la présence de métabolites dans des fluides et tissus biologiques. La plupart des recherches qui ont été effectuées en NIRS ont eu pour objet de déterminer des changements dans les liquides biologiques et les tissus dans la "fenêtre medicale" dans la region située entre 600 et 1100 nm du spectre électromagnétique. Toutefois, peu d'études ont, jusqu'à présent, porté sur l'analyse de fluides biologiques dans les longueurs d'onde de la région NIR de la région 1100 à 2500nm et l'analyse des alternances métaboliques qui pourraient survenir pendant la grossesse et l'allaitement. L'objectif principal de cette thse était de déterminer si l'analyse NIRS de deux fluides biologiques, le sérum et le liquide amniotique, à la région entre 1600 et 2400 nm pourrait donner un aperçu des conditions materno-fœtales comme par exemple: infections maternelles et carences en alimentation pendant la grossesse et l'allaitement et la taille de l'enfant. Donc, une méthode de chémoinformatique pour l'analyse des signaux spectraux NIR à ondes longues a été développée en utilisant une analyse statistique multivariée (ANOVA) et a été appliquée à des échantillons du sérum et de liquide amniotique soit animal ou humain. Nos résultats démontrent que la spectroscopie NIR à ondes longues peut être utilisée pour identifier des régions spectrales potentiellement associés à divers traits de la santé materno-fœtale comme les infections, l'alimentation et la croissance périnatale. Ces résultats contribuent à fournir une plate-forme pour la découverte de nouveaux marqueurs spectraux pour deux fluides biologiques pouvant être utilisés

pour le développement d'instrumentation clinique servant à l'évaluation de la santé maternelle et fœtale.

CONTRIBUTION OF THE AUTHORS

This thesis comprises the full text and figures of three scientific articles, all of them in preparation for publication. These articles are listed below in the order they appear in the thesis. I am the first author of each of them, except the paper that is in the Appendix in which I am a second author.

- **Chapter 4:**

  Javier E. Sanchez-Galan, Maurice R. Odiere, Marilyn E. Scott, David H. Burns and Kristine G. Koski. "Metabolic Profiling of Maternal and Pup Serum using Near-Infrared Spectroscopy Identifies Distinct Patterns for Nematode Infection and Dietary Protein Deficiency". *Manuscript in preparation.*

  The design and implementation of the computational tools and analysis in this publication was done by me under supervision of Dr. David Burns and Dr. Koski. The mouse serum samples were collected by Maurice Odiere and analyzed using a near-infrared spectrometer by me under supervision of Dr. David Burns. The biological implication of the results was written by me under supervision of Dr. Kristine Koski and Dr. Marilyn Scott.

- **Chapter 5:**

  Javier E. Sanchez-Galan, Gary Luskey, David H. Burns and Kristine G. Koski. "NIR Metabolic Profiling of 2nd Trimester Amniotic Fluid Distinguishes Between Infants Subsequently Born Appropriate-for-Gestational Age (AGA)

and Large-for-Gestational Age (LGA)". *Manuscript in preparation.*

The design and implementation of the computational tools and analysis in this publication was done by me under supervision of Dr. David Burns and Dr. Koski. The amniotic fluid samples were collected by Dr. Gary Luskey and analyzed using a near-infrared spectrometer by me under supervision of Dr. David Burns. The biological implication of the results was written by me under supervision of Dr. Kristine Koski.

- **Chapter 6:**

  Javier E. Sanchez-Galan, Doris Gonzalez-Fernandez, Marilyn E. Scott, David H. Burns and Kristine G. Koski. "Use of Near-Infrared Spectroscopy for the Analysis of Infection, Nutrition and Fetal Growth for Pregnant Women in a Rural Community in Panama". *Manuscript in preparation.*

  The design and implementation of the computational tools and analysis in this publication was done by me under supervision of Dr. David Burns and Dr. Koski. The serum samples were collected by Doris Gonzalez-Fernandez and analyzed using a near-infrared spectrometer by me under supervision of Dr. David Burns. The biological implication of the results was written by me under supervision of Dr. Kristine Koski and Dr. Marilyn Scott.

- **Appendix :**

  Kristin M. Power, **Javier E. Sanchez-Galan**, Gary W. Luskey, Kristine G. Koski and David H. Burns. "Use of Near-Infrared Spectroscopic Analysis of

Second Trimester Amniotic Fluid to Assess Preterm Births". *Journal of Pregnancy, vol. 2011, Article ID 980985, 6 pages, 2011.*

Manuscript was written by Kristin M. Power under the supervision of Dr. David Burns and Dr. Kristine Koski. The implementation of the computational tools and the second analysis of the data was done by me under supervision of Dr. David Burns and Dr. Koski. The amniotic fluid samples were collected by Dr. Gary Luskey and analyzed using a near-infrared spectrometer by me under supervision of Dr. David Burns. Part of the biological implication of the results was written by me under supervision of Dr. Kristine Koski.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

λ Wavelength

AGA Appropriate for Gestational Age

ANOVA ANalysis Of VAriance

CCD Charge-Coupled Device

CRL Crown-to-Rump-Length

FGR Fetal Growth Restriction

FH:GA Fundal Height over Gestational Age ratio

FIR Far-Infrared

FT-IR Fourier Transform IR

GA Gestational Age

GI Gastrointestinal

IN Infected

IR Infrared

IUGR Intrauterine Growth Restriction

LBW Low Birth Weight

LED Light Emitting Diode

LGA Large for Gestational Age

MIR Mid-Infrared

MLR Multi Linear Regression

mRNA Messenger Ribonucleic Acid

MSC Multiplicative Scatter Correction

NIRS Near-Infrared Spectroscopy

NIR Near-Infrared

nm nanometers

PCA Principal Component Analysis

PD Protein Deficient

PLS Partial Least Squares

PS Protein Sufficient

SGA Small for Gestational Age

SOM Self-Organizing Maps

UV Ultraviolet

VIF Variance Inflation Factor

CHAPTER 1
Introduction: Overview, Rationale and Methodology

The field of molecular biology has changed considerably since the completion and publication of the first draft of human genome [148, 20]. That single event opened the doors for the study of all the constituents of the cell and the interactions between the proteins and nucleic acids within the nucleus. It also gave birth to the discipline of *genomics* or the study of the genetic make-up of an organism. Along with the birth of genomics, other research disciplines were created to focus in other aspects of the flow of genetic information that happens from DNA to RNA. For instance *transcriptomics*, puts its interest in the transcribed gene information into mRNA products and *proteomics* look at the proteins or the finally translated mRNA molecules. These three disciplines have been encompassed by the umbrella of *-omic* sciences.

Another discipline that has emerged as complement to the three main *-omic* sciences is *metabolomics*. Metabolomics is a diagnostic discipline that provides a snapshot of an organism's metabolic states through the systematic analysis of its metabolites. [58]. It centers around the study of small molecules <1500 Da, that are found in a cells coming from tissues and more importantly in biological fluids [162].

In general, metabolomics assesses the metabolic difference between biological systems that are either perturbed (case) and unperturbed (control or reference). It is based on the premise that different disease states will be reflected in differences in the concentration of small molecules or metabolites in biological fluids [86].

To some extent metabolomics can be seen as a first building blocks towards personalized medicine, in which correlations between metabolic perturbations which will not be done between an individual current state and a pooled sample from a study population, but between the individual's current state and a previous metabolic state. This four *-omic* sciences has been described as the main building blocks of *systems biology* [162, 13].

Traditionally, metabolomic studies of biological fluids rely on the use of Nuclear Magnetic Resonance spectroscopy (NMR) and Mass Spectrometry (MS), because these technologies provide a greater dynamic range and thus a more specific molecular fingerprint of the samples they analyze. However the equipment is prohibitively expensive, requires specialised training for its use, they are destructive methods that requires the use of reagents for fragmentation (denaturation or lysis) and are mostly used as a second step after a chromatography separation. These technologies also have specific downsides, for instance, the problem of reproducibility that is found for peaks in MS [142] and the technical fineness of NMR equipment, in which sample volumes determine the size of the magnets and the intensity of the magnetic field it requires to separate the molecules [134]. Although Near-Infrared Spectroscopy (NIRS) is not usually regarded as a metabolic profiling methodology as these other techniques described above, it provides a fairly stable signal, especially in a temperature controlled instrument, and it provides a wealth of information about the samples analyzed in a single experiment and even when working with very small volumes.

1.1 Use of Near-Infrared Spectroscopy

Clinical chemistry analysis of biological fluids can be divided into one of two categories: quantitative analysis or diagnostic analysis. The quantitative analysis is the measurement of the concentration of an analyte of interest in a given biological

fluid, whereas diagnostic analysis is the comparison of a known biomarker for which properties are related to a medical condition (sufficiency or insufficiency state). Near-infrared spectroscopy (NIRS) provides interesting advantages in the measurement of metabolic perturbations over other techniques, mainly because it provides a global view of all the analytes present in a biological fluid in a single scan of the sample. Also, it is valuable because the differences in its spectral signal, collected either as absorbance or as transmittance, at specific wavelength regions, can be related back to the chemical functional groups (moieties) that conform to the functional groups present in the biological fluids or samples we are analyzing.

1.1.1 Characteristic of the Near-Infrared Region

NIRS signals provides a global picture of the important functional groups related to metabolism. However, the task of determining these functional groups in experimental conditions is limited by three major factors:

1. Spectral vibrations measured in the NIR region are overtones of the fundamental rotational vibrations which are perceived in the Mid-IR spectral region (2500nm-25000nm).
2. Functional groups in the NIR region overlap highly among each making it hard to describe a single characteristic moiety to distinguish a condition in a group of samples.
3. Absorbance signals resulting from NIR analysis are rather small and measurements in biological samples are prone to error due to light scattering effects and temperature effects. Moreover, spectral signals can be partially attributed to changes in water structure due to solvent-biomolecule interactions and variation in pH. This changes are more evident in liquid samples when NIR measurements

are made around water absorbance peaks (the most prominent being around 1940nm).

1.2 Research Methodology

After doing a background literature search to asses the state of the art of the use of NIR for the analysis of biological fluids we noticed that only a handful of studies have looked at pregnancy or maternal-fetal health status by measuring NIR spectral signals biological fluids. Moreover, there was a smaller number that used amniotic fluid collected from pregnant mothers [82, 83]. In this regard, we took advantage of serum and amniotic fluid samples previously collected in our laboratory and to analyze them using NIR spectrometer setting we had available.

The main goal of this research project was to identify candidate spectral regions based on the NIR profiling of two biological fluids, that in our case: serum and amniotic fluid (AF), to be used for the assessment of maternal-fetal health. For this matter we proposed the use of computational and statistical approaches, namely, chemoinformatic techniques and the use of statistics for patterns that might help us differentiate between infectious, nutritional, growth and health conditions in pregnant women and their offspring.

The decision of using NIRS is based on five major premises:

1. Diseases states or abnormal conditions can change the metabolic composition of biological fluids. These metabolic changes can be related to changes in the molecular composition of biological fluids.

2. Molecular changes or shifts can be measured using NIR spectroscopy. More, specifically they can be measured as shifts of C-H, N-H and O-H functional groups present in the molecules of the biological fluids.
3. Molecular shifts can be uniquely quantified, thus being able to be represented by spectral fingerprints for each metabolic conditions.
4. Spectral fingerprints can be separated from one another by the use of multivariate statistics, with the help of chemoinformatic analyses. These separations could be later used to build an instrument that could be greatly beneficial as a point-of-care solution.

1.2.1 NIR Spectral Regions of Interest

We propose to analyze the changes in the NIR spectral signals in serum and amniotic fluid samples, and by taking into account the different metabolic conditions and metabolic systems from where the samples were collected, to the develop a statistical framework to determine differences among these conditions.

For the elucidation of these spectral profiles we will focus our analysis to the first overtone region and combination bands regions in the 1600 to 2400nm wavelength region of the NIR region of the electromagnetic spectrum, where 1st overtone CH, SH, 2nd overtone C=O and combinations of OH, NH, NH+OH, CH+CH and CH+CC are known to have higher absorbance.

These regions were selected because it may play an important biological role in fetal development that could be later associated with one or more functional groups between 1600-2400nm our region of interest according to the 'Atlas of interpretative NIR Spectroscopy' [163]. In terms of the specific regions we selected 3 in the NIR 1st overtone bands: Region 1 (1650nm-1700nm, spanning 6 wavelengths), Region

2 (1701nm-1750nm, spanning5 wavelengths) and Region 3 (1850nm-1899nm, spanning 6 wavelengths). Also, 5 regions were selected in the combination bands: Region 4 (2050nm-2099nm, spanning 6 wavelengths), Region 5 (2100nm-2130nm, spanning 4 wavelengths), Region 6 (2130nm- 2200nm, spanning 8 wavelengths), Region 7 (2275nm-2310nm, spanning 5 wavelengths), Region 8 (2311nm-2370nm, spanning 8 wavelengths).

1.2.2 Context of the Proposed Research

Our research will address questions in the context of maternal-fetal health. In metabolic terms, during the pregnancy both the mother and the fetus are part of a single system and the health of one directly affects the health of the other. At the moment of birth, the infant becomes its own system and the metabolic relationship between infant and mother is mostly via the breast milk in the lactation phase the maternal. Once lactation is done the relationship between mother and infant is indirect and reflected through the environment and diet (both mother and child share living spaces and most likely will eat the same food). In this thesis we are not taking into account other influences that might have an impact in pregnancy, such as: genetics of the parents, nor

1.2.3 Analytical Methodology

For this thesis we developed an analytical methodology, shown in Figure 1–1, that we used for the NIR spectroscopy analysis of biological samples and the interpretation of the results. In this methodology we first collected the spectra for each sample using NIR spectroscopy in the 1600nm to 2400nm region. Then we applied *signal processing* techniques for cleaning up the spectral signals from systematic errors by smoothing and mean-centering. Moreover, we applied an *spectral region integration*

approach analyzing the 8 spectral regions previously identified. We made use of non-parametric statistics to determine the differences in these 8 spectral regions and their ratios, namely Mann-Whitney U-Test and Kruskal-Wallis ANOVA test. Finally, we interpreted the significant difference using an *Interpretive NIR spectroscopy* using information provided in the Atlas of Near-Infrared Spectroscopy [163] to relate our findings to biological facts regarding maternal-fetal health.

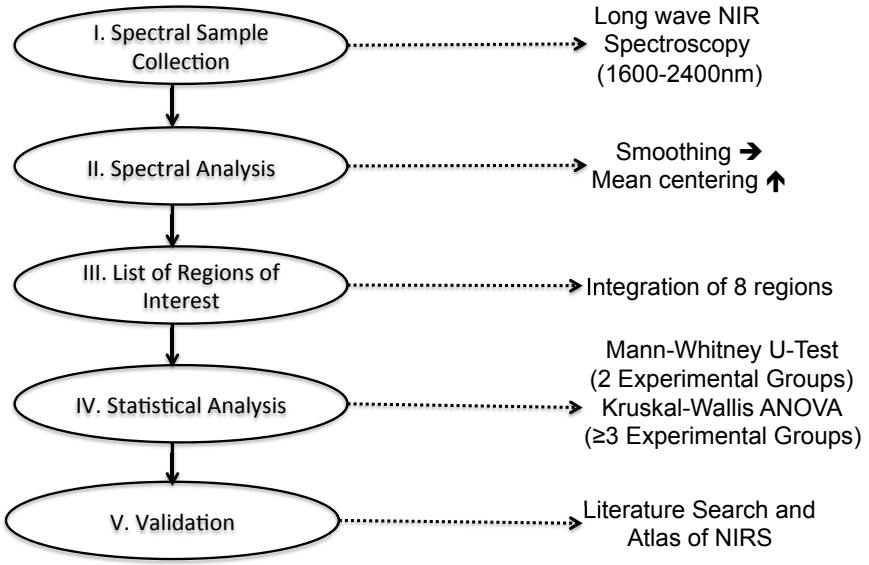


Figure 1–1: *NIR spectral analysis framework*

1.3 Proposed Structure of the Thesis

Chapter 2 introduces some general NIR spectroscopy concepts and the methodologies used for the analysis of biological fluids. These concepts apply to all four of the following chapters, which constitute the research contributions of this thesis. Each chapter corresponds to one of the projects developed and includes the background material and the research question that motivated the project itself.

This thesis as a whole has been divided into studies that have a maternal and/or fetal health component and is based on the near-infrared (NIR) spectroscopy analysis

of a biological fluid, either blood (serum) or amniotic fluid. We start with a feasibility study described in **Chapter 3 (Feasibility Study)** that describes in detail the analysis methodology of NIR spectral signals coming from the serum and amniotic fluid of non-pregnant and pregnant mice, with an emphasis on the metabolic impacts different levels nematode infection on their NIR profiles. This study presents the basis for the other studies in the subsequent Chapters of the thesis.

Chapter 4 (Study 1) describes the analysis of lactating mouse serum with an emphasis on the metabolic impacts of diets, nematode infection and nutrition on dams and their nursing pups. **Chapter 5 (Study 2)** shows the analysis of human amniotic fluid in an urban population in Montreal; it explores the association of 2nd trimester amniotic fluid with infant size at birth. **Chapter 6 (Study 3)** presents the NIR spectral analysis of human serum collected in an indigenous community in western Panama and extends our research to the impact of multiple infections and multiple nutrient deficiencies and presents a methodology to assess impact of nutrient and infections to measure the nutritional status and fetal growth in marginalised populations.

Other than NIR spectra collected for samples in each project of this thesis, there is a wealth of information regarding the samples we used, including anthropometric, nutritional and infection status. Medical assessments were made on each one of the patients and an array of clinical test (blood work and cytokines measurements) was done on these samples. Known medical conditions were identified and samples were labeled accordingly.

The conclusions and general discussion in **Chapter 7** summarises the research contributions of this thesis and presents a discussion on future works and directions that results from these thesis findings.

1.3.1 Feasibility Study (Chapter 3)

This pilot study explored the possibility that NIR spectral analysis could identify metabolomic differences in maternal and fetal serum and in amniotic fluid between mice subjected to three different doses of a *Heligmosomoides bakeri* nematode infection. Three hypothesis were formulated around these samples. We explored whether a maternal infection changed the NIR metabolomic profile in serum and amniotic fluid of pregnant mice and in serum of non-pregnant mice. We hypothesised that 1) the differences would be found in NIR spectral samples in non-pregnant and pregnant dams (serum and amniotic fluid samples coming from the same metabolic environment), 2) significant spectral differences would exist between serum and amniotic fluid and 3) different infection levels would change the NIR spectral profiles in a significant way.

Figure 1–2 represents the biological questions and the experimental framework of this study. The experimental treatment, three levels of gastrointestinal (GI) infections were applied to non-pregnant and pregnant mice and their impacts were measured in the serum of non-pregnant and pregnant mice and in the amniotic fluid of pups born from the infected pregnant dams. The basic research question asked if we could identify the different levels of infection in the NIR spectral signals.

1.3.2 Study 1 (Chapter 4)

The overall objective of Study 1 was to assess the possibility of identifying distinct NIR spectral regions that could be associated with a nematode infection and/or

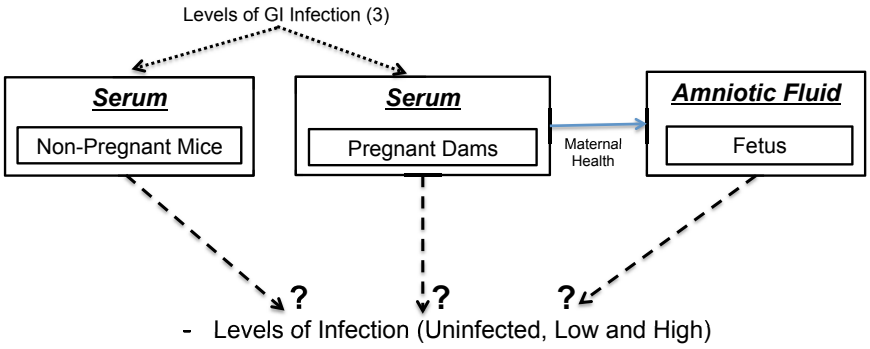


Figure 1–2: *Experimental Framework for the Feasibility Study*

dietary protein deficiency. In this study we investigated dietary protein and a gastrointestinal (GI) nematode infection, on the metabolomic profile in serum of lactating dams, as well as their nursing pups. We hypothesised that 1) diet and/or infection would result in significant differences in the one or more of our eight identified NIR spectral regions and/or their ratios and that these NIR metabolomic profiles would be similar in the serum of lactating dams and their nursing pups.

Figure 1–3 represents the biological questions and the experimental framework of Study 1. The experimental treatment, levels of diet (protein deficient-PD and protein sufficient-PS) and levels of infection (uninfected and infected) status applied to lactating dams. Its impact was measured in the serum of the lactating dams and their nursing pups. We were interested in knowing if there were metabolic differences between the 4 combinations of protein and infections, namely PS-Uninfected, PS-Infected, PD-Uninfected, PD-Infected).

1.3.3 Study 2 (Chapter 5)

The objective of this study was to assess the possibility of identifying early in pregnancy differences in the NIR spectral regions of interest that could be associated with the programming of infant size, defined in our specific case for two categories:

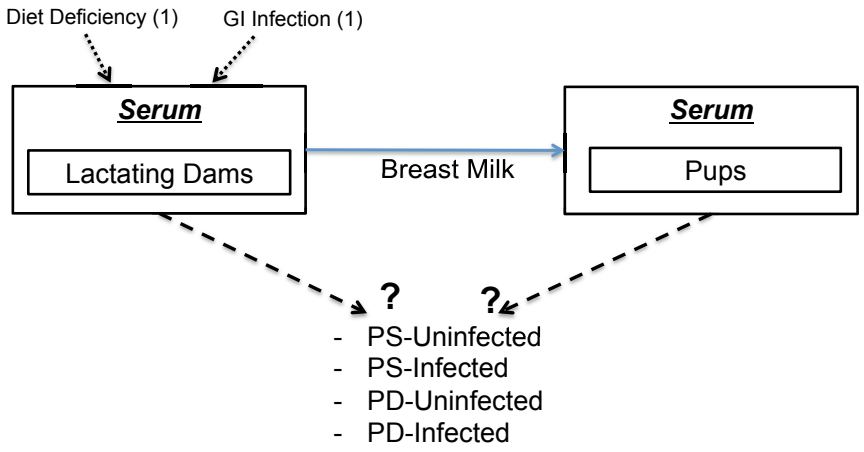


Figure 1–3: *Experimental Framework for Study 1*

appropriate for gestational age (AGA) and large for gestational age (LGA). Figure 1–4 represents the biological questions and the experimental framework of Study 2. We were also interested in exploring if maternal body-mass-index (BMI) differentially alters the spectral profiles of maternal amniotic fluid for the AGA and LGA early in pregnancy. To do this a 4-way comparison among AGA and LGA infants from mothers with normal weights (BMI $\geq$18.5 and $\leq$24 kg/m^2) were compared with 2nd trimester amniotic fluid profiles of overweight and obese mothers (BMI $\geq$ 25 kg/m^2).

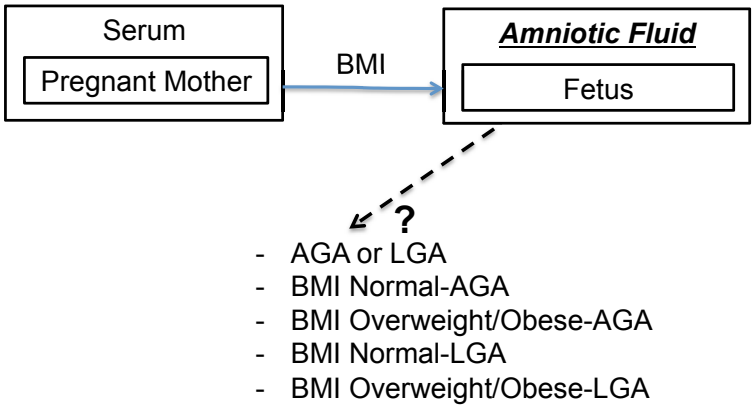


Figure 1–4: *Experimental Framework for Study 2*

1.3.4 Study 3 (Chapter 6)

The objective of Study 3 was to assess the possibility of identifying differences in the NIR spectral regions of interest that could be associated associated with micronutrients deficiencies (vitamin A, vitamin D, B12 and folic acid) and presence of infections (oral and skin) in pregnant women belonging to the Ngabe-Bugle, indigenous community in rural Panama. We hypothesised that 1) by the use of near-infrared spectroscopy we might identify changes in the spectral profiles of serum samples collected from pregnant women with multiple vitamin and infections and that we would be able to use these NIR measurements to assess the maternal-fetal status, in particular via a fundal height measurement of fundal heigh to gestational age (FH:GA).

Figure 1–5 represents the biological questions and the experimental framework of Study 3. This experimental framework is very similar to Study 1, but in this case we measured the impact of multiple infections and vitamin deficiencies in pregnant human mothers.

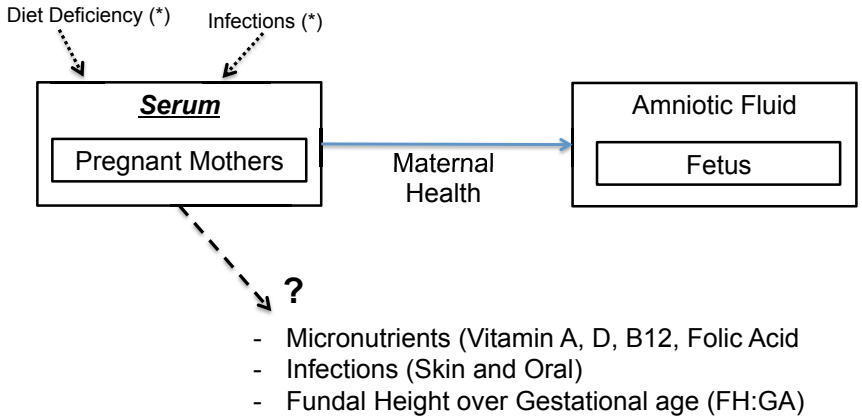


Figure 1–5: *Experimental Framework for Study 3*

1.4 One Step Beyond: Taking Near-Infrared Spectroscopy into New Arenas

In this thesis in general we will work in newer directions where we can use NIRS to differentiate among several maternal-fetal conditions, such as: infections levels, different diets and growth aspects given that we are trying to expand NIRS, out of the laboratory settings and making it a suitable technology that could used in a applied setting for exploring maternal-fetal health relationships. Moreover, it will hopefully provide a basis for the development of a portable instrument that could be used in the field and in remote rural settings, translating into better health status of those who need it the most.

CHAPTER 2
Literature Review: Using Near-Infrared Spectroscopy for the Identification of Molecular Species in Biological Fluids

In terms of the electromagnetic spectrum, the Infrared (IR) spectral range that corresponds to wavelengths from 700nm up to 1mm. It is divided in three main regions: 1) the Far-IR (FIR), from 25μm to 1mm, 2) Mid-IR (MIR), from 2.5μm to 25μm and 3) the Near-IR region that spans from the end of the visible light region at 700 to 2500nm (2.5μm), as depicted in Figure 2–1. By convention, when referring to vibrational frequencies in the MIR region, the ranges are usually given in wavenumber or cm^{-1} and when referring to frequencies the wavelength or nm (nanometers) is used.

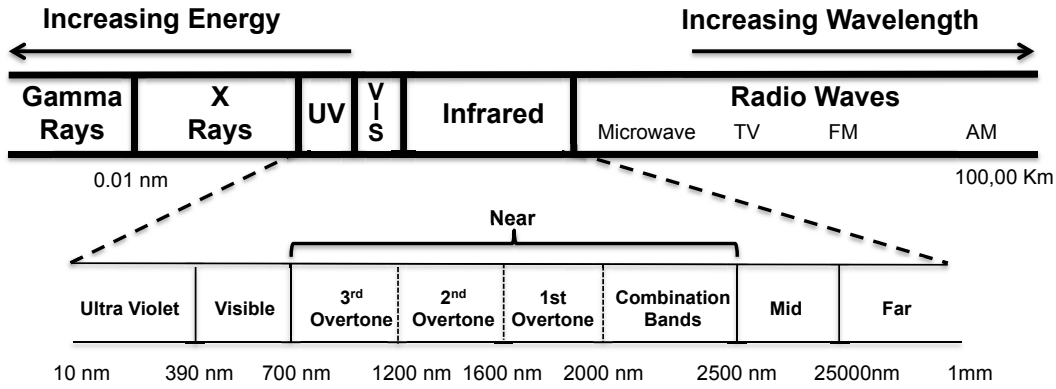


Figure 2–1: *Electromagnetic Spectrum: from Gamma Rays to Radio Waves. Source: Adapted from [11].*

Figure 2–1, depicts the relationship between frequency, wavelength and energy. Frequency is the number of complete wavecycles to pass a fixed point in one second (it is usually measure in Hz) and wavelength (symbolised λ) is the distance that exist between any two peaks of a wave. The electromagnetic spectrum encompass all the frequencies from gamma and X-rays to radio waves. The first having high frequency, lower wavelength and high energy that can cause ionisation of molecules and the

latter have high wavelength, low frequency and low energy, that can cause nuclear transitions [11].

The importance of understanding this division is due to the fact that in the MIR region, specifically between 8500 to 12500 nm lies the 'finger print region' where molecular species have their characteristic or 'fundamental' transitions (vibrations), that are used to make structural analysis of samples. In that regards, these fundamental molecular vibrations will experience oscillations, in the same way that a mass-spring system will experience harmonics in time, but in the case of molecular vibrations they are called *overtones*.

While the fundamental vibrations will occur in the MIR region, their overtones and combinations of overtones will be experienced in the NIR region. These overtones occur approximately $\frac{1}{2}$ or $\frac{1}{3}$ of the fundamental absorption wavelength and 2 or 3 times the frequency, with the intensity getting weaker in each overtone. In terms of mode of vibration a molecule of n atoms will experience $3n$-6 functional modes of vibration, while a linear molecules will experience $3n$-5 modes.

2.1 Characteristics of the Infrared Region of the Electromagnetic Spectrum

When molecules are exposed NIR energy (light), the chemical bonds present in the molecule will experience mechanical vibrations in two forms: 1) stretching, which is a continuos change in the interatomic distance, that could be symmetrical and anti-symmetrical and 2) bending, which is a change in bond angle that could be symmetrical in-plane (scissoring), anti-symmetrical in-plane (rocking), symmetrical out-of-plane (waging) and anti-symmetrical out-of-plane (twisting). NIR energy will only cause the atoms in the molecules to vibrate and it will not cause electronic

transitions as happens in the ultraviolet (UV) region of the electromagnetic spectrum. That is why in UV spectroscopy we can get information about the electron structure of the molecules we are analysing.

Figure 2–2 shows an example of a simple water molecule H_2O (water) molecule (on the left) and the type of vibrations it might experience when NIR light is passed through it (on the right). This molecule could have an angle bending or have a symmetric or anti-symmetric stretching.

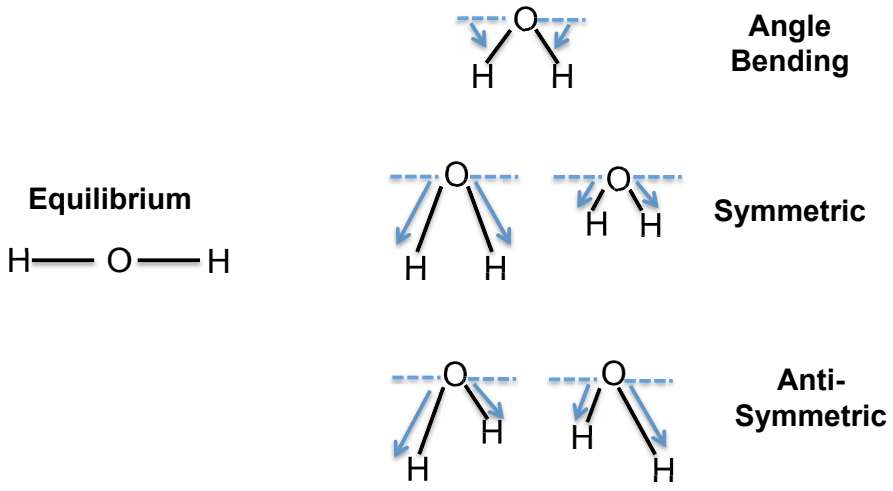


Figure 2–2: *NIR Vibration of water molecule. Source: The author.*

The NIR region only contains overtones and combination bands, mostly overtones of functional groups that absorb strong NIR energy; for instance functional groups that are attached to an hydrogen molecule (X-H) as: O-H, N-H, C-H and S-H. Also, R-H groups have the strongest overtones in this region since the dipole moment that could be experience is high, experiencing R-H stretch and R-H stretch/bend in most of the wavelengths.

The NIR region can be sub-divided into three regions, as shown in Figure 2–3: the short wave NIR region (700nm to 1200nm), the middle NIR region (1200nm to

1800nm), and long-wave NIR region (1800nm to 2500nm) [164, 107]. The importance of this division is due to the fact that these regions encompass the overtones of the fundamental vibration. For instance the short wave NIR region will encompass third and second overtones, while the middle NIR region will encompass the second and first overtones, and lastly the long wave NIR region will experience the first and combination bands oscillations [133].

The occurrence of overtones is not random and they usually happens at integer multiples of the fundamental vibration. For example the fundamental frequency of C-H stretch vibration is near the 3000 cm^{-1} (3333 nm) frequency, so the first overtone is around 6000 cm^{-1} (1600 nm) and the second overtone near 9000 cm^{-1} (1100 nm). In the case of N-H, the fundamental frequency for its stretch vibration is near 3400 cm^{-1} (2942 nm), so the second overtone is found near 7000 cm^{-1} (1400 nm), and the third near 10000 cm^{-1} (1000 nm). The fundamental vibration for O-H is roughly at 3550 cm^{-1}. There are cases in which the overtones don't follow the 'geometric' oscillation pattern described earlier and are called to be anharmonic, as is the case of the water molecule which varies significantly with temperature. In general the intensity of the vibration diminishes by a factor of 10 in each overtone.

The combination bands refer to the fact that in these frequencies there exist overlaps in the overtones of fundamental vibrations. For instance the C-H combination will occur near 44260 cm^{-1} (2242 nm) because that is given by the sum of the C-H stretch that occurs near the 3000 cm^{-1} (3333 nm) and the C-H bend that occurs at 1460 cm^{-1} (6849 nm).

Figure 2–4 shows the anharmonic vibrations (not equally spaced) of a H_2O (water) molecule. The fundamental vibration happens at 2865nm in the MIR region, while its oscillations happen in the NIR region. The first oscillation happens at 1940

Figure 2–3: *Near-Infrared Absorption Bands Chart. Source: ASD Inc. website.*

nm in the combination band region, the first overtone happens at 1430nm, the second overtone at 960nm and the third at 740nm [28, 22].

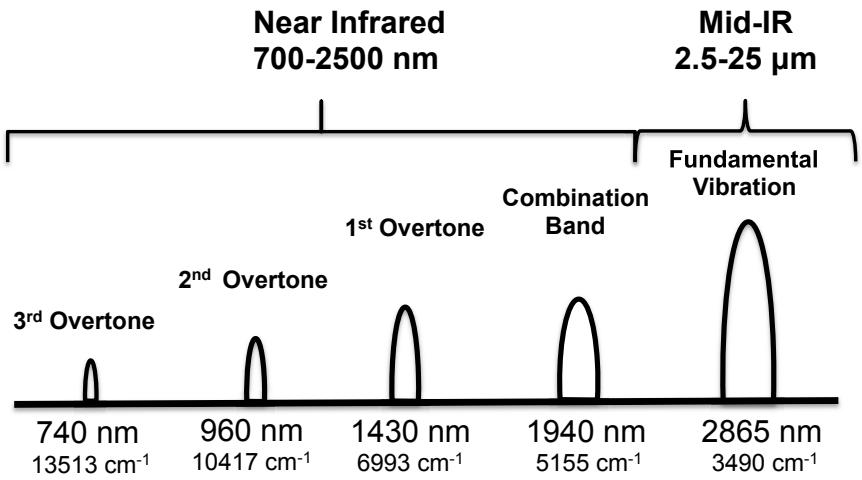


Figure 2–4: *Fundamental vibration and overtones of a water molecule. Source: The Author.*

Since NIRS is based of combination and overlapping between of vibrating modes the molar absorptivity in the NIR is quite small, making the spectra appear to be confusing and even uninteresting, full of peaks and valleys (due to absorption effects) and it is difficult or impossible to arrive to specific functional groups assignments for individual absorption peaks [133], making the task of analysing NIR spectral signals to some extent an interpretative one. Specific band assignment can be done via spiking experiments. In this type of experiments known amounts of an analyte are added (a spike) to the sample, and the sample is re-measured to determine if the spiked amount added could be recovered or the its effect in the signal can be measured. This type of corrections are made in order to build more robust calibration models in NIR spectroscopy [125].

Despite these facts, NIR spectroscopy is still commonly used since it is a versatile vibrational spectroscopy technology that is suited for analyzing many types of samples (solids and liquids) and has many diverse applications. NIRS is an accurate and

fast technique that has been used extensively in monitoring and quality control in the bioprocessing and fermentation industries [127], and also applied to the food processing industry [106, 112], and pharmacology [19, 70]. It is of high relevance in biomedical and clinical contexts because it is a fast, reliable and inexpensive method for biomarker discovery in biological fluids and tissues [89, 133].

2.2 Near-Infrared Spectroscopy in a Nutshell

Since NIR energy will cause specific molecular vibrations in specific molecules and functions groups, then if we measure the light absorption pattern of a sample we can have a notion of the molecular structure of a sample. We can do this measurement toward the visible part of the electromagnetic spectrum with a $UV - VIS - NIR$ spectrometer or in the short, middle and long wave regions a normal NIR spectrometer or a FT-IR spectrometer can be used on the long wave over 2000nm and into the MIR region. Each one of these spectrometers have been designed and constructed with appropriate optical elements, gratings, beam splitters and charge-coupled device (CCD), to measure changes in specific wavelengths regions.

There exist various ways to measure NIR absorption, for instance: transmittance, applied to solutions (often liquids and aqueous solutions) and (diffuse) reflectance, applied to solids (powders and materials). Transmission spectroscopy where lights passes through the sample typically in a configuration similar to the one shown in Figure 2–5, with a thermal light source, typically a lamp Tungsten-Halogen or Xenon or light emitting diodes (LED's), lenses for collimating light into and out of the sample cell, an spectrometer that will receive the light signal passed through the sample and a detector that will save the transmittance information of the spectrometer. In reflectance spectroscopy the light measured is the light reflected from the sample and captured by the CCD detectors. In general the spectra produced from

a NIRS experiment is known to fluctuate with temperature (specially in the case of water and aqueous samples), thus a temperature controlled instrumental setting is recommended.

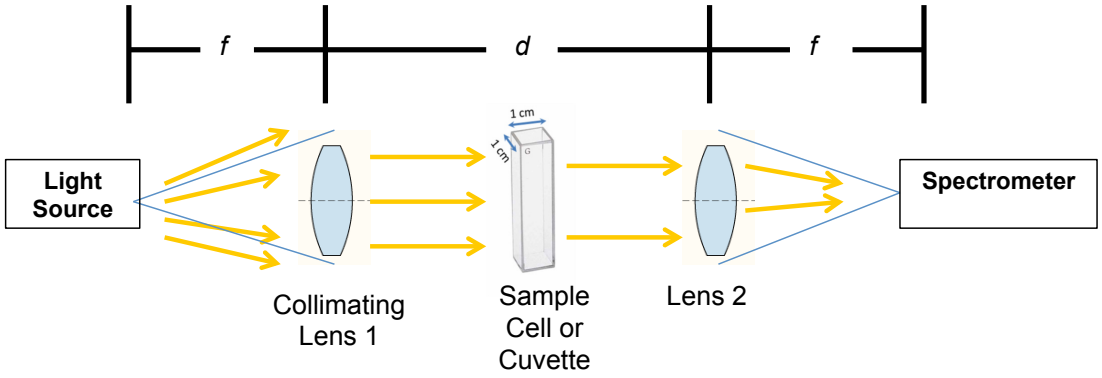


Figure 2–5: *Typical NIR Experimental Setting, in which the distance f, the focal point of the lenses, is the distance from the light source to the lenses at the beginning of the setting and the same distance from the lens to the spectrometer, and the distance d is the distance between the lenses. Source: The author.*

2.2.1 Sample Handling

NIRS measurement is specially used for aqueous samples since absorption of substances present can be confirmed if a 1 mm path length cell is used in combination with a glass or quartz sample cell since absorption in these materials is barely measurable.

Measurements in a NIRS are usually given as relative measurements of transmittance (T) which can be represented as $T = \frac{I_r}{I_s}$ also expressed $T\% = 100*T$. where I_r denotes a reference measurement is obtained by letting pass light into the detector, and I_s measures the light transmitted out of the sample cell and perceived in the spectrometer. This amount of light is usually transformed to relative absorbance values that are equivalent to transmittance and are calculated using the following formulas: $A = -\log_{10}(\frac{l_r}{l_s})$ also expressed $A = -\log_{10}(\frac{1}{T})$ or $A = 2 - \log_{10}(\%T)$.

Finally, after all these transformations, the NIR spectral measurement is presented for inspection as a plot that will have absorbance (in relative absorbance units) in the Y-axis and wavelength (one value per wavelength inspected, in the wavelength region defined by the settings of the spectrometer) as a X-axis.

2.3 Uses of Near-Infrared Spectroscopy

Near-infrared spectroscopy (NIRS) produces a spectral profile of analytes and small molecules present in them without the need of pre-processing or the use of reagents to promote coloured reactions or soaps for fragmentation of peptides or lysis of cell membranes reducing analysis time from hours to to just a few seconds [133]. It is also well suited when working with small liquid samples volumes and irregular material surfaces [11]. But, in general when NIR is used to measure samples that have weak light absorption the signals themselves are going to be low.

As is common to other spectroscopy techniques data acquired with NIRS requires the expertise of a data analyst with knowledge of chemo-informatic methods to understand the data it generates and sufficient knowledge of chemistry associated with molecular vibrations that might be causing the peaks and variations in the signals. This concerns will be addressed later in this Chapter 3.

2.3.1 Medical and Clinical uses of Near-Infrared Spectroscopy

Common clinical chemistry laboratory techniques include quantitative and qualitative analysis of body fluids as whole blood, serum (blood without red or white cells), urine, saliva, amniotic fluid, cerebrospinal fluid (CSF). Also it includes the analysis of feces and in some cases such as biopsies the analysis of tissue [133]. NIRS is a suitable technique that can replace or enhance most of the commonly used methodologies and

equipment used for these analyses, providing a way to quantify the amount of the constituent of the samples. One problem of using NIR in these settings is that one of most abundant species found in biological fluids is water, which is known to dominate and may out-shadows other NIR absorption signals [133].

NIRS is very attractive as a technology in the fields of clinical chemistry and medical settings due to the possibility of developing non-invasive instruments that can provide in vivo measurements. One of the most important advances in this regards has been the measurement of glucose in blood.

2.3.2 Analysis of Serum

Serum might be one of the most studied biological fluids using NIRS. Probably because of its abundance, since is routinely in medical examinations and collected everywhere. Methods for acquisition, calibration, analysis of serum protein, triglycerides and glucose using a NIRS system have been previously described by Hall and Pollard [46]. Other of their studies have also focused in measuring total protein, albumin, globulins, has been identified using a multiple linear regression (MLR) and a Partial Least Squares regression for urea [47], suggesting a partial assignment to peaks for albumin at 2178 nm, for globulins at 2240 nm and 2200 nm for urea. Furthermore, a complete set of analyses including total protein, albumin protein, globulin protein, triglycerides, cholesterol, urea, glucose, and lactate is described in [53].

Another study was set to study human serum albumin (HSA), globulin and glucose and also proposed a method for selecting the best spectral regions as factors for partial least squares algorithm [67]. A more recent study done in the Valencia region of spain [38] used 447 serum samples to determine models for total protein, albumin, total cholesterol and triglycerides using NIR spectral profiles. This last

study, as many others, had as a main objective, using NIR methodologies to be a candidate substitute for normal clinical blood work examinations recommended by physicians.

2.3.3 Glucose Monitoring

Glucose is a monosaccharide sugar with molecular formula $C_6H_{12}O_6$, it is used in the body as a source of energy. Insulin is an hormone that regulate its intake and is in charge of transporting it from the bloodstream to the muscle, fat and liver cells to be used as energetic fuels. Elevated serum glucose is the main cause for diabetes and the development of its complication including: atherosclerosis, coronary artery disease, diabetic kidney disease, urinary tract infections, stroke, peripheral artery disease, blindness and in extreme cases amputation [35, 87]. Therefore the monitoring of this molecule in patients and the early detection of its elevation in blood is critical.

NIR instruments have been designed for the analysis of blood glucose levels, not only a rapid way to test for glucose in blood samples, but more importantly many non-invasive instrumentation have been designed taking advantage of the fact that under steady steady conditions sub-cutaneous glucose concentrations are at least $\tilde{7}5\%$ similar to intra corporal blood concentrations [33, 140].

Nowadays glucose is measured on the skin and index fingers prominently [115]. However other places of the body such as: lips [90] ear lobe [63] and others have been extensively studied have been studied with diverse results [14]. Also some have applied secondary measurement to detect glucose levels, for instance using the patient saliva [126].

2.3.3.1 Commercial Instruments for Diabetics

Despite the technical difficulties caused by overlapping signals, weak NIR glucose signals and the fact that it is complex to isolate glucose signals from the signals of other constituent with signal-to-noise ratio. The 1550nm wavelength has been isolated and chosen as a candidate for measuring glucose signals [3]. Based on this and other findings commercial build monitoring instrument has been made and are currently being sold for diabetic patients. This type of instrument take two forms:

1. The Invasive form (Blood Glucose meters): Here a droplet of blood is placed in a film strip and inserted in an automatic machine designed to measure the blood glucose levels in the tissue, but has the disadvantage of having to buy the test strips and having to prick the finger every time and proceed to inject the amount to insulin needed.
2. The Non-Invasive form (Instant Blood Glucose): Here glucose monitoring is based on a sensor that is placed under the skin and is connected to an insulin pump and a receiver that controls the glucose concentrations and shows real-time updates on the screen

2.3.4 Near-Infrared Spectroscopy for Clinical Settings

Although most of efforts in NIRS have been dedicated to the task of determining clever non-invasive ways to measure glucose in blood and tissues due of its importance in the production of metabolic energy and more importantly because its role in diabetes, which is the 7th most common cause of death in the United States and with increasing prevalence in Canada. We think that near infrared spectroscopy can be used to investigate a wealth of other biological fluids, and tissues and provide an

incredible amount of information in relation to other metabolic alterations and conditions. Some areas we are interested are: NIRS measurements of neonatal medicine, the rapid determination of infection levels or nutrition status.

2.3.5 Near-Infrared Measurements for Determining Infections in Biological Fluids

There exist a growing interest at looking at infection using spectral techniques. Want *et al* provides a comprehensive list of metabolites and metallic fragments that are affected by nematode infections biological fluids measured by MIR [156]. However, in the case of NIRS, few studies have investigated at infection and how they change the spectral profile of a biological fluid. Sakudo *et al* [122] provided an overview of the possibility of using visible and near infrared (Vis-NIR; 600-1100nm) spectroscopy on biological fluids for the determination of viral infections in human patients. This group has worked in the detection of spectral differences found in the nasal fluids of infected and non-infected human patients for influenza around the 970nm water peak [121, 120] and for human immunodeficiency virus (HIV) [123, 124], finding shifts in the -OH vibration in water at 950 and 1030nm, respectively.

2.3.6 Applications of Near-Infrared Spectroscopy to Nutrition in Biological Fluids

Most of the NIRS studies found related to nutrition were in the field of animal science. The fecal NIRS field (*F. NIRS*) has mostly focused in the study of diet chemistry, and through it, the physiological state of the animal, just by looking the fecal content in both caged and free animals, for instance the case of giant pandas [160] and white-tailed deer [62]. However, no work has been done in quantifying protein contents in serum or any other nutrition parameters through a biological fluid or skin NIRS measurement.

2.3.7 Applications of Near-Infrared Spectroscopy to Tissue Analysis

Lately, NIRS have been successfully used for measurement of changes and metabolic differences in tissue, for instance the case for different types cancer tissues [99, 77, 72, 73, 79]. Here NIRS has provided information for the determination of Alzheimer [43]. Moreover, there has been more work on measuring infections in muscle through muscle oxygenation saturation (SmO_2), after dengue infection [136] and in wound infections after surgery [60].

2.4 Near-Infrared Spectroscopy applied to Pregnancy and Neonatal Medicine

There exist a very limited number of NIRS studies in pregnancy and neonatal medicine. These studies have focused on blood plasma, embryo culture, oxygenation monitoring and amniotic fluid analysis.

NIRS have been successfully used for the the determination of metabolic profiles for culture media for in-vitro fertilization [151, 149, 50, 150, 152]. While there are important advances in this field, they are only concerning embryo cultured in-vitro and are not related to fetal health.

In terms of fetal health and monitoring another area in which NIRS has been used is fetal development for the non-invasive determination of placental oxygenation in pregnant women [65, 52]. A recent review on the development of an analysis methodology for the screening of the growth restricted fetus through placental tissue oxygenation index in simulated data was published [2]. This study used previously collected data in the 700-900nm region to simulated both well oxygenated and not oxygenated experimental groups.

One biological fluid that has been little investigated, despite the wealth of information it provide is amniotic fluid. While amniocentesis, the extraction of amniotic fluid from the amniotic sac for testing purposes, is a relatively simple and common method performed on pregnant women, it is surprising that the amount of papers using NIRS for the analysis amniotic fluid is quite low. One author has used amniotic fluid as a proxy measurement to lung maturity [84, 82, 83] and our research group has used it for prediction of early preterm births in human pregnancies [110]. One probable reason for this phenomenon might be the complexity of the amniotic fluid matrix itself and most likely the great amount of water that it has, making it hard to be measured in the NIR region, that has prominent water peaks, that could mask real measurements.

This field is still developing and there exist limited amount of studies applying NIR to the area of maternal-fetal. Moreover, after literature review we have identified a number of relevant articles, none of which uses the NIR wavelength range we would explore in this thesis, in the 1600-2400nm region.

CHAPTER 3
Feasibility Study

A starting point for this thesis was the work we previously published in our laboratory by Powers, Sanchez-Galan *et al* [110], in Appendix A. That paper presents the analysis of NIR spectral profiles of amniotic fluid samples collected at 15 weeks of gestation for the identification of distinct of prematurity profiles before 35 weeks of gestational age compared to term samples at 37 weeks of gestation. The conclusion was that two wavelet regions, selected by the genetic algorithm, corresponding to third overtone CH_3 and second overtone NH group vibrations distinguished premature births from term births. The NIR wavelengths ranged from 700-1050nm and two different approaches for selection of wavelengths of interest for describing the condition were considered. The first approach used was a *discovery driven approach* where 'spectral data was compressed then optimized by multilinear regression to create a calibration model', for the second approach we focused in *spectral region integration*, summing and ratioing the values of specific NIR spectral regions. This approach was further refined in the feasibility study, described below.

For the feasibility study, serum and amniotic fluid samples previously collected in our laboratory were available from a study that measured serum and amniotic fluid in pregnant mice subjected to three levels of infection compared to non-pregnant mice. Distinct to our first study, we decided to use a different instrument, in this case a long-wave NIR spectrometer (1600-2400nm) and a *spectral region integration* approach where eight regions of interest were selected to assess if we could determine NIR spectral changes in serum and amniotic fluid in pregnant mice.

3.1 Pilot Study Protocol

This pilot study explored the possibility that NIR spectral analysis could identify metabolomic differences in maternal and fetal serum and in amniotic fluid between mice subjected to three different doses of a *Heligmosomoides bakeri* nematode infection. Serum from non-pregnant (n=43) and pregnant (n=24) CD1 mice subjected to three levels of trickle nematode infection (sham or uninfected: 3 x 0 L3; low: 3 x 50 3 L3; high: 3 x 100 3 L3) was collected at day 19 of pregnancy (day of necropsy). Amniotic fluid samples were also collected on the same date but only for pregnant mice. Since the volume of amniotic fluid collected was scarce, samples that had 2 vials of 5μL were pooled resulting in 10μL samples bringing the final number of amniotic fluid samples to n=14.

The detailed description of this trickle infection protocol in mice has been previously described [101, 102]. Briefly, previously uninfected, non-pregnant mice and primiparous pregnant mice (day 4 of pregnancy, defined as experimental day 0) were received from Charles River (Quebec, Canada). Mice were randomized into 3 infection doses, and infected on experimental days 2, 7 and 12 (days 6, 11 and 16 of pregnancy) to allow 3 infections prior to parturition. *H. Bakeri* L3 were obtained by faecal culture of stock parasites maintained in outbred CD1 mice (Charles River) in our laboratory for over 15 years. Larvae were counted in 10 sham doses for accuracy before infecting the mice. The same larval culture was used for all infections within each set of mice. Animal ethics approval was obtained from the McGill (Macdonald Campus) Animal Ethics Committee and complied with the regulations of the Canadian Council on Animal Care [104].

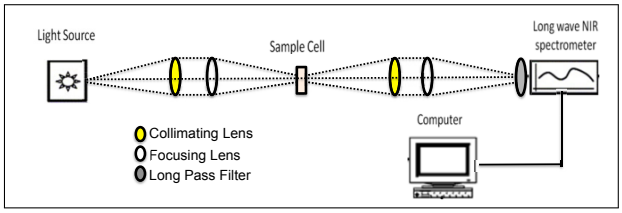
3.1.1 Hypothesis

Hypotheses were formulated using our biobanked mouse serum and amniotic fluid samples. Our intent was to explore whether a maternal infection changed the NIR metabolomic profile in serum and/or amniotic fluid of pregnant mice and in serum of non-pregnant mice. We hypothesised that:

1. Differences will be found in NIR spectral samples between non-pregnant and pregnant dams.
2. Significant spectral differences will be found between serum and amniotic fluid.
3. Different infection levels (sham or uninfected, low and high) will change the NIR spectral profiles of biological fluids.
4. Different infection levels (sham or uninfected, low and high) will change the following pre-selected NIR spectral regions. The following eight regions of interest were targeted:
    - Three regions in the first NIR overtone bands: Region 1 (1650nm-1700nm, 6 wavelengths), Region 2 (1701nm-1750nm, 5 wavelengths), Region 3 (1850nm-1899nm, 6 wavelengths).
    - Four regions in the NIR combination bands: Region 4 (2050nm-2099nm, 6 wavelengths), Region 5 (2100nm-2130nm, 4 wavelengths), Region 6 (2130nm-2200nm, 8 wavelengths), Region 7 (2275nm-2310nm, 5 wavelengths), Region 8 (2311nm-2370nm, 8 wavelengths).
5. That ratios of these eight regions, could be used to quantify the relative change in NIR spectral signals.

3.1.2 Instrumental Settings

The instrumental setting for NIR spectra collection consisted of a LABPOD NIR spectrometer manufactured by Polychromix (Wilmington, Mass.) in the 1600-2400nm wavelength range with a 24-bit resolution. A 6.5 Watt power output Tungsten-Halogen lamp (manufactured by BWTEK BPS101, B&W TEK, Newark, DE) was used as a light source. For serum samples of dams and pups, a spectrometer system in which light was collimated using a lens system into the sample cell was used. The schematic and a photo instrumental setting are show in Figure 3–1. For both experimental settings the sample cell used was an U-shaped flow cell with 1 mm path length (Starna Cell, Inc)

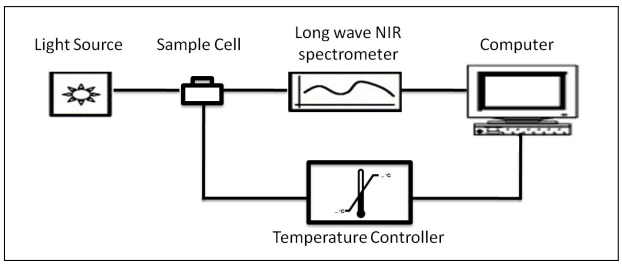


(a) *Schematic of the instrumental setting*

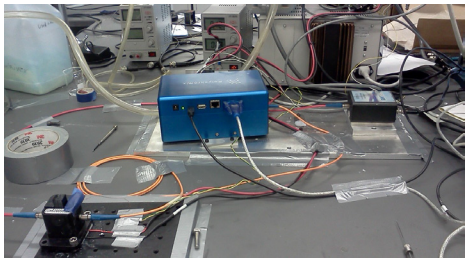
(b) *Photo of the instrumental setting*

Figure 3–1: *Instrumental Settings with Collimating Lenses. Source: Adapted from [16]*

For the amniotic fluid samples light was transmitted from the light source to the temperature controlled sample holder via a fiber optics cable (38mm) and into the spectrometer using via fiber optics (30mm). A schematic and a photo of the instrument is shown in Figure 3–2.



(a) *Schematic of the instrumental setting*



(b) *Photo of the instrumental setting*

Figure 3–2: *Instrumental Settings with Fiber Optics. Source: The author.*

3.1.3 Spectral Collection Protocols

Prior to spectrometric measurements, frozen samples were thawed on ice at room temperature and spectra were acquired at room temperature for serum samples and at 23°C for amniotic fluid samples. A flow sample cell was filled with 10μL of serum and 40μL in the case of amniotic fluid samples (diluted of 1:3 in water). The sample cell was cleaned between measurements by rinsing it with 1mL of 0.1M NaOH followed by 5mL of distilled, deionized water. NIR spectra of deionized water samples (40μL) were taken as references before processing of each sample and saved for posterior spectral subtraction. NIR spectra for all experiments were collected as transmittance values in the 1600nm to 2400nm wavelength range using the SpectralCodePlus v3.5 software. Each single spectra represented the average of 10 scans at an integration time of 100ms over 100 wavelengths.

The choice of using deionized water was a reference sample since it matches the main structures of biological fluids we set to investigate in terms of consistency, with biological fluids being mostly composed by water . Moreover, the the acquisition of water references served as a way to monitor the stability and reproducibility of the NIR instrument over the course of the experiment, for instance to monitor the temperature differences in the sample cell. If there is a substantial difference in the

water spectra collected at the beginning of the experiment versus the water spectra collected at the end of the experiment, this can indicate that an instrumental drift (change in lamp intensity, saturation of the spectrometer, displacement of the sample cell) happened through the experiment. In case these differences arrive the signals can be manipulated with the use of detrending, that is used to remove linear trends from a vector structure.

3.2 Spectral Data Processing

Analysing the NIR Signals: Matlab programming package (ver. R2012a; The Math Works Inc., MA) was used for all steps of data processing and analysis described in this section. Sample absorbance was calculated as a ratio between the intensity of light absorbed by the sample (Is) and the intensity absorbed by the water reference (Ir), using the following formula: $Absorbance = -log_{10}(\frac{l_r}{l_s})$

Values between 1900nm-2050nm were set to zero to minimize the effect of the high absorption water peak found around 1940nm [91, 163].

3.2.1 Pre-processing Techniques

The aim of pre-processing was to enhance features of the NIR spectra that could carry important spectral information and to eliminate regions that are uninteresting or full of noise [133]. For that cause we first smoothed and then mean-center the signal. Figure 3–3 describes an example of in which a group of mock signals for experimental group "red", that will be used for explaining purpose.

Absorbance signals were smoothed for the existence of errors introduced by fluctuations of the signal that might happen in-between samples by applying a 3-point recursive boxcar-averaging algorithm [158], applying the following formula $S_j =$

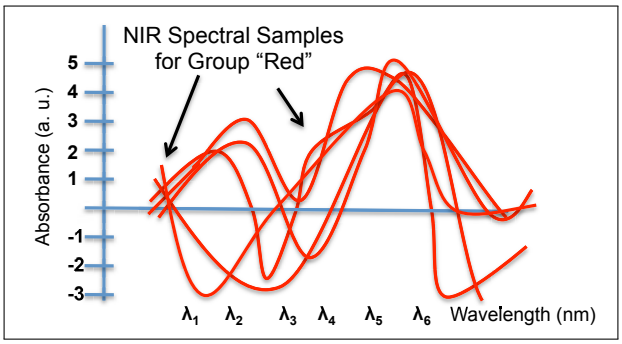


Figure 3–3: *Conditioning of Mock NIR Spectral Samples. Source: The author.*

$\frac{Y_{j-1}+Y_j+Y_{j+1}}{3}$, where Yj is the point we are smoothing at wavelength j and Y_{j-1} and Y_{j+1} are the previous and following point in the series, respectively.

For this normalization we selected a window size of 3, which means each point was averaged from 3 points, the point itself, the previous and the posterior points, thus becoming recursive, using parts of its output as input for the next stage of the algorithm. Figure 3–4 describes a working example in which a mock signal is processed into windows of size 3 for each of the six wavelengths presented (λ_1 to λ_6), in a recursive manner.

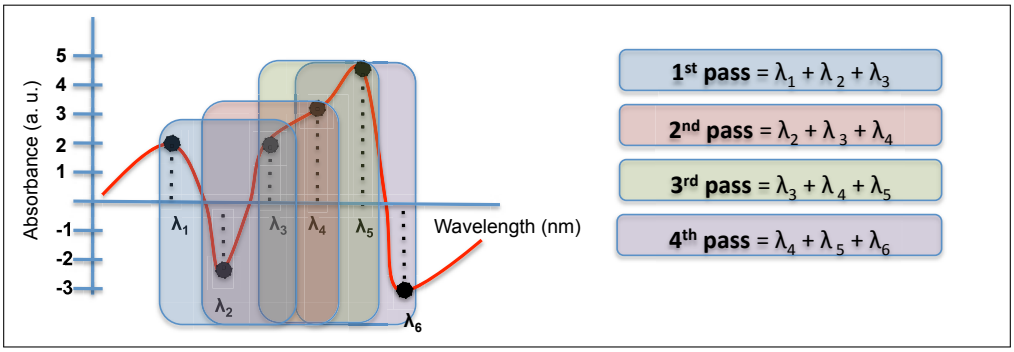


Figure 3–4: *Smoothing of Mock NIR Spectral samples. Windows size of three points. Source: The author.*

3.2.2 Mean Centering

After the NIR spectral signal was smoothed then it had to be mean centered [147]. Mean centring calculates the average spectrum of the NIR signals and subtracts that average from each spectrum. It shifts the values proportionally so they are translated to be centered around zero (absorbance units) for each wavelength [133], which in turn make more evident the greatest variations on the data.

It uses the following equation $Mean\ Centered = x_i - mean(x)$, where x is a given wavelength and x_i is a sample value for wavelength x. Figure 3–5 describes a working example of in which a group of mock signals for experimental group "red", are processed using mean centring procedures.

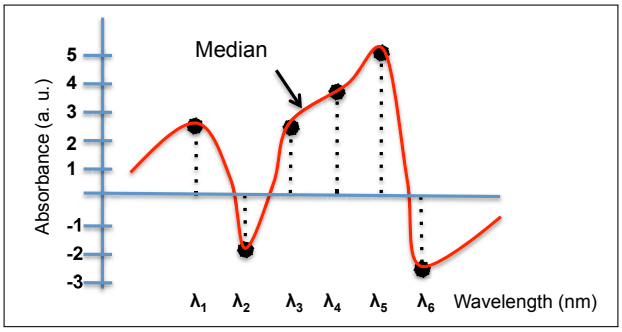


Figure 3–5: *Mean Centering of Mock NIR Spectral Samples. Source: The author.*

3.2.3 Integration of Region of Interest

We focused our analysis to the first overtone region and combination bands regions in the 1600 to 2400nm wavelength region of the NIR region of the electromagnetic spectrum, where 1st overtone CH, SH, 2nd overtone C=O and combinations bands of OH, NH, NH+OH, CH+CH and CH+CC are known to have higher absorbance.

In terms of the specific regions we selected 3 in the 1600-1900nm region (1st overtone region): Region 1 (1650nm-1700nm, spanning 6 wavelengths), Region 2 (1701nm-1750nm, spanning5 wavelengths) and Region 3 (1850nm-1899nm, spanning 6 wavelengths). Also, 5 regions were selected in the 2000-2400nm region (combination bands region): Region 4 (2050nm-2099nm, spanning 6 wavelengths), Region 5 (2100nm-2130nm, spanning 4 wavelengths), Region 6 (2130nm- 2200nm, spanning 8 wavelengths), Region 7 (2275nm-2310nm, spanning 5 wavelengths), Region 8 (2311nm-2370nm, spanning 8 wavelengths).

Figure 3–6, shows a visual interpretation of the integration of a mock NIR spectral signal. The absorbance values for different wavelengths were divided into bins representing regions of interest and summed up to represent one value for each NIR spectral signal in the set. Wavelengths λ_1 and λ_2 were summed up and its value represented the value of Region 1, the same happened for λ_4 and λ_5 for Region 3 and λ_7 and λ_8 for Region 8.

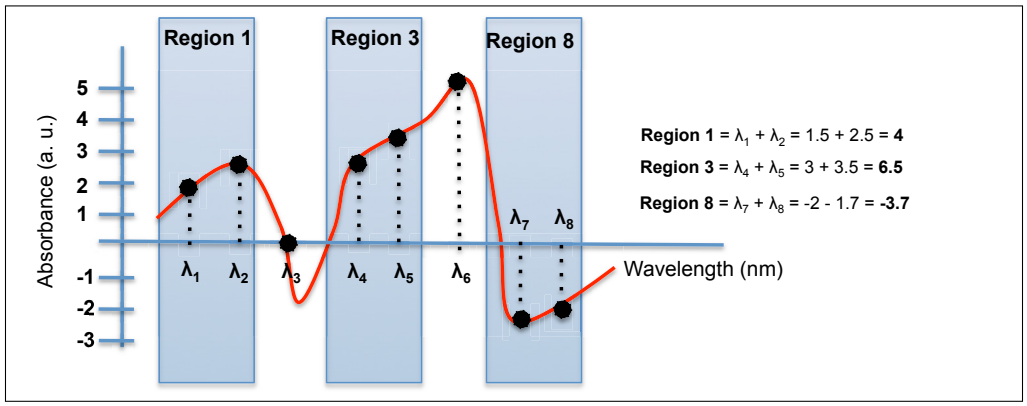


Figure 3–6: *Integration of NIR spectral signals. Source: The author.*

3.2.4 Ratios as Interactions

Immunology provides an interesting use of ratios that we set to explore with our NIR spectral data. In immunology the ratio of two groups of T-helper cells,

Type 1 (Th1) and Type 2 (Th2) cells, provides an idea of the status of depletion states, balance and finally dominance that might happen (Th1 or Th2 dominance) in a specific metabolic state or disease [71]. This ratio of Th1/Th2 (often written Th1:Th2), depicted in Figure 3–7, is very important since if provides a global view of the interaction of Type 1 (Th1) cells, which are responsible of protective responses against intracellular pathogens (viruses) and cell-mediated immunity and Type 2 (Th2) cells, that are known to antibody production and create responses against extracellular pathogens (multi cellular parasites) [116].

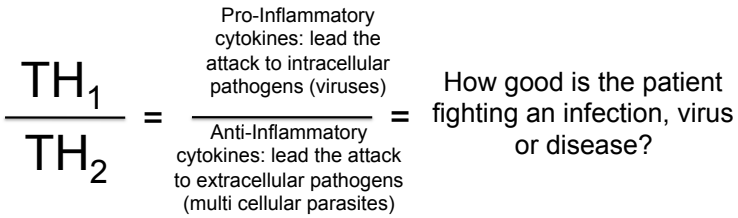


Figure 3–7: *Ratio of Th1/Th2 Cells Explained. Source: The author.*

We thought that this notion of ratios was very useful since we were dealing with different experimental and clinical conditions. Moreover, we set to interpret ratios as interactions [6, 76] between the 8 NIR regions described in the previous section.

For instance, using a ratio between absorbance values Region 7 and Region 8, regions that we might associate with lactate and glucose respectively, we get a magnitude that represents the amount of change in Region 7 relative to Region 8, a description is shown in Figure 3–8.

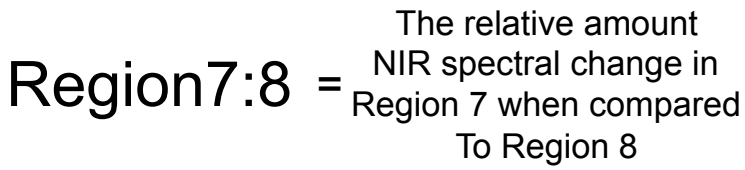


Figure 3–8: *NIR Spectral Region Ratio. Source: The author.*

When calculating we first determined which were all the possible combination. This was made using the expressed as the binomial coefficient, expressed as the number of combination of regions that we would get if from the 8 regions we pick 2 at a time, which is expressed in mathematical terms as $\binom{8}{2}$, an operation which yields 28 combinations. After, each of the values corresponding for each of those 28 combinations was calculated by simply dividend the absorbance values for a region, between the values of another region.

3.2.5 Statistical Analysis: Determining differences between experimental conditions

To determine difference between experimental conditions, all the NIR spectral samples were first put together into a matrix, called *Data*, of size 'n X 36', where n is the number of samples and 36 the number of columns composed from the values of the 8 original regions plus the 28 region ratios. Also, a vector of labels, called *Label*, of size 'n X 1', that contains labels or classes for the k experimental groups was selected. In the case of the feasibility study $k = 3$ since we were exploring the differences among 3 experimental groups. Figure 3–9, depicts these data structures. Finally, a matching between the Labels vector with the rows in the Data matrix provided the values for a given experimental group. So, if we have a Data matrix of 60 NIR samples, and only samples 1, 3, 45, 57 are assigned to experimental group 1 and the rest to experimental group 2, then the values for those 4 rows are statistically compared against the remaining 56 rows.

Statistical Tests. As the values obtained for the regions and ratios were not normally distributed among the experimental groups (confirmed via Shapiro-Wilk test), spectral medians for each experimental group for each of the spectral regions and regions ratios were compared using Kruskal-Wallis one-way analysis of variance.

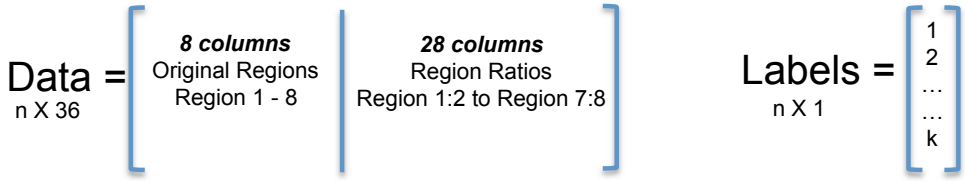


Figure 3–9: *Data Matrix with integrated values for Regions and Ratios. Source: The author.*

Experimental group differences that were statistically significant after the Kruskal-Wallis were also tested in a pairwise two-sided Mann-Whitney U test and a Bonferroni correction was performed as a post-hoc test. In all instances a value of $P<0.05$ was considered to be statistically significant.

3.2.6 Calculation of interquartiles regions for the NIR Spectral Signal

For the purpose of having a visual understanding of the distribution of the values for each of the experimental groups, the upper and lower quartiles of the median signal were calculated, and points for each wavelength were determined as borders for the signal, as shown in Figure 3–10. Also, the interquartile region was calculated as the area in between the upper minus the lower quartile and shaded for distinction, as shown in Figure 3–11.

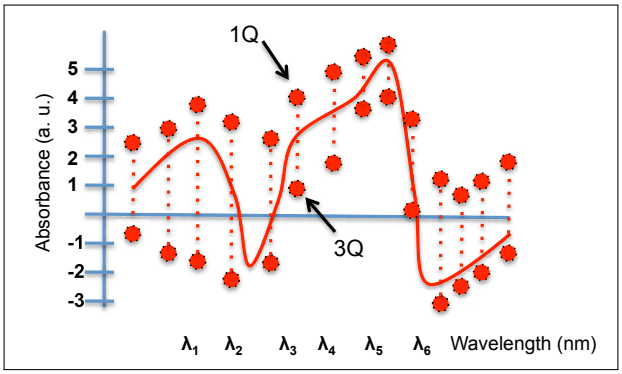


Figure 3–10: *Upper and lower Quartiles from Median Signal. Source: The author.*

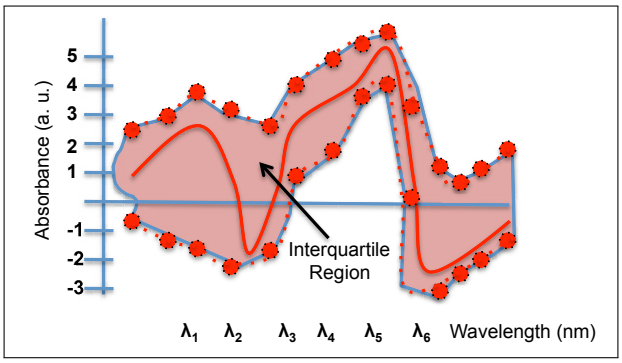


Figure 3–11: *Calculation of interquartile regions for Experimental Groups. Source: The author.*

3.3 Observations from the Feasibility Study

In this section the main observations resulted from the analysis of serum of non-pregnant and pregnant mouse, and of amniotic fluid samples using the methods described in the previous section.

3.3.1 Non-Pregnant Mouse Serum by Infection Status

Pre-processed NIR spectra for individual non-pregnant mice are shown in Figure 3–15-A and the spectra divided among the three experimental groups (uninfected, low and high) in black, blue and red in Figure 3–15-B, respectively. The spectra showed minimal variation among experimental groups with the exception of a peak in Region 3 near the beginning of the water absorption band at 1900nm where there was no overlap in their median signal (interquartile shaded area). Kruskal-Wallis test of the Regions (Table 3–1) showed that the only significant difference was for Region 3 in non-pregnant serum (P=0.02), where post-hoc analyses showed a significant difference between low and high infection groups, but neither differed from sham or uninfected groups, see Figure 3–12. Interestingly, no ratios or interactions among regions were found statistically significant.

3.3.2 Pregnant Mouse Serum by Infection Status

As with the non-pregnant serum, NIR spectra for individual pregnant mice are shown in Figure 3–15-C and NIR spectra from the pregnant mice, divided by experimental categories, is shown in Figure 3–15-D. Variation could be noted among the experimental groups, with the sham or uninfected group behaving very differently than the low and high infected groups, especially in the 1600-1800nm region. Statistically significant differences are shown in Table 3–1 where four ratios were found to be significant Region 1:2 (P=0.03); Region 1:4 (P=0.03); Region 4:8 (P=0.04); and Region 5:7 (P=0.05). Post hoc analyses revealed that the difference in the Region 1:2 ratio was between sham and low infection, presented in Figure 3–13 and the difference in the Region 1:4 ratio was between sham and high infection, presented in Figure 3–14. The other two ratios were not significant between experimental groups.

Table 3–1: *Comparisons among infection categories for Non-Pregnant and Pregnant Mice*

Region/Ratio	**Uninfected**	**Low**	**High**	**P Value**
Non-Pregnant	n=12	n=15	n=16	
Region 3	-0.003^{AB} (-0.04, 0.01)	0.005^{A} (-0.02, 0.03)	-0.027^{B} (-0.03, -0.00)	0.017
Pregnant	n=7	n=9	n=8	
Region 1:2	1.104^{B} (0.88, 1.15)	1.245^{A} (1.15, 1.81)	1.261^{AB} (1.13, 1.50)	0.026
Region 1:4	0.361^{A} (-4.52, 0.46)	1.025^{AB} (0.33, 1.42)	0.740^{B} (0.57, 0.97)	0.031
Region 4:8	0.372^{AB} (0.17, 0.82)	0.933^{AB} (0.52, 1.49)	0.279^{AB} (-1.30, 0.65)	0.039
Region 5:7	1.589^{AB} (1.32, 2.30)	1.572^{AB} (1.19, 1.78)	0.756^{AB} (0.14, 1.35)	0.046

3.3.3 Pregnant Mouse Amniotic Fluid by Infection Status

Amniotic fluid showed no variability in the shorter wavelengths between 1600 to 2100nm, but differences between experimental conditions were evident in the amniotic fluid in the 2200 to 2400nm region shown in Figure 3–16-B. The spectral pattern of

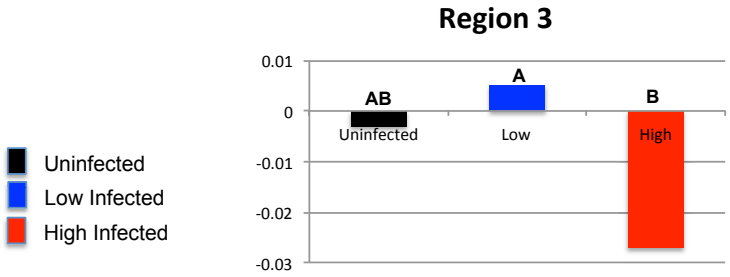


Figure 3–12: *Difference in Region 3 for Non-Pregnant Dams Serum*

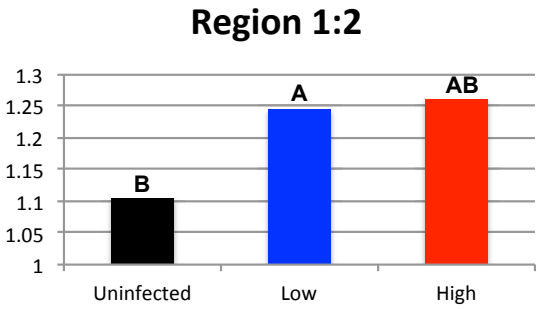


Figure 3–13: *Difference between Uninfected and Low Infected Pregnant Dams*

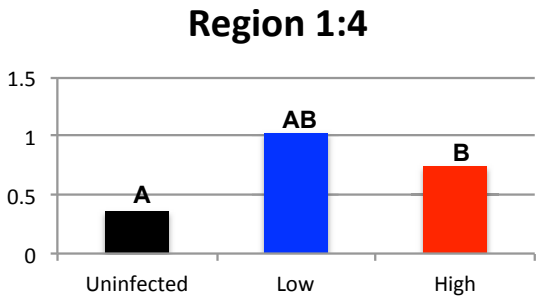


Figure 3–14: *Difference between Uninfected and High Infected Pregnant Dams*

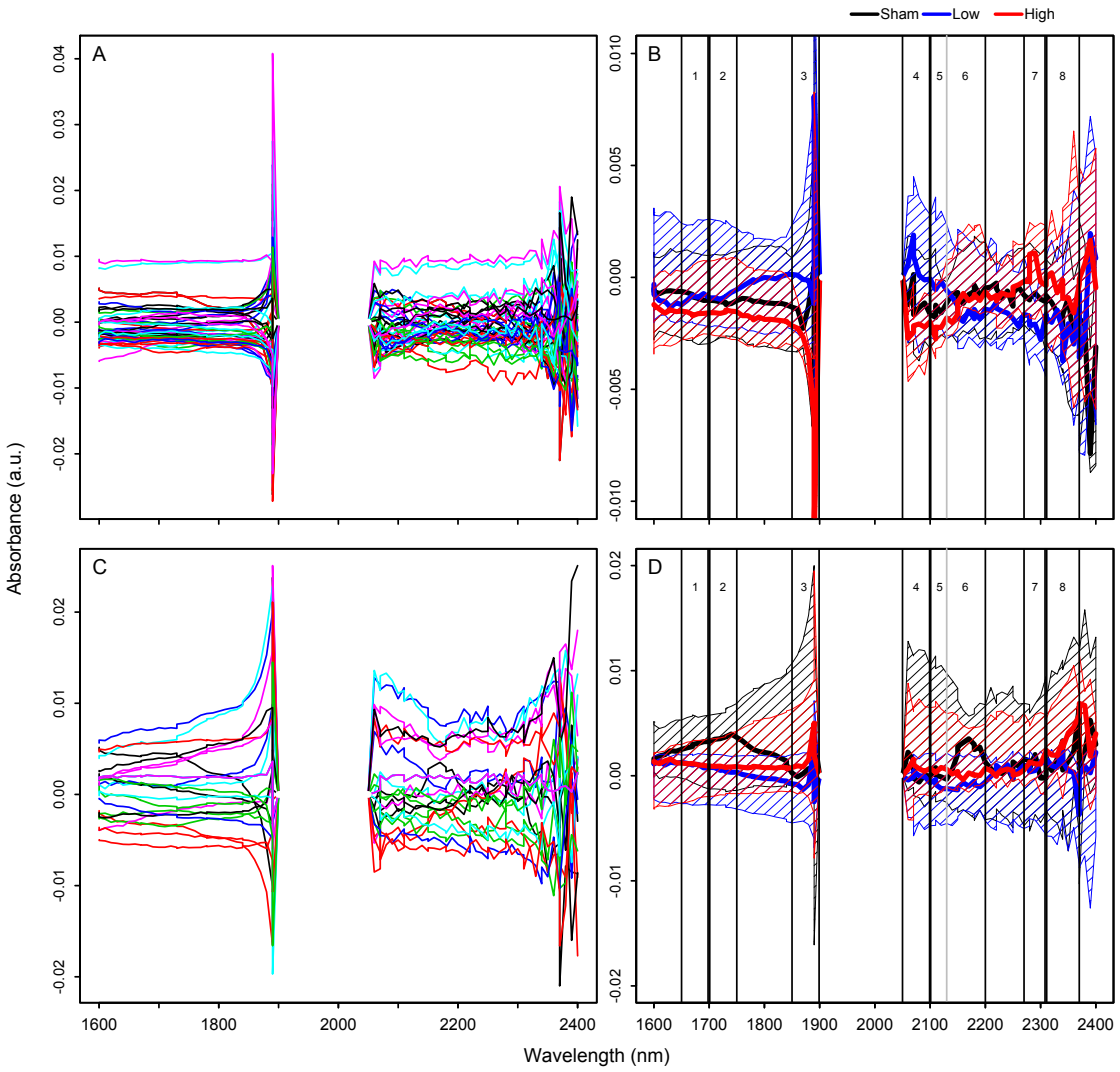


Figure 3–15: *NIR spectra of 43 non-pregnant samples (A, B), 24 pregnant samples (C,D). Median spectral absorbance was divided by experimental groups (solid black line represents the sham group, a blue line represents the low infection and a red line represents the high infection groups) for serum from non-pregnant mice (B), serum from pregnant mice (D). Shaded regions, with respective colors, represent the interquartile region (1Q-3Q).*

the high infection group had an absorbance pattern towards the negative values. In contrast, both sham and low infected categories show positive absorbance in the same region. Kruskal-Wallis test revealed significant differences for Region 6 (P=0.04); Region 3:5 (P=0.05); Region 3:7 (P=0.01); Region 4:7 (P=0.03); Region 6:7 (P=0.03); and Region 7:8 (P=0.04). Interestingly, post-hoc analysis revealed no significant differences among experimental groups.

Table 3–2: *Comparisons among infection categories for using Amniotic Fluid*

Region/Ratio	**Uninfected**	**Low**	**High**	**P Value**
Amniotic Fluid	n=5	n=5	n=4	
Region 6	0.002^{AB} (-0.03, 0.02)	0.019^{AB} (0.01, 0.02)	-0.023^{AB} (-0.03, -0.01)	0.026
Region 3:5	0.561^{AB} (-0.82, 3.28)	-0.824^{AB} (-1.84, 1.68)	4.233^{AB} (1.42, 7.44)	0.050
Region 3:7	-0.040^{AB} (-0.27, -0.03)	0.043^{AB} (-0.03, 0.41)	-0.195^{AB} (-0.35, -0.10)	0.028
Region 4:7	-0.032^{AB} (-0.21, 0.08)	0.077^{AB} (0.00, 0.46)	-0.112^{AB} (-0.27, -0.03)	0.026
Region 6:7	0.138^{AB} (-0.09, 0.28)	0.276^{AB} (0.18, 0.58)	0.065^{AB} (-0.08, 0.14)	0.032
Region 7:8	0.356^{AB} (0.27, 0.41)	0.266^{AB} (-3.60, 0.34)	0.382^{AB} (0.34, 1.52)	0.042

3.4 Interpretation of the Observations

Our feasibility study showed that maternal infection modified Region 3 in non-pregnant mice whereas these same infection burdens produced differences in two ratios, Region 1:2 and Region 1:4 during pregnancy. Interestingly, more NIR spectral region differences were identified in amniotic fluid, but post hoc tests did not allow detection of these differences by infection status. We can suggest that this may have resulted from the small sample size.

In the case of pregnant mice, we found that the NIR spectral profile for serum differed from that of maternal amniotic fluid, implying that these are distinct metabolic

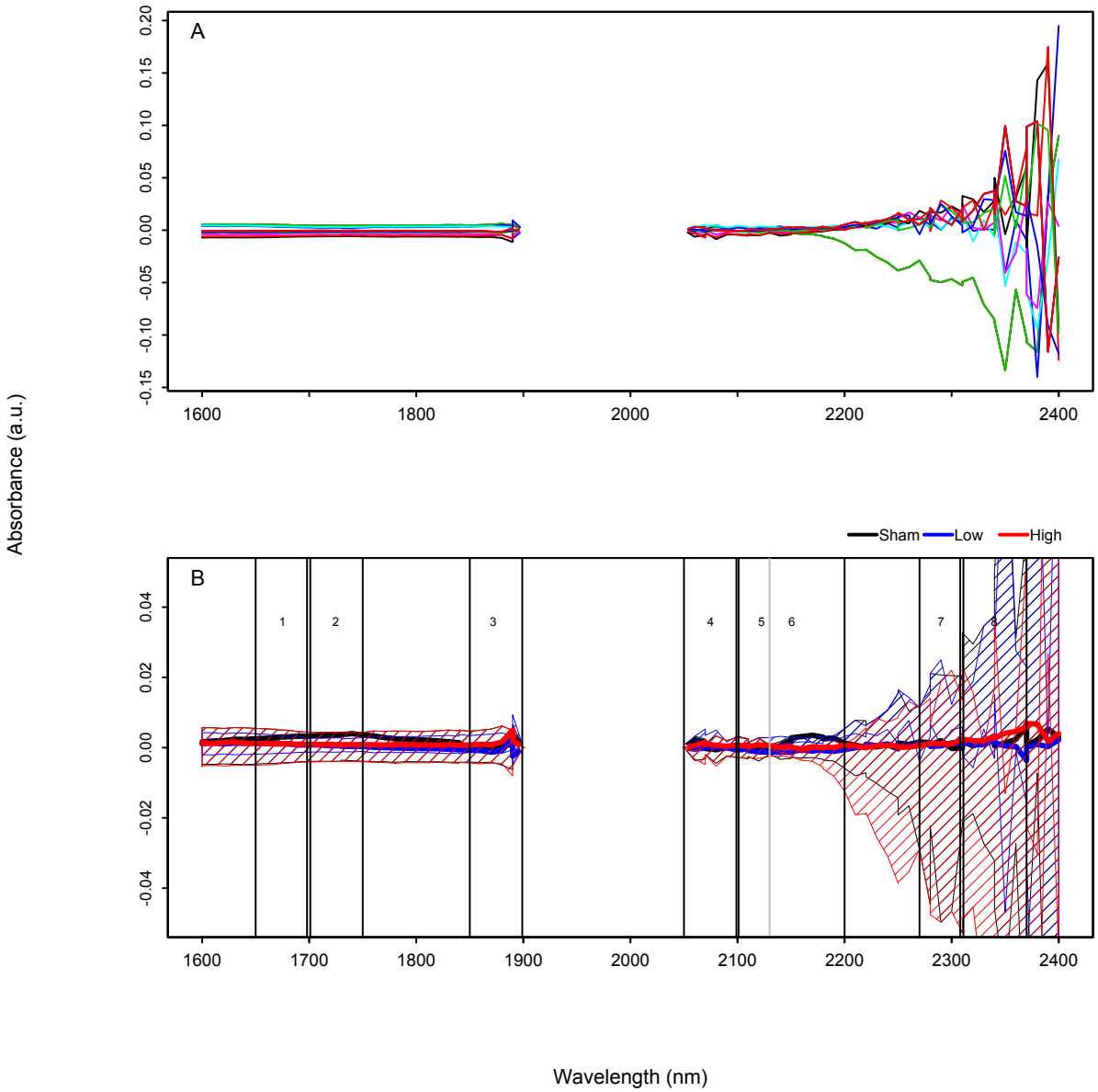


Figure 3–16: *NIR spectra of 14 amniotic fluid samples (A, B). Median spectral absorbance was divided by experimental groups (solid black line represents the sham group, a blue line represents the low infection and a red line represents the high infection groups) for amniotic fluid from pregnant dams (B). Shaded regions, with respective colors, represent the interquartile region (1Q-3Q).*

pools. It is reasonable to suggest that since the experimental treatments were imposed on the dams both compartments perceive the insult and react in metabolically different ways. However, this is in contrast to literature that suggests that there could be correlation between the maternal and fetal compartments.

An interesting observation was noted when looking at Region 5 and Region 6 in the 2100-2200nm a pattern emerged for pregnant dams, shown in Figure 3–17, for uninfected samples median spectral values for Region 5 were low and were higher on Region 6.

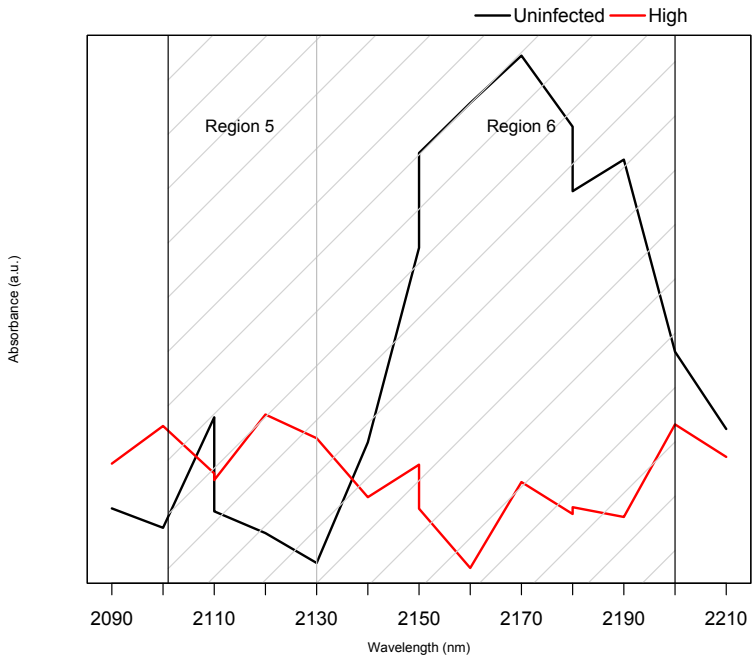


Figure 3–17: *Close up of the 2100nm-2200nm Spectral Region for Pregnant Dams*

3.5 Conclusion from the Feasibility Study

After developing a framework for the analysis of NIR spectral signals and applying this method to an specific set of mouse biological fluids we can summarise our accomplishments as follows:

- We move from a previous study on serum where the 700 to 1100nm region was analysed to a study involving both serum and amniotic fluid and in a different wavelength region in the 1600nm to 2400nm, with most of the variability happening the later portion of the region (2000-2400nm), in wavelengths that are be closer to the primary vibrations measured in the mid-IR region.
- Differences in NIR signals were noticed in two different fluids and in two conditions (pregnancy and non-pregnancy), conditions which are the main basis of the rest of the thesis
- We can measure the differences between experimental groups by first processing the NIR spectra, integrating them in specific regions, making use of ratios and then making use of statistics and analysis of variance test and post-hoc comparisons.
- Despite the variations in the NIR spectral signals, which can be attributed to concentration changes or bimolecular interactions in the sample, the integration into regions of interest and the use of these ratios as a measurement of variation provides an interesting picture of the metabolic impact of the experimental condition, which in this case was three levels of infection.

Taken together these results provide a sufficient evidence that we could use NIRS in the 1600-2400nm to explore NIR spectral signal variations in *serum* and *amniotic fluid* that could be associated with infection. This feasibility study as a whole forms the basis and provides the methodological chemo-informatic framework that would

be used in Study 1 (in Chapter 4), for the effects of diets and infections in mouse *serum*; for the Study 2 (in Chapter 5), in which human *amniotic fluid* is used for the determination of size at birth and lastly for Study 3 (in Chapter 6) for the analysis of *serum* of pregnant indigenous women with multiple nutrition deficiencies and multiple infections.

CHAPTER 4
Study 1: Metabolic Profiling of Maternal and Pup Serum using Near-Infrared Spectroscopy Identifies Distinct Patterns for Nematode Infection and Dietary Protein Deficiency

Javier E. Sanchez-Galan[1], Maurice R. Odiere[2], Marilyn E. Scott[2], David H. Burns[1,3] and Kristine G. Koski[1,4]

[1]Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, Canada H3A 1A3

[2]Institute of Parasitology, McGill University (Macdonald Campus), Ste-Anne de Bellevue, QC, Canada H9X 3V9

[3]Department of Chemistry, McGill University, Montreal, QC, Canada H3A 2K6

[4]School of Dietetics and Human Nutrition, McGill University (Macdonald Campus), Ste-Anne de Bellevue, QC, Canada H9X 3V9

4.1 Abstract

Recent advances in Near-Infrared (NIR) spectroscopy make it a suitable and inexpensive candidate technique for the analysis of biological fluids and the determination of metabolic profiles. This study explored the possibility that NIR spectral analysis could: 1) identify metabolomic differences in lactating mice ($n = 22$) that were subjected to two levels of dietary protein (PD 6% and PS 24%) and two doses of nematode infection, uninfected (UN) and infected (IN) and 2) determine if these differences were similarly reflected in the serum of their nursing pups. Maternal serum and serum pooled from the pups in each litter ($n = 25$) were collected on day 21 postpartum. A 4-way comparison of the four experimental groups (PS-UN, PS-IN, PD-UN and PD-IN) was conducted for both dams and pups using 8 NIR spectral regions: 1650-1700nm (Region 1), 1701-1750nm (Region 2), 1850-1899nm (Region 3), 2050-2099nm (Region 4), 2100-2130nm (Region 5), 2130-2200nm (Region 6), 2275-2310nm (Region 7) and 2311-2370nm (Region 8). These regions and their ratios were compared among the experimental groups using Kruskal-Wallis non-parametric tests, followed by pair wise comparisons using Mann-Whitney U test. Overall significant differences ($P<0.05$) were found in lactating dams between diets for three ratios Region 1:2, Region 1:7 and Region 2:7, reflecting a diet effect that was only observed in the infected dams. In their pups infection resulted in differences for one single region, Region 5, and 6 spectral region ratios: Region 1:4, Region 2:3, Region 2:3, Region 3:5, Region 3:6, Region 6:7, Region 7:8 whereas protein deficiency resulted in differences for 2 ratios: Region 2:6 and Region 2:7. In contrast to lactating dams, both maternal PD and infection were reflected in pup serum, with the NIR spectral profile of PS-UN pups differing from PS-IN pups (Region 1:4 and Region 3:6) and the NIR spectral profile of PS-IN pups differing from PD-IN pups (Region 2:6). Taken together these results suggest the the impact of infection and/or dietary protein deficiency affected

the NIR spectral profiles not only of the mother, but especially the nursing pups, even though they were not subjected directly to the experimental interventions.

4.2 Introduction

Near-infrared spectroscopy (NIRS) is a vibrational spectroscopy technique used for the identification of molecules present in a sample by analyzing their constituent chemical bonds [89, 164, 107]. NIRS is an accurate technique that has been used extensively in monitoring and quality control in the bioprocessing and fermentation industries [127], and also applied to the food processing industry [106, 112], and pharmacology [19, 70].

It is of high relevance in biomedical and clinical contexts because it is a fast, reliable and inexpensive method for biomarker discovery in biological fluids and tissues [133]. It has also been used for the determination of metabolic profiles of culture media for in vitro fertilization [50, 150, 151, 129, 132] and whole blood [44], serum [46, 53] and amniotic fluid for lung maturity [82, 83], and for pre-term pregnancies [110]. NIR spectroscopy has also been used in the context of pregnancy to monitor fetal status associated with placental and cerebral oxygenation in humans [65, 165] and in animals [17].

There is growing application of NIR spectroscopy to study the metabolic effect of infections. However, most of spectroscopic efforts have focused on the utilization of FT-IR [103] and 1H nuclear magnetic resonance (NMR) spectroscopy techniques [81]. For instance, Wang *et al* [156] reviewed of NMR metabolic fingerprints of mice infected by three intestinal parasitic worms (*Echinostoma caproni*, *Trichinella spiralis*

and *Schistosoma mansoni*). Sakudo *et al* [122] provided an overview of the possibility for the application of visible and near infrared (Vis-NIR; 600-1100nm) spectroscopy to biological fluids for the determination of viral infections in human patients. This group has worked in the detection of spectral differences found in the nasal fluids of infected and non-infected human patients for influenza around the 970nm water peak [121, 120] and for human immunodeficiency virus (HIV) [123, 124], finding shifts in the-OH vibration in water at 950 and 1030nm. Other researchers have used NIRS to examine viral infections in an endophytic fungus [109] and bacterial infection in Drosophila [7].

The relationship between malnutrition and the increased susceptibility to infection has been very well established and studied [21, 74, 68, 131]. Nutritional deficiencies are known to have negative effects on pregnancy and lactation, by restricting the growth and development of the fetus and more importantly causing a poor immune response to infection [139]. Moreover, protein deficiencies in pregnancy are known to affect both the maternal health status and fetal development of muscle, joints and bone, cause brain damage or severe birth defects [128]. Protein deficiencies have an amplified negative effect in the presence of nematode or helminthic infections [94]. However, most studies that have evaluated the interrelationship between infection and nutrition have focused on stress markers, such as cytokines, and free single molecules in bodily fluids and tissues. Few have used NIR spectroscopy to study of the NIR spectral changes in serum that are experienced by the nematode affected mice during pregnancy.

The overall objective of this study was to identify distinct NIR spectral profiles in serum that could be associated with a nematode infection and protein deficiency during pregnancy and lactation in mice and to their metabolic consequences on the offspring. To the best of our knowledge this is a novel approach that can be used for understanding infection and diets in the context of NIRS in the 1st overtone and combination bands region (1600 to 2400nm) using serum samples.

4.3 Materials and Methods

Experimental Design. Using samples previously collected by our group [101, 102], we investigated both the impact of maternal protein deficiency and nematode infection on the metabolomic profile in serum of lactating dams, as well as serum from their nursing pups. We hypothesised that 1) diet and/or infection would result in significant differences in one or more of our eight identified regions and/or their interactions [54, 6], measured as ratios; and 2) that the NIR metabolomic profiles would differ between the diet and infection levels in serum of lactating dams and their nursing pups.

Experimental Design and Protocols. Samples from a previous animal study, which received ethical approval from the McGill (Macdonald Campus) Animal Ethics Committee and complied with the regulations of the Canadian Council on Animal Care [104] were used to investigate both the impact of maternal diet and infection on the NIR spectral profile in serum of lactating dams, as well as serum from their nursing pups. Serum was collected from CD1 lactating dams (n=22) and serum pooled from the their pups (n=25) at 21 days post-parturition. We hypothesized that 1) two levels of dietary protein (protein sufficient, PS, 24% and protein deficient, PD, 6%)

and two nematode infection levels of a trickle nematode infection (uninfected, UN and infected, IN) would result in significant differences in one or more of our eight identified regions and/or their interactions [6, 76], measured ratios; and 2) that the NIR metabolomic profiles would differ between the diet and infections in serum of lactating dams and their nursing pups.

NIR Spectral Integration into regions of interest. For the analysis of the NIR spectral signals, a spectral integration approach was used. Instead of using all the wavelengths for which spectra were collected, only specific wavelength regions, where specific moieties are known to have a higher absorbance of light were considered. Therefore, the 1600-2400 nm was selected because of the important higher energy transitions in the 1st overtone region (1600-2000nm) and the combinations bands region (2000-2400nm) where more intense absorption are found, and thus providing the possibility of using for analytical purpose.

The 1600nm to 2400nm wavelength range was divided into eight regions: Region 1 (1650nm-1700nm, spanning 6 wavelengths), Region 2 (1701nm-1750nm, spanning 5 wavelengths) and Region 3 (1850nm-1899nm, spanning 6 wavelengths). Also, 5 regions where selected in the combination bands: Region 4 (2050nm-2099nm, spanning 6 wavelengths), Region 5 (2100nm-2130nm, spanning 4 wavelengths), Region 6 (2130nm-2200nm, spanning 8 wavelengths), Region 7 (2275nm-2310nm, spanning 5 wavelengths), Region 8 (2311nm-2370nm, spanning 8 wavelengths). The absorbance values for each of these eight regions were integrated into one value per sample. A similar approach, in which spectral regions or *spectral windows* are integrated and related back to a set of specific moieties, has been used in the spectral of analysis of

other NIR [5] and FT-IR [108] data.

Near-infrared spectroscopy instrument and spectral collection. Prior to analysis, frozen samples were thawed on ice at room temperature and spectra were acquired at room temperature. A flow sample cell was filled with 10 μL of serum. The sample cell was cleaned between measurements by rinsing it with 1mL of 0.1M NaOH followed by 5mL of distilled, deionized water. NIR spectra of water samples (40 μL) were taken as references before processing of each sample and saved for posterior spectral subtraction.

All NIR spectra were acquired using a NIR spectrometric system consisting of a LABPOD NIR spectrometer (Polychromix, Wilmington, Mass.), a Tungsten-Halogen lamp (BWTEK BPS101, B&W TEK, Newark, DE) as a light source and a U-shaped flow sample cell with 1 mm path length (Starna Cell, Inc). NIR spectra for all experiments were collected as transmittance values in the 1600nm to 2400nm wavelength range using the SpectralCodePlus v3.5 software. Each single spectra represented the average of 10 scans at an integration time of 100ms over 100 wavelengths.

Spectral Data Processing. Matlab programming package (ver. R2012a; The Math Works Inc., MA) was used for data processing and analysis. Sample absorbance was calculated as a ratio between the intensity of light absorbed by the sample (Is) and the intensity absorbed by the water reference (Ir), using the following formula:- log10 (Is/Ir). Values between 1900nm-2050nm were set to zero to minimize the effect of the high absorption water peak found around 1940nm [91, 163]. Raw absorbance signals were smoothed for the existence of systematic errors, first, by applying a 5-point recursive boxcar-averaging algorithm [158] and later by mean-centering to

remove constant instrumental background from all the samples.

Statistical Analysis. To analyze the relative changes in these eight regions, the notion of ratios or interactions was used to compare one region to another. Ratios were calculated by dividing the mean centered absorbance values from different NIR spectral regions. Previously, this approach has been used by our group in the determination of spectral regions associated with pre-term pregnancies in human amniotic fluid [110].

As the values obtained from the integrations of the 8 regions and their ratios were not normally distributed among the experimental groups (confirmed via Shapiro-Wilk test), spectral region medians were compared among experimental conditions using Kruskal-Wallis one-way analysis of variance. These medians were presented as values and the values for the 1st and 3rd quartile, which were values above and below the median of an integrated spectral region or region ratio.

Kruskal-Wallis one way analysis of variance was used to determine the combinatorial effect of two levels of infection (uninfected and infected) and two levels of diet (protein sufficient and protein deficient) thus creating a four-way comparison for the PS-UN, PS-IN, PD-UN and PD-IN experimental conditions. In contrast with a normal ANOVA test, no interaction term between diet and infection (Protein levels $\times$ Infection levels) was analyzed, due to the fact that we were using integrated spectral regions instead of continuous measurements or concentrations for a specific analyte and ratios capture the interaction between moieties that was of interest to us.

Experimental group differences that were statistically significant after the Kruskal-Wallis were also tested in a pairwise two-sided Mann-Whitney U test and a Bonferroni correction was performed as a multiple comparison post-hoc test. In all instances a value of $P<0.05$ was considered statistically significant.

4.4 Results

Maternal Serum. NIR spectral profiles for maternal serum are shown in Figure 4–1-A. Visualization of NIR spectral medians of the experimental groups, shown in Figure 4–1-B, revealed distinct spectral profiles, but also overlapping variability of the interquartile region (shaded) of the spectral signals.

Diet Impact on Lactating Dams. A comparison of the experimental groups, shown in Table 4–1, showed that three ratios were significantly different between the PS-IN and the PD-IN for: Region 1:2 ($P<0.01$), Region 1:7 ($P=0.02$) and Region 2:7 ($P=0.02$). In the three cases the ratios had higher median values for Region 1 in the PS and higher absorbance values for Region 7 in the PD. Figure 4–2-A shows, in the form of a bargraph, the difference between the PS-IN and PD-IN, with PS-IN having higher values for the Region 1:2 ratio and lower values for PD-IN, but no diet difference was found between the uninfected groups. Two other ratios Region 1:6 ($P=0.03$) and Region 2:6 ($P=0.03$) differed, however, post-hoc analysis did not characterize the group differences.

There was no significant infection impact in the maternal serum. However, investigating the NIR spectral signal of the 4 experimental groups, we found an interesting pattern in the 2100-2200nm range (Region 5 and Region 6). For the lactating dams (Figure 4–3), those mice exposed to the most extreme condition, for instance those with the combined impact of PD and infection had higher values for Region 5 and

lower values of Region 6 compared to those with the control group in the PS-UN experimental group.

Another pattern emerged in the 2100-2200nm wavelength range (Region 7 and Region 8) for lactating dams. In this case we saw a graded response where PS-UN >PD-UN >PS-IN >PD-IN, with the uninfected (PS-UN and PD-UN) groups having higher absorbances values shown in Figure 4–4-A. In contrast, for pups, shown in Figure 4–4-B, the infected (PS-IN and PD-IN) groups showed higher absorbance than their uninfected counterparts (PS-UN and PD-UN).

Pups Serum. NIR spectral profile for nursing pups is shown Figure 4–1-C. Visualization of medians of the experimental groups (Figure 4–1-D) revealed distinct spectral profiles, but also overlapping variability of the interquartile region (shaded) of the spectral signals for the PS-IN, PD-UN and PD-IN, experimental groups. However, the PS-UN category showed a very distinct pattern, with lower absorbance values than for the other 3 experimental groups.

Infection Impact on Nursing Pups. Maternal infection had a much broader impact on the NIR spectral profile of pup serum, despite the fact that the infection burden was applied to the dams and not the pups (shown in Table 4–1). For Region 5 (P=0.01), PS-UN pups had higher values than the PS-IN pups. The same differences were not found for PD pups as there was no difference in spectral profiles between PD-UN and PD-IN. In contrast PS-UN differed from PS-IN in one region: Region 5 (P=0.01) and six region ratios: Region 1:4 (P=0.014), Region 2:3 (P=0.014), Region 3:5 (P<0.01), Region 3:6 (P<0.01), Region 6:8 (P<0.01) and Region 7:8 (P<0.01). In four of this cases PS-UN had higher ratios than the PS-IN pups, this pattern of

difference is illustrated with a bargraph for Region 1:4 in Figure 4–2-B. For the remaining two cases PS-IN pups had higher values than the PS-UN pups, this difference is illustrated for Region 3:6 in Figure 4–2-C.

Diet Impact on Nursing Pups. A comparison among the experimental groups, shown in Table 4–1, showed that two ratios, Region 1:7 (P=0.02) and Region 2:7 (P<0.01), were significantly different between the PS-IN and the PD-IN groups. Results showed higher median values for Region 1 in the PS and higher Region 7 in the PD, whereas PS-IN and PD-IN differed in two ratios: Region 2:6 (P=0.021) and Region 2:7 (P=0.01), with PS-IN having higher ratios than the PD-IN group. These differences are presented as a bargraph for Region 2:6 ratio is shown in Figure 4–2-D.

4.5 Discussion

Infection and diet can modify the NIR spectral profile of biological fluids, in general, infection accounted for the spectral changes in the metabolic profiles of maternal serum, whereas both maternal diet and infection impacted the pups who were not exposed directly to the nutritional deficiency and/or nematode infection. In particular, diet, but not infection, modified the spectral profiles of the maternal serum between PS-IN and PD-IN experimental groups for Region 1:2, Region 1:7 and Region 2:7.

In the pup's serum there were two differences between PS-IN and PS-IN, for Region 2:6 and Region 2:7, with Region 2 having higher values. However, most of the differences were found in the NIR spectral profile of PS-UN and PS-IN, for one single region, Region 5 and six region ratios: Region 1:4, Region 2:3, Region 3:5, Region 3:6, Region 6:8 and Region 7:8 In general, PS-UN had higher ratios for Region 5, Region 1:4, Region 2:3, Region 6:8 and Region 7:8, and lower ratios for ratios for Region 3:5 and Region 3:6, than their PS-IN counterparts.

Taken together these results suggest that the growing fetus was experiencing the maternal insults even though the diet and infection treatments were applied to the dams, suggesting that the lactating dams, which received the experimental treatments beginning in pregnancy, continue to transfer the infection and dietary insults to their nursing pups offspring post-partum. Our results are important in the context of the study biological fluids using NIR studies since they demonstrate that NIR spectral regions change between *healthy* and *unhealthy* mice suffering a chronic nematode (*H.bakeri*) infection and a single nutrient deficiency, in our case protein deficiency.

PS-UN vs. PS-IN. Pups in the control group (PS-UN) showed higher absorbance values in Region 1, Region 2, Region 5, Region 6 and Region 7. In contrast their infected counterparts showed higher values in Region 3, Region 4 and Region 8. The increase in values for Region 8 in PS-IN mice might be associated with glucose which is a molecule that is known absorb in this region, and more importantly is a metabolic fuel for developing and for the growing fetus [95].

Diet Effects. Interestingly, two ratios were similar in both maternal and pup serum. Both, Region 1:7 and Region 2:7 were higher in PS infected compared to PD infected dams and nursing pups. Given that the diet differences emerged in infected animals, we can suggest that adequate dietary protein is particularly important in nematode infected mice. However, whether this similarity is related to metabolic imprinting in utero, to transfer of nutrients through milk from the mother to the pup, or to a non-biological association arising with NIR spectral profiles and data processing, can not be fully determined.

Infection Effects. Our results suggests that infection had an impact in the spectral changes seen in Region 7 and Region 8 for both lactating dams and pups.

For lactating dams, higher absorbance values were found in these regions, in the non-infected mice (PS-UN and PD-UN). In contrast, the response in the pups was inverse, with both infected groups (PS-IN and PD-IN) having higher absorbance values for Region 7 and Region 8.

Strengths of the Study. We employed a spectral integration approach in which we measure the differences resulted in the NIR signal of serum of lactating dams and their pups subjected to either protein deficiency and/or infection burden. The use of ratios was proposed as an analysis methodology that was used to explain metabolic changes, that is, how certain spectral regions, stabilize or decrease in relation to others. Despite the small NIR absorbance signals resulting from our analysis, by smoothing, zeroing the water peak, mean centering and more importantly the use of ratios, we were able to detect significant statistical differences in eight targeted spectral regions. Taken together, these results show that metabolomic responses of the foetus and the neonate to either a gastrointestinal nematode infection and/or maternal dietary protein deficiency differ, implying that the fetal compartment as well as the nursing pup continue to alter their metabolism throughout the duration of the maternal exposure to either stressor. It is clear that the long-term consequences of these adaptations need further exploration.

Limitations of the study. The small number of samples and small sample volume was a limitation, as we were not able to build strong predictive models, or models based on regression or classification techniques. However we were able to determine regions of interest for specific diet and infection conditions. The use of non-parametric statistics, Kruskal-Wallis ANOVA and Mann-Whitney Test, were of great benefit for the assessment of experimental group differences, they make the

interpretation of NIR absorbance signals harder since it translate the measurements to a ranks metrics. Moreover, in terms of presentation it might confuse a reader, since the NIR spectral signals are presented as figures and next to their statistics as tables, this is specially true for the bargraphs that we used to represent the pair-wise differences. Use of a spectroscopic instrument based on collimated lenses and the lack of temperature introduced more scattering thus more variability in the NIR spectral samples, and therefore impacted on our results, given the small sample size. In contrast to the PS experimental groups, we did not pick distinct spectral profiles between the PD experimental groups, possibly due to small sample size, or that biochemical adaptation to protein deficiency overrides infection status and is likely similar in both experimental groups. This requires further investigation.

Conclusion. We showed that NIRS could be used to identify specific differences resulted from a nematode infection and distinct protein deficiency during pregnancy and lactation. We found that even when the infection and diet burden where the subjected to the dams, the effects were visible in both the dams and their nursing pups. In general terms maternal serum differed from pup serum, and there was a diet effect only in the infected dams. For pups differences were found for both nutrition and infection effects that were present in the pups, with the NIR spectral profile of PS-UN pups being different from PS-IN pups and the NIR spectral profile of PS-IN pups differing from PD-IN pups. Our approach using NIR spectral profiles of maternal and with neonatal serum explored the degree to which these in which a chronic (trickle) nematode infection and protein deficiency would result in *lingering* metabolic consequences for the fetus and/or nursing pups. No one to date has considered that exposure during lactation could impact the offspring not directly exposed to either stressor and possibly be transferred through milk.

Table 4–1: *Comparisons among Protein and Infection Categories for Maternal and Pups serum* [a]

Region/Ratio	PS-UN	PS-IN	PD-UN	PD-IN	P Value
Lactating Dams	n=5	n=8	n=4	n=5	
Region 1:2	1.153^{AB} (1.09,1.36)	1.180^{A} (1.16,1.19)	1.086^{AB} (1.01,1.13)	1.102^{B} (0.92,1.12)	0.0001
Region 1:7	1.177^{AB} (-0.29,1.32)	1.341^{A} (1.21,2.47)	0.955^{AB} (0.60,1.17)	0.763^{B} (0.41,0.99)	0.017
Region 2:7	1.045^{AB} (-0.09,1.13)	1.133^{A} (1.05,2.10)	0.874^{AB} (0.59,1.04)	0.688^{B} (0.44,0.89)	0.020
Pups Serum	n=6	n=9	n=3	n=7	
Region 5	-0.029^{A} (-0.04,-0.03)	0.010^{B} (-0.02,0.02)	-0.003^{AB} (NA)	0.005^{AB} (0.00,0.04)	0.014
Region 1:4	1.391^{A} (0.96,2.17)	1.262^{B} (0.72,1.35)	-0.363^{AB} (NA)	0.985^{AB} (-0.07,1.23)	0.014
Region 2:3	1.070^{A} (0.91,1.20)	0.784^{B} (0.59,0.87)	-0.227^{AB} (NA)	1.044^{AB} (-0.07,1.10)	0.0001
Region 2:6	0.690^{AB} (0.47,0.76))	0.705^{A} (0.70,0.82)	0.655^{AB} (NA)	0.597^{B} (-0.09,0.64)	0.021
Region 2:7	1.028^{AB} (0.73,1.17)	1.041^{A} (0.86,1.30)	0.600^{AB} (NA)	0.881^{B} (-0.11,0.94)	0.001
Region 3:5	1.576^{B} (1.07,1.83))	1.985^{A} (1.89,2.25)	5.444^{AB} (NA)	1.258^{AB} (0.90,2.14)	0.001
Region 3:6	0.590^{B} (0.55,0.64)	0.906^{A} (0.89,1.21)	-2.884^{AB} (NA)	0.576^{AB} (0.40,1.25)	0.001
Region 6:8	1.054^{A} (1.01,1.20)	0.775^{B} (0.23,0.83)	-5.840^{AB} (NA)	0.849^{AB} (0.23,0.93)	0.001
Region 7:8	0.690^{A} (0.66,0.73)	0.523^{B} (0.27,0.53)	-6.380^{AB} (NA)	0.575^{AB} (0.21,0.63)	0.001

[a] NA: not a sufficient number of values to calculate interquartiles

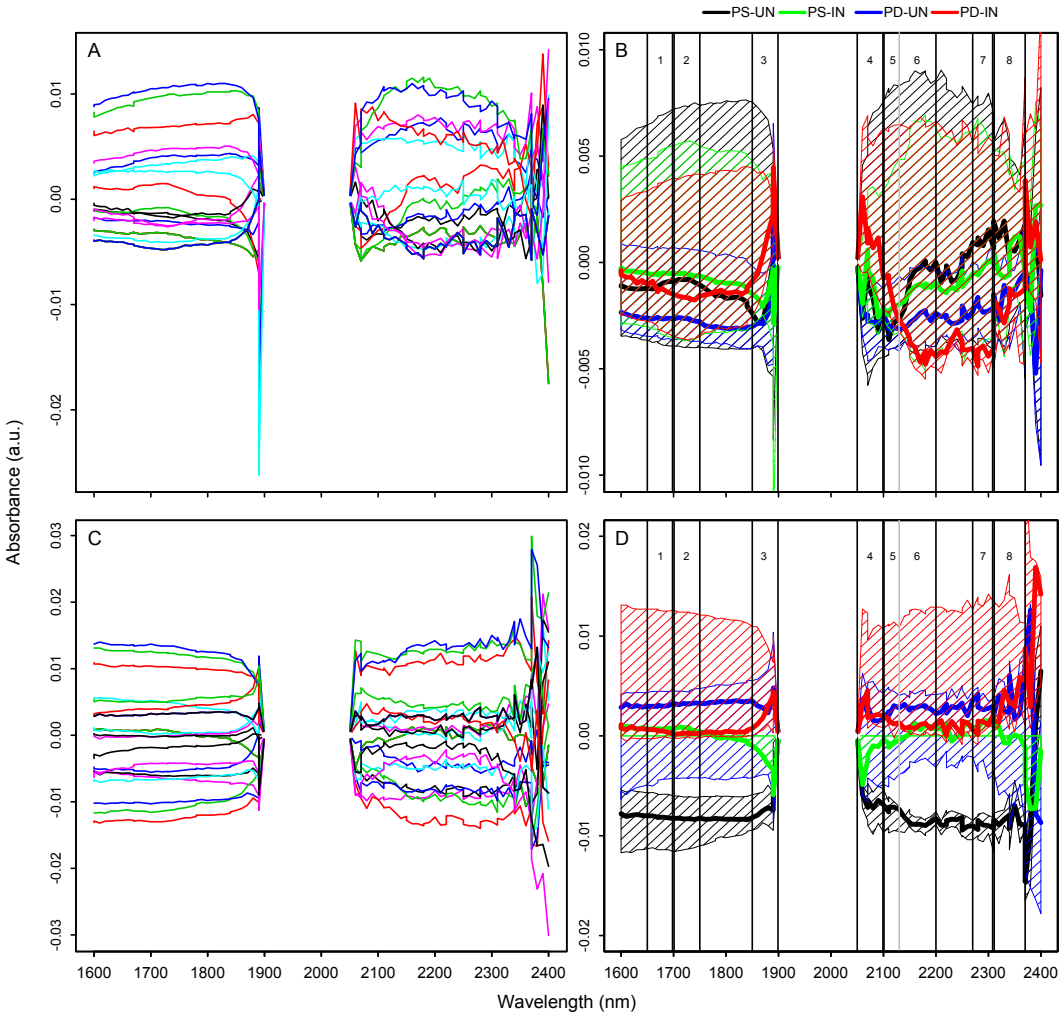


Figure 4–1: *NIR spectra for lactating dams (A, B) and pups (C, D). Median spectral absorbance divided by infection burdens and diets groups (solid black line represents the (PS-UN), a green line represents the (PS-IN), a blue line represents (PD-UN) and a red line represents the (PD-IN) for dams (B) and pups (D). Shaded regions, with respective colors represent the area between the first and third quartiles (1Q-3Q).*

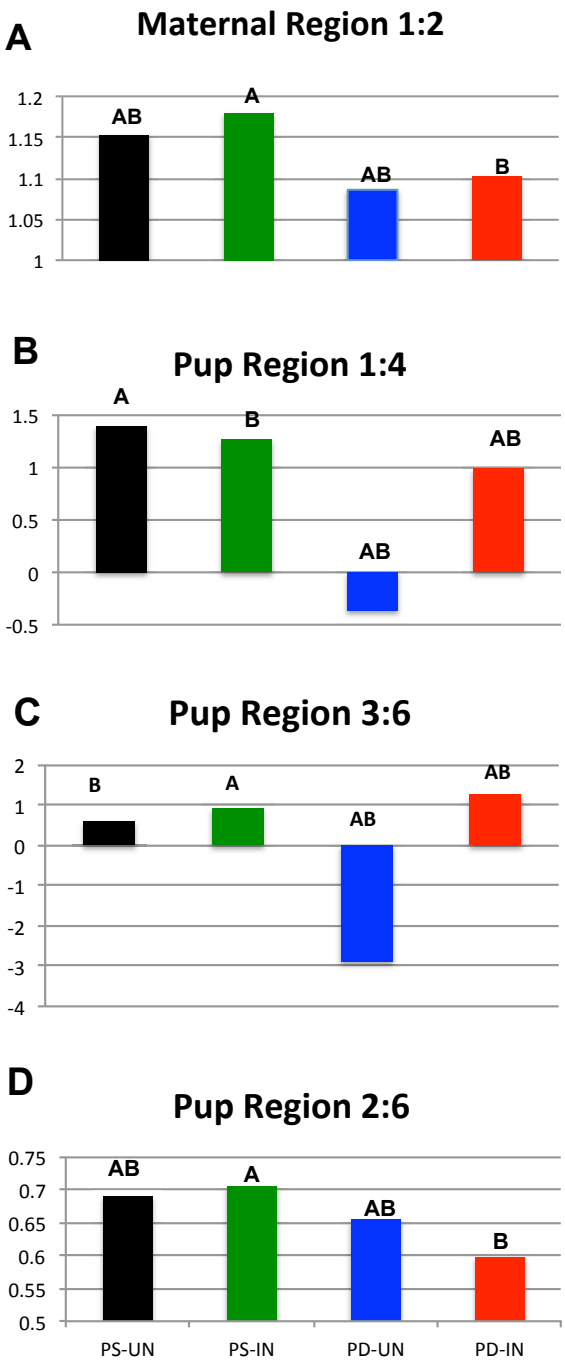


Figure 4–2: *Kruskal-Wallis Comparison among PS-UN, PS-IN, PD-UN and PD-IN for Lactating Dams and Nursing Pups. Pairwise Comparison between PS-IN and PD-IN for Region 1:2 in Lactating Dams (A). Pairwise Comparison between PS-UN and PS-IN for Nursing Pups for Region 1:4 (B) and Region 3:6 (C). Pairwise Comparison PS-IN and PD-IN for Nursing Pups for Region 2:6 (D)*

Figure 4–3: *Close up of the 2100nm-2200nm Spectral Region for Lactating Dams.*

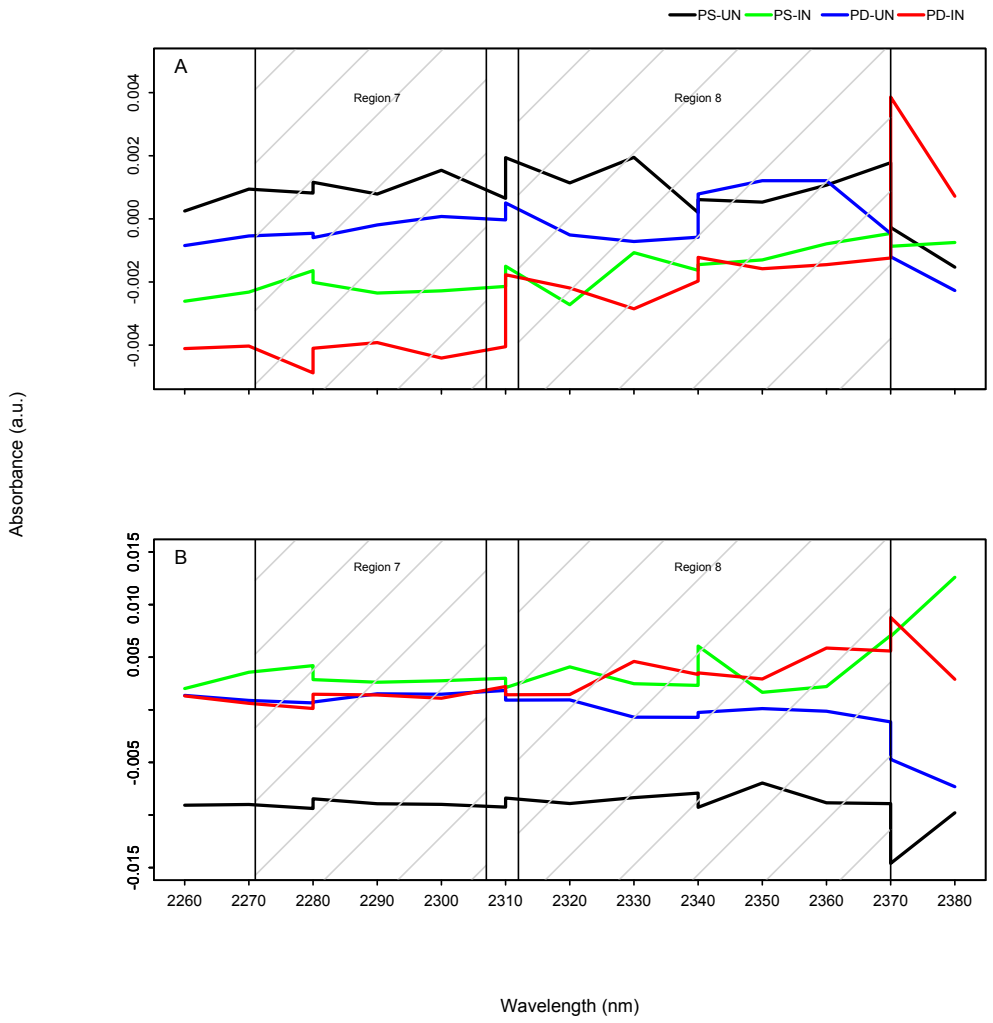


Figure 4–4: *Close up of Region 7 and Region 8 (2260nm-2380nm) for maternal serum (A) and pup serum (B).*

BRIDGING STATEMENT: Study 1 to Study 2

In the previous chapter we analyzed near-infrared spectral profiles in lactating dams and their pups and were able to distinguish different profiles as a result of dietary interventions and infection burdens. We observed differences in the spectral profiles of serum from pregnant and lactating mice and importantly in their offspring where there was a greater number of significant differences in NIR spectral profiles, indicating that fetal programming could be occurring in the pups despite experimental treatment not being applied directly to them.

However, this was a controlled experiment using a mouse model of a human hookworm with controlled doses of a *H. bakeri* nematode infection and some of these differences will be further explored in a larger human study where individuals have multiple nutrient deficiencies and multiple infections in Chapter 6.

In the next study we wanted to further explore, using a larger human population (n=590) and in a different biological fluid, the possibility of the early metabolic imprinting influencing fetal growth, in particular in the case of large-for-gestational age infant. Recently others have shown using 2nd trimester amniotic fluid that there exist early metabolic changes that are associated in the programming of large-for-gestational age infant [146, 9, 10].

We approach this serum analysis using the same methodological framework, that is, the integration of eight NIR spectral regions and use of non-parametric statistics.

Also, we were interested in investigating if maternal BMI had an impact on fetal development early in gestation.

CHAPTER 5
Study 2: NIR Metabolic Profiling of 2nd Trimester Amniotic Fluid Distinguishes Between Infants Subsequently Born Appropriate-for-Gestational Age (AGA) and Large-for-Gestational Age (LGA)

Javier E. Sanchez-Galan[1], Gary Luskey[2], David H. Burns[1,3] and Kristine G. Koski[1,4]

[1]Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, Canada H3A 1A3

[2]Division of Perinatal/Fetal Medicine, St. Mary's Hospital Center, Montreal, QC, Canada H3T 1H53

[3]Department of Chemistry, McGill University, Montreal, QC, Canada H3A 2K6

[4]School of Dietetics and Human Nutrition, McGill University (Macdonald Campus), Ste- Anne de Bellevue, QC, Canada H9X 3V9

Preliminary results from this Study were presented at the **14th Annual McGill Biomedical Graduate Conference (AMBGC)**, under the title*NIR analysis of 2nd Trimester Amniotic Fluid for the Determination of Gestational Length and Infant Birth Weight*. Montreal, Canada. March, 2014.

5.1 Abstract

Near-infrared (NIR) spectroscopy of amniotic fluid has been used to characterize fetal lung maturity and for the identification of prematurity <35 weeks. However, this approach has been seldom used to classify infants according to birth weight for gestational age. We explored the possibility that NIR spectral analysis of 2nd trimester amniotic fluid (AF) would identify early metabolic differences in birth weight among appropriate- (AGA, n=494) and large- (LGA, n=51)-for-gestational age infants. A secondary objective explored whether AF profiles of AGA and LGA infants differed by maternal pre-pregnancy BMI classifications normal (BMI<24.9 kg/m^2, n=310) versus overweight/obese (BMI>25 kg/m^2, n=168). Amniotic fluid samples were analyzed using NIR spectroscopy (1600-2400nm). The resulting spectra were pre-processed and eight regions of interest were selected: Region 1- 1650-1700nm, Region 2 -1701-1750nm, Region 3 1850-1899nm, Region 4 2050-2099nm, Region 5 - 2100-2130nm, Region 6 - 2130-2200nm, Region 7 - 2275-2310nm and Region 8 - 2311-2370nm. These regions and their ratios were compared among the experimental groups using Mann-Whitney U-test and Kruskal-Wallis one way ANOVA test, followed by post-hoc comparisons. By 15 wks gestation, fifteen AF spectral region ratios differed between AGA and LGA infants. LGA infants had (1) higher AF values for Region 8 relative to Region 1, Region 2, Region 3 and Region 4, (2) higher values for Region 6 relative to values found in Region 3 and Region 4, Region 8 and Region 5 and (3) higher values for Region 5 and Region 2 relative to Region 3 and Region 4. In contrast, AGA infants had (1) higher values for the ratio of Region 3:8, Region 3:7, Region 3:6 and Region 3:5 and (2) higher values for the ratios of Region 4:8, Region 4:6, Region4:5. In addition, AGA infants had lower values for Region 2:4 and higher values for Region 1:2, Region 1:8 and Region 2:8 compared with LGA infants. Further subdivision by maternal BMI categories showed that only AGA infants of overweight/obese

(BMI>25kg/m^2) mothers had higher values for Region 8 compared to normal weight mothers (BMI<24.9 kg/m^2; surprisingly this maternal BMI-based difference did not exist for LGA infants. Neither did AF differences exist for other integrated spectral nor ratios between AGA and LGA infants of overweight/obese mothers. In contrast, in normal weight mothers, 2nd trimester AF of these LGA infants had higher ratios of Region 7 relative to both Region 5 and Region 4 and a higher ratio of Region 3 relative to Region 4 when compared to their AGA counterparts. NIR spectral profiles of 2nd trimester AF differed between LGA and AGA infants and these AF differences emerge early in pregnancy in normal weight but not overweight/obese mothers. These results suggest that maternal BMI does not underscore these perturbations.

5.2 Introduction

Near-infrared spectroscopy (NIRS) is a vibrational spectroscopy technique that identifies small molecules present in a sample by relying on vibrational properties of their constituent chemical bonds. NIRS is an accurate and fast technique that has been proposed and used and in the context of clinical and medical diagnostics of biological fluids [46, 30, 45, 133]. It has been used a diagnostic tool for metabolic fingerprinting of cell culture media [114], for assessing viability of embryo cultures [50, 150, 151, 129, 132], for the calculation of total protein [98] and for identifying viral infections in biological fluids [122, 121, 120] and for fetal lung maturity using amniotic fluid [82, 83].

The influence of maternal health during pregnancy and its long-term consequences for fetal health is well known. Recently researchers have been examining the possibility that metabolic imprinting begins much earlier in pregnancy [111, 41, 36, 37, 2] and that there are important early metabolic changes that can be associated with the development of gestational diabetes mellitus [10, 144], or are associated with

infant birth weight [9, 145] and differences in gestational age [34, 42, 40, 88, 146, 37]. Moreover, the influence of maternal BMI on fetal macrosomia and ultimately on birth weight is also well known [57, 130] and is based on the presumption that obesity is the major cause of fetal programming for LGA infants. However there is a growing concern about the increasing prevalence of LGA births in normal weight mothers [23].

Previously we compared NIR differences between 2nd trimester amniotic fluid samples from mothers delivering at term vs. late pre-term and concluded that there were distinct metabolomic profiles between these conditions. Several wavelength regions that might be associated to proteins, carbohydrates, fats, polyols, and water were identified [110]. The overall objective of this human study was to assess the possibility of identifying, in 2nd trimester amniotic fluid (AF), distinct NIR spectral profiles associated with births of appropriate-for-gestational-age (AGA) and large-for-gestational-age (LGA) infants. Also, we were interested in determining if maternal body-mass-index (BMI) would further alter the spectral profiles of amniotic fluid (AF) of AGA and LGA infants early in pregnancy. To achieve this goal, a 4-way comparison among the median NIR profile of amniotic fluid of AGA and LGA infants from mothers with normal weights (BMI $<$24.9 kg/m^2) and obese mothers (BMI $\geq$ 25 kg/m^2) was conducted by integrating the absorbance values of eight NIR wavelength regions: 1650-1700nm, 1701-1750nm, 1850-1899nm, 2050-2099nm, 2100-2130nm, 2130-2200nm, 2275-2310nm and 2311-2370nm.

5.3 Materials and Methods

Recruitment of Study Population: From 1998 to 2004, pregnant women (n=790) undergoing routine, age-related amniocentesis, at St. Mary‘s Hospital Center in Montreal, Canada, were approached to participate in this prospective study. Signed consents allowed researchers to collect frozen amniotic fluid samples, following age-related

genetic testing, for subsequent biochemical analyses and to review maternal medical records. Ethics approval was obtained from Institutional Review Boards of McGill, Montreal Children's Hospital, and St Mary's Hospital Centre. Medical record information collected following delivery included pre-pregnancy weight, BMI, age, parity, ethnicity, infant gestational age, gender, and birth weight. Infant gender, gestational age, and birth weight were then used to calculate birth-weight-for-gestational-age [75]. Gestational age was based on physicians estimates using last menstrual period and uniform hospital protocols.

Exclusion criteria included the existence genetic abnormalities, multiple pregnancies, gestational diabetes mellitus or pre-eclampsia (n=73), mothers with BMI of <18.5 kg/m^2, and mothers giving birth to small-for-gestation age (n=45). In addition, other exclusion criteria were having incomplete clinical files, and lastly the presence of abnormal NIR spectral profile or spectral profiles collected outside the acceptable temperature range of 24.3 $\pm$1 °C (n=127). The final total number of NIR spectral samples included in the study was 545.

Experimental Design. Two retrospective analyses of 2nd trimester AF were performed. First, we compared 2nd trimester amniotic fluid NIR spectral profiles between the AGA (n=494) and LGA (n=51) birth categories. Secondly, these NIR spectral profiles were further subdivided by maternal BMI using the clinical cut-offs of normal ($\geq$18.5-<24.9 kg/m^2, n=310) and overweight/obese ($\geq$BMI 25 kg/m^2, n=168).

NIR Spectral Integration into regions of interest. For the analysis of the NIR spectral signals, a spectral integration approach was used. Instead of using all the wavelengths for which spectra were collected, only specific wavelength regions, where specific moieties are known to have a higher absorbance of light were considered.

Therefore, the 1600-2400 nm was selected because of the important higher energy transitions in the 1st overtone region (1600- 2000nm) and the combinations bands region (2000-2400nm) where more intense absorption are found, and thus providing the possibility of using for analytical purpose. The 1600nm to 2400nm wavelength range was divided into eight regions: Region 1 (1650nm-1700nm, spanning 6 wavelengths), Region 2 (1701nm-1750nm, spanning 5 wavelengths) and Region 3 (1850nm-1899nm, spanning 6 wavelengths). Also, 5 regions where selected in the combination bands: Region 4 (2050nm-2099nm, spanning 6 wavelengths), Region 5 (2100nm-2130nm, spanning 4 wavelengths), Region 6 (2130nm- 2200nm, spanning 8 wavelengths), Region 7 (2275nm-2310nm, spanning 5 wavelengths), Region 8 (2311nm-2370nm, spanning 8 wavelengths). The absorbance values for each of these eight regions were integrated into one value per sample. A similar approach, in which *spectral windows* are integrated and related back to a set of specific moieties, has been used in the spectral of analysis of other NIR [5] and FT-IR [108] data.

Near-infrared spectroscopy instrument and spectral collection. Prior to analysis, frozen samples were thawed on ice at room temperature and NIR spectra were acquired at room temperature (around 24°C) for amniotic fluid samples. A flow sample cell was filled with 10 μL of amniotic fluid. The sample cell was cleaned between measurements by rinsing it with 1mL of 0.1M NaOH followed by 5mL of distilled, deionized water. NIR spectra of water samples (40 μL) were taken as references before processing of each sample and saved for posterior spectral subtraction.

All NIR spectra were acquired using a NIR spectrometric system consisting of a LABPOD NIR spectrometer (Polychromix, Wilmington, Mass.), a Tungsten-Halogen

lamp (BWTEK BPS101, B&W TEK, Newark, DE) as a light source and a U-shaped flow sample cell with 1 mm path length (Starna Cell, Inc). NIR spectra for all experiments were collected as transmittance values in the 1600nm to 2400nm wavelength range using the SpectralCodePlus v3.5 software. Each single spectra represented the average of 10 scans at an integration time of 100ms over 100 wavelengths.

Spectral Data Processing. The Matlab programming package (ver. R2012a; The Math Works Inc., MA) was used for data processing and analysis. Sample absorbance was calculated as a ratio between the intensity of light absorbed by the sample (Is) and the intensity absorbed by the water reference (Ir), using the following formula: -log10 (Ir/Is). Values between 1900nm-2050nm were set to zero to minimize the effect of the high absorption water peak found around 1940nm [163, 91] Raw absorbance signals were smoothed for the existence of systematic errors, first, by applying a 3-point recursive boxcar-averaging algorithm [158] and later by mean centering to remove constant background from all the samples.

To analyze the relative changes in these eight regions, the notion of ratios or interactions was used to compare one spectral region to another. Ratios were calculated by dividing the mean centered absorbance values from different spectral regions. Also, it should be noted that ratios represent interactions between two chosen integrated spectral regions [6, 76]. Previously, this approach has been used by our group in the determination of spectral regions associated with pre-term pregnancies in human amniotic fluid [110].

Statistical Analysis. As the values obtained from the integrations of the 8 regions and their ratios were not normally distributed among the experimental groups (confirmed via Shapiro-Wilk test). Spectral region medians for AGA and LGA experimental groups were compared between this two experimental conditions using a

Mann-Whitney U test. The results are presented as medians and their values for the 1st and 3rd quartile. A 4-way comparison among the median NIR profile of amniotic fluid of AGA and LGA infants from mothers with normal weights (BMI $<$24.9 kg/m^2) and obese mothers ($\geq$BMI 25 kg/m^2) was also conducted. For those spectral regions that showed significant differences, a two-sided Mann-Whitney U test was applied with a Bonferroni test applied for post-hoc multiple comparisons.

A regression model with birth weight as dependent variable was built to determine which integrated spectral regions and/or ratios might be associated with birth weight when LGA infants were combined with AGA infants. This model was built using Stata Statistical Software: Release 11. (StataCorp. 2009. College Station, TX: StataCorp LP.). This regression model included a variable representing the intercept, knowing that birth weight cannot be zero and thus the regression never will pass through the origin [119, 29]. Moreover, since the spectra were mean centered and then summed, we anticipated that values for integrated spectral regions might indeed be zero. Also, the resulting model was checked after for multicollinearity by looking at the variance inflation factor (VIF) [100]. If the VIF factor for a specific spectral region was less than 5, it was eliminated from the model and the model was then re-calculated.

5.4 Results

Population Characteristics. Key population characteristics are summarized in Table 5–1 for age, pre-pregnancy weight, gestational age, and parity. On average, this population consisted of mostly healthy mothers (39$\pm$6ys), on their second or third pregnancy and giving birth AGA (91%) or LGA (9%) infants. The AGA and LGA experimental groups were further subdivided among normal BMI normal and

overweight and obese BMI categories, with the AGA group having 62% in the normal BMI category and 29% in the overweight and obese category and 3% and 6% respectively for the LGA.

Integration of spectral region Regions for Amniotic Fluid (AF). Processed NIR spectra for the 2nd trimester amniotic fluid samples of pregnant women (n=545) is shown in Figure 5–1-A. A simple inspection of the spectral profiles for all AF samples revealed that variability of the absorption values were evenly distributed across all wavelengths.

Samples were also classified according to intrauterine growth conditions: AGA (n=494), and LGA (n=51). These experimental groups are presented as spectral medians for each wavelength as solid red, black and blue line, respectively, in Figure 5–1-B. In this figure, AGA and LGA experimental groups, showed the same variability across all wavelengths.

Fifteen ratios differed between AGA and LGA infants. These were divided into two subgroups (Table 5–3); 1) those ratios that were higher in LGA infants compared to AGA infants and 2) the 13 ratios that were lower in LGA infants compared to AGA infants. In the first case, two Region 4 ratios - Region 1:4 (P=0.01) and Region 2:4 (P<0.01) showed that higher Region 1 and Region 2 relative to Region 4 were associated with birth of LGA infants. For the second case the following ratios - Region 1:2 (P<0.01), Region 1:8 (P<0.01), Region 2:8 (P<0.01), Region 3:8, (P<0.01), Region 4:8 (P<0.01), Region 5:8 (P<0.01) and Region 6:8 (P=0.02), were lower for LGA compared to AGA. Other ratios were lower in LGA compared to AGA infants.There were: Region 3:5 (P<0.01), Region 4:5 (P=0.01), Region 3:6 (P=0.01), Region 4:6 (P=0.02), Region 5:6 (P=0.04) and Region 3:7 (P<0.01), suggesting that

higher values for Region 3 and Region 4 were associated with birth of an AGA infant.

Comparison of AGA and LGA by Maternal BMI Categories. As shown in Figure 5–2 the median for AGA and LGA infants subcategorized by BMI classifications (normal and overweight/obese) are represented in black, red, green, and blue solids lines and their inter quartile values as shaded areas of the same colors. The spectral signals for the four categories were similar from 1600nm to 1900nm, with LGA-overweight/obese having in general a lower absorbance. In the 2050-2400 region, the AGA experimental group shows a U-shaped pattern, with high absorbance values around 2050nm, lower around 2200nm and high again towards the 2300nm. The LGA groups showed lower absorbance values, around Region 7 and Region 8, regions associated with Region 7 and Region 8.

Comparison of the overall AF spectral patterns for the four subgroups. We noticed that the AF of all LGA infants were metabolically similar in the NIR spectral regions from 2300-2400nm despite different maternal pre-pregnancy BMIs. Furthermore, NIR spectral signals of AGA infants, regardless of maternal BMI, were more similar between 2300-2400nm; these spectral regions might be associated with Region 7 and Region 8, suggesting that early pregnancy perturbations may be associated with fetal growth trajectories. Moreover, the spectral profile representative of the LGA infants from mothers with normal pre-pregnancy BMIs parallels that for AGA infants and only begins to diverge in the spectral region characterized by Region 7, again suggesting changes in Region 8 metabolism that is reflected in 2nd trimester AF. Finally the mean spectral profile of LGA infants for both overweight/obese mothers

differed from the other 3 experimental groups for most NIR spectral regions, suggesting that many more shifts in metabolism have occurred early in pregnancy for these individuals.

Median values and confidence intervals for our four experimental groups are shown in Table 5–3. Differences were found in one single region, Region 8 (P=0.03); otherwise there was a difference in the Region 3:8 ratio (P=0.02). Also two ratios to region 4, Region 1:4 (P=0.01), Region 2:4 (P<0.001) and two ratios to Region 5: Region 3:5, (P=0.05) and Region 4:5 (P=0.04) differed as did three ratios to Region 7: Region 3:7 (P=0.02), Region 4:7 (P=0.02) and Region 5:7 (P=0.02). Post-hoc comparisons revealed five types of differences: 1) Differences between AGA infants regardless of maternal BMI status, which were only found for Region 8, shown in Figure 5–3-A; 2) Differences between normal AGA infants and overweight/obese LGA infants for Region 3:8, shown in Figure 5–3-B; 3) Differences among normal BMI infants: Region 5:7, shown in Figure 5–3-C; 4) Differences between normal LGA infants and AGA infants regardless of BMI status: Region 1:4, Region 2:4 and Region 4:7, shown in Figure 5–3-D; and 5) Global differences that did not yield post-hoc differences for Region 3:5, Region 3:7 and Region 4:5.

Regression model for infant size at birth. Results from the overall regression model for birth weight are described in Table 5–4. The overall model was statistically significant ($R^2 = 0.03$, P<0.001) with three significant spectral region ratios Region 5:7 (P=0.02), Region 3:7 (P=0.02), and Region 4:8 (P=0.01) entering the model. The first ratio was positively related to birth weight whereas the remaining were negatively associated with birth weight. There were other integrated spectral

regions that were significant after the stepwise regression, but were removed due to collinearity (VIF >5): specifically: Region 1 (VIF=11.10); Region 7 (VIF=50.64) and Region 6 (VIF=79.33). All of the other ratios did not emerged as significant differences.

5.5 Discussion

Our study population consisted of mothers undergoing age related amniocentesis (age 37 ± 6 ys). The population consisted of 9% of LGA births, which are values comparable to the Canadian population at large [157]. One of our objectives was to determine if there were significant spectral differences between AGA and LGA infants. We hypothesized that NIR spectral profiles of 2nd trimester AF might reflect metabolic differences in fetal growth between AGA and LGA infants in second trimester amniotic fluid, if indeed early differences in growth trajectories might emerge early in pregnancy (as early as 2nd trimester) as suggested by other studies [9, 10, 37, 146, 26, 42, 110, 69]. Several important observations emerged: 1) First, distinct AF NIR spectral profiles were found for LGA and AGA infants by the 2nd trimester; 2) values for Region 8 were lower for AGA infants of normal weight mothers compared to obese/overweight mothers but spectral profiles for this region did not different among LGA infants, despite changes in maternal BMI; 3) in general comparisons of AGA and LGA in normal weight mothers revealed higher Region 7 to both Region 4 and Region 5 and higher Region 2 to Region 4 for LGA infants and higher Region 3 and Region 4 for AGA infants and 4) a final multiple regression model for ratios for birth weight for all infants combined showed 3 ratios emerging. Region 5:7 positively associated and Region 3:7 and Region 4:8 negatively correlated with larger infant birth weights, suggesting that more amides, possibly urea [163] or ceramides [93] and higher Region 8 and Region 7 were associated with birth of

LGA infants. Taken together these results suggest the early emergence of distinct metabolic profiles for LGA and AGA infants that are reflected in amniotic fluid by second trimester.

BMI. The effects of maternal BMI on infant size are very well documented, and it is especially well known that mothers who are overweight or obese are more likely to have LGA infants [96, 153]. Contrary to this, our results suggested the early emergence LGA infants in normal weight mother and that metabolic differences existed between AGA and LGA infants within the normal BMI categories but the 2nd trimester of pregnancy. For mothers with normal BMI Region 1:4, Region 2:4, Region 3:7, Region 4:5, Region 4:7 and Region5:7 differed between AGA and LGA, suggesting that different spectral profiles emerge early in pregnancy and that it can be detected in amniotic fluid. In contrast in mothers with BMI ≥ 25, no differences were found in the amniotic fluid. Only Region 3:5 and Region 3:8 differed across the AGA normal and the overweight/obese LGA, suggesting a trend for these spectral regions. For both ratios the AGA had higher values for Region 3, which is associated with Region 3 and lower Region 5/Region 8 absorbance values compared with NIR spectral profile of the LGA overweight/obese. Moreover, the suggestion that maternal overweight or obesity might influence metabolic profiles and can be reflected in NIR spectra of amniotic fluid was supported only by differences found in Region 8. There were no differences between the two LGA categories by maternal BMI, suggesting that there were no differences and more similarities in fetal compartment early in pregnancy. Whereas a number of studies have described the association higher maternal weight to LGA, there is an emerging growing concern about normal weight mothers giving birth to LGA infants [66, 166, 138, 57]. This study points to distinct early metabolic perturbations occurring in LGA infants of normal weight mothers. Taken together these data suggest that the mechanism of fetal programming of a LGA infant occurs in

mothers with a normal BMI and may be distinct from LGA associated with mothers with BMIs >25. This will require further investigation

Region 8. Disturbances in Region 8 were evident by the beginning of the second trimester, and were influenced by maternal pre-pregnancy BMI. We saw a relative increase in Region 8 in our LGA infants compared to our AGA infants. Overall, we saw an increase in Region 3:7 when we compared across our BMI categories, where the amniotic fluid of overweight and obese mothers that gave birth to LGA infants had higher values relative to Region 3. We also saw that AGA infants of normal mothers differed from AGA infants from overweight and obese mothers, whereas the same difference did not emerged among LGA counterparts, suggesting that Region 8 alone is not sufficient to explain the metabolic perturbations that underscore the birth of an LGA infant.

Region 4:8. Another ratio that emerged in relation to birth weight was Region 4:8, where higher absorbance values of Region 4 was associated with birth of a LGA infant. One of the many molecules that could be associated with the Region 4 might be related to polyphenols. There is evidence that associates higher polyphenols with less diet induced obesity and with increase anti-oxidant capacity. Interestingly higher values in Region 8 in AF have been associated with both LGA and GDM [146, 9, 10].

Region 3:7. Interestingly AGA infants had higher Region 3 relative to Region 6, Region 5, Region 7 and Region 8. Higher Region 3 could be associated with phosphate (ATP) production. However the only one that emerged in the overall model was Region 3:7, suggesting less Region 3 and more Region 7 in amniotic fluid of LGA infants. One suggestion that could be related to increased amniotic fluid Region 7 might be hypoxia. However, a recent study [56] showed that there were no differences in lactic production among SGA, AGA and LGA infants and they

concluded that hypoxia could not explain differences in birth weight. Other causes of increased Region 7 in LGA infants need further exploration.

Region 5. A third important ratio was Region 5:7. In our results, it was associated with LGA infants, suggesting that birth of an LGA infant was associated with higher abundance n 2nd trimester AF of amides (urea) and amines (purine, pyrimidines and indoles, where tryptophan is an example of the latter). This in line with results relating higher amides, the primary one being urea, with earlier kidney development in the LGA infant [31] and with more ceramide for brain development [93] and with tryptophan is reportedly higher concentrations in the AF of 2nd trimester [59], which may to be the case in our LGA populations.

Finally the results of the regression model for birth weight suggest that Region 5:7 and are positively related with birth weight, while Region 3:7, Region 4:8 were negatively associated with birth weight. These results show that higher Region 5 is associated with increasing birth weights between (1.2 to 5.2 kg) and increased organ development of the LGA infants. Lower values of Region 3 and Region 4 were associated with birth of an LGA infant, moreover higher values of Region 8 and Region 7 relative to Region 4 and Region 3 start to indicate positive fetal growth.

Strengths and Limitations. This study presents evidence for the existence of early programming of an LGA infant regardless of maternal BMI using a large biobanked library of frozen AF samples. Although these findings were measured in older women undergoing age related amniocentesis, they raise the possibility that a similar outcome might be occurring in the general population. Despite the fact our NIR spectral absorbance signals were small when compared to our previous studies [110], the use of the median of the groups and the new application of ratios helped us uncover differences among the experimental groups. The use of non-parametric

statistics, Kruskal-Wallis ANOVA and Mann-Whitney Test, were of great benefit for the assessment of experimental group differences, they make the interpretation of NIR absorbance signals harder since it translate the measurements to a ranks metrics. Moreover, in terms of presentation it might confuse a reader, since the NIR spectral signals are presented as figures and next to their statistics as tables, this is specially true for the bargraphs that we used to represent the pair-wise differences. The differences between LGA infants from BMI normal and overweight/obese divided are still not evident and will require further investigation.

Conclusion. Our results are important because they provide evidence for the early emergence in amniotic fluid of distinct metabolic profiles between AGA and LGA infants. They also show important metabolic differences by the 2nd trimester between AGA and LGA infants, with notable difference between AGA versus LGA in normal weight mothers than in overweight or obese mothers. This suggests that the mechanism of fetal programming of a LGA infant is not only determined by maternal BMI but may already be occurring and affecting other metabolic pathways. Moreover, we were able to show that the integration of integrated spectral regionand their ratios can be used to separate between birth outcomes, but still cannot be used to distinguish between classes of LGA infants in clinical settings as a more robust discriminatory model would.

Table 5–1: *Maternal Characteristics*

General Characteristics (N=545)	
Age (ys) (N=534)	37 ± 6
Gestational Age (wks) (N=543)	39.5 ± 2
Infant size by Gestational Age (N=545)	
AGA	494 (91%)
LGA	51 (9%)
BMI (kg/m^2) by Gestational Age Category (N= 478)	
AGA	
Normal ($\geq$18.5 and $\leq$ 24)	294 (62%)
Overweight and Obese ($\geq$25)	140 (29%)
LGA	
Normal ($\geq$18.5 and $\leq$24)	16 (3%)
Overweight and Obese ($\geq$25)	28 (6%)

Table 5–2: *Comparison between AGA and LGA*

Region/Ratio	AGA	LGA	P Value
	n=494	n=51	
	Lower in AGA and Higher in LGA		
Region 1:4	0.918 (-0.02, 0.00)	1.167 (-0.02, -0.00)	0.010
Region 2:4	0.772 (-0.02, 0.01)	0.976 (-0.02, -0.00)	0.006**
	Higher in AGA and Lower in LGA		
Region 1:2	1.198 (-0.02, 0.01)	1.187 (-0.02, -0.00)	0.005**
Region 1:8	0.458 (-0.02, 0.00)	0.218 (-0.01, 0.00)	0.001**
Region 2:8	0.384 (-0.02, 0.00)	0.208 (-0.02, -0.00)	0.005**
Region 3:5	1.096 (-0.04, 0.00)	0.875 (-0.04, 0.00)	0.004**
Region 3:6	0.480 (-0.03, 0.01)	0.301 (-0.03, -0.00)	0.014
Region 3:7	0.720 (-0.05, 0.02)	0.355 (-0.05, 0.02)	0.001**
Region 3:8	0.389 (1.16, 1.24)	0.109 (1.02, 1.21)	0.001**
Region 4:5	1.387 (0.79, 1.65)	1.119 (0.84, 2.04)	0.014
Region 4:6	0.603 (0.61, 1.29)	0.445 (0.73, 1.64)	0.018
Region 4:8	0.449 (1.14, 1.60)	0.172 (0.96, 1.51)	0.003**
Region 5:6	0.464 (0.48, 0.80)	0.419 (0.29, 0.67)	0.041
Region 5:8	0.342 (0.54, 1.21)	0.260 (0.33, 0.99)	0.010
Region6:8	0.812 (0.22, 0.75)	0.632 (-0.17, 0.52)	0.019

Table 5–3: *Mann-Whitney Comparisons between AGA and LGA infants divided by maternal BMI (normal and overweight/obese)*

	BMI Categories [a]				
	BMI <24.9 kg/mm^2		**BMI $\geq$ 25 kg/m^2**		
Region/Ratio	**AGA**	**LGA**	**AGA**	**LGA**	**P Value**
	n=294	n=1	n=140	n=28	
Region 8	-0.0297^{A} (-0.05, 0.01)	-0.0324^{AB} (-0.06, 0.01)	-0.0210^{B} (-0.04, 0.05)	-0.0222^{AB} (-0.05, 0.01)	0.03
Region 1:4	0.9398^{A} (0.64, 1.34)	1.2855^{B} (0.98, 2.47)	0.9332^{A} (0.60, 1.27)	1.1341^{AB} (0.69, 1.56)	0.01
Region 2:4	0.7990^{A} (0.51, 1.13)	1.0906^{B} (0.82, 2.46)	0.7817^{A} (0.49, 1.02)	0.9511^{AB} (0.65, 1.43)	0.01
Region 3:8	0.3972^{A} (0.15, 0.60)	0.1961^{AB} (0.02, 0.41)	0.4024^{AB} (0.08, 0.57)	0.1192^{B} (-0.13, 0.38)	0.02
Region 4:7	0.8696^{A} (0.45, 1.23)	0.4416^{B} (0.09, 0.70)	0.8700^{A} (0.31, 1.14)	0.6416^{AB} (0.37, 1.18)	0.02
Region 5:7	0.6643^{A} (0.48, 0.83)	0.4283^{B} (0.20, 0.63)	0.6558^{AB} (0.40, 0.78)	0.6304^{AB} (0.43, 0.91)	0.02
Region 3:5 [b]	1.0986^{A} (0.81, 1.43)	0.9104^{A} (0.67, 1.29)	1.0919^{A} (0.61, 1.35)	0.7401^{A} (0.34, 1.18)	0.05
Region 3:7	0.7143^{A} (0.40, 1.00)	0.3942^{A} (0.04, 0.67)	0.7702^{A} (0.21, 0.95)	0.3992^{A} (0.15, 0.77)	0.02
Region 4:5	1.3753^{A} (1.07, 1.60)	1.0042^{A} (0.43, 1.29)	-1.4014^{A} (1.01, 1.55)	1.1009^{A} (0.69, 1.58)	0.04

[a] AGA and LGA underweights were excluded from this analysis

[b] were found to be significant in the overall comparison, but significant only at P(¡0.06) after bonferroni post-hoc tests.

Table 5–4: *Multiple Linear Regression for birth weight in kg for LGA and AGA infants combined*

	Birth weight (in kg) (n=545)	
Region/Ratio	β Coefficient	P value
Intercept	3.46	<0.001
Region 5:7	0.05	0.02
Region 3:7	-0.03	0.004
Region 4:8	-0.02	0.001
	R^2=0.03	P<0.001

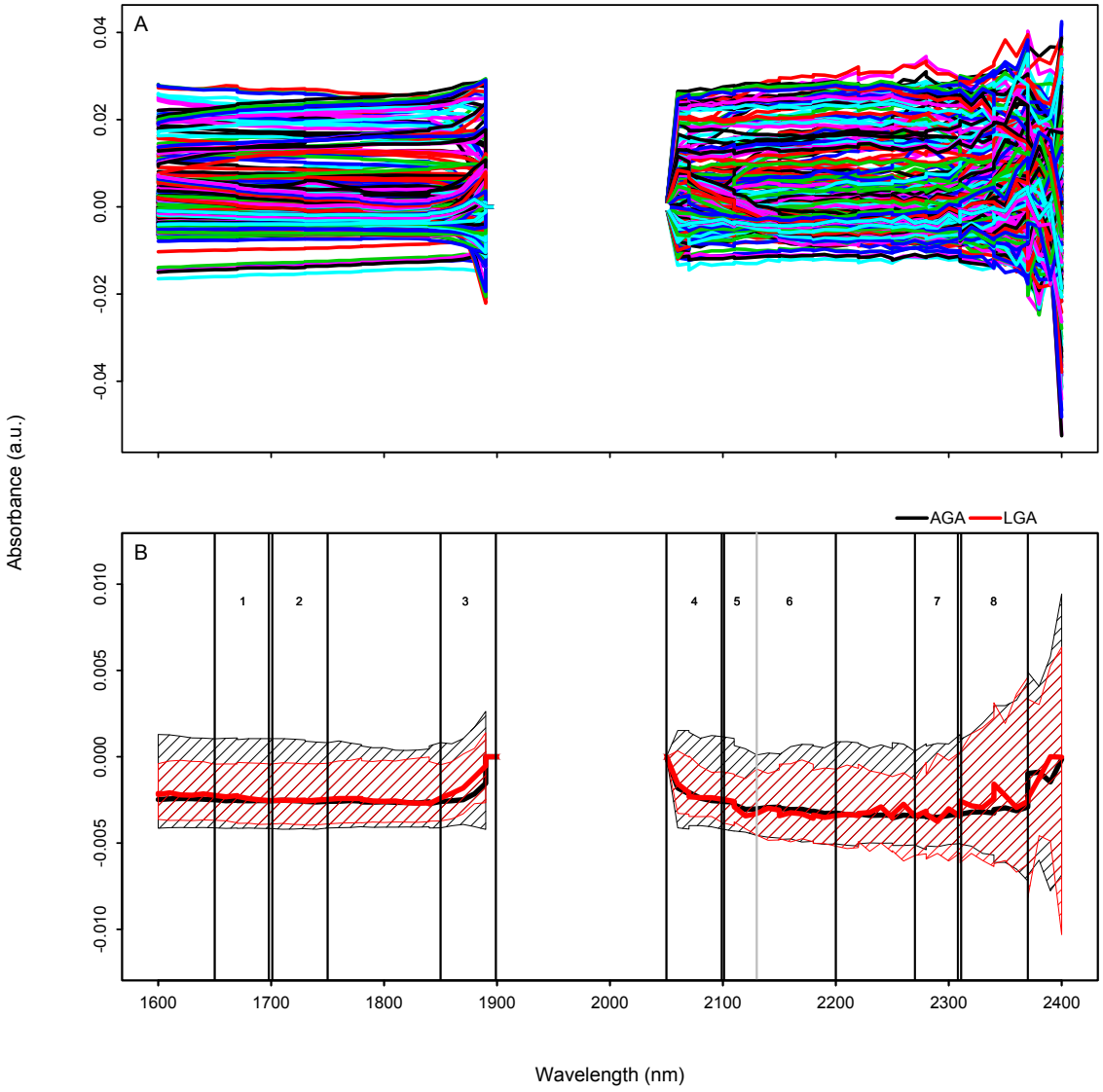


Figure 5–1: *Smoothed and mean centered NIR spectra of 590 amniotic fluid samples (A). Median spectral absorbance grouped by size corrected for gestational age and area comprised between the 1st and 3rd quartiles (B). AGA and LGA groups and areas are represented in red, black and blue solids lines and shared areas respectively.*

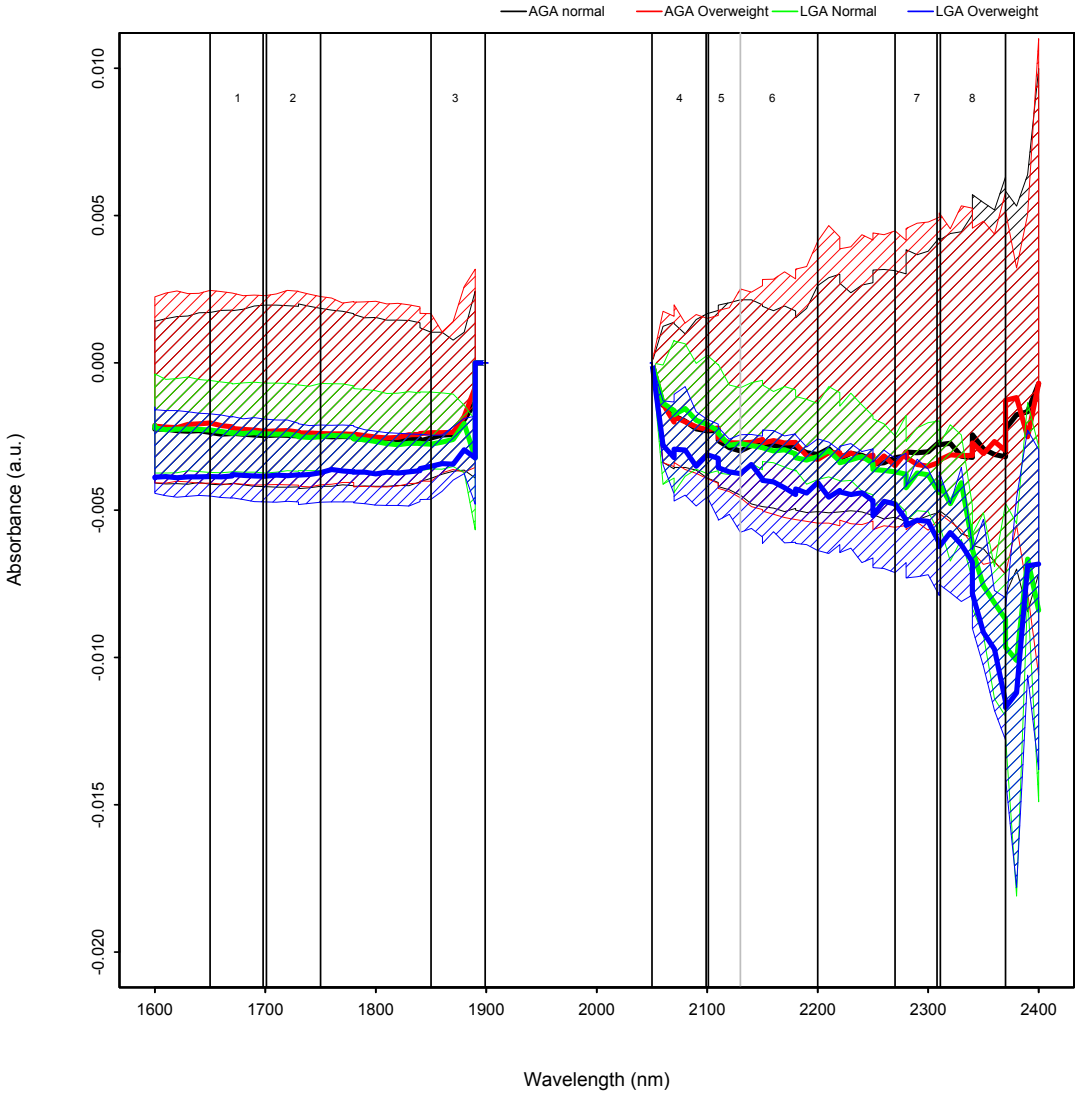


Figure 5–2: *Median spectral absorbance and area comprised between the 1st and 3rd quartiles for AGA and LGA infants grouped by maternal body mass index (A), BMI groups are represented as solid lines, normal AGA, overweight AGA, normal LGA and overweight LGA, obese groups and areas are represented in black, red, green, and blue solids lines and shared areas respectively.*

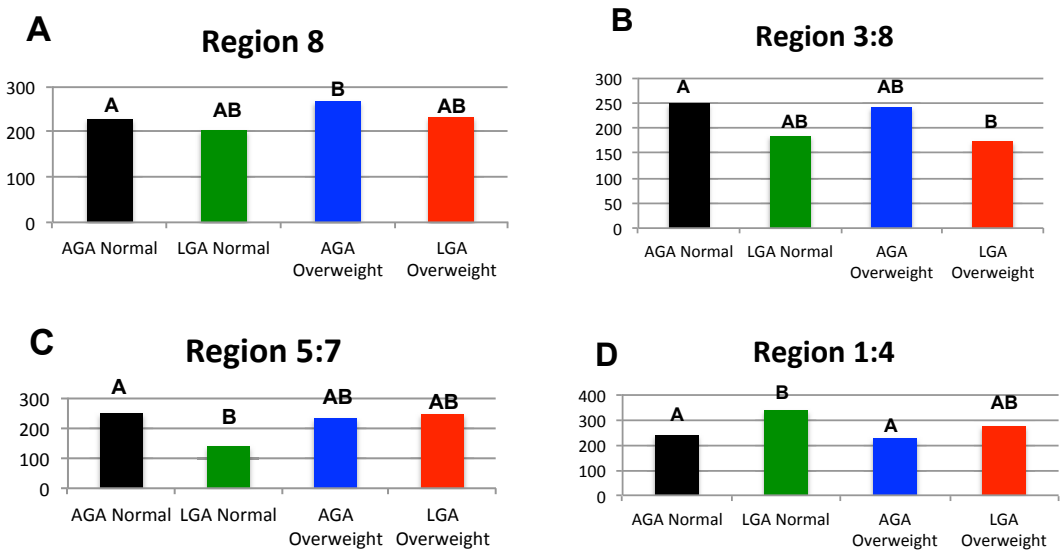


Figure 5–3: *Kruskal-Wallis Comparison by BMI and Birth Weight Categories. Difference between AGA Normal and AGA Overweight (A and B), Difference between AGA Normal and LGA Normal (C) and Difference between LGA Normal and AGA groups (D)*

BRIDGING STATEMENT: Study 2 to Study 3

In the Study 1 we showed that NIR spectroscopy was able to used to identify differences in the levels of nematode infection and single protein deficiency for mice. In Study 2, we focused on two pregnancy outcomes (LGA and AGA) and the ways it might change the signals of amniotic fluid. Moreover, we linked maternal BMI information to further classify the outcomes.

In this next chapter we propose to investigate if NIR spectroscopy could distinguish between multiple infections and multiple micronutrient deficiencies in a human population. For this we analyzed the NIR spectral signals of serum samples (n=190) from pregnant indigenous women in Panama who have a range of micronutrient deficiencies (vitamin A, vitamin D, vitamin B12 and folic acid) and multiple infections (oral and skin). Moreover we set to investigate if any NIR spectral signals changes could be associated with fetal growth status using the fundal height to gestational age (FH:GA) ratio, a notion of health status that is often used in pregnancy in rural environments.

The importance of this last study is threefold: first, we expand the infection and nutrition paradigm from Study 1 that explored a controlled macronutrient deficiency and single nematode infection on mice and put to test in a larger scale since, we are going to look into human samples from a rural setting (an indigenous community) were both, multiple infections and multiple nutrition deficiencies are abundant. Second, our results might be used as a suggestion for an important policy change regarding the supplementation of pregnant women in rural Panama requiring both vitamin D

and folic acid throughout pregnancy. Third, the results from this study can help elucidate question regarding what type of infection or which micronutrient could be further explored in human serum. Moreover, this study set the base for further investigations in the development of a portable NIR instrument that could be used to detect these deficiencies in a rural setting.

We approach this serum analysis using the same methodological framework of Study 1 and Study 2, that is, the integration of eight NIR spectral regions and use of non-parametric statistics.

CHAPTER 6
Study 3: Use of Near-Infrared Spectroscopy for the analysis of infection, nutrition and fetal growth for pregnant women in a rural community in Panama

Javier E. Sanchez-Galan[1], Doris Gonzalez-Fernandez[2], Marilyn E. Scott[2], David H. Burns[1,3] and Kristine G. Koski[1,4]

[1]Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, Canada H3A 1A3

[2]Institute of Parasitology, McGill University (Macdonald Campus), Ste-Anne de Bellevue, QC, Canada H9X 3V9

[3]Department of Chemistry, McGill University, Montreal, QC, Canada H3A 2K6

[4]School of Dietetics and Human Nutrition, McGill University (Macdonald Campus), Ste- Anne de Bellevue, QC, Canada H9X 3V9

Preliminary results from this Study were presented at **Experimental Biology 2014**, under the title: *Near-infrared spectroscopy characterizes vitamin deficiencies and infection status during pregnancy.* San Diego, California. April 2014.

6.1 Abstract

Metabolomic approaches using near-infrared (NIR) spectroscopy have been used to identify metabolic differences in uninfected and infected hosts, however, they have not been applied to identification of multiple nutritional deficiencies nor to multiple infections during pregnancy. We explored the possibility that NIR spectral analysis would identify differences in a) maternal micronutrients deficiencies (A, D, B12, folic acid), b) presence/absence maternal infections (skin and oral) and c) whether the combination of these micronutrients deficiencies and infections would have an impact on fundal height, a measurement of in utero growth. Maternal serum samples (n=190) were analyzed using NIR spectroscopy, their spectral signals were divided into 8 NIR spectral regions: 1650-1700nm (Region 1), 1701-1750nm (Region 2), 1850-1899nm (Region 3), 2050-2099nm (Region 4), 2100-2130nm (Region 5), 2130-2200nm (Region 6), 2275-2310nm (Region 7) and 2311-2370nm (Region 8). These regions and their ratios were compared among the experimental groups using Mann-Whitney U test. Also, a similar comparison was done for 2nd and 3rd trimester on the fundal height to gestational age (FH:GA) variable. Overall significant differences were found for vitamin D, B12 and folic acid. Oral infections were significant for 2nd trimester, however not for the 3rd trimester. In terms of growth, FH:GA was associated with vitamin D and folic acid in the 2nd and 3rd trimesters; B12 emerged only for the 2nd trimester only. In conclusion our results suggest that NIR spectral changes is able to discriminate between micronutrient deficiency or sufficiency stares, however the presence or absence of infections were non-trivial to determine. These results are important, because they become the starting point for the analysis of biological fluids in rural context. Finally, this is a first attempt to examine at both multi micronutrients deficiencies and infections in the same pregnant women using NIR spectroscopy.

6.2 Introduction

Near-infrared spectroscopy (NIRS) is a vibrational spectroscopy technique that has been previously used for identifying viral infections in biological fluids including serum [122] and nasal fluids [121, 120], but there is still much left to be done in other types of infections such as those caused by nematodes and other infections that are pertinent in rural areas lacking water systems and latrines. These populations are known to be highly vulnerable to various infections and have prevalence for poor growth and other conditions enhanced by malnutrition [97, 25]. Little work has been done for the simultaneous characterization of the impact of dietary deficiencies and infections burdens during pregnancy on fetal growth [68, 74].

The influence of maternal health during pregnancy and its long-term outcomes on fetal growth is known [1]. It is a main concern in developed countries but it is especially critical in developing countries and in rural areas where extreme poverty exists and where nutrition deficiencies and infections coexist and are known to affect women in childbearing age [155]. A major cause for concern in these type of environments is intrauterine growth restriction (IUGR), also known as fetal grow restriction (FGR), often observed prenatally by low fundal height over gestational age and postnatally by low birth weight (LBW), or when an infant is born small-for-gestational age (SGA). Both conditions are long term predictors of childhood stunting [18]. Another factor that affects normal infant growth are micronutrient deficiencies [32, 12].

The overall objective of this study was to assess the possibility of identifying distinct NIR spectral profiles and spectral regions representing molecules and integrated spectral of interest that could be associated with the prevalence of micronutrient

deficiencies (vitamin A, vitamin D, B12 and folic acid) and presence or absence of infections (oral and skin) in pregnant women belonging to the Ngabe-Bugle , indigenous community in rural Panama. We hypothesized that 1) by using near-infrared spectroscopy we could see changes in the spectral profiles of serum samples from pregnant women with multiple infections and multiple micronutrient deficiencies and that we would be able to associate these NIR spectral profiles with fetal growth using the FH:GA measurement.

6.3 Materials and Methods

Samples Collection and Study Population. This study was conducted in the Ngabe-Bugle indigenous *Comarca*, an indigenous administrative area in Western Panama in the province of Chiriqui. Over 90% of the population lives in extreme poverty and households rely on small-scale agriculture (corn, tubers and pif), or seasonal labor (coffee, sugar cane, livestock). Untreated river or stream water is delivered to 35% of homes by gravity through plastic tubes, 30% of homes have latrines, and homes lack electricity. Cooking is done on wood stoves and organic waste is used as fertilizer [48].

From this population n=213 pregnant women were selected to participate in the study and 10mL of venous blood samples were collected over a four month period, mostly via medical consultations. Also, relevant medical information such as: anthropometric measurements, including height and weight, clinical exam for infections including skin and oral (caries and gingivitis) was also gathered either from medical records or directly by a member of the research team during the medical consultation. As a measure of fetal growth, fundal height (FH) was measured using a measuring tape. Fundal height is described as the distance from the pubic symphysis to the top

of the uterine fundus. It is a well-defined measure that can be used in rural settings and is an alternative to the ultrasound measurements done in hospitals when compared to gestational age (GA) [159]. This measurement is based on the assumption that fundal height is more or less an indicative measurement that can be related to gestational age within an error margin of 1 to 3 cm.

Serum Analysis Protocols. Analysis for vitamin A (n=209) was done on-site in the San Felix Hospital, at the Gorgas Memorial Institute, at University of Panama and Institute of Scientific Research and High Technology Services INDICASAT (Panama City). Vitamin D, folic acid and vitamin B12 were assayed in Montreal, Canada, on sera stored at -20°C. Vitamin D (total 25(OH)D) was measured in Montreal using the LIAISON direct competitive chemiluminescence immunoassay (Agilent Technologies, Waldbronn, Germany; minimum detectable concentration 9.98 nmol/L). Folic acid and vitamin B12 (minimum detectable concentrations 1.45 nmol/L and 22.14 pmol/L respectively) were measured using immunoelectrochemiluminescence with a ROCHE 241 Modular E170, Elecsys Module (Basel, Switzerland). The definition of sufficiency and deficiency groups among the samples was done by using the clinical cut-offs for micronutrient deficiencies as follows: for water soluble vitamins: <10nmol/L for folic acid, and <150pmol/L for vitamin B12 [24] and for fat soluble vitamins: <30μg/dL for vitamin A [27] and <50nmol/L for vitamin D [105]. In the case of infections, classes were defined using a binary variable taking as values presence or absence of a given infection.

Exclusion Criteria. A detailed description of the collection methodology, inclusion and exclusion criteria, procedure of the medical consultation and interview and respective ethical approvals are described in the materials and methods sections

in previous studies carried in our group [161, 141, 39, 167]. The participants in their first trimester (gestational age less or equal to 12 weeks) were excluded because fundal height is an appropriate measurement only for the late second (gestational age between 13 and 26) and third (gestational age greater or equal to 27 weeks) trimesters of pregnancy [159]. This resulted in a final sample size n=192. Since 3 samples did not have fundal height measurements the number was reduced to n=189. Another exclusion criterion was set for the number of samples per clinical category. Infections and nutrition conditions with less than 10 samples in a category (sufficiency or deficiency state, or presence or absence of an infection), were excluded.

Calculation of the fundal height to gestation age ratio. Part of the objectives of this study was to find a way to relate NIR spectra, or spectral regions, to growth outcomes of fetal health. A variable describing the growth relationship between fundal height to gestational age, FH:GA was created for this purpose. It relates the fundal height measurement to that of gestational age (which ranged from 5 to 42 weeks among participants) and can be used as a rough estimate of fetal growth and intra uterine development. A value of FH:GA ≥ 1 indicates a healthy growth and a value of FH:GA <1 indicate an individual with compromised growth.

NIR Spectral Integration into regions of interest. For the analysis of the NIR spectral signals, a spectra integration approach was used. Instead of using all the wavelengths for which spectra were collected, only specific wavelength regions, where specific moieties are known to have a higher absorbance of light were considered. Therefore, the 1600-2400 nm was selected because of the important higher energy

transitions in the 1st overtone region (1600- 2000nm) and the combinations bands region (2000-2400nm) where more intense absorption are found, and thus providing the possibility of using for analytical purpose. The 1600nm to 2400nm wavelength range was divided into eight regions: Region 1 (1650nm-1700nm, spanning 6 wavelengths), Region 2 (1701nm-1750nm, spanning 5 wavelengths) and Region 3 (1850nm-1899nm, spanning 6 wavelengths). Also, 5 regions where selected in the combination bands: Region 4 (2050nm-2099nm, spanning 6 wavelengths), Region 5 (2100nm-2130nm, spanning 4 wavelengths), Region 6 (2130nm- 2200nm, spanning 8 wavelengths), Region 7 (2275nm-2310nm, spanning 5 wavelengths), Region 8 (2311nm-2370nm, spanning 8 wavelengths). The absorbance values for each of these eight regions were integrated into one value per sample. A similar approach, in which *spectral windows* are integrated and related back to a set of specific moieties, has been used in the spectral of analysis of other NIR [5] and FT-IR [108] data.

Near-infrared spectroscopy instrument and spectral collection. Prior to analysis, frozen samples were thawed on ice at room temperature and spectra were acquired with a temperature controller set for 23°C. A flow sample cell was filled with 40 μL of serum. The sample cell was cleaned between measurements by rinsing it with 1mL of 0.1M NaOH followed by 5mL of distilled, deionized water. NIR spectra of water samples (40 μL) were taken as references before processing of each sample and saved for posterior spectral subtraction.

All NIR spectra were acquired using a NIR spectrometric system consisting of a LABPOD NIR spectrometer (Polychromix, Wilmington, Mass.), a Tungsten-Halogen lamp (BWTEK BPS101, B&W TEK, Newark, DE) as a light source and a U-shaped

flow sample cell with 1 mm path length (Starna Cell, Inc). NIR spectra for all experiments were collected as transmittance values in the 1600nm to 2400nm wavelength range using the SpectralCodePlus v3.5 software. Each single spectra represented the average of 10 scans at an integration time of 100ms over 100 wavelengths.

Spectral Data Processing. Matlab programming package (ver. R2012a; The Math Works Inc., MA) was used for data processing and analysis. Sample absorbance was calculated as a ratio between the intensity of light absorbed by the sample (Is) and the intensity absorbed by the water reference (Ir), using the following formula: -log10 (Is/Ir). Values between 1900nm-2050nm were set to zero to minimize the effect of the high absorption water peak found around 1940nm [91, 163]. Raw absorbance signals were smoothed for the existence of systematic errors, first, by applying a 5-point recursive boxcar-averaging algorithm [158] and later by mean centering to remove constant background from all the samples.

Statistical Analysis. To analyze the relative changes in these eight regions, the notion of ratios or interactions was used to compare one spectral region to another. Ratios were calculated by dividing the mean centered absorbance values from different spectral regions. Previously, this approach has been used by our group in the determination of spectral regions associated with pre-term pregnancies in human amniotic fluid [110].

As the values obtained from the integrations of the 8 regions and their ratios were not normally distributed among the experimental groups (confirmed via Shapiro-Wilk test), spectral regions medians were compared among experimental conditions using Mann-Whitney U-test. These medians were presented as values and the values for

the 1st and 3rd quartile, which represented values above and below the median of an integrated spectral region or region ratio.

Some medians and confidence intervals were negative in value due to a mathematical artifact of the calculation of ratios that arise when dividing positive values above the mean between negative values below the mean or vice-versa. Because we wanted to compare differences among experimental groups using mean-centered NIR spectral values of our selected regions, the sign component of our medians was not as relevant as the absolute value to our interpretation. We interpret this comparison of positive or negative medians to say the average value was above or below the median. For this we present the confidence intervals for comparison purposes among the groups.

6.4 Results

Population Characteristics. From the maternal information files there were n=192 mothers in total with n=83 in their second trimester and n=109 in their third trimester. The mean maternal age was 24±7 years old and parity was 3.7 for both groups. As expected for fundal height, mothers in the second trimester had a lower (18.3 ±0.7 cm) than those on the third trimester (31.1 ± 0.36 cm). In terms of micronutrient deficiencies, 8% of the mothers were vitamin A deficient in the second trimester and 1% in the third; 59% of the mothers were vitamin D deficient in the second trimester and 66% in the third; 80% of the mothers were vitamin B12 deficient in the second trimester and 92% in the third and 23% of the mothers were folic acid deficient in the second trimester and 27% in the third. For infections we found that 22% of the mothers had a skin infection in the second trimester and 20% in the third; also, 20% of the mothers had oral infection in the second trimester and 27% in the

third trimester.

Processed NIR spectra for pregnant women samples from the second trimester (n=80) and from the third (n=109) trimester are shown in Figure 6–1-A and Figure 6–1-B, respectively. An inspection of the both spectrum revealed greater absorption and variability in the longer wavelengths of 2050-2400nm which corresponded to the integrated regions: Region 5, Region 6, Region 7 and Region 8.

Second trimester. Spectra divided by experimental conditions for second trimester are shown in Figure 6–2. In this figure the spectral medians of the deficient/infected group in red and the median of the sufficient/non infected group in black. The spectra for serum divided by micronutrient deficiencies are shown from top to bottom: vitamin A (n=36 and n=46, sufficient and deficient, respectively, Figure 6–2-A), vitamin D (n=31 and n=52, Figure 6–2-B), vitamin B12 complex (n=12 and n=71, Figure 6–2-C), and folic acid (n=70 and n=13, Figure 6–2-D). A general common trend that was observed in the median spectral signal, in Figure 6–2, was that the micronutrient deficient groups had a lower absorbance for Region 7 and Region 8, respectively. Also it was possible to observe that there was a significant overlap between in the inter-quartile regions (depicted in Figure 6–3 as red and black shaded regions) for sufficient and deficient mothers.

In examining the integrated NIR spectral regions (Table 6–1), statistical differences were found for vitamin D, B12 and folic acid. For vitamin D there were differences between sufficient and deficient for 6 spectral region, with values for Region 3 and Region 7 being higher in the sufficient group. In contrast, vitamin D deficient women had higher values for Region 1, Region 2, Region 4 and Region 5 relative to

Region 6 and Region 7. Figure 6–6-A describes, in the form of a bargraph, the difference between Vitamin D sufficient and deficient groups for the spectral region ratio of Region 5:7, with the sufficient group having higher values, than their deficient counterparts. A similar patter was observed for Region 1:6 is shown in Figure 6–6-C, where the sufficient group had higher values when compared to their deficient counterparts. For B12, just one ratio differed with B12 sufficient having higher Region 3 relative to Region 4. Interestingly, no NIR spectral region differences where found for vitamin A.

For folic acid we observed 3 single spectral group regions that were significant, with Region 6, Region 7 and Region 8 higher in deficient mothers. In addition 4 ratios differed, with Region 3 relative Region 1 and Region 2, Region 6 relative to Region 2 and Region 5 higher in the deficient groups. For folic acid sufficient mothers higher values were found for Region 1 and Region 2 relative to Region 3 and Region 6, and Region 5 relative to Region 6. Figure 6–6-B shows the difference for Region 1:3, where the deficient group had higher values compared to the sufficient group. The similar pattern was observed for Region 2:6 shown in Figure 6–6-D.

In the case of infections, skin (n=68 and n=15, Figure 6–2-E) and oral (n=73 and n=10, Figure 6–2-F), showed higher absorbance in Region 8, when compared to the non-infected group (black and red respectively). When looking at the spectral regions for infection, Table 6–1, only one statistical differences was found for oral infection in Region1:2 (P=0.01).

Third trimester. NIR spectral patterns divided by micronutrient deficiencies and infection conditions for third trimester are shown in Figure 6–3. As with Figure 6–2, the panels on the left represent micro nutrients, from top to bottom: vitamin

A (n=65 and n=43, sufficient and deficient, respectively, Figure 6–3-A), vitamin D (n=39 and n=70, Figure 6–3-B), vitamin B12 complex (n=18, and n=91, Figure 6–3-C), and folic acid (n=88 and n=21, Figure 6–3-D). A visual examination of the spectral patterns showed that for Vitamin A and D the spectra for sufficient and deficient groups have very small signal showing minimal absorbance, while in the case of B12 and folic acid the profiles for the deficient group has in both cases a lower absorbance. As with Figure 6–2, for 2nd trimester the plots in Figure 6–3 for 3rd trimester showed significant overlap between the inter-quartiles for both sufficient and deficient groups. One notable difference can be seen in the shaded region for folic acid, where a very distinct peak is found in Region 8.

In examining the differences between integrated spectral (Table 6–2) several differences were found for vitamin D for four spectral regions and for eight ratios. For single integrated spectral regions: Region 1 (P=0.03), Region 2 (P=0.03), Region 3 (P=0.03), Region 4 (P=0.03), differed with the deficient group having higher absorbance than the sufficient group. Two distinct ratio patterns occurred, for ratios involving Region 6 and ratios involving Region 8. In the case of ratios to Region 6, where Region 1:6 (P=0.02), Region 2:6 (P=0.02), Region 5:6 (P=0.03) had lower medians in the deficient group when compared to the sufficient one. This difference is shown for Region 5:6 in Figure 6–7-A. The ratios involving Region 8: Region 1:8 (P=0.03), Region 2:8 (P=0.03), Region 3:8 (P=0.02), Region 4:8 (P=0.02), and Region 5:8 (P=0.03), had higher values for Region 8 in the deficient group when compared to the sufficient vitamin D group. This difference is shown for Region 3:8 in Figure 6–7-B.

For folic acid there were three single regions, Region 6 (P<0.01), Region 7 P<0.01), and Region 8 (P<0.01), in which the folic acid sufficient group had higher values. There were also two region ratios: Region 1:6 (P=0.05) and Region 2:6 (P=0.04), where Region 6 had higher median values in the deficient group. This difference is shown for Region 1:6 in Figure 6–7-C.

For NIR spectral profiles, the following infections were characterised for skin (n=88 and n=21, Figure 6–3-E) and oral infections (n=92 and n=30, Figure 6–3-F) for the infected mothers it was observed that the spectral values had lower absorbance when compared to non-infected mothers. As with the other conditions a significant overlap existed between in the inter-quartile regions (depicted in Figure 6–3 as red and black shaded regions) for sufficient and deficient mothers.

Analysis of the FH:GA as Proxy Variable of Fetal Growth. The population distribution for FH:GA is depicted in Figure 6–4-A with mothers in the second trimester in red circles and mothers in the third trimester as blue squares. For the second trimester there were 46 (57%) of the mothers that had a FH:GA ratio under the 0.95 cut-off point, whereas in the third trimester there were 97 (89%) of the mothers with FH:GA <1 (Table 6–3). Boxplots of the FH:GA ratio variable for second and third trimester are shown in Figure 6–4-B. In this boxplot we see that the mothers in our study on average are below a value of 1 for the FH:GA ratio with medians of 0.92 and 0.89, for 2nd and 3rd trimester respectively. Despite the fact that were more mothers in the third trimester than the second there was more variability in the distribution in the latter, where extreme values ranged from 0.4 up to 1.5. In

contrast, for the third trimester values clustered around the median.

Definition of cut-offs for the Fundal Height. Based on the values and the overall distribution of the FH:GA variable, samples were divided by trimester. Different cut-offs values were investigated (Table 6–3), but only those cut-offs that created an appropriate sample size in each experimental group for comparison, were selected for each trimester. A final cut-off of 0.95 for 2nd trimester and of 1.00 for 3rd trimester were chosen for NIR spectral divisions for FH:GA, resulting in n=80 mothers in their second trimester and n=109 in their third trimester. The spectral medians are shown in Figure 6–5-A and Figure 6–5-B, respectively.

NIR spectral medians for both the 2nd and 3rd trimesters showed a similar pattern with the median for samples below the cut-off, depicted as a solid red line, having lower absorbance than the median for samples above the cut-off, depicted as a solid black line; differences were most prominent in Region 8. Also we observed that 2 spectral regions (Table 6–4) Region 3 (P=0.05) and Region 4 (P=0.04), and one region ratio, Region 3:7 (P=0.04) differed between those above and below the medians for the second trimester. For third trimester only one ratio differed Region 6:8 (P=0.04) with Region 8 being higher in those samples with FH:GA < 1.00.

6.5 Discussion

Maternal serum is a complex biological fluid that is usually analyzed during pregnancy for determination of pregnancies at increased risk of Down‘s syndrome [117] or physical malformations such as neural tube defects [117]. Less is done for the detection of malnutrition or defects caused by maternal infections. Near-infrared

spectroscopy is a technology that might be used to provide a rapid way to identify metabolic changes in biofluids. In the present study we analyzed NIR spectra of serum collected from pregnant women living in conditions of extreme poverty in rural Panama and investigated how multiple micronutrient deficiencies and infections might affect the NIR spectral profiles of maternal serum samples and how these NIR spectral profiles might be related to a fetal growth. This is a first attempt to assess both multiple micronutrients deficiencies and infections using NIR spectroscopy in the same serum samples. Major findings included: 1) The impact of vitamin deficiencies were similar for both trimesters with vitamin D and folic acid being important modulators in both the 2nd trimester and 3rd trimester. However, the number of significant integrated spectral regions differed between 2 nutrients. For Vitamin D significant differences were found for NIR spectral ratios to Region 6 and Region 7 for 2nd trimester and Region 6 and Region 8 for 3rd trimester, whereas for folic acid differences were found for ratios to Region 3 and Region 6 for 2nd trimester and ratios to Region 6 in the 3rd trimester. 2) a single region ratio, Region 3:4, was found to be a significant for B12 and only in the 2nd trimester but vitamin A did not appear as a factor for either trimester; 3) Finally, there were differences in the NIR spectral regions related to FH:GA for each trimester. In the second trimester Region 3 and Region 4 were associated with FH:GA ≥ 0.95, whereas in the third trimester increased Region 6:8 was associated with FH:GA ≤ 1.

Vitamin D. Differences in NIR spectral regions were found for Vitamin D for both 2nd and 3rd trimester. In the 2nd trimester 6 spectral region ratios were significantly different between sufficient and deficient, whereas in the 3rd trimester 4 single spectral regions and 8 spectral region ratios differed. However, three of them: Region 1:6, Region 2:6 and Region 5:6, emerged for both trimesters and were in the same direction, with higher values for Region 6 associated with vitamin D sufficiency.

Interestingly, in the 3rd trimester, four single spectral regions (Region 1, Region 2, Region 3, Region 4) differed with higher values in the deficient group. However when they were ratioed to other groups, higher values were associated with the sufficient group; only Region 8 and Region 5 were associated with the deficient group. Taken together this data suggest that vitamin D deficiency may play a more important role both in the second and third trimester of pregnancy. As suggested by our results, vitamin D may play an important role in Region 8 and Region 5 metabolism relative to Region 3, Region 4, Region 6 and Region 2, which is in line with other reports that show that vitamin D deficiency is associated with the development GDM [64], poor bone growth [78], brain development [4, 49] and pre-eclampsia [51], which prevalent in this Panamanian population [154].

Folic Acid. Similar to vitamin D results, folic acid deficiency was related to distinct metabolic perturbation in both the second and third trimesters. In both the second trimester and third trimester, Region 6, Region 7 and Region 8 were higher in folic acid deficient mothers in both trimesters. However, in the 2nd trimester 4 other ratios differed, whereas the 3rd trimester only 2 spectral region ratios differed, only one of which (Region 1:6) was the same for both trimesters. In the later case, higher values of Region 2 was associated with deficiency. For folic acid higher values for Region 6 were associated with deficiency. Given that folic acid emerged in both trimesters, this finding would emphasize the importance of continued folic acid supplementation in both trimesters, an idea that has been recently suggested by others [92].

FH:GA. In our study, we explored the possibility that micronutrient deficiencies might be related with specific NIR spectral regions that in turn would be associated with FH:GA. In the second trimester, we found that FH:GA was associated with Region 4 and both being associated with a sufficient B12 status. In the third trimester, the relation between NIR spectral region with FH:GA was associated with increased Region 6:8 and higher values for Region 6 in turn were associated with both vitamin D and folic acid deficiency.

Strengths and Limitations. In this study we were able to find distinct metabolic profiles for several vitamins in the serum of indigenous Panamanian women. Moreover, we were able to find specific spectral profiles despite the presence of other infections burdens and micronutrient deficiencies. Interestingly, several integrated spectral and some ratios, were sufficient to determine difference between the sufficiency/deficiency and presence/absence of multiple conditions.

A weakness of this study was that there were no controls as most women had multiple infections and multiple deficiencies. To create a *control* we would need to reexamine our data to see if we could identify a sufficient sized subset that might be infection free and might not have multiple nutrient deficiencies. Currently only 3% of the population is healthy, which still too small a sample for statistical analyses.

Conclusion. This study provides a methodology for the application NIRS analysis to serum samples. Our results also suggest that NIR could be used to determine differences between sufficient and deficient states for Vitamin D and folic acid for women in for both 2nd and 3rd trimesters. NIRS was not able to identify distinct

metabolic profiles for presence/absence of infections; only oral infection revealed distinct NIR profile and only in 2nd trimester. Taken together suggesting that NIRS might be used in a portable instrument to screen for vitamin deficiencies in marginalized pregnant women in rural populations, but not their infection status. Finally, our study using NIRS has yet to provide a clear picture of the candidate spectral regions that might be used to study fetal growth, nevertheless our results suggest some interesting areas for further exploration.

Table 6–1: *Comparisons between Spectral Group Regions divided by Micronutrients and Infection for 2nd Trimester*

Micronutrients			
Region/Ratio	**Sufficient**	**Deficient**	**P Value**
Vitamin D	n=46	n=36	
Region 1:6	0.465 (-0.06, 0.11)	1.006 (-0.03, 0.22)	0.012
Region 2:6	0.381 (-0.05, 0.10)	0.796 (-0.02, 0.18)	0.011
Region 3:4	1.254 (-0.11, 0.09)	1.126 (-0.04, 0.19)	0.038
Region 5:6	0.260 (-0.07, 0.07)	0.475 (-0.01, 0.17)	0.011
Region 5:7	0.015 (-0.03, 0.07)	−0.046 (-0.02, 0.11)	0.045
Region 6:7	0.190 (-0.11, 0.13)	−0.035 (-0.09, 0.18)	0.016
B12	n=12	n=71	
Region 3:4	1.458 (-0.07, 0.09)	−0.106 (-0.05, 0.21)	0.025
Folic Acid	n=70	n=13	
Region 6	0.070 (-0.05, 0.21)	1.674 (-0.05, 0.09)	0.005
Region 7	−0.202 (-0.04, 0.17)	−0.704 (-0.04, 0.07)	0.002
Region 8	0.232 (-0.07, 0.21)	−0.782 (-0.10, 0.07)	0.019
Region 1:3	1.118 (-0.04, 0.18)	0.432 (-0.07, 0.06)	0.024
Region 2:3	0.911 (-0.02, 0.10)	0.313 (-0.04, 0.03)	0.025
Region 2:6	0.648 (-0.08, 0.18)	0.186 (-0.22, 0.01)	0.038
Region 5:6	0.428 (-0.48, 0.13)	0.129 (-1.18, -0.24)	0.050
Infection			
Region/Ratio	**Not-Present**	**Present**	**P Value**
Oral	n=73	n=10	
Region 1:2	1.253 (-0.05, 0.16)	1.230 (0.01, 0.21)	0.014

Table 6–2: *Comparisons between Spectral Group Regions divided by Micronutrients and Infection for 3rd Trimester*

Micronutrients			
Region/Ratio	**Sufficient**	**Deficient**	**P Value**
Vitamin D	n=39	n=70	
Region 1	0.040 (-0.08, 0.14))	0.078 (-0.03, 0.25)	0.03
Region 2	0.027 (0.22, 0.75)	0.063 (0.22, 0.75)	0.03
Region 3	−0.011 (-0.06, 0.11)	0.066 (-0.02, 0.20)	0.03
Region 4	−0.012 (-0.08, 0.13)	0.055 (-0.03, 0.18)	0.03
Region 1:6	0.510 (-0.04, 0.07)	0.916 (-0.02, 0.11)	0.02
Region 1:8	0.046 (-0.11, 0.13)	0.009 (-0.08, 0.20)	0.03
Region 2:6	0.385 (-0.66, 0.66)	0.737 (-0.65, -0.02)	0.02
Region 2:8	0.034 (-1.55, 1.46)	0.006 (-1.26, 0.77)	0.03
Region 3:8	0.035 (1.23, 1.29)	−0.011 (1.23, 1.29)	0.02
Region 4:8	0.030 (0.70, 1.28)	−0.007 (0.74, 1.37)	0.02
Region 5:6	0.283 (0.84, 1.71)	0.438 (1.02, 1.81)	0.03
Region 5:8	0.023 (1.65, 2.39)	0.004 (1.73, 2.41)	0.03
Folic Acid	n=18	n=91	
Region 6	0.065 (-0.05, 0.21)	−0.094 (-0.05, 0.17)	0.009
Region 7	−0.198 (-0.04, 0.17)	−0.685 (-0.04, 0.14)	0.002
Region 8	−0.038 (-0.05, 0.19)	−1.234 (-0.08, 0.12)	0.008
Region 1:6	0.849 (-0.04, 0.15)	0.306 (-0.07, 0.10)	0.05
Region 2:6	0.685 (-0.02, 0.10)	0.242 (-0.04, 0.07)	0.04

Table 6–3: *Number of samples divided by different Cut-offs divided by Trimesters*

2nd Trimester		
FH:GA (n=80)	# Below (%)	# Above (%)
0.85	27 (34)	53 (66)
0.90	35 (44)	45 (56)
0.95	46 (57)	34 (42)
1.00	62 (77)	18 (22)
1.05	66 (82)	14 (17)
1.10	71 (89)	9 (11)
3rd Trimester		
FH:GA (n=109)	# Below (%)	# Above (%)
0.85	32 (29)	77 (71)
0.90	57 (52)	52 (48)
0.95	77 (71)	32 (29)
1.00	91 (84)	18 (16)
1.05	102 (94)	7 (6)
1.10	105 (96)	4 (4)

Table 6–4: *Mann-Whitney Tests for Fundal Height divided by Trimesters*

2nd Trimester	**FH:GA < cut-off**	**FH:GA ≥ cut-off**	**P Value**
Region/Ratio	**Group**	n=46	n=36
Region 3	0.026 (-0.08, 0.14)	−0.059 (-0.13, 0.04)	0.05
Region 4	0.026 (-0.07, 0.11)	−0.050 (-0.10, 0.03)	0.04
Region 3:7	−0.073 (-0.06, 0.13)	0.061 (-0.11, 0.05)	0.04
3rd Trimester	**FH:GA < cut-off**	**FH:GA ≥ cut-off**	**P Value**
Region/Ratio	n=91	n=18	
Region 6:8	0.112 (-0.17, 0.13)	−0.034 (-0.12, 0.06)	0.04

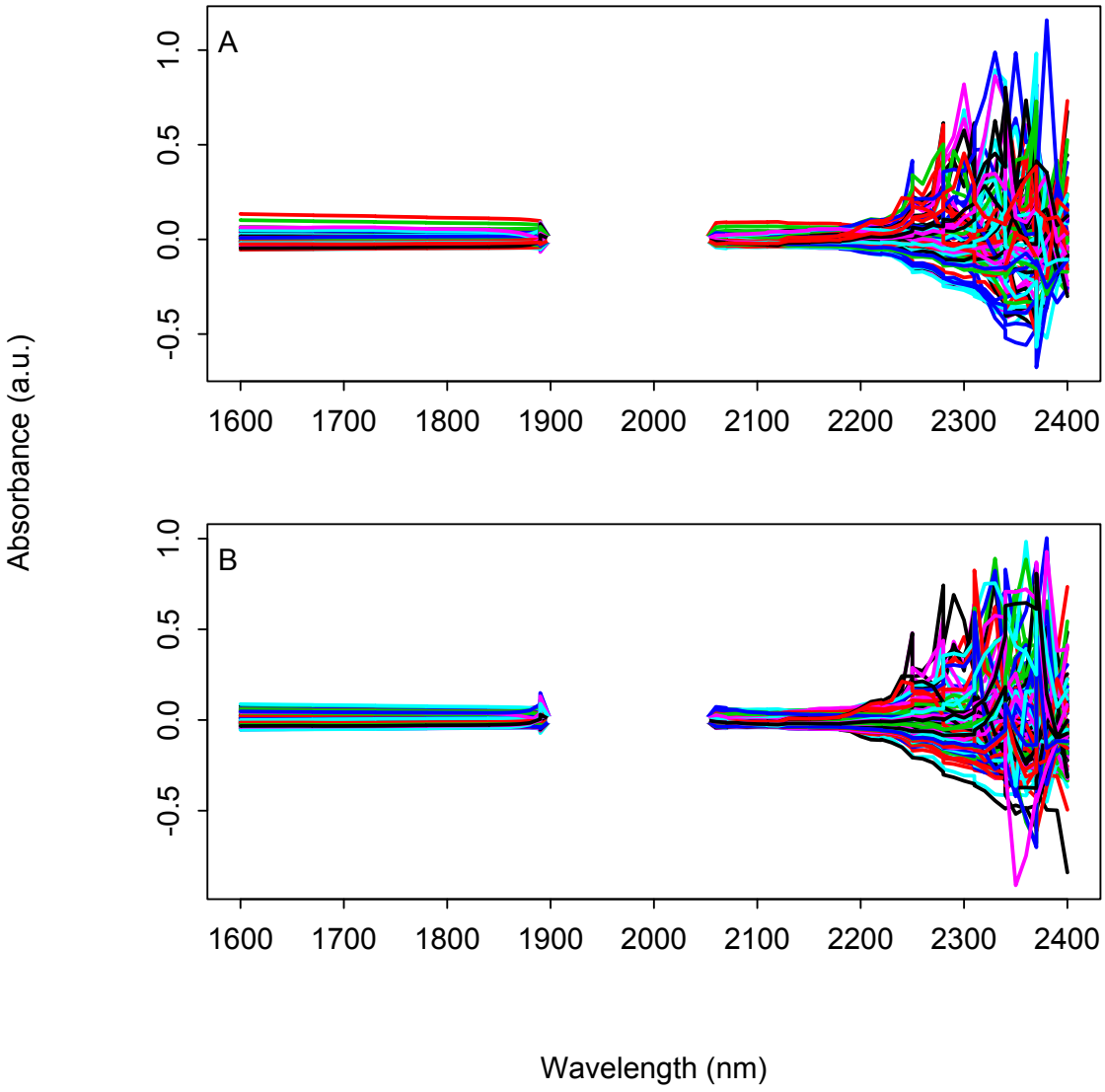


Figure 6–1: *Smoothed and mean centered NIR spectra of serum for n=80 second trimester serum samples (A) and n=109 third trimester serum samples (B)*

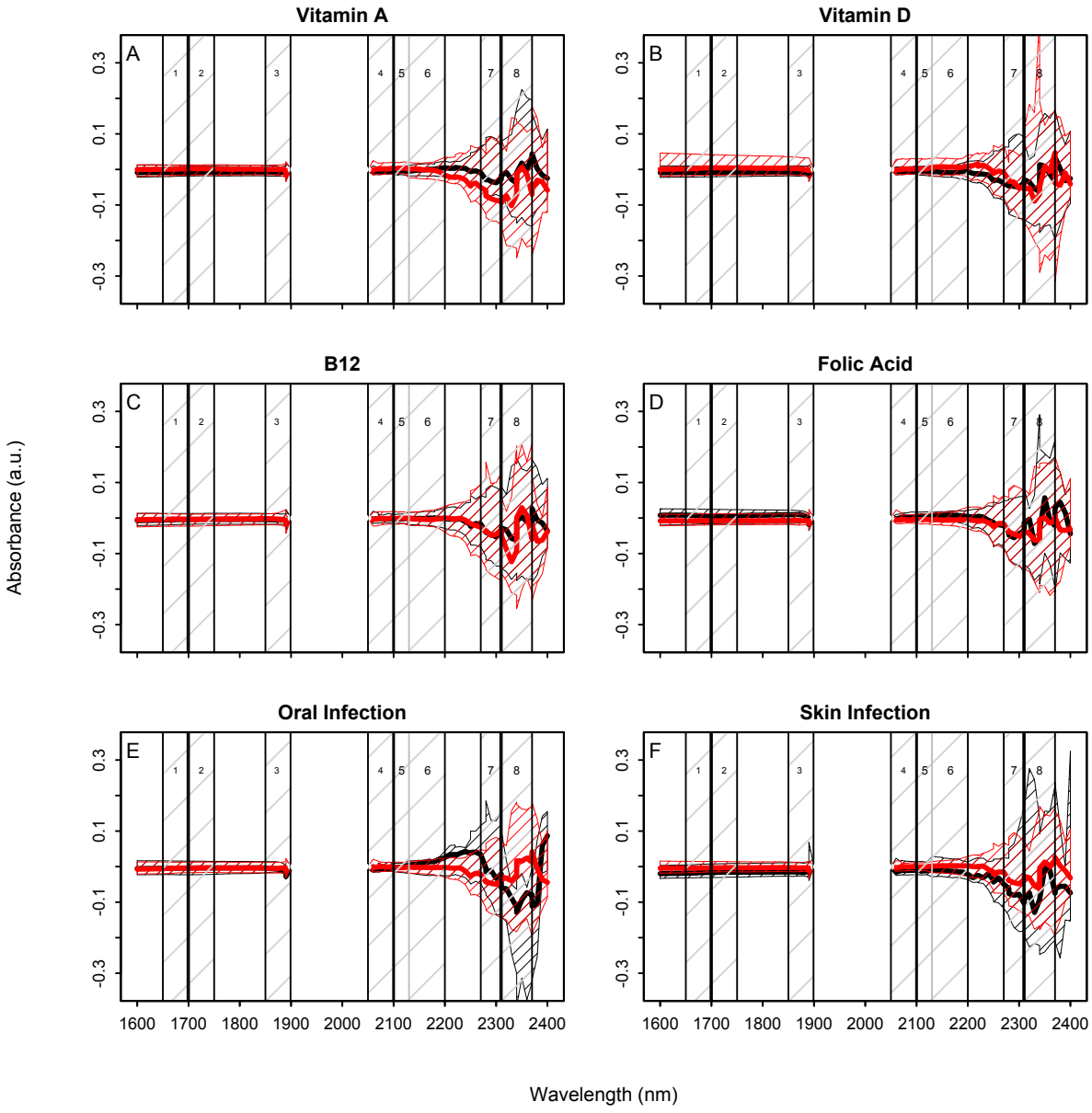


Figure 6–2: *Median spectral absorbance for maternal serum for mothers on their second trimester divided by micro-nutrients: Vitamin A (A), Vitamin D (B), B12 (C), Folic Acid (D) and infection types: Skin (E), Oral (F). A solid black line represents the control group (samples which are sufficient or do not have the infection) and a red solid line represents the unhealthy group (nutrient deficient or compromised by the presence of an infection).*

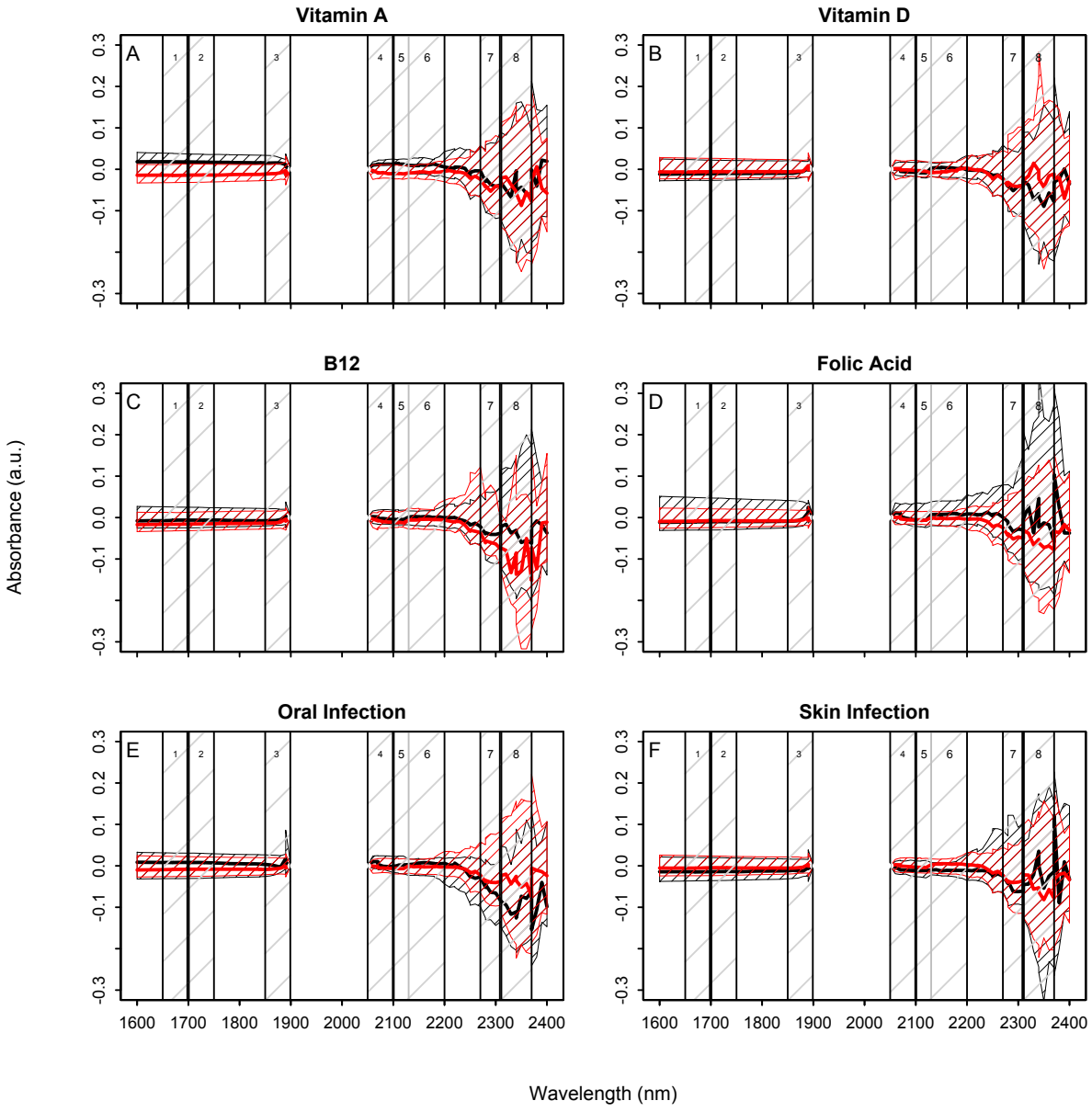


Figure 6–3: *Median spectral absorbance for maternal serum for mothers on their third trimester divided by micronutrients: Vitamin A (A), Vitamin D (B), B12 (C), Folic Acid (D) and infection types: Skin (E), Oral (F). A solid black line represents the control group (samples which are sufficient or do not have the infection) and a red solid line represents the unhealthy group (nutrient deficient or compromised by the presence of an infection).*

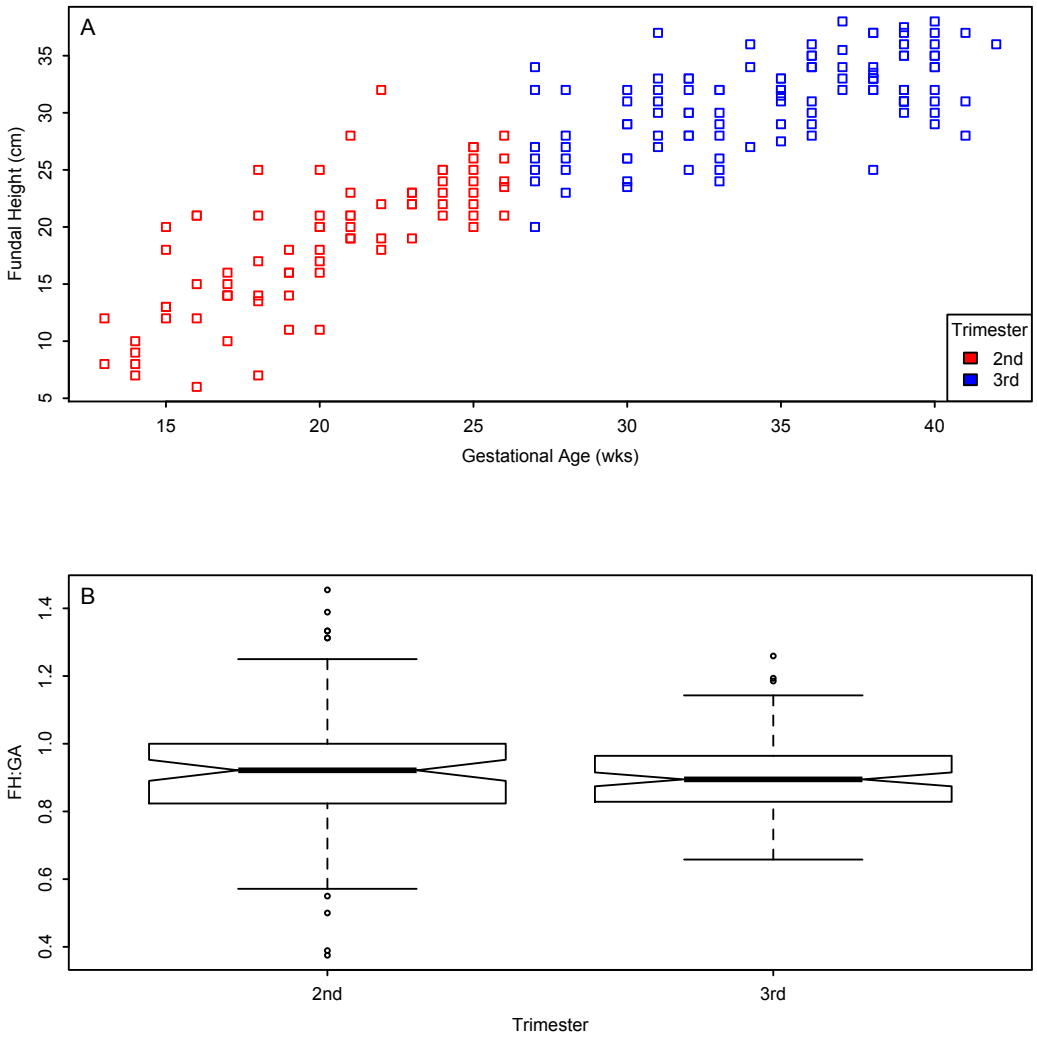


Figure 6–4: *(A) Plot of fundal height vs. gestational age. (B) Boxplot for fundal height variable by gestational age for second and third trimester.*

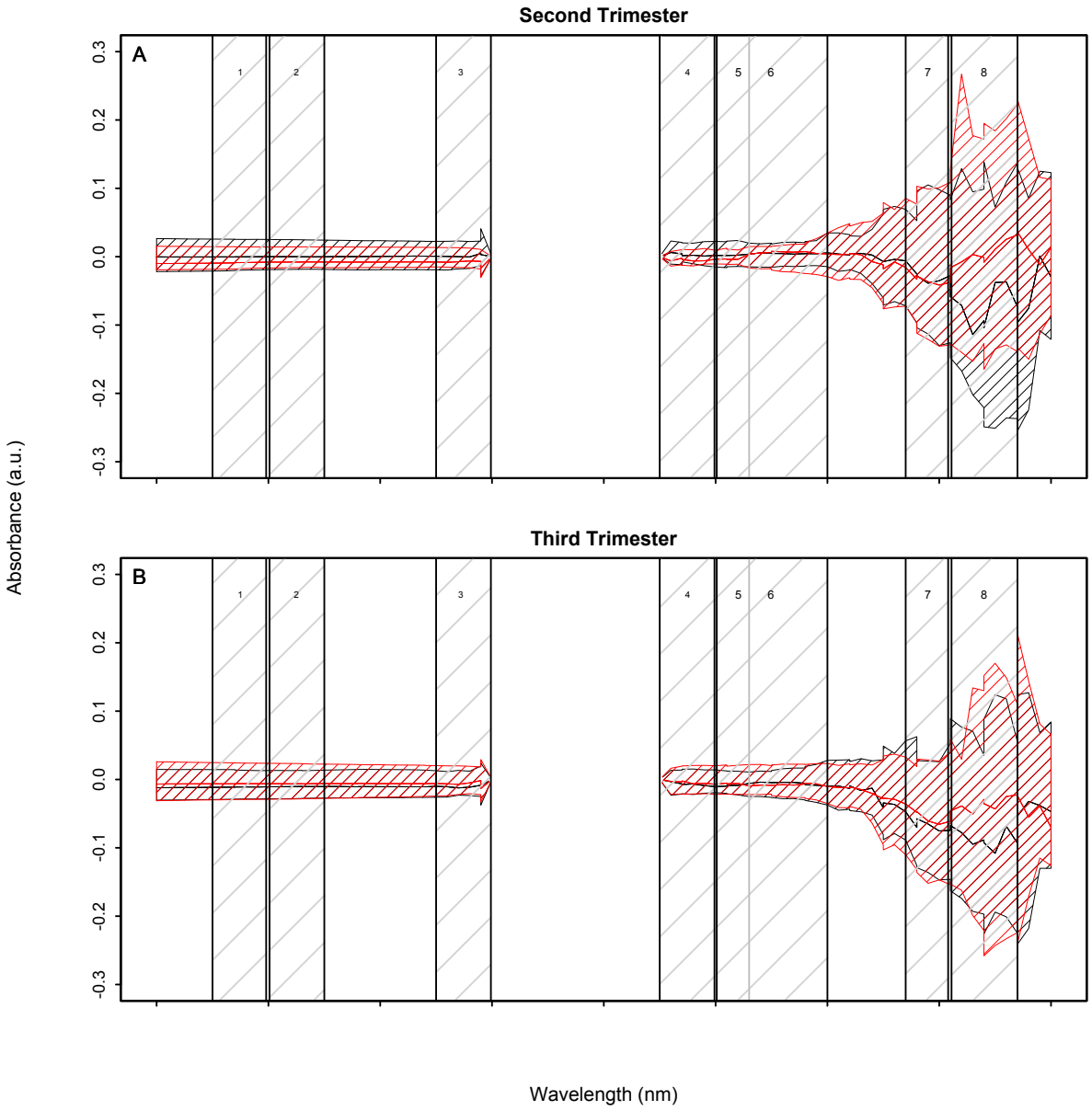


Figure 6–5: *Median spectral absorbance for maternal serum for fundal height over gestational age ratio divided by semester. (A) For mothers in second trimester. (B) For mothers in third trimester.*

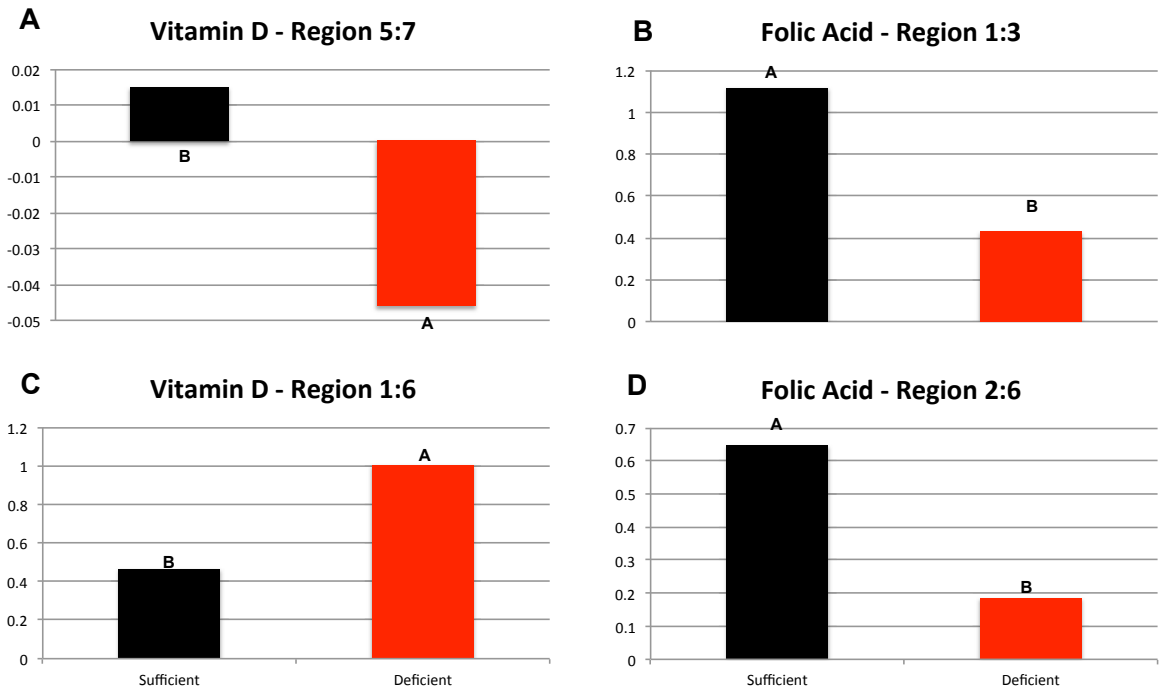


Figure 6–6: *Comparison between Sufficient and Insufficient Status by Micronutrient for 2nd Trimester. Differences for Vitamin D (A and C). Differences for Folic Acid (B and D).*

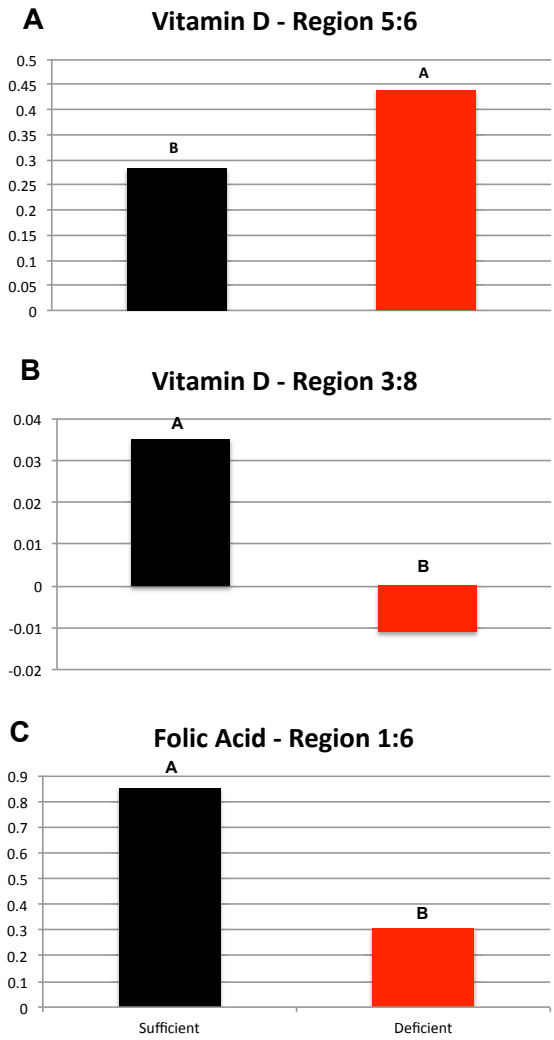


Figure 6–7: *Comparison between Sufficient and Insufficient States by Micronutrient for 2nd Trimester. Differences for Vitamin D (A and B). Differences for Folic Acid (C).*

CHAPTER 7
General Discussion: Major Findings and Future Directions

7.1 Major Findings

In this thesis we explored the application of NIR in the 1600-2400nm region, and applied it to two biological fluids in three studies. We chose maternal and fetal serum and amniotic fluid in order to see if distinct metabolic profiles emerged with different nutritional status and infections. In addition we wanted to explore whether NIR would distinguish between patterns for both nutritional and infection status. We started with a controlled animal study where we analysed serum samples to determine if protein deficiency (PD) and/or nematode infection changed the NIRS profiles in mice exposed to a fixed diet (PD 6% vs PS 24%) and/or nematode infection (Uninfected (0%), Infected dose of *H. bakeri*). From there we moved into a human study where we analyzed NIR spectral profiles of amniotic fluid in a much larger sample set to determine if in-utero NIR spectral profiles differed between LGA and AGA infants. We concluded with a third study in which we analyzed NIR spectral profile of serum samples collected from pregnant indigenous women, that experienced both multiple infections and multiple nutrient deficiencies to assess if any of these profiles could be associated with in utero growth, as measured by fundal height or more specifically a FH:GA ratio. Our principle major findings from these studies were as follows:

1. In our feasibility study, we were able to successfully show that we could use NIR in a controlled laboratory setting, in a nematode infection model. We focused

on a pregnancy model, achieving results regarding the presence of infection, however not the dose-response effect expected *a priori*.

2. In our first study, we observed differences in the spectral profiles of serum from pregnant and lactating mice and even more importantly in their offspring, where we found a greater number of significant differences in NIR spectral profiles, indicating that fetal programming could be occurring in the pups despite experimental treatment not being applied directly to them.
3. In our second study, we determined that there were NIR spectral differences in the 2nd trimester amniotic fluid of mother delivering infants in AGA and LGA categories. Moreover, these difference were evident in the amniotic fluid of normal BMI mothers, and not overweight/obese mothers, suggesting distinct metabolic profiles in the 2 populations.
4. In our third study, we concluded that NIR spectroscopy was able find specific differences between multiple micronutrient deficiencies by 2nd and 3rd trimester but less so for infection in pregnant indigenous women in Panama.

Taken together these results demonstrate that our approach, using NIR spectral profiles in the 1600-2400nm regions, is able to capture part of the metabolic perturbations that occur in maternal and fetal systems during pregnancy. We can further suggest that distinct NIR profiles emerged in this spectral region, that distinguished between clinical conditions including nutrient status (sufficient/deficient), infection status (presence/absence), growth (below and above a certain cutoff, i.e: AGA vs LGA and FH:GA above and below a population mean). This allowed us to distinguish one condition from another, as has been successfully done for lung maturity [82, 83], embryo [132, 149, 151], infections [122, 121, 120].

The research effort presented in this thesis is far from being exhaustive, and applies to the specific context of maternal-fetal medicine, a context in which a non-invasive intervention could be extremely beneficial. However we developed an analysis methodology that could be easily modified and applied to other types of spectral data. More than a biomarker discovery method we think this thesis presents a proof of concept for the use of long wave NIRS for the determination of spectral regions related to clinical conditions.

In the larger context the results presented in this thesis might be later used as the basis for the development of a non-invasive portable measurement device that could be either used in the clinical or rural settings applied in the context of maternal-fetal medicine for measurements of macro and micronutrients.

7.2 Biological Conclusions: What we did learn? What remains to be explored?

This thesis gave us the opportunity to learn about maternal and fetal conditions and candidate predictors of their outcomes (such as: preterm birth, gestational age and growth) and the biological implications that might be related. In general in our studies we were only able to characterize the presence or absence of a deficiency, so our results might be used as a screening method, but not as a diagnostic tool.

In our feasibility study (Chapter 3) we expected to observe a dose response difference in NIR spectral profiles related to infection levels (sham, low and high), however it was not evident. Despite this we can suggest based on Chapter 6 that our approach could be used to distinguish between different levels, for instance, of micronutrient deficiencies for vitamin D (sufficient, deficient or insufficient) [85], and the same for vitamin A, folic acid and B12.

In our mouse serum study (Chapter 4) we were able to determine differences regarding diet and infections and this was despite the limitation of having small sample size ($n = 22$ for lactating dams and $n = 25$ for pups). However these findings were not further tied to a growth outcome, so further investigations are required. Possibilities include the introduction of crown-to-rump-length (CRL) measurements and therefore exploring the impact of PD and/or nematode infection on in-utero linear growth.

When we started the analysis for our amniotic fluid study (Chapter 5), we started comparing SGA, AGA and LGA. However, after examining at the results of the multiple comparison test, we observed that the SGA category was not significantly different from either AGA or LGA. As recent studies have suggested [80, 113], we believe that SGA category has metabolic implications that should be studied on its own, but controlling for gestational age and prematurity including maternal BMI may be required. The reasons for this could be many, since SGA is a multifactorial condition. As well, in a previously published study by our groups we determined that prematurity could also be determinant of SGA, something that we did not control for in this population [110] (see Appendix A). Thus we chose to report the difference between AGA and LGA and in addition studied the impact of BMI and maternal characteristics on the NIR spectral signals. Moreover since NIR spectral changes were evident in the LGA and AGA infants coming from normal weight mothers, but not for LGA infants coming overweight/obese mothers. These results suggest that maternal BMI does not underscore these perturbations, which will need further investigation.

For our pregnant serum fluid study (Chapter 6) we were able to determine NIR spectral differences in micronutrients (vitamin D and folic acid for Both 2nd and 3rd

trimester), and oral infection for 2nd trimester. Also, differences were evident for FH:GA in the 2nd and 3rd trimesters. We observed quite different spectral profiles for each trimester, something that should be taken into consideration if we were to build a portable instrument.

7.3 Strength and Limitations

Near-infrared spectroscopy analysis of biological fluids is an area of research that requires knowledge of very diverse subjects from chemistry, engineering, multivariate statistics and the life sciences. This technique has the advantages of being a rapid analysis method, especially if one compares it to other quantitative and qualitative separation and analysis procedures used for candidate biomarker discovery and the determination of influential molecules in clinical and health conditions and perturbations. The strengths and limitations of our studies are many, but we feel they can be summarised into categories related to the nature of NIR signals, the experimental setting used for their measurement, the NIR signal processing and the statistics applied to the determination of differences in NIR spectral data.

7.3.1 Targeting Specific Moieties and Molecules in the Near-Infrared Region

The results of the studies presented in this thesis used a combination of NIR spectral information and clinical information related to maternal and infant conditions for the definition of experimental groups and therefore for the assessment of the changes in the NIR spectral signal that these conditions might cause. However, one of the big limitations of our methodology is that our results can not be used to infer strongly about the biology or metabolic mechanism that may cause them. To

overcome this caveat we should assign specific moieties and molecules to our integrated spectral regions.

One way to do this will be to look closely to the regions we are integrating over and to look to the specific moieties that have been assigned by previous studies in literature. In this regards looking at both the NIR presented in Figure 2–3 and the 'Atlas of interpretative NIR Spectroscopy' [163], we see that Region 1 (1650-1700nm) and Region 2 (1701-1750nm), could be associated with first overtone CH, CH_2, CH_3, $ArCH$ and SH. Region 3 (1850-1899nm), could be associated with RCO_2H. Region 4 (2050-2099nm), could be associated with $R-OH$ and $CONH_2(R)$. Region 5 (2100-2130nm) and Region 6 (2130-2200nm), could be associated with $R-NH_2$, $CONH_2(R)$, $C-C$ and $CH-O$. Region 7 (2275-2310nm) and Region 8 (2311-2370nm), could be associated with combination band CH, CH_2, CH_3 and H_2O. With this information we could look at the structure of known molecules that have this moieties present in their structure and that are realised to maternal-fetal health and are present in the biological fluids we are measuring. Using this list of molecules we can spike our samples and measure the spectral changes that the NIR signal experience relative to concentration of the known molecules that we added.

Spiking the samples with known concentrations might be possible for serum samples which are denatured and therefore no matrix interactions could arise. However, the complexity of amniotic fluid, being a viscous fluid made of biochemically active mixture of water, electrolytes, proteins, carbohydrates, lipids, phospholipids, urea (among many other components), will make the interpretation of the spiking of these samples more difficult, mostly because we would not be able to control the multiple interactions that could arise in this dynamically changing fluid.

Another way to make specific assignments that would not need the spiking of samples but could be to use the information previously collected for both amniotic fluid and serum samples for glucose, lactate, amino acids, urea and fatty acids that was made available to us with the samples. These measurements could be binned and associated with the spectral regions that we integrated in our methodology, for instance glucose measured can be matched to absorbance values in Region 8, the same for lactate in Region 7, amino acids (amines) for Region 7, urea for Region 6 and fatty acids to Region 1. By doing this we could build calibration curves relating the known concentration to the measured concentrations as it is commonly done for spectral measurements using chemometrics.

7.3.2 Experimental Setting Differences

In our studies we used two different instrumental settings (described in Chapter 3), therefore we feel that we must comment about the insights that the use of each instrument provided. For instance, the use of lenses (used in feasibility study, study 1 and study 2) versus the use of fiber optics (used in study 3) for light guidance. On the one hand building an instrument with lenses, was advantageous, because it is easier to tweak and fix the spectrometric system to accommodate for an specific set of samples (by adjusting the optical path length in the sample cell). But, at the same time, it is easier to get error signals because of water droplets adhering to the lenses and making light reflect scatter in many more directions, thus not collimating uniformly into the spectrometer. On the other hand, fiber optics ensure that the light passes from the light source to the sample cell and then to the detector system, which is crucial in absorbance measurements. Another crucial change in the instrumental setting was the introduction of a temperature controller, which kept the sample cell temperature stable within a 0.5 degrees of variation.

7.3.3 Wavelength Range Selected

Despite the instrumental differences in the two settings, NIR spectral signals from both indicated that there is less variability in the 1600 to 1900nm region in the 2 biological fluids (amniotic fluid and serum) and more variability in the 2100 to 2400nm region, regardless of the experimental condition. This starts to suggest the use of longer wavelengths in the combination region (2100-2400nm), where molecular groups such as: $C-H$, CH_2, CH_3, H_2O, $C-C$, $CH-O$ and $R-NH_2$ are known to have higher absorbance of NIR light, suggesting that further studies can target on this region for the analysis of biological fluids.

7.3.4 Pre-Processing of NIRS Spectral Signals

Since NIRS absorbance signals are small, pre-processing of the signal is highly recommended for the elimination of noise, since it helps to get the most information possible from a single sample and at the same time get rid of the errors (instrumental drift, operator error). In our studies we used smoothing and mean-centering for preprocessing. We also tried using the raw signal (instead of absorbance values) and used auto-scaling for pre-processing but with different results. Raw signals did not provided any separation of signals and auto-scaling augmenting the variability of the signals, which in turn yielded no separation between experimental groups.

We suggest that other pre-processing methods could be also explored. For instance, pre-processing using the derivatives of the signals which is used to separate overlapping peaks by making background constant in the first derivative and zero in the second derivative. However, there are inherent disadvantages of applying derivatives, for instance the decrease in the signal-to-noise ratio as higher derivatives are used since noise always is present as broad peaks in the signal and the difficult of

interpretation of negative signals. There are also methods that correct for scattering effects that could be useful. Examples of this are: 1) Standard normal variate (SNV) transformation, used to minimize slope variation effects; or 2) Multiplicative scatter correction (MSC), which corrects all samples in a group, based on the average spectrum of a target or ideal sample spectrum, could be used [15].

In terms of reduction of noise there are a few statistical methods that are commonly applied to NIR that can be used in our spectral data, such as adaptive filtering, in which the systems captures not only the signal but also senses the environment in which it was produced, adapting the filter applied to the signal before its final acquisition to reduce noise. Moreover, a different implementations of Wiener filtering, where the goal is to estimate the spectral signal of interest from a *real* one that is noisy [158, 164, 61].

After signals were preprocessed, we integrated (summed) the absorbance values in the regions we set to investigate. One idea was to calculate the average of the region, as this could eliminate the sum of negative absorbance which produces negative absorbance values for the spectral regions we investigated and which are quite difficult to interpret since they do not represent depletion of a specific molecule in the samples, but just happen as an artifact of the preprocessing.

One correction that was suggested in the early stages of this project was to only work with the positive part of the spectral signals by adding a constant offset value. This offset was defined by first finding the minimum (negative) value of each spectral signal, calculating the absolute value of the signal and adding that value to all the values of the signal. Another idea was to implement a normalization function to make the signals to be between 0 and 1 absorbance units and thus just work with the positive part of the signal, by calculating the normalized version of a signal, using

the following formula:$Normalized(S_i) = \frac{S_i - Smin}{Smax - Smin}$, where $Smin$ is the minimum absorbance value found for the spectra S_i, $Smax$ is the maximum absorbance value found for S_i. However, both of these approaches did not make the separations between groups bigger, nor facilitated the interpretation of the results.

7.3.4.1 Working with Ratios

The use of ratios while analysing NIRS signals is not new. A major commercial non-invasive clinical application of NIR ratios is used to separate between conditions in pulse oximetry [135, 137, 118]. In the pulse oximetry measurements a ratio is calculated between the pulsatile components (DC) and the nonpulsatile (AC) or steady-state parts of the signal at 660nm and 940nm, as: $S = \frac{AC_{\lambda=660nm}/DC_{\lambda=660nm}}{AC_{\lambda=940nm}/DC_{\lambda=940nm}}$ S is then related to an O_2 saturation value.

In our implementation, ratios resulted an useful way to investigate the differences between experimental conditions by dividing an integrated region of the signal and comparing it to another region. However, the problem of working with ratios was that our signals had both negative and positive parts and in general signals were rather small (less than ± 0.25 absorbance units). Having negative values becomes confusing when calculating a ratio between two regions that have negative median value. Mathematically, the division of two negative values is a positive number (a.e Ratio $= \frac{-1}{-1} = 1$), which in biological terms is hard to interpret.

Another issue with the mathematical nature of ratios, happened when calculating a ratio and the denominator is *big*, as the value for our ratios become small, for example, $\frac{1}{10} = 0.1$ and in the general case $\lim_{x\to\infty} \frac{1}{x} = 0$. When the denominator is *small*, the values become big, for example, $\frac{1}{0.1} = 10$ and in the general case the smaller the denominator the bigger the value, which gave rise to very high dispersion

of values (positive and negative) and not normal distributions. As a consequence we used of non-parametric statistics in the calculation of difference among experimental groups using Kruskal-Wallis and Mann-Whitney tests, respectively.

7.3.4.2 Building higher order models

Part of the abnormalities of the NIR spectral signals we measured is that they came from highly multicollinear variables [15], calculated from regions of interest. In our case even highly efficient dimensionality reduction algorithms as: Principal Component Analysis (PCA) or even a powerful clustering and visualisation methods as Self-Organized Maps (SOM's) were not able to find meaningful correlations in our datasets.

A few other methods has been suggested to deal with this problem, for instance the use of Bayesian methods (Bayesian regressions, Bayesian clustering methods), in which the correlation can be introduced into the models as a form of *priors*, thus weighting in favour of regions that are less correlated. In the same vein other approaches that could be applied are the ones based on *penalization methods*, in which a penalty could be imposed to those statistical significant models that favour correlated spectral regions or their ratios.

Both multiple linear and logistic regressions were used to group together the spectral regions ratios as factor in order to build a more robust model to distinguish between the experimental categories, but did not yield significant models. Moreover, 2D Bayesian probability plots were made to use the factors found significantly different to distinguish between maternal BMI categories and AGA and LGA categories, however the results were limited and the models had a low sensitivity and specificity and were not able to distinguish the LGA infants from the AGA, which might be

caused by a power issue related to minimum sample size, due to the fact that there were almost 10 times more samples in the AGA experimental group than in the LGA experimental group.

7.4 Implications and Future Directions

7.4.1 Instrumental

In this thesis we explored the 1600-2400nm region and provided interesting results in terms of the biology of metabolic profiles related to maternal and fetal health conditions. In terms of specific results for micronutrients, we could foresee the development of a portable instrument, for measuring NIR spectral signal on patients on site. In order for this to happen we would have to extrapolate our findings, identifying the third overtones of those same spectral regions we found to be significant, but in the 700-1100nm region, where the NIR light absorption on the skin is higher.

Based on the results of this thesis, we think that a micro-nutrient instrument could be developed to monitor vitamin D and folic acid deficiencies. However, it is known that for this type of development we would need to spend a considerable amount of time in characterisation of light penetration on different skin tones, and since it might be targeted to be used in a developing country setting, it might be important other cultural factors, as well as a better understanding of how micronutrient deficiencies interact with other infections at the metabolic level, and how they might relate to the NIR spectral readings one might need to analyse.

In our results we were able pick up some spectral differences in amniotic fluid, however in our pregnancy study we were not able to determine any difference between SGA and AGA. We believe this differences could be studied more successfully using a

different wavelength range or even a difference methodology, such as Raman or FT-IR spectroscopy.

7.4.2 Statistics and NIR spectral signal analysis

Differences in the NIR spectral signals of the biological fluids analysed in this thesis were determined by non-parametric statistics (some of the reasons were described in the previous subsection), using either Mann-Whitney U-test for binary comparisons and Kruskal-Wallis one-way analysis of variance for multiple group comparisons. In the case of multiple comparisons P-values were adjusted to account to the effect of multiple comparisons using a Bonferroni correction. The Bonferroni method creates a new acceptable tolerance value by dividing the set P-value (α) between the number, n of binary comparisons possible among the m groups, calculated by the binomial coefficient $\binom{m}{2}$. It is a very "conservative" method, that has strong assumptions about the independences of the test it makes, however it is prone to determine a high number of false negatives.

There exist other multiple comparison methods that are less restrictive and have less independence assumptions that could be suitable for the analysis of NIR spectral absorbance values, especially if we are working with ratios from integrated regions which are highly dependent, for instance: Holm [55], which uses a stepwise (iterative) method to modify the rejection criteria while keeping a strong control of the family-wise error rate. A newer method called False-Discovery Rate method introduced by Benjamini & Hochberg [8], defines a ratio of the number of Type I errors calculated by the number of group comparisons. Since NIR spectral data is highly correlated and the multiple comparison test highly dependent this approach could be beneficial.

7.5 Conclusion

Taken together the findings from our three studies suggest that the long wave NIR spectroscopy can be used to identify several candidate NIR spectral regions. Moreover, we provide a first list of candidate regions that could be further explored ior the elucidation of the metabolic changes related to maternal-fetal health. After these studies, it is not clear, how could these regions could be related to infections in mice nor in human biological fluid, with the exception that some candidate spectral regions could be important to study the potential impact of infection transfer that can occur through lactation as a lingering effect that could influence growth and development.

We conclude this thesis knowing the important implications of our results. Our approach opens new avenues into the exploration of different instrumental settings, with the suggestion of using different statistical strategies and using or extrapolating our results to other wavelength range (either Mid-IR or shorter wavelengths). Also, it provides new data analysis challenges that we foresee addressing in the near future, for instance the use of *bayesian probability plots* and gaussian mixture models as has been proposed elsewhere [143, 168] to provide a visual interpretation of the data, that would quantify the difference and not merely state of their existence. Also, providing a full probability model that quantify the separations as a distribution more than just saying where the separation happens. Moreover, it opens doors to think about other ways to combine clinical and NIR spectral data to provide more robust candidate biomarkers.

References

[1] Kathleen Abu-Saad and Drora Fraser. Maternal Nutrition and Birth Outcomes. *Epidemiologic Reviews*, 32(1):5–25, April 2010.

[2] Angus L. Acton, Tarun Goswami, and David S. McKenna. Utility of near infrared spectroscopy for the screening of the growth restricted fetus. *Congenital anomalies*, 53(3):109–114, September 2013.

[3] M. N. Anas, A. Syafirah, A. N. Norali, and M. Normahira. Non-invasive blood glucose measurement Application of near infrared optical measurement. In *Sustainable Utilization and Development in Engineering and Technology (STUDENT), 2012 IEEE Conference on*, pages 258–261. IEEE, October 2012.

[4] C. Annweiler and O. Beauchet. Maternal vitamin D status in pregnancy and offspring brain development: the forgotten (but essential) needs of vitamin D era. *Osteoporosis international*, 25(4):1419–1420, April 2014.

[5] Douglas D. Archibald and Danny E. Akin. Use of spectral window preprocessing for selecting near-infrared reflectance wavelengths for determination of the degree of enzymatic retting of intact flax stems. *Vibrational Spectroscopy*, 23(2):169–180, June 2000.

[6] William R. Atchley, Charles T. Gaskins, and Dwane Anderson. Statistical Properties of Ratios. I. Empirical Results. *Systematic Biology*, 25(2):137–148, June 1976.

[7] Wen C. Aw, Floyd E. Dowell, and J. William Ballard. Using near-infrared spectroscopy to resolve the species, gender, age, and the presence of Wolbachia infection in laboratory-reared Drosophila. *G3 (Bethesda, Md.)*, 2(9):1057–1065, September 2012.

[8] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[9] Michel R. Boisvert, Kristine G. Koski, David H. Burns, and Cameron D. Skinner. Early prediction of macrosomia based on an analysis of second trimester amniotic fluid by capillary electrophoresis. *Biomarkers in medicine*, 6(5):655–662, October 2012.

[10] Michel R. Boisvert, Kristine G. Koski, David H. Burns, and Cameron D. Skinner. Prediction of gestational diabetes mellitus based on an analysis of amniotic fluid by capillary electrophoresis. *Biomarkers in medicine*, 6(5):645–653, October 2012.

[11] L. Bokobza. Origin of Near-Infrared Absorption Bands. pages 11–41, 2007.

[12] F. Branca and M. Ferrari. Impact of micronutrient deficiencies on growth: the stunting syndrome. *Annals of nutrition & metabolism*, 46 Suppl 1:8–17, 2002.

[13] Frank J. Bruggeman and Hans V. Westerhoff. The nature of systems biology. *Trends in microbiology*, 15(1):45–50, January 2007.

[14] Jason J. Burmeister and Mark A. Arnold. Evaluation of Measurement Sites for Noninvasive Blood Glucose Sensing with Near-Infrared Transmission Spectroscopy. *Clinical Chemistry*, 45(9):1621–1627, September 1999.

[15] A. Candolfi, R. De Maesschalck, D. Jouan-Rimbaud, P. A. Hailey, and D. L. Massart. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *Journal of pharmaceutical and biomedical analysis*, 21(1):115–132, October 1999.

[16] Steven Cassin. Development of a near-infrared detection system for oxidative stress analysis of pregnant and non-pregnant mouse serum samples with parasite loading. Master's thesis, 2011.

[17] Regine Choe, Turgut Durduran, Guoqiang Yu, Mark J. M. Nijland, Britton Chance, Arjun G. Yodh, and Nirmala Ramanujam. Transabdominal near infrared oximetry of hypoxic stress in fetal sheep brain in utero. *Proceedings of the National Academy of Sciences*, 100(22):12950–12954, October 2003.

[18] Parul Christian. Fetal growth restriction and preterm as determinants of child growth in the first two years and potential interventions. *Nestlé Nutrition Institute workshop series*, 78:81–91, 2014.

[19] Emil W. Ciurczak. *Pharmaceutical and medical applications of near-infrared spectroscopy*. Marcel Dekker, 2002.

[20] Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, February 2001.

[21] R. L. Coop and I. Kyriazakis. Nutrition-parasite interaction. *Veterinary parasitology*, 84(3-4):187–204, August 1999.

[22] Paul J. Curran. Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3):271–278, December 1989.

[23] S. Das, M. Irigoyen, M. B. Patterson, A. Salvador, and D. L. Schutzman. Neonatal outcomes of macrosomic births in diabetic and non-diabetic women. *Archives of disease in childhood. Fetal and neonatal edition*, 94(6), November 2009.

[24] Bruno de Benoist. Conclusions of a WHO Technical Consultation on folate and vitamin B12 deficiencies. *Food and nutrition bulletin*, 29(2 Suppl), June 2008.

[25] Mercedes de Onis and Monika Blössner. The World Health Organization Global Database on Child Growth and Malnutrition: methodology and applications. *International Journal of Epidemiology*, 32(4):518–526, August 2003.

[26] Sílvia O. Diaz, António S. Barros, Brian J. Goodfellow, Iola F. Duarte, Eulália Galhano, Cristina Pita, Maria Almeida, Isabel M. Carreira, and Ana M. Gil. Second Trimester Maternal Urine for the Diagnosis of Trisomy 21 and Prediction of Poor Pregnancy Outcomes. *J. Proteome Res.*, 12(6):2946–2957, April 2013.

[27] M. L. Dreyfuss, R. J. Stoltzfus, J. B. Shrestha, E. K. Pradhan, S. C. LeClerq, S. K. Khatry, S. R. Shrestha, J. Katz, M. Albonico, and K. P. West. Hookworms, malaria and vitamin A deficiency contribute to anemia and iron deficiency among pregnant women in the plains of Nepal. *The Journal of nutrition*, 130(10):2527–2536, October 2000.

[28] David S. Eisenberg. *The structure and properties of water [by] D. Eisenberg and W. Kauzmann.* Oxford University Press, 1969.

[29] Joseph G. Eisenhauer. Regression through the Origin. *Teaching Statistics*, 25(3):76–80, September 2003.

[30] David I. Ellis and Royston Goodacre. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *The Analyst*, 131(8):875–885, August 2006.

[31] Robert A. Fenton and Mark A. Knepper. Urea and Renal Function in the 21st Century: Insights from Knockout Mice. *Journal of the American Society of Nephrology*, 18(3):679–688, March 2007.

[32] Francesc Figueras and Jason Gardosi. Intrauterine growth restriction: new concepts in antenatal surveillance, diagnosis, and management. *American journal of obstetrics and gynecology*, 204(4):288–300, April 2011.

[33] U. Fischer, K. Rebrin, T. von Woedtke, and P. Abel. Clinical usefulness of the glucose concentration in the subcutaneous tissue–properties and pitfalls of electrochemical biosensors. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et métabolisme*, 26(11):515–522, November 1994.

[34] Christina Fotopoulou, Sarah Kyeyamwa, Mattea Linder, Daniela Thieme, Sebastian Hartenstein, Oliver Klein, Joachim W. Dudenhausen, Wolfgang Henrich, Karim D. Kalache, and Christian Bamberg. Proteomic analysis of midtrimester amniotic fluid to identify novel biomarkers for preterm delivery. *The journal of maternal-fetal & neonatal medicine*, 25(12):2488–2493, December 2012.

[35] Michael J. Fowler. Microvascular and Macrovascular Complications of Diabetes. *Clinical Diabetes*, 26(2):77–82, April 2008.

[36] Tao Gao, Nadine Zablith, David H. Burns, Kristine G. Koski, and Cameron D. Skinner. Identification and quantitation of human amniotic fluid components using capillary zone electrophoresis. *Analytical biochemistry*, 388(1):155–157, May 2009.

[37] Tao Gao, Nadine R. Zablith, David H. Burns, Cameron D. Skinner, and Kristine G. Koski. Second trimester amniotic fluid transferrin and uric acid predict infant birth outcomes. *Prenatal diagnosis*, 28(9):810–814, September 2008.

[38] J. L. Garcia-Garcia, D. Perez-Guaita, J. Ventura-Gayete, S. Garrigues, and M. de la Guardia. Determination of biochemical parameters in human serum by near-infrared spectroscopy. *Anal. Methods*, 6(12):3982–3989, 2014.

[39] Doris Gonzalez. Combined impact of infections and nutritional deficiencies on maternal health and fetal parameters in a rural community in Panama. Master's thesis, 2012.

[40] Francesca Gotsch, Francesca Gotsch, Roberto Romero, Offer Erez, Edi Vaisbuch, Juan Pedro P. Kusanovic, Shali Mazaki-Tovi, Sun Kwon K. Kim, Sonia Hassan, and Lami Yeo. The preterm parturition syndrome and its implications for understanding the biology, risk assessment, diagnosis, treatment and prevention of preterm birth. *The journal of maternal-fetal & neonatal medicine*, 22 Suppl 2:5–23, 2009.

[41] Gonçalo Graça, Ana Sofia S. Moreira, Ana João J. Correia, Brian J. Goodfellow, António S. Barros, Iola F. Duarte, Isabel M. Carreira, Eulália Galhano, Cristina Pita, Maria do Céu d. o. . C. Almeida, and Ana M. Gil. Mid-infrared (MIR) metabolic fingerprinting of amniotic fluid: a possible avenue for early diagnosis of prenatal disorders? *Analytica chimica acta*, 764:24–31, February 2013.

[42] Goncalo Graca, Iola F. Duarte, Antonio S. Barros, Brian J. Goodfellow, Silvia O. Diaz, Joana Pinto, Isabel M. Carreira, Eulalia Galhano, Cristina Pita, and Ana M. Gil. Impact of Prenatal Disorders on the Metabolic Profile of Second Trimester Amniotic Fluid: A Nuclear Magnetic Resonance Metabonomic Study. *J. Proteome Res.*, 9(11):6016–6024, September 2010.

[43] Martin Griebe, Michael Daffertshofer, Mark Stroick, Magdalena Syren, Parviz Ahmad-Nejad, Michael Neumaier, Juergen Backhaus, Michael G. Hennerici, and Marc Fatar. Infrared spectroscopy: A new diagnostic tool in Alzheimer disease. *Neuroscience Letters*, 420(1):29–33, June 2007.

[44] David M. Haaland, M. Ries Robinson, Gary W. Koepp, Edward V. Thomas, and R. Philip Eaton. Reagentless Near-Infrared Determination of Glucose in Whole Blood Using Multivariate Calibration. *Appl. Spectrosc.*, 46(10):1575–1578, October 1992.

[45] H. Hahn, J. D. Pallua, C. Pezzei, V. Huck-Pezzei, G. K. Bonn, and C. W. Huck. Infrared-spectroscopy: a non-invasive tool for medical diagnostics and drug analysis. *Current medicinal chemistry*, 17(26):2956–2966, 2010.

[46] J. W. Hall and A. Pollard. Near-infrared spectrophotometry: a new dimension in clinical chemistry. *Clinical chemistry*, 38(9):1623–1631, September 1992.

[47] Jeffrey W. Hall and Alan Pollard. Near-infrared spectroscopic determination of serum total proteins, albumin, globulins, and urea. *Clinical Biochemistry*, 26(6):483–490, December 1993.

[48] Carli M. Halpenny, Kristine G. Koski, Victoria E. Valdés, and Marilyn E. Scott. Prediction of child health by household density and asset-based indices in impoverished indigenous villages in rural Panamá. *The American journal of tropical medicine and hygiene*, 86(2):280–291, February 2012.

[49] Sarah Hanieh, Tran T. Ha, Julie A. Simpson, Tran T. Thuy, Nguyen C. Khuong, Dang D. Thoang, Thach D. Tran, Tran Tuan, Jane Fisher, and Beverley-Ann A. Biggs. Maternal vitamin D status and infant outcomes in rural Vietnam: a prospective cohort study. *PloS one*, 9(6), 2014.

[50] T. Hardarson, A. Ahlström, L. Rogberg, L. Botros, T. Hillensjö, G. Westlander, D. Sakkas, and M. Wikland. Non-invasive metabolomic profiling of Day 2 and 5 embryo culture medium: a prospective randomized trial. *Human reproduction (Oxford, England)*, 27(1):89–96, January 2012.

[51] Nicholas C. Harvey, Christopher Holroyd, Georgia Ntani, Kassim Javaid, Philip Cooper, Rebecca Moon, Zoe Cole, Tannaze Tinati, Keith Godfrey, Elaine Dennison, Nicholas J. Bishop, Janis Baird, and Cyrus Cooper. Vitamin D supplementation in pregnancy: a systematic review. *Health technology assessment (Winchester, England)*, 18(45):1–190, July 2014.

[52] Junichi Hasegawa, Masamitsu Nakamura, Ryu Matsuoka, Takashi Mimura, Kiyotake Ichizuka, Akihiko Sekizawa, and Takashi Okai. Evaluation of placental function using near infrared spectroscopy during fetal growth restriction. *Journal of perinatal medicine*, 38(1):29–32, 2010.

[53] Kevin H. Hazen, Mark A. Arnold, and Gary W. Small. Measurement of glucose and other analytes in undiluted human serum with near-infrared transmission spectroscopy. *Analytica Chimica Acta*, 371(2-3):255–267, October 1998.

[54] Hoffer, Arlene Taveroff, and Mazen J. Hamadeh. Dietary protein restriction alters glucose but not protein metabolism in noninsulin-dependent diabetes mellitus. *Metabolism*, 47(9):1145–1151, September 1998.

[55] Sture Holm. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

[56] M. Holzmann, S. Cnattingius, and L. Nordström. Lactate production as a response to intrapartum hypoxia in the growth-restricted fetus. *BJOG : an international journal of obstetrics and gynaecology*, 119(10):1265–1269, September 2012.

[57] Holly R. Hull, Mary K. Dinger, Allen W. Knehans, David M. Thompson, and David A. Fields. Impact of maternal body mass index on neonate birthweight and body composition. *American journal of obstetrics and gynecology*, 198(4), April 2008.

[58] Philip Hunter. Reading the metabolic fine print. *EMBO reports*, 10(1):20–23, January 2009.

[59] R. Imbesi, V. Mazzone, and P. Castrogiovanni. Is tryptophan 'more' essential than other essential aminoacids in development? A morphologic study. *Anatomia, histologia, embryologia*, 38(5):361–369, October 2009.

[60] CharlotteL Ives, D. K. Harrison, and G. S. Stansby. Prediction Of Surgical Site Infections After Major Surgery Using Visible And Near-Infrared Spectroscopy. In DavidJ Maguire, DuaneF Bruley, and DavidK Harrison, editors, *Oxygen Transport to Tissue XXVIII*, volume 599 of *Advances in Experimental Medicine and Biology*, pages 37–44. Springer US, 2008.

[61] Meltem Izzetoglu, Prabhakar Chitrapu, Scott Bunce, and Banu Onaral. Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. *Biomedical engineering online*, 9(1):16+, March 2010.

[62] Pierre-Olivier Jean, Robert L. Bradley, Marie-Andrée Giroux, Jean-Pierre Tremblay, and Steeve D. Côté. Near Infrared Spectroscopy and Fecal Chemistry as Predictors of the Diet Composition of White-Tailed Deer. *Rangeland Ecology & Management*, 67(2):154–159, March 2014.

[63] Jingying Jiang, Da Zou, Xiaolin Min, Zhenhe Ma, and Kexin Xu. Investigation on how to choose measurement sites for non-invasive near-infrared blood glucose sensing. volume 8222, pages 822216–822216–11, 2012.

[64] Jan S. Joergensen, Ronald F. Lamont, and Maria R. Torloni. Vitamin D and gestational diabetes: an update. *Current opinion in clinical nutrition and metabolic care*, 17(4):360–367, July 2014.

[65] Jun Kakogawa, Kazuhiro Sumimoto, Takakazu Kawamura, Shigeki Minoura, and Naohiro Kanayama. Noninvasive monitoring of placental oxygenation by near-infrared spectroscopy. *American journal of perinatology*, 27(6):463–468, June 2010.

[66] P. Kalk, F. Guthmann, K. Krause, K. Relle, M. Godes, G. Gossing, H. Halle, R. Wauer, and B. Hocher. Impact of maternal body mass index on neonatal outcome. *European journal of medical research*, 14(5):216–222, May 2009.

[67] S. Kasemsumran, Y. P. Du, K. Murayama, M. Huehne, and Y. Ozaki. Near-infrared spectroscopic determination of human serum albumin, -globulin, and glucose in a control serum solution with searching combination moving window partial least squares. *Analytica Chimica Acta*, 512(2):223–230, June 2004.

[68] Peter Katona and Judit Katona-Apte. The Interaction between Nutrition and Infection. *Clinical Infectious Diseases*, 46(10):1582–1588, May 2008.

[69] Anneloes Kerssen, Harold W. de Valk, and Gerard H. A. Visser. Increased Second Trimester Maternal Glucose Levels Are Related to Extremely Large-for-Gestational-Age Infants in Women With Type 1 Diabetes. *Diabetes Care*, 30(5):1069–1074, May 2007.

[70] P. R. Khan, R. D. Jee, R. A. Watt, and A. C. Moffat. The Identification of Active Drugs in Tablets Using Near Infrared Spectroscopy. *Pharmacy and Pharmacology Communications*, 3(9):447–453, 1997.

[71] Parris Kidd. Th1/Th2 balance: the hypothesis, its limitations, and implications for health and disease. *Alternative medicine review : a journal of clinical therapeutic*, 8(3):223–246, August 2003.

[72] Venkata R. Kondepati, Michael Keese, Ralf Mueller, Bernd C. Manegold, and Juergen Backhaus. Application of near-infrared spectroscopy for the diagnosis of colorectal cancer in resected human tissue specimens. *Vibrational Spectroscopy*, 44(2):236–242, July 2007.

[73] VenkataRadhakrishna Kondepati, Heise, and Juergen Backhaus. Recent applications of near-infrared spectroscopy in cancer diagnosis and therapy. *Analytical and Bioanalytical Chemistry*, 390(1):125–139, January 2008.

[74] K. G. Koski and M. E. Scott. Gastrointestinal nematodes, nutrition and immunity: breaking the negative spiral. *Annual review of nutrition*, 21:297–321, 2001.

[75] Michael S. Kramer, Robert W. Platt, Shi W. Wen, K. S. Joseph, Alexander Allen, Michal Abrahamowicz, Béatrice Blondel, Gérard Bréart, and for the Fetal/Infant Health Study Group of the Canadian Perinatal Surveillance System. A New and Improved Population-Based Canadian Reference for Birth Weight for Gestational Age. *Pediatrics*, 108(2):e35, August 2001.

[76] Richard A. Kronmal. Spurious Correlation and the Fallacy of the Ratio Standard Revisited. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 156(3):379–392, 1993.

[77] Shwayta Kukreti, Albert Cerussi, Bruce Tromberg, and Enrico Gratton. Intrinsic near-infrared spectroscopic markers of breast tumors. *Disease markers*, 25(6):281–290, 2008.

[78] S. A. Lanham, C. Roberts, A. K. Habgood, S. Alexander, T. H. Burne, D. W. Eyles, C. N. Trueman, M. Cooper, J. J. McGrath, and R. O. Oreffo. Effect of vitamin D deficiency during pregnancy on offspring bone structure, composition and quality in later life. *Journal of developmental origins of health and disease*, 4(1):49–55, February 2013.

[79] Dae-weon Lee, K. W. Seo, and B. R. Min. Discrimination between Cancer and Normal Tissue using Near Infrared Spectroscopy. In R. Magjarevic and J. H. Nagel, editors, *World Congress on Medical Physics and Biomedical Engineering 2006*, volume 14 of *IFMBE Proceedings*, pages 1341–1344. Springer Berlin Heidelberg, 2007.

[80] Claire Levy-Marchal and Delphine Jaquet. Long-term metabolic consequences of being born small for gestational age. *Pediatric diabetes*, 5(3):147–153, September 2004.

[81] Jia Li, Jasmina Saric, Yulan Wang, Jennifer Keiser, Jurg Utzinger, and Elaine Holmes. Chemometric analysis of biofluids from mice experimentally infected with Schistosoma mansoni. *Parasites & Vectors*, 4(1):179+, 2011.

[82] K. Z. Liu, M. K. Ahmed, T. C. Dembinski, and H. H. Mantsch. Prediction of fetal lung maturity from near-infrared spectra of amniotic fluid. *International Journal of Gynecology & Obstetrics*, 57(2):161–168, May 1997.

[83] K. Z. Liu, M. K. Ahmed, T. C. Dembinski, and H. H. Mantsch. Prediction of fetal lung maturity from near-infrared spectra of amniotic fluid. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*, 57(2):161–168, May 1997.

[84] K. Z. Liu, T. C. Dembinski, and H. H. Mantsch. Rapid determination of fetal lung maturity from infrared spectra of amniotic fluid. *American journal of obstetrics and gynecology*, 178(2):234–241, February 1998.

[85] Chuanyi M. Lu. Pathology consultation on vitamin D testing: clinical indications for 25(OH) vitamin D measurement. *American journal of clinical pathology*, 137(5):831–832, May 2012.

[86] Rasmus Madsen, Torbjörn Lundstedt, and Johan Trygg. Chemometrics in metabolomics–a review in human disease diagnosis. *Analytica chimica acta*, 659(1-2):23–33, February 2010.

[87] Stephen F. Malin, Timothy L. Ruchti, Thomas B. Blank, Suresh N. Thennadil, and Stephen L. Monfre. Noninvasive Prediction of Glucose by Near-Infrared Diffuse Reflectance Spectroscopy. *Clinical Chemistry*, 45(9):1651–1658, September 1999.

[88] F. Mannino. Neonatal complications of postterm gestation. *The Journal of reproductive medicine*, 33(3):271–276, March 1988.

[89] Henry H. Mantsch, Lin-P'ing Choo-Smith, and Shaw. Vibrational spectroscopy and medicine: an alliance in the making. *Vibrational Spectroscopy*, 30(1):31–41, September 2002.

[90] R. Marbach, Th Koschinsky, F. A. Gries, and H. M. Heise. Noninvasive Blood Glucose Assay by Near-Infrared Diffuse Reflectance Spectroscopy of the Human Inner Lip. *Applied Spectroscopy*, pages 875–881, July 1993.

[91] Paul Martinsen, Jean-Luc L. Charlier, Tim Willcox, Guy Warman, Andrew McGlone, and Rainer Künnemeyer. Temperature dependence of near-infrared spectra of whole blood. *Journal of biomedical optics*, 13(3), 2008.

[92] Breige McNulty, Helene McNulty, Barry Marshall, Mary Ward, Anne M. Molloy, John M. Scott, James Dornan, and Kristina Pentieva. Impact of continuing folic acid after the first trimester of pregnancy: findings of a randomized trial of Folic Acid Supplementation in the Second and Third Trimesters. *The American journal of clinical nutrition*, 98(1):92–98, July 2013.

[93] Chiara Mencarelli and Pilar Martinez-Martinez. Ceramide function in the brain: when a slight tilt is enough. *Cellular and molecular life sciences : CMLS*, 70(2):181–203, January 2013.

[94] E. Michael and D. A. P. Bundy. Nutrition, immunity and helminth infection: effect of dietary protein on the dynamics of the primary antibody response to Trichuris muris (Nematoda) in CBA/Ca mice. *Parasite Immunology*, 14(2):169–183, March 1992.

[95] D. K. Miller and J. Lenard. Antihistaminics, local anesthetics, and other amines as antiviral agents. *Proceedings of the National Academy of Sciences*, 78(6):3605–3609, June 1981.

[96] Subarna Mitra, PrasantaK Nayak, Sujata Misra, and JayaPrakash Sahoo. Effect of maternal anthropometry and metabolic parameters on fetal growth. *Indian Journal of Endocrinology and Metabolism*, 16(5):754+, 2012.

[97] Olaf Müller and Michael Krawinkel. Malnutrition and health in developing countries. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 173(3):279–286, August 2005.

[98] Koichi Murayama, Keiichi Yamada, Roumiana Tsenkova, Y. Wang, and Y. Ozaki. Determination of human serum albumin and -globulin in a control serum solution by near-infrared spectroscopy and partial least squares regression. 362(1):155–161, 1998.

[99] S. Nioka and B. Chance. NIR spectroscopic detection of breast cancer. *Technology in cancer research & treatment*, 4(5):497–512, October 2005.

[100] RobertM O'brien. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality and Quantity*, 41(5):673–690, October 2007.

[101] M. R. Odiere, K. G. Koski, H. A. Weiler, and M. E. Scott. Concurrent nematode infection and pregnancy induce physiological responses that impair linear growth in the murine foetus. *Parasitology*, 137(06):991–1002, 2010.

[102] Maurice R. Odiere, Marilyn E. Scott, Hope A. Weiler, and Kristine G. Koski. Protein Deficiency and Nematode Infection during Pregnancy and Lactation Reduce Maternal Bone Mineralization and Neonatal Linear Growth in Mice. *The Journal of Nutrition*, 140(9):1638–1645, September 2010.

[103] Julian Ollesch, Steffen L. Drees, H. Michael Heise, Thomas Behrens, Thomas Brüning, and Klaus Gerwert. FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification. *The Analyst*, 138(14):4092–4102, July 2013.

[104] Canadian Council on Animal Care. *Guide to the care and use of experimental animals*. Canadian Council on Animal Care, 1993.

[105] World Health Organization. *Vitamin and mineral requirements in human nutrition*. World Health Organization ; Food and Agriculture Organization of the United Nations, 2004.

[106] Brian G. Osborne. Near-Infrared Spectroscopy in Food Analysis. 2000.

[107] Yukihiro Ozaki. Near-infrared spectroscopy–its versatility in analytical chemistry. *Analytical sciences*, 28(6):545–563, 2012.

[108] C. Petibois, V. Rigalleau, A. M. Melin, A. Perromat, G. Cazorla, H. Gin, and G. Déléris. Determination of glucose in dried serum samples by Fourier-transform infrared spectroscopy. *Clinical chemistry*, 45(9):1530–1535, September 1999.

[109] Cristina Petisco, Balbino Garcia-Criado, Iñigo Zabalgogeazcoa, Beatriz R. Vázquez-de Aldana, and Antonia Garcia-Ciudad. A spectroscopy approach to the study of virus infection in the endophytic fungus Epichloë festucae. *Virology journal*, 8, 2011.

[110] Kristin M. Power, Javier E. Sanchez-Galan, Gary W. Luskey, Kristine G. Koski, and David H. Burns. Use of near-infrared spectroscopic analysis of second trimester amniotic fluid to assess preterm births. *Journal of pregnancy*, 2011, 2011.

[111] Pierre-Alain A. Queloz, David Crettaz, Lynne Thadikkaran, Vincent Sapin, Denis Gallot, Jacques Jani, Jan Deprest, Didier Lémery, Stefano Barelli, and Jean-Daniel D. Tissot. Proteomic analyses of amniotic fluid: potential applications in health and diseases. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 850(1-2):336–342, May 2007.

[112] Linda M. Reid, Colm P. O'Donnell, and Gerard Downey. Recent technological advances for the determination of food authenticity. *Trends in Food Science & Technology*, 17(7):344–353, July 2006.

[113] Thomas Reinehr, Michaela Kleber, and Andre Michael M. Toschke. Small for gestational age status is associated with metabolic syndrome in overweight children. *European journal of endocrinology*, 160(4):579–584, April 2009.

[114] Martin Rhiel, Michael B. Cohen, David W. Murhammer, and Mark A. Arnold. Nondestructive near-infrared spectroscopic measurement of multiple analytes in undiluted samples of serum-based cell culture media. *Biotechnology and bioengineering*, 77(1):73–82, January 2002.

[115] M. R. Robinson, R. P. Eaton, D. M. Haaland, G. W. Koepp, E. V. Thomas, B. R. Stallard, and P. L. Robinson. Noninvasive glucose monitoring in diabetic patients: a preliminary evaluation. *Clinical Chemistry*, 38(9):1618–1622, September 1992.

[116] S. Romagnani. Th1/Th2 cells. *Inflammatory bowel diseases*, 5(4):285–294, November 1999.

[117] N. C. Rose and M. T. Mennuti. Maternal serum screening for neural tube defects and fetal chromosome abnormalities. *The Western journal of medicine*, 159(3):312–317, September 1993.

[118] T. L. Rusch, R. Sankar, and J. E. Scharf. Signal processing methods for pulse oximetry. *Computers in biology and medicine*, 26(2):143–159, March 1996.

[119] Thomas P. Ryan. *Modern regression methods*. Wiley, 2009.

[120] Akikazu Sakudo, Koichi Baba, and Kazuyoshi Ikuta. Analysis of Vis-NIR spectra changes to measure the inflammatory response in the nasal mucosal region of influenza A and B virus-infected patients. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 55(4):334–338, December 2012.

[121] Akikazu Sakudo, Koichi Baba, and Kazuyoshi Ikuta. Discrimination of influenza virus-infected nasal fluids by Vis-NIR spectroscopy. *Clinica chimica acta; international journal of clinical chemistry*, 414C:130–134, August 2012.

[122] Akikazu Sakudo, Yoshikazu Suganuma, Takanori Kobayashi, Takashi Onodera, and Kazuyoshi Ikuta. Near-infrared spectroscopy: Promising diagnostic tool for viral infections. *Biochemical and Biophysical Research Communications*, 341(2):279–284, March 2006.

[123] Akikazu Sakudo, Yoshikazu Suganuma, Rina Sakima, and Kazuyoshi Ikuta. Diagnosis of HIV-1 infection by near-infrared spectroscopy: analysis using molecular clones of various HIV-1 subtypes. *Clinica chimica acta; international journal of clinical chemistry*, 413(3-4):467–472, February 2012.

[124] Akikazu Sakudo, Roumiana Tsenkova, Taisuke Onozuka, Kazuo Morita, Shuming Li, Jiranan Warachit, Yukie Iwabu, Guimei Li, Takashi Onodera, and Kazuyoshi Ikuta. A novel diagnostic method for human immunodeficiency virus type-1 in plasma by near-infrared spectroscopy. *Microbiology and immunology*, 49(7):695–701, 2005.

[125] Marko Sandor, Ferdinand Rudinger, Dorte Solle, Roland Bienert, Christian Grimm, Z. Sven GroS, and Thomas Scheper. NIR-spectroscopy for bioprocess monitoring & control. *BMC Proceedings*, 7(Suppl 6):P29+, 2013.

[126] Hiroki Sato, Akiko N. Obata, Ichiro Moda, Kazutaka Ozaki, Takaomi Yasuhara, Yukari Yamamoto, Masashi Kiguchi, Atsushi Maki, Kisou Kubota, and Hideaki Koizumi. Application of near-infrared spectroscopy to measurement of hemodynamic signals accompanying stimulated saliva secretion. *Journal of Biomedical Optics*, 16(4):047002–047002–8, 2011.

[127] Matthew Scarff, S. Alison Arnold, Linda M. Harvey, and Brian McNeil. Near infrared spectroscopy for bioprocess monitoring and control: current status and future trends. *Critical reviews in biotechnology*, 26(1):17–39, March 2006.

[128] P. A. Schoknecht, W. G. Pond, H. J. Mersmann, and R. R. Maurer. Protein restriction during pregnancy affects postnatal growth in swine progeny. *The Journal of nutrition*, 123(11):1818–1825, November 1993.

[129] Richard Scott, Emre Seli, Kathy Miller, Denny Sakkas, Katherine Scott, and David H. Burns. Noninvasive metabolomic profiling of human embryo culture media using Raman spectroscopy predicts embryonic reproductive potential: a prospective blinded pilot study. *Fertility and sterility*, 90(1):77–83, July 2008.

[130] R. Scott-Pillai, D. Spence, C. R. Cardwell, A. Hunter, and V. A. Holmes. The impact of body mass index on maternal and neonatal outcomes: a retrospective study in a UK obstetric population, 2004-2011. *BJOG : an international journal of obstetrics and gynaecology*, 120(8):932–939, July 2013.

[131] N. S. Scrimshaw and J. P. SanGiovanni. Synergism of nutrition, infection, and immunity: an overview. *The American Journal of Clinical Nutrition*, 66(2):464S–477S, August 1997.

[132] Emre Seli, Denny Sakkas, Richard Scott, Shing C. Kwok, Scott M. Rosendahl, and David H. Burns. Noninvasive metabolomic profiling of embryo culture media using Raman and near-infrared spectroscopy correlates with reproductive potential of embryos in women undergoing in vitro fertilization. *Fertility and Sterility*, 88(5):1350–1357, November 2007.

[133] R. Anthony Shaw and Henry H. Mantsch. Infrared Spectroscopy in Clinical and Diagnostic Analysis. 2000.

[134] Maria V. Silva Elipe. Advantages and disadvantages of nuclear magnetic resonance spectroscopy as a hyphenated technique. *Analytica Chimica Acta*, 497(1-2):1–25, November 2003.

[135] J. E. Sinex. Pulse oximetry: principles and limitations. *The American journal of emergency medicine*, 17(1):59–67, January 1999.

[136] Babs Soller, Anon Srikiatkachorn, Fengmei Zou, Alan Rothman, In K. Yoon, Robert Gibbons, Siripen Kalayanarooj, Stephen Thomas, and Sharone Green. Preliminary evaluation of near infrared spectroscopy as a method to detect plasma leakage in children with dengue hemorrhagic fever. *BMC Infectious Diseases*, 14(1):396+, July 2014.

[137] A. O. Soubani. Noninvasive monitoring of oxygen and carbon dioxide. *The American journal of emergency medicine*, 19(2):141–146, March 2001.

[138] Unni Mette M. Stamnes Koepp, Lene Frost Andersen, Knut Dahl-Joergensen, Hein Stigum, Oyvind Nass, and Wenche Nystad. Maternal pre-pregnant body mass index, maternal weight change and offspring birthweight. *Acta obstetricia et gynecologica Scandinavica*, 91(2):243–249, February 2012.

[139] Richard W. Steketee. Pregnancy, nutrition and parasitic diseases. *The Journal of nutrition*, 133(5 Suppl 2), May 2003.

[140] Fabio Sternberg, Carsten Meyerhoff, Franz J. Mennel, Friederike Bischof, and Ernst F. Pfeiffer. Subcutaneous Glucose Concentration in Humans: Real estimation and continuous monitoring. *Diabetes Care*, 18(9):1266–1269, September 1995.

[141] Karine Suissa. mpact of maternal Vitamin A and D deficiency and infection status on fetal and infant growth in Panama. Master's thesis, 2013.

[142] David L. Tabb, Lorenzo Vega-Montoto, Paul A. Rudnick, Asokan Mulayath M. Variyath, Amy-Joan J. Ham, David M. Bunk, Lisa E. Kilpatrick, Dean D. Billheimer, Ronald K. Blackman, Helene L. Cardasis, Steven A. Carr, Karl R. Clauser, Jacob D. Jaffe, Kevin A. Kowalski, Thomas A. Neubert, Fred E. Regnier, Birgit Schilling, Tony J. Tegeler, Mu Wang, Pei Wang, Jeffrey R. Whiteaker, Lisa J. Zimmerman, Susan J. Fisher, Bradford W. Gibson, Christopher R. Kinsinger, Mehdi Mesri, Henry Rodriguez, Stephen E. Stein, Paul Tempst, Amanda G. Paulovich, Daniel C. Liebler, and Cliff Spiegelman. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of proteome research*, 9(2):761–776, February 2010.

[143] Stefania Tentoni, Paola Astolfi, Antonio De Pasquale, and Laura A. Zonta. Birthweight by gestational age in preterm babies according to a Gaussian mixture model. *BJOG : an international journal of obstetrics and gynaecology*, 111(1):31–37, January 2004.

[144] Daniel K. Tisi, David H. Burns, Gary W. Luskey, and Kristine G. Koski. Fetal exposure to altered amniotic fluid glucose, insulin, and insulin-like growth factor-binding protein 1 occurs before screening for gestational diabetes mellitus. *Diabetes care*, 34(1):139–144, January 2011.

[145] Daniel K. Tisi, Jessica J. Emard, and Kristine G. Koski. Total protein concentration in human amniotic fluid is negatively associated with infant birth weight. *The Journal of nutrition*, 134(7):1754–1758, July 2004.

[146] Daniel K. Tisi, Xu-Jing J. Liu, Linda J. Wykes, Cameron D. Skinner, and Kristine G. Koski. Insulin-like growth factor II and binding proteins 1 and 3 from second trimester human amniotic fluid are associated with infant birth weight. *The Journal of nutrition*, 135(7):1667–1672, July 2005.

[147] Robert A. van den Berg, Huub C. Hoefsloot, Johan A. Westerhuis, Age K. Smilde, and Mariët J. van der Werf. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics*, 7(1):142+, June 2006.

[148] J. Craig Venter and et al. The Sequence of the Human Genome. *Science*, 291(5507):1304–1351, February 2001.

[149] C. G. Vergouw, L. L. Botros, P. Roos, J. W. Lens, R. Schats, P. G. Hompes, D. H. Burns, and C. B. Lambalk. Metabolomic profiling by near-infrared spectroscopy as a tool to assess embryo viability: a novel, non-invasive method for embryo selection. *Human reproduction (Oxford, England)*, 23(7):1499–1504, July 2008.

[150] Carlijn G. Vergouw, Lucy L. Botros, Kevin Judge, Mark Henson, Pieter Roos, E. Hanna Kostelijk, Roel Schats, Jos W. Twisk, Peter G. Hompes, Denny Sakkas, and Cornelis B. Lambalk. Non-invasive viability assessment of day-4 frozen-thawed human embryos using near infrared spectroscopy. *Reproductive biomedicine online*, 23(6):769–776, December 2011.

[151] Carlijn G. Vergouw, Dorit C. Kieslinger, E. Hanna Kostelijk, Lucy L. Botros, Roel Schats, Peter G. Hompes, Denny Sakkas, and Cornelis B. Lambalk. Day 3 embryo selection by metabolomic profiling of culture medium with near-infrared spectroscopy as an adjunct to morphology: a randomized controlled trial. *Human reproduction (Oxford, England)*, 27(8):2304–2311, August 2012.

[152] Carlijn G. Vergouw, E. Hanna Kostelijk, Els Doejaaren, Peter G. Hompes, Cornelis B. Lambalk, and Roel Schats. The influence of the type of embryo culture medium on neonatal birthweight after single embryo transfer in IVF. *Human reproduction (Oxford, England)*, 27(9):2619–2626, September 2012.

[153] Anjana Verma. Maternal Body Mass Index and Pregnancy Outcome. *Journal of Clinical and Diagnostic Research*, 2012.

[154] P. Vigil-De Gracia. Pregnancy complicated by pre-eclampsia-eclampsia with HELLP syndrome. *International journal of gynaecology and obstetrics*, 72(1):17–23, January 2001.

[155] José Villar, Mario Merialdi, A. Metin Gülmezoglu, Edgardo Abalos, Guillermo Carroli, Regina Kulier, and Mercedes de Onis. Nutritional interventions during pregnancy for the prevention or treatment of maternal morbidity and preterm delivery: an overview of randomized controlled trials. *The Journal of nutrition*, 133(5 Suppl 2), May 2003.

[156] Yulan Wang, Jia V. Li, Jasmina Saric, Jennifer Keiser, Junfang Wu, Jürg Utzinger, and Elaine Holmes. *Advances in Metabolic Profiling of Experimental Nematode and Trematode Infections*, volume 73, pages 373–404. Elsevier, 2010.

[157] Shi Wu W. Wen, Michael S. Kramer, Robert Platt, Kitaw Demissie, K. S. Joseph, Shiliang Liu, and Reg Sauve. Secular trends of fetal growth in Canada,

1981 to 1997. *Paediatric and perinatal epidemiology*, 17(4):347–354, October 2003.

[158] Peter D. Wentzell and Christopher D. Brown. Signal Processing in Analytical Chemistry. 2000.

[159] Lisa J. White, Sue J. Lee, Kasia Stepniewska, Julie A. Simpson, Saw Lu Mu L. Dwell, Ratree Arunjerdja, Pratap Singhasivanon, Nicholas J. White, Francois Nosten, and Rose McGready. Estimation of gestational age from fundal height: a solution for resource-poor settings. *Journal of the Royal Society*, 9(68):503–510, March 2012.

[160] Erin E. Wiedower, Andrew J. Kouba, Carrie K. Vance, Rachel L. Hansen, Jerry W. Stuth, and Douglas R. Tolleson. Fecal Near Infrared Spectroscopy to Discriminate Physiological Status in Giant Pandas. *PLoS ONE*, 7(6):e38908+, June 2012.

[161] Brock Williams. The interrelationship among maternal infection, folate and vitamin B12 status, and offspring growth in Indigenous Panamanians. Master's thesis, 2012.

[162] David S. Wishart. Computational approaches to metabolomics. *Methods in molecular biology (Clifton, N.J.)*, 593:283–313, 2010.

[163] Jerry Workman. *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*. CRC Press, 2012.

[164] Zou Xiaobo, Zhao Jiewen, Malcolm J. W. Povey, Mel Holmes, and Mao Hanpin. Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, 667(1-2):14–32, May 2010.

[165] Kaori Yamazaki, Kazunao Suzuki, Hiroaki Itoh, Keiko Muramatsu, Kotomi Nagahashi, Naoaki Tamura, Toshiyuki Uchida, Kazuhiro Sugihara, Hideki Maeda, and Naohiro Kanayama. Cerebral oxygen saturation evaluated by near-infrared time-resolved spectroscopy (TRS) in pregnant women during caesarean section a promising new method of maternal monitoring. *Clin Physiol Funct Imaging*, 33(2):109–116, March 2013.

[166] Zhangbin Yu, Shuping Han, Jingai Zhu, Xiaofan Sun, Chenbo Ji, and Xirong Guo. Pre-Pregnancy Body Mass Index in Relation to Infant Birth Weight and Offspring Overweight/Obesity: A Systematic Review and Meta-Analysis. *PLoS ONE*, 8(4):e61627+, April 2013.

[167] Zsofia Zambo. Vitamin deficiency, infection and cytokine interactions in pregnant Panamanian women: impact on cortisol-IGF-1-fundal height pathway. Master's thesis, 2014.

[168] Guangyu Zhang, Nathaniel Schenker, Jennifer D. Parker, and Dan Liao. Identifying implausible gestational ages in preterm babies with Bayesian mixture-models. *Statist. Med.*, 32(12):2097–2113, May 2013.

Appendix A

Kristin M. Power, Javier E. Sanchez-Galan, Gary W. Luskey, Kristine G. Koski, and David H. Burns. Use of near-infrared spectroscopic analysis of second trimester amniotic fluid to assess preterm births. *Journal of pregnancy*, 2011, 2011

Hindawi Publishing Corporation
Journal of Pregnancy
Volume 2011, Article ID 980985, 6 pages
doi:10.1155/2011/980985

Research Article

Use of Near-Infrared Spectroscopic Analysis of Second Trimester Amniotic Fluid to Assess Preterm Births

Kristin M. Power,[1] Javier E. Sanchez-Galan,[2] Gary W. Luskey,[3] Kristine G. Koski,[2,4] and David H. Burns[1,2]

[1] *Department of Chemistry, McGill University, Montreal, QC, Canada H3A 2K6*
[2] *Division of Experimental Medicine, Faculty of Medicine, McGill University, Montreal, QC, Canada H3A 1A3*
[3] *Division of Perinatal/Fetal Medicine, St. Mary's Hospital Center, Montreal, QC, Canada H3T 1H5*
[4] *School of Dietetics and Human Nutrition, McGill University, Hacdonald Campus, Montreal, QC, Canada H9X 3V9*

Correspondence should be addressed to David H. Burns, david.burns@mcgill.ca

Received 4 April 2011; Accepted 21 June 2011

Academic Editor: Yves Jacquemyn

This pilot study investigated the possibility that metabolomic differences exist in second trimester of women delivering at term (≥37 weeks, $n = 216$) and preterm (≤35 weeks, $n = 11$). For this retrospective study, biobanked AF samples underwent near-infrared (NIR) spectral analysis using wavelengths from 700 to 1050 nm. Spectral data was compressed then optimized by multilinear regression to create a calibration model. The resultant model was able to classify term and preterm births based on differing AF metabolomic profiles with a sensitivity and specificity of 100%. When groups were classified using a prematurity index (PI), there was a statistical difference ($P < 0.001$) between the predicted preterm group (PI 0.77 ± 0.08) and the term group (PI 1.00 ± 0.02). In conclusion, the 2nd trimester AF samples showed distinct differences in metabolomic profiles between patients delivering preterm as compared to those at term in functional groups related to proteins, carbohydrates, fats, polyols, and water.

1. Introduction

Preterm birth, defined as birth prior to 37 weeks of gestation, is a leading cause of infant morbidity and mortality worldwide [1] and has been increasing [2, 3]. Since 1990, the number of preterm births has risen by 20% [4] and is even more significant when the increased costs associated with preterm pregnancies are considered [5]. However, despite advances in identifying some of the causes of preterm birth [6], our understanding of the physiologic process leading to preterm labor is poorly understood [7].

A growing body of literature has been examining possible proteomic markers for preterm labor and for intrauterine infection, a leading cause of preterm labor [6], in maternal serum [8, 9] and in amniotic fluid [9, 10]. None has produced significant positive predictive value for preterm births. On the other hand, metabolomics, which is the summation of ongoing cellular activity and downstream of protein metabolism [9, 11], may provide another interesting approach.

Metabolomics is the measurement of multiple small molecules in various tissues and fluids. These small molecules are the products of protein metabolism and cellular function within an organism. When examined as a whole, these metabolites can be viewed as biomarkers of a functional phenotype [12]. In the case of preterm labor, differences in metabolomic profiles found in amniotic fluid are thought to be possible since biological processes of the fetus and the mother both impact on its biochemical composition.

One successful approach to measure metabolomic profiles uses near-infrared (NIR) spectroscopy. The use of NIR vibrational spectroscopy preserves the matrix of constituent metabolites and provides important information about the interactions among the various constituents in situ. This can provide insight into metabolism, based on relational properties that cannot be captured when individual components are measured. Metabolite profiling using NIR spectroscopy has been used to detect disease in different

Table 1: Population Demographics.

Maternal/Fetal Characteristics[a]	≤35 weeks	≥37 weeks	P
Gestational Age (wks)	34.8 ± 2.4*	39.3 ± 1.9	N/A
Maternal Age (yrs)	37.5 ± 1.7	37.6 ± 2.4	0.4
Pre-pregnancy BMI (kg/m^2)	23.8 ± 4.3	23.8 ± 4.9	0.6
Amniocentesis Week	15.5 ± 0.8†	15.2 ± 1.0‡	0.8
Birth weight (g)	2555 ± 540	3429 ± 623	<0.05
Parity	1.2 ± 0.6	1.1 ± 1.1	<0.05

[a]Data are reported as means ± standard deviation. Population characteristics for mothers delivering in term ($n = 216$) and preterm ($n = 11$). *Range is 28.5–35.1 weeks. †Range is 14–17 weeks. ‡Range is 12–20 weeks.

scenarios where discrimination between groups is an objective [13]. Specifically, the application of NIR spectroscopy to amniotic fluid has been used to predict fetal lung maturity [14, 15]. Differences in these metabolomic profiles obtained by NIR spectroscopy also employ multivariate regression models and optimization functions [16, 17].

This pilot study was undertaken to test the hypothesis, using NIR spectroscopy, that differences in the metabolomic profile exist in second trimester amniotic fluid samples for term (≥37 weeks) compared to preterm births (≤35 weeks). We propose that identifying the existence of a metabolic fingerprint for preterm labor early in pregnancy could be of major importance in the appropriate ongoing monitoring of at-risk pregnancies and the development of a better understanding of the biologic basis of preterm labor.

2. Materials and Methods

This was a retrospective cohort study, approved by both the McGill Institutional Review Board and St. Mary's Hospital Center (Montreal, Canada). The population included 227 subjects recruited between 2000 and 2003 who provided a small volume of amniotic fluid for spectral profiling using NIR. Women were subdivided into term and preterm categories. Inclusion criteria for term births ($N = 216$) included age-related amniocentesis for genetic testing and a singleton pregnancy with no fetal complications. The preterm group included only patients with premature rupture of members (PROM) and/or preterm labour and a spontaneous vaginal delivery; patients who were induced or had a C-section were excluded. Demographics of the study participants are listed in Table 1. The two study groups were similar with respect to both maternal age and maternal BMI. The AF samples were obtained following genetic testing and stored at −80°C; there is minimal source of biochemical error resulting from repeated freezing and thawing of amniotic fluid [18, 19].

To determine the feasibility of estimating true premature births using spectral analysis of 2nd trimester amniotic fluid collected at 12–20 wks gestation, a calibration model was constructed using a set of AF samples with known birth outcomes in the NIR region of the spectra (700–1050 nm). This spectral region is known to contain functional group information on overtone bands of CH, NH, and OH moieties [20]. Glucose, proteins, fatty acids, oils, and myoglobin have been identified as contributing to the absorbance in this NIR region [21–23].

Prior to analysis, frozen AF samples were thawed at room temperature (25°C) for 30 minutes. NIR profiles were analyzed using a reflective spectrograph with a CCD detector (B&W TEK, Newark, DE) in randomized order. A flow sample cell with 10 mm path length was filled with 15 μl of sample media. Spectra were recorded from 700–1050 nm at room temperature (25°C ± 1°C). The spectrophotometer was set to measure absorbance relative to air, and a signal average of 200 measurements with an integration time of 100 ms for each measurement was used. This measurement procedure involved rinsing the sample cell with 1 ml of 0.1 M NaOH followed by 5 ml of distilled, deionized water. The last injection of water was used to record a reference spectrum.

Quantification of the sample properties from NIR spectra consisted of determining the most parsimonious combination of variables in selected wavelength domains using a genetic algorithm optimization [16]. In this method preprocessing based on Haar wavelets, which is similar to jpeg compression of images, was used to objectively select wavelength regions. A combination of wavelength regions that most parsimoniously estimated prematurity index was determined by inverse least-squares regression, using a genetic algorithm optimization. The model investigated is of the form

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n, \quad (1)$$

where Y is the dependent variable or prematurity index (PI), $x_1, x_2, \ldots, x_n$ are independent variables (i.e., integrated wavelength region), and $\alpha_0, \alpha_1, \ldots, \alpha_n$ are the coefficients determined from a set of calibration x's. Many models were screened using the GA which is based on genetic principles such as mating, crossover, and mutation, to select the wavelength region that best separates the term and preterm groups [16].

Each sample was estimated independently using a leave-one-out cross-validation approach in a continuous multilinear model. For each individual in the population, the coefficients α_1 to α_n of (1) were calculated by inverse least-squares regression using an independent calibration. Estimates of prematurity index were obtained by applying (1) with the determined α_n parameters and the x-values of a monitoring set. Fitness of the model for each wavelength region selected was calculated as the squared difference between the mean of the term and preterm groups divided by the sum of the pooled variance. A higher fitness corresponds to better separation between the two groups as calculated by a Student's t-test. Subsequent to the estimated prematurity index (PI) optimization, notched box plots, Student's t-test results, and a receiver operator curve (ROC) were used to separately determine the statistical characteristics of the groups. The optimum value for the sensitivity and specificity was determined from the ROC [24, 25].

Additionally, spectra were examined using the major chemical groups present in the 700–1050 nm region of the NIR spectra. Molecular absorbance regions in the NIR related to H_2O, ROH, CH_2, and NH_3 were defined using

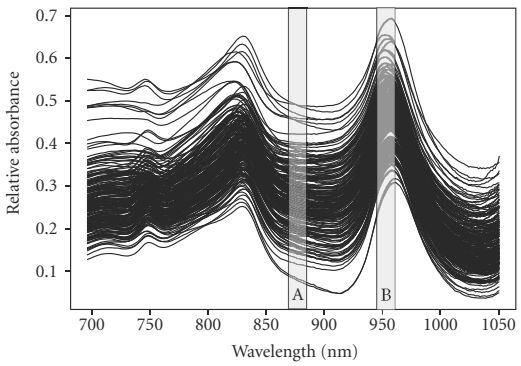


Figure 1: Raw data from 227 NIR spectral profiles of AF samples plotted relative to air. The boxes labeled A and B show the regions selected by a genetic algorithm, which give the best separation of term and preterm groups.

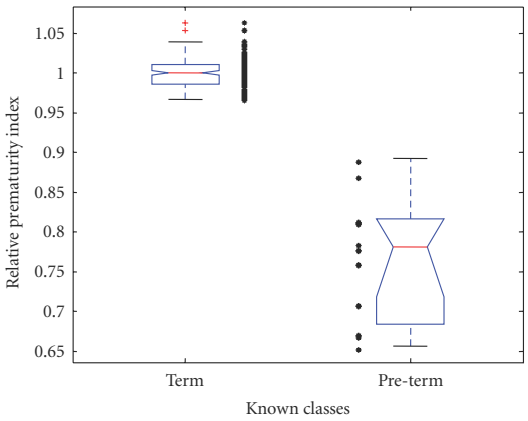


Figure 2: Box plots comparing the prematurity index for term and preterm births.

known standards [20]. Means and standard deviations were then calculated for normal (≥37 wks) and preterm (≤35 wks) using these integrated regions. Likewise, ratios of pairs of selected functional groups (NH_3/CH_2, NH_3/ROH, CH_2/ROH CH_2/ROH, NH_3/H_2O, and CH_2/H_2O) were determined to characterize concentration shifts in glucose, proteins fats and oils relative to water as well as polyols, which may play important roles in maintaining ATP concentrations, cellular redox potential, and in drawing water and solutes across the placenta [26].

All statistical analysis was done using the MATLAB (the Math Works Inc., MA) programming package.

3. Results

Population characteristics are described in Table 1. As expected, gestational age and birth weight differed. Figure 1 describes NIR spectral differences for each individual in the range of 700–1050 nm. Of all of the possible wavelet combinations, only two wavelet regions were needed to develop the best model to distinguish preterm from the term births. Those wavelet regions selected by the genetic algorithm were at 872–879 nm (region A) and 943–954 nm (region B), respectively. The first wavelength region corresponded to third overtone CH_3 and second overtone NH group vibrations [21]. This wavelet was negatively correlated with concentrations of chemicals absorbing in this region. Wavelet B selected an absorbance region characterized by aliphatic alcohol functional groups and third overtones from CH vibrations in the 943 to 954 nm wavelength range. These were negatively correlated with preterm births.

A parsimonious calibration model was constructed with the 2 wavelets selected by the genetic algorithm. Using these regions, the prematurity index (PI) was calculated. Each data point represented a blind estimation of the optimal separation value using the rest of the data for calibration. Results are represented in Figure 2 as notch box plots showing the statistical distribution for each group of classified samples [25]. A relative prematurity index (RPI) was calculated as a percentage of the median of the normals. The size of each box was determined by the quartile distribution about the median of the data. The median of the term group was 1.00 ± 0.02 and was 0.78 ± 0.08 for the preterm group. Notches of box plots, which do not overlap, have different medians at the 5% confidence level [25].

In addition, the mean and standard deviations of preterm and term prematurity indices were 0.77 ± 0.08 and 1.00 ± 0.02, respectively ($P < 0.001$). Table 2 summarizes the results of the model's validation. Using only the 2 components for the calibration model, AF samples were classified into preterm and term groups with 100% sensitivity and specificity determined by a ROC curve. Positive and negative predictive values were 100% (Table 2).

To further understand the relationship of the measured spectra to prematurity, differences in the amniotic fluid metabolomic profile between preterm and term infants arising from changes in selected spectral regions were examined. Regions selected for integration of the various functional groups as well as normalized spectra for the means for both the normal and preterm groups are shown in Figure 3. There were substantial metabolomic differences in the spectra between term and preterm infants. As well, ratios of the integrated signals from the different functional groups (Table 3) suggested nonsignificant trends in the concentrations of functional groups relative to water and/or polyols. There was a relative increase in concentration of protein as compared to carbohydrates and fats (NH/CH ratio) even though both decrease in total concentration as reflected in theNH/H_2O and CH/H_2O ratios. There was also nonsignificant increase in the NH/ROH ratio and a decrease in the CH/ROH ratio related to preterm births.

4. Discussion

Previous studies have shown gestational length is difficult to predict [9]. Genomics, proteomics, mass spectroscopy, and

Table 2: Results for the cross-validation of the calibration model and diagnostic statistics.

Preterm pregnancies		Control term pregnancies	
No. of group	11	No. of group	216
Prematurity index (PI) ($\bar{x} \pm$ sd)	0.77 ± 0.08*	Prematurity index ($\bar{x}$s $\pm$ sd)	1.00 ± 0.02
True positives	11	True negatives	216
False negatives	0	False positives	0
Sensitivity	100%	Specificity	100%
Positive predictive value	100%	Negative predictive value	100%

* P value < 0.001.

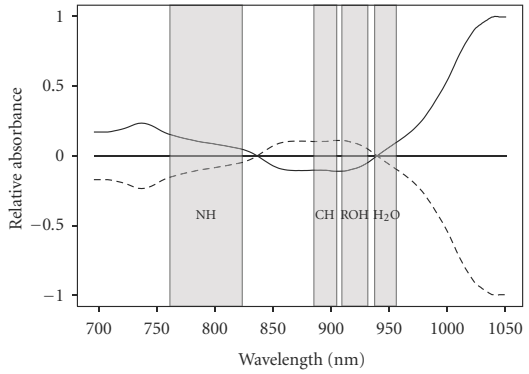


Figure 3: Normalized spectral absorbance for term and preterm groups. A solid line represents the term group (≥37 weeks), and the dashed line represents the pre-term group (≤35 weeks). This region of the spectrum consists of the 2nd and 3rd overtone absorption from CH, NH, and OH functional groups. Differing concentrations of proteins, fats, and carbohydrates in the matrix lead to increased or decreased absorbance in the spectral region corresponding to the functional groups. The result is essentially a metabolome fingerprint for preterm and term births using 2nd trimester amniotic fluid.

Table 3: Means and standard deviations by functional groups for term and preterm births[a].

Functional group ratios	Gestation period[b]	
	Preterm < 35 weeks[c]	Term ≥ 37 weeks
NH/CH*	3.49 ± 0.5	3.43 ± 0.8
NH/ROH†	2.38 ± 0.3	2.36 ± 0.4
CH/ROH‡	0.69 ± 0.0	0.69 ± 0.0
NH/H_2O§	2.28 ± 0.2	2.29 ± 0.2
CH/H_2O¶	0.67 ± 0.1	0.69 ± 0.1

[a] Data are reported as means ± standard deviation. Birth outcomes and second trimester amniotic fluid functional groups ratios for mothers delivering in term ($n = 216$) and preterm ($n = 10$).

[b] All the differences between term and preterm groups were nonsignificant at a confidence level of 10%, with the exception of NH/CH which had a P value of 0.090.

[c] It is important to mention that one sample that had a gestational age of 35.1 was discarded as an outlier. Its exclusion from this ratio analysis was primarily because it had a gestational age 2 weeks above the mean and also was 500 g heavier than the average of the group. *Relative amount of protein to carbohydrates/fats. †Relative amount of protein to modified polyalcohols. ‡Relative amount of carbohydrates/fats to modified polyalcohols. §Relative amount of protein to water. ¶Relative amount of carbohydrates/fats to water.

other assays have been used for analysis of maternal and fetal biofluids with poor success rates [8, 9, 14, 15]. In contrast, the results of this pilot study showed that NIR spectral waveforms differed significantly in the range 700–1050 nm between term and preterm births. Varying concentrations of several functional groups contributed to the variation in these spectral analyses, confirming our hypothesis that metabolomic differences exist early in pregnancy in the amniotic fluid of second trimester pregnancies between subjects delivering preterm and those delivering at term.

Although the spectral features are broad, absorbance is still directly related to the concentration of CH, NH, and OH functional groups within the matrix. These differences seen in the preterm group corresponded to CH_3 and NH (region A), aliphatic alcohols and CH functional groups (region B). Presence of proteins, water, and polyalcohols in AF are well known and can differ in concentration from sample to sample [26]. Our findings are intriguing in that they suggest the underlying processes ultimately leading to preterm delivery are present between 12 and 20 weeks of gestation. Prior studies have focused on identifying biomarkers present in amniotic fluid at the onset of preterm labor [8, 9, 14, 15]. Likewise, limitations such as AF samples taken after 22 weeks gestation, destructive analytical methods, and small populations plague many existing techniques [6, 9, 10, 14, 15]. Others have cited inability to distinguish preterm from term groups [10] and low positive predictive values [9].

Techniques from proteomics and genomics also have characterized few biomarkers in AF, thus limiting the amount of information about bioprocesses that might be associated with the amniotic fluid matrix [9–11]. The differing metabolomic profiles seen here in early pregnancy may represent an abnormal metabolic process in the fetus that ultimately predisposes the pregnancy to insults resulting in preterm labor and delivery. In addition, many recent technologies using proteomics and genomics rely on invasive procedures, expensive equipment, and technical expertise [6, 8–10]. The use of metabolomics with NIR spectroscopy is relatively inexpensive, easy to use and can characterized a metabolomic fingerprint resulting from multiple biological processes.

These results, however, should be considered preliminariy since the number of patients in the preterm delivery group is small. Despite this, our data using spectral analysis were able to distinguish between preterm and term deliveries with 100% sensitivity and specificity, which supports the potential diagnostic possibilities of this technique. Even though we were not able to identify specific differences in the ratios of the functional groups, probably owing to our small sample size, the strength of our results lies in the power of the spectral analysis to extract meaningful information about the metabolic profile from the 2nd trimester AF matrix that was associated with distinct metabolomic fingerprints early in pregnancy in our preterm and term deliveries. Moreover, the suggested metabolomic profiles were consistent with previous studies that show increased protein in amniotic fluid of premature infants [27], alterations in polyols in IUGR infants [26] and with a higher incidence of oligohydramnios [28].

Our results raise the question of whether obtaining metabolomic profiles of amniotic fluid in early pregnancy will help obstetricians identify those pregnancies destined to deliver preterm. Additionally, it may be possible to develop a noninvasive probe to analyze the NIR spectra of amniotic fluid on an ongoing basis. If future studies using larger sample sizes confirm these findings, this information would indicate that a "problematic metabolomic profile" emerges very early in pregnancy and could lead to much earlier identification of prematurity and earlier decision for potential therapies.

Conflict of Interests

None of the authors have any conflict of interests regarding this work.

Acknowledgments

The project was funded by grants from the Natural Science and Engineering Research Council of Canada and the Canadian Institute for Health Research. J. E. Sanchez-Galan thanks SENACYT and IFARHU, government agencies of the Republic of Panama, for sponsoring his Ph.D. studies at McGill University.

References

[1] W. A. Engle, "A recommendation for the definition of "late preterm" (near-term) and the birth weight-gestational age classification system," *Seminars in Perinatology*, vol. 30, no. 1, pp. 2–7, 2006.

[2] J. Lumley, "Defining the problem: the epidemiology of preterm birth," *The British Journal of Obstetrics and Gynaecology*, vol. 110, no. 20, pp. 3–7, 2003.

[3] M. S. Kramer, R. Platt, H. Yang et al., "Secular trends in preterm birth: a hospital-based cohort study," *Journal of the American Medical Association*, vol. 280, no. 21, pp. 1849–1854, 1998.

[4] B. E. Hamilton, J. A. Martin, and S. J. Ventura, "Births: preliminary data for 2005," *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, vol. 55, no. 11, pp. 1–18, 2006.

[5] M. L. Wang, D. J. Dorer, M. P. Fleming, and E. A. Catlin, "Clinical outcomes of near-term infants," *Pediatrics*, vol. 114, no. 2, pp. 372–376, 2004.

[6] A. Y. El-Bastawissi, M. A. Williams, D. E. Riley, J. Hitti, and J. N. Krieger, "Amniotic fluid interleukin-6 and preterm delivery: a review," *Obstetrics and Gynecology*, vol. 95, no. 6, part 2, pp. 1056–1064, 2000.

[7] D. N. M. Papatsonis, "Prepregnancy counseling: preterm birth," *International Congress Series*, vol. 1279, pp. 251–270, 2005.

[8] R. Shankar, N. Gude, F. Cullinane, S. Brennecke, A. W. Purcell, and E. K. Moses, "An emerging role for comprehensive proteome analysis in human pregnancy research," *Reproduction*, vol. 129, no. 6, pp. 685–696, 2005.

[9] R. Romero, J. Espinoza, F. Gotsch et al., "The use of high-dimensional biology (genomics, transcriptomics, proteomics, and metabolomics) to understand the preterm parturition syndrome," *The British Journal of Obstetrics and Gynaecology*, vol. 113, supplement 3, pp. 118–135, 2006.

[10] P. D. Harrington, N. E. Vieira, P. Chen et al., "Proteomic analysis of amniotic fluids using analysis of variance-principal component analysis and fuzzy rule-building expert systems applied to matrix-assisted laser desorption/ionization mass spectrometry," *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1-2, pp. 283–293, 2006.

[11] W. B. Dunn and D. I. Ellis, "Metabolomics: current analytical platforms and methodologies," *Trends in Analytical Chemistry*, vol. 24, no. 4, pp. 285–294, 2005.

[12] E. Seli, D. Sakkas, R. Scott, S. C. Kwok, S. M. Rosendahl, and D. H. Burns, "Noninvasive metabolomic profiling of embryo culture media using Raman and near-infrared spectroscopy correlates with reproductive potential of embryos in women undergoing in vitro fertilization," *Fertility and Sterility*, vol. 88, no. 5, pp. 1350–1357, 2007.

[13] D. I. Ellis and R. Goodacre, "Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy," *The Analyst*, vol. 131, no. 8, pp. 875–885, 2006.

[14] K. Z. Liu, T. C. Dembinski, and H. H. Mantsch, "Rapid determination of fetal lung maturity from infrared spectra of amniotic fluid," *The American Journal of Obstetrics and Gynecology*, vol. 178, no. 2, pp. 234–241, 1998.

[15] K. Z. Liu, M. K. Ahmed, T. C. Dembinski, and H. H. Mantsch, "Prediction of fetal lung maturity from near-infrared spectra of amniotic fluid," *International Journal of Gynecology and Obstetrics*, vol. 57, no. 2, pp. 161–168, 1997.

[16] C. E. Gributs and D. H. Burns, "Parsimonious calibration models for near-infrared spectroscopy using wavelets and scaling functions," *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 1, pp. 44–53, 2006.

[17] E. Turner, J. A. Brewster, N. A. Simpson, J. J. Walker, and J. Fisher, "Plasma from women with preeclampsia has a low lipid and ketone body content—a nuclear magnetic resonance study," *Hypertension in Pregnancy*, vol. 26, no. 3, pp. 329–342, 2007.

[18] R. C. Keniston, G. H. Prescott, and M. L. Pernoll, "Effects of freezing and thawing on certain properties of early gestation amniotic fluid," *Obstetrics and Gynecology*, vol. 46, no. 3, pp. 279–281, 1975.

[19] T. R. Nelson, R. J. Gillies, D. A. Powell, M. C. Schrader, D. K. Manchester, and D. H. Pretorius, "High resolution of

proton NMR spectroscopy of human amniotic fluid," *Prenatal Diagnosis*, vol. 7, no. 5, pp. 363–372, 1987.

[20] Analytical Spectral Devices Inc., "Near-IR Absorption Bands Chart," http://www.asdi.com/nir-chart_grid_rev-3.pdf.

[21] C. E. W. Gributs and D. H. Burns, "In vivo near-infrared spectrometry," in *Handbook of Vibrational Spectroscopy*, 2006.

[22] L. S. Greci, G. J. Gilson, B. Nevils, L. A. Izquierdo, C. R. Qualls, and L. B. Curet, "Is amniotic fluid analysis the key to preterm labor? A model using interleukin-6 for predicting rapid delivery," *The American Journal of Obstetrics and Gynecology*, vol. 179, no. 1, pp. 172–178, 1998.

[23] R. Romero, Y. K. Wu, M. Mazor, E. Oyarzun, J. C. Hobbins, and M. D. Mitchell, "Amniotic fluid arachidonate lipoxygenase metabolites in preterm labor," *Prostaglandins, Leukotrienes and Essential Fatty Acids*, vol. 36, no. 2, pp. 69–75, 1989.

[24] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clinical Chemistry*, vol. 39, no. 4, pp. 561–577, 1993.

[25] R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of box plots," *The American Statistician*, vol. 32, no. 1, pp. 12–16, 1978.

[26] E. Jauniaux, J. Hempstock, C. Teng, F. C. Battaglia, and G. J. Burton, "Polyol concentrations in the fluid compartments of the human conceptus during the first trimester of pregnancy: maintenance of redox potential in a low oxygen environment," *The Journal of Clinical Endocrinology and Metabolism*, vol. 90, no. 2, pp. 1171–1175, 2005.

[27] E. Bujold, R. Romero, J. Pedro et al., "Proteomic profiling of amniotic fluid in preterm labor using two-dimensional liquid separation and mass spectrometry," *The Journal of Maternal-Fetal and Neonatal Medicine*, vol. 21, no. 10, pp. 697–713, 2008.

[28] E. F. Magann, D. A. Doherty, M. A. Lutgendorf, M. I. Magann, S. P. Chauhan, and J. C. Morrison, "Peripartum outcomes of high-risk pregnancies complicated by oligo- and polyhydramnios: a prospective longitudinal study," *The Journal of Obstetrics and Gynaecology Research*, vol. 36, no. 2, pp. 268–277, 2010.