ASSESSMENT OF ACCULTURATION AND ITS ASSOCIATIONS WITH TYPE 2 DIABETES, IMPAIRED GLUCOSE TOLERANCE AND OBESITY IN AN ISOLATED CANADIAN ABORIGINAL COMMUNITY

By

Joshua Paul Feuer, B.Sc.

A THESIS SUBMITTED IN CONFORMITY
WITH THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE,
GRADUATE DEPARTMENT OF
COMMUNITY HEALTH,
IN THE
UNIVERSITY OF TORONTO
TORONTO, ONTARIO
2001

0-612-62930-9

Abstract

Title: Assessment of acculturation and its associations with type 2 diabetes, impaired glucose tolerance and obesity in an isolated native Canadian Aboriginal community
Degree: Master of Science
Year: 2001
Name: Joshua Paul Feuer
Department: Graduate Department of Community Health, University of Toronto

One approach used to identify causal risk factors for disease in First Nations has been a biocultural model. This integration allows for an analysis of cultural risk factors which still retains real biologic plausibility. For this study, subscales of acculturation were developed and test of associations were made with health outcomes in Sandy Lake, Ontario. Factor analysis was performed on the data and tests for association then examined the relationship between subscales of acculturation and disease. The results indicated that a) the risk of becoming obese or of developing IGT is lower among individuals with a diet high in traditional food, b) the risk of developing diabetes and IGT is lower among those who report having items used for traditional rather than modern activities in their home, and c) diabetes is less common among those proficient in Oji-Cree. These results are important in designing prevention and health promotion strategies in the community.

Contents

List of Tables

Acknowledgements

"Try not to have a good time…this is supposed to be educational."

- Charles Shulz

My first debt of gratitude is to Gail Eyssen and the other members of my committee: Bernard Zinman and Joel Gittelsohn. To each, I owe immeasurable thanks. In particular, to Dr. Zinman for allowing me to join the Sandy Lake Health and Diabetes Research team and for inviting me to present my findings at the 1st National Conference on Diabetes in Aboriginal Peoples in Winnipeg, Manitoba. Also to Dr. Eyssen, whose direction and careful and detailed commentaries on each chapter of this thesis proved to be an invaluable source of insight. Without her help, support and encouragement, this thesis could not have been written.

Much of my learning has also come through conversations with fellow students. I am especially indebted to Tony Hanley, not only for his conceptual advice on all matters related to the Sandy Lake Health and Diabetes Project, but also for his friendship during some of the most anxious and trying times. He has helped make this process fun and has been my *nee-gee* throughout. *Mee-gwitch*!

And before and after all, there are my parents, whose love and support are indispensable, and whose personal sacrifices for me border on the obscene.

Dedication

To the memory of my grandfather, Harvey Herstein.

To him, learning was holiness.

Chapter 1

Background and Literature Review

1.0 Introduction

Type 2 diabetes is a major health concern for native populations all over the world. In North America, prevalence rates of diabetes have been reported as high as 49.5% among the Pimas of Arizona[1] and 26.1% among the Ojibway-Cree of Sandy Lake, Ontario.[2] Some believe that this is due to the dissolution of the traditional way of life as it has paralleled the rise in the prevalence of type 2 diabetes and the 'westernization' of these communities.[3-5] Over the past 50 years, native cultures in general have become modernized; their traditional lifestyles have become an amalgam of native culture and western-minded ideologies. Although it would be useful to design a study to assess the effects of these cultural changes on the health of traditional-minded communities, historically, there has been limited success in integrating cultural and biological research perspectives. However, a more recent focus on ethnomedical models of disease has yielded studies that integrate the ethnomedical approach of the cultural anthropologist with the epidemiological approach of the biological researcher into one biocultural model. This approach considers the interaction between the population, the context in which the disease is occurring and other dynamic processes such as the rapidly changing culture.

"Acculturation" describes the adoption of a dominant culture's characteristics within a less dominant culture or community. An epidemiologic analysis of acculturation in Sandy Lake, Ontario and its association with diabetes, IGT and obesity will help illustrate how a biocultural approach to diabetes in a First Nations community can help to connect the biology with the cultural risk factors for diabetes and obesity.

This first chapter will define type 2 diabetes, describe its epidemiology, especially in relation to native communities; it will also define acculturation and describe the history and culture of the study population of Sandy Lake, Ontario.

1.1 Diabetes Mellitus: Subtypes and Characteristics

Diabetes mellitus (DM) is a chronic disease found all over the globe. It is not a single disease but a group of heterogeneous disorders all characterized by a high concentration of blood glucose (hyperglycemia).[6] This is either due to a deficiency of insulin, to a decreased sensitivity to the action of insulin or both.[7] Insulin is a hormone which is produced by the pancreas and its secretion is fundamental to the normal metabolism of carbohydrates, lipids and protein. Diabetes is known to be associated with long term complications including retinopathy, neuropathy, nephropathy as well as coronary heart disease and stroke.[6, 8, 9]

The most common forms of diabetes are termed type 1 and type 2 diabetes mellitus.[6] A third, less common subtype is known as gestational diabetes mellitus (GDM) and develops in women during pregnancy.[6] Type 1 diabetes, previously referred to as insulin-dependent diabetes, occurs most commonly in children and young adults and persists throughout adult life.[6, 10] It is a chronic disorder that results from an autoimmune process which destroys the insulin-producing beta cells of the pancreas.[11] A patient with type 1 diabetes may present with weight loss, insatiable thirst, frequent urination and elevated fasting blood glucose levels due to the body's inability to produce and secrete insulin.[6, 12] Type 1 diabetes is treated with daily insulin injections in conjunction with an appropriate diet in an effort to achieve satisfactory blood glucose control, with the ultimate goal of maintaining levels as close to normal as possible. Insulin is life-saving for people with type 1 diabetes mellitus.[13]

Type 2 diabetes, previously known as non-insulin dependent diabetes, is also a chronic disorder which involves the inability to properly metabolize blood glucose. This form of DM differs significantly from type 1 diabetes. First, individuals most at risk of type 2 diabetes include the elderly[14] and those that are overweight.[15] In the United States, prevalence rates of type 2 diabetes increase from 1.3% from the age of 18 – 44 years to 10.4% for those aged ≥ 65 years[6] – particularly in those with a high degree of 'central adiposity' or obesity localized to the mid-section of the body.[16] Second, whereas with type 1 diabetes the problem lies solely in the secretion of insulin, for patients with type 2 diabetes, the problem is an insensitivity to the action of insulin, termed 'insulin resistance' (IR) as well as the development of a defect in insulin secretion.[7, 17]

The prevalence of type 2 diabetes is lower among physically active persons than in those who are sedentary.[7] Studies have shown that increased levels of physical activity improve insulin sensitivity.[18,19] Therefore, therapy for type 2 diabetes focuses on diet and exercise and aims to improve insulin activity.[7] Type 2 diabetes may also be treated with oral drugs, which improve insulin sensitivity, increase insulin secretion, decrease excessive hepatic glucose production or slow the absorption of carbohydrates from the gastrointestinal tract.[7] In addition, 30-40% of patients with type 2 diabetes may ultimately require insulin for optimal treatment. Poorly controlled type 2 diabetes may increase the risk of cardiovascular complications such as heart attack and stroke as well as microvascular complications such as diabetic retinopathy and neuropathy.[20]

A test used to determine a body's ability to metabolize glucose is the oral glucose tolerance test (OGTT). It involves consuming a sugar drink and testing blood glucose levels over a 2-hour period. Elevated blood glucose levels as determined by this test categorize individuals as either having diabetes or a 'pre-diabetes' condition known as impaired glucose tolerance (IGT).[14, 21, 22] This test was used to help diagnose patients in this study with IGT and will be described in more detail in subsequent sections.

1.2 The Epidemiology of Type 2 Diabetes

1.2.1 DISTRIBUTION

A) REGIONAL DIFFERENCES IN DIABETES PREVALENCE

The most recent global prevalence for type 2 diabetes was estimated in 1997 at > 120 million cases.[23] Between continents, the age-adjusted prevalence of diabetes ranges from 0.682% in Africa to 4.69% in North America.[20] Prevalence rates have been estimated at 3.14% in Europe, 2.34% in Latin America, 2.64% in Oceania, 2.01% in the former USSR, and 1.4% in Asia.[20] These values were based on population figures and published prevalence rates for diabetes.[23] Actual numbers of cases were greatest in Asia with more than 46 million cases, and least in Oceania with 742,000 cases due to differences in total population.[20] Research in the United States indicates that for every 2 diagnosed cases of diabetes, there is one that remains undiagnosed.[6] In Canada, where there is an estimated 5% prevalence rate,[24] this translates into a 7.5% prevalence rate or 2 million cases of diabetes. For developing countries, in which inadequate screening and surveillance systems result in inadequate healthcare and generally poor diabetic care,[25] the true impact of the disease and its complications is unknown.

B) PREVALENCE RATE PROJECTIONS

By the year 2010, it is predicted that 215 million people will be diagnosed with type 2 diabetes worldwide.[23] This represents an increase of 44% in the current worldwide prevalence of type 2 diabetes.[20] Projections are made based on the current age distribution of the disease as well as changes expected in lifestyle and economic development.[23] Although the rate of type 2 diabetes is expected to rise on all continents, the worldwide estimate is most affected by the incidence of type 2 diabetes in high population areas like Asia.[23] For example, by 2010, the prevalence of type 2 diabetes in Asia is expected to increase by 169%,[20] representing almost 80 million new cases in just 16 years. Other continents expected to demonstrate a dramatic rise in the prevalence of diabetes include Africa, with an estimated increase of 256%, or more than 12 million new cases; the former USSR is expected to increase by 108%, or more than 6 million cases; Latin America is expected to increase by 61%, or almost 6.8 million new cases; Europe and Oceania are each expected to increase by 52%, or 8.3 million and 360 thousand, respectively; and North America's prevalence rate is projected to increase by 25%, or approximately 3.3 million new cases.[20] These projections are predicated on the theory that the global risk of diabetes will continue to rise due to the persistence of environmental factors and their potential interaction with a population's genetic susceptibility.[20, 26] Environmental factors suspected to contribute in the pathogenesis of diabetes include advancing age, nutritional factors, obesity, physical inactivity, and modernization.[20]

c) Differences in Diabetes Prevalence Among Ethnicities and Within Countries

The prevalence of diabetes is not evenly distributed between or even within countries.[27] It has been observed that in developing countries, the risk of diabetes is higher among newly prosperous groups while in industrialized countries diabetes is a disease of economic disadvantage.[5] Among these high prevalence populations, "there is a high disease burden, and a changing burden suggesting preventability, and fear that things are unknown and out of control" – which, according to Vinicor, satisfies the three criteria for classification of a public health disorder.[28]

King and Rewers' landmark report to the WHO Ad Hoc Diabetes Group illustrated the emerging pattern of diabetes in these susceptible populations.[27] This large study investigated the prevalence of type 2 diabetes among 75 diverse population groups in 32 countries. It was determined that indigenous and migrant groups were disproportionately affected by the prevalence and social impact of diabetes.[20, 27] There was enormous variation in diabetes prevalence between populations, and exceptionally high rates were documented in populations that have changed from a traditional to modern lifestyle.

The rates of diabetes are extremely high among recent generations of indigenous and migrant groups around the world. For instance, the highest recorded prevalence rate of diabetes is among the Pima Indians of Arizona and the Micronesians in Nauru. In contrast to a 3.1% general prevalence rate for all of the United States[7], Pima and Tohono O'odham American Indians exhibit a prevalence rate of ~50%;[1, 27] the Micronesians of Nauru have a prevalence rate over 40%;[27] the rate among the Melanesians of Papua New Guinea is over 30%;[27, 30] the Urban Indians in Fiji,[29] and the Australian Aborigines have reported prevalence rates in excess of 20%.[31, 32] Migrant populations have also demonstrated a high prevalence of diabetes. In particular, South Asian migrants to Britain have a prevalence of 20–30% and those who migrated to South Africa have a prevalence of 11–20%.[27, 33]

In some communities, diabetes has had little social or health impact.[34] For example, the rural Bantu in Tanzania and also the Chinese of mainland China are two communities in which the prevalence of diabetes is less than 2%.[20, 27]

While some attribute the cause of higher prevalence areas to disadvantageous genetic profile or to a rapidly changing culture, one may extrapolate that the virtual absence of the disease in these areas may be the result of favorable genetics or cultural stability. For example, these communities have maintained many traditions over the past few generations such as a traditional diet low in saturated fat. Moreover, they have not been exposed to the pressures of resettling in a new country which beset migrant populations.

The public health impact and the economic cost of diabetes and its complications are enormous in communities exhibiting extremely high prevalence rates for a single disease.[20] Diabetes is usually diagnosed after the age of 50 years,[20] however, within high prevalence populations like the Pacific Islanders, Native Americans, and migrant Asian Indian and Chinese, diabetes is prevalent at much younger ages.[23] This phenomenon will likely result in major health and social implications in the future as complications from the disease begin to emerge. Improved longevity and improved control measures for diabetes will ultimately result in a greater burden to the health system of these populations. These factors as well as the predominance of diabetes in these communities has prompted researchers to investigate the potential determinants of the disease.

1.2.2 DETERMINANTS

A) GENETIC ETIOLOGIC FACTORS

There is no single definitive cause of type 2 diabetes. Three hypotheses offered in the literature suggest that either a genetic or environmental etiology alone or some interaction of the two causes diabetes. Evidence for a genetic contribution to the pathogenesis of diabetes comes from studies of familial aggregation, as well as a high concordance rate among monozygotic twins (60 – 100%).[20, 35] However, genetic research has been unable to definitively identify the gene(s) involved in diabetes.[20,35] Researchers have postulated that the search for diabetes genes has been difficult because type 2 diabetes is a complex, heterogenous, multifactorial disease.[35] It is commonly believed that environmental determinants interact with genetic susceptibility to cause diabetes making the discovery of a single genetic mutation unlikely. This argument is reinforced by the rapid increase in the rate of diabetes among indigenous populations in just a few decades.

B) ENVIRONMENTAL ETIOLOGIC FACTORS

Some risk factors for diabetes have been linked to the social environment. Elements of culture which influence health through recognized biological pathways (e.g. diet, exercise) are identified as part of "the inter-related triad of obesity", defined by reduced physical activity and inappropriate diet.[36] These 'biologically-active' elements are considered strong and widely accepted; decreased exposure is associated with a decreased risk of diabetes.

There are also aspects of the social environment which may be associated with biologically active factors, but without any direct biological activity themselves; for example, the possession of certain household items which reduce daily physical activity, or adherence to certain customs which reduce activity or lead to poor nutritional intake. Medical anthropology has suggested that biocultural models of disease causation which include these non-biologically active risk factors can demonstrate strong associations between non-biologically active factors and disease and therefore provides a more complete picture of disease causation in communities with a very high prevalence of diabetes and obesity.[37]

The following is a list of documented risk factors for type 2 diabetes mellitus around the world:

Age

Perhaps the most well known risk factor for diabetes without regard for nationality, ethnicity or social status is advancing age. Age-specific prevalence rates for diabetes increase with advancing age in almost every study population reported.[27] Even among the rural Bantu of Tanzania and the Mainland Chinese where the prevalence of diabetes is lowest in the world, there was an almost linear increase in prevalence from age 20 to 74.[27] For high prevalence populations like the Pimas of Arizona, where susceptible persons develop diabetes at earlier ages, rates of diabetes increase from as low as 5.5 and 7.3% among males and females in the 20 – 24 year age range respectively to 66.7 and 63.6% among males and females in the 70 – 74 year, respectively.[27]

Physical Activity

Epidemiologic studies have demonstrated the role of physical activity in preventing diabetes; and regular exercise has consistently shown to improve long-term glucose control. Cross-sectional studies of aboriginal people in the South Pacific,[38] a retrospective longitudinal study of female college athletes[39] and a prospective study of men[40] and women[41] all suggest a protective effect of exercise in relation to diabetes. The benefit is believed to stem from a reduction in blood glucose levels due to exercise and can persist for hours and even days in patients with diabetes. This is due to a) an insulin-independent increase in glucose uptake by muscle, b) an increase in insulin sensitivity, and c) in persons with diabetes, the inhibition of hepatic glucose production by insulin.[42]

Physical inactivity often emerges in epidemiological studies as an independent risk factor for the development of diabetes. Cross-sectional studies have shown a two- to fourfold increase in diabetes prevalence in least-active versus most-active individuals.[26] Other prospective studies showed a positive association between level of physical activity and reduction in the risk of diabetes.[43] Exercise regimens are often prescribed for patients with diabetes with the goal of decreasing obesity – another primary risk factor for diabetes – and improve insulin sensitivity.[30, 28, 44, 45] However, evidence also shows that physical activity has a direct effect on insulin's action as well. Studies have found that peripheral insulin sensitivity is improved among individuals that are physically active.[17, 46] Therefore, active individuals are found to have a reduced risk of diabetes not only as a result of avoiding the ill effects of obesity, but through improved insulin sensitivity as well.[20, 35] This affirms the suggestion that exercise improves the body's sensitivity to insulin and will therefore decrease one's chances of developing diabetes.[40]

Diet

The most consistent dietary risk factors reported for diabetes are total calories and the percentage of calories from fat.[47] However, there is no universal association between a component of diet (for example, glucose content) and the risk of diabetes.[42] Certain populations with a high rate of diabetes have been found to consume diets higher in fat and lower in fiber than those with lower diabetes prevalence rates.[47] Yet, no conclusive association has been established between any consumption patterns and diabetes risk. The direct effect of a diet high in total calories and fat on the risk of obesity,[48] a well-established risk factor for diabetes and insulin resistance, may account for the consistent association of diet and diabetes risk.[15, 49]

A common misconception is that an improper diet is the primary reason for the rise in the prevalence rate of diabetes in Native North Americans and also that a proper diet may be protective of developing diabetes. In fact, it is likely not diet itself but rather a more complex combination of genetic and environmental risk factors which has lead to diseases such as diabetes in these groups.[47] Some believe that a traditional diet characterized by animal protein and low complex carbohydrates may protect against diabetes and obesity.[50] However, researchers have speculated that the composition of the diet is less influential on diabetes risk than the effect of drastically changing a community's eating habits over a short period of time.[51] For example, although the current diet of the Pima Indians is similar in composition to that of the general U.S. population, they have the highest rate of diabetes in the world.[47] The fact that a dramatic change in Pima culture – including dietary practices – is evident over several decades may explain why many aboriginal groups display high rates of diabetes while consuming diets similar to those of their neighbors. This may explain why only limited research exists which shows any connection between particular components of diet and the incidence of diabetes that have been reproduced across communities undergoing similar rapid modernization processes.[42]

Obesity

Both physical inactivity and consumption of a diet high in fat and low in fiber contribute to the risk of obesity. Obesity is not only viewed as a risk factor for diabetes, but similar to diabetes, is also considered to be closely linked with the inability to respond to the action of insulin, or 'insulin resistance'.[36] Among the Pima Indians, obesity has had little effect on the rate of diabetes in the absence of a family history of diabetes,[49] and in monozygotic twins, the genetic component of diabetes has produced results which indicate that diabetes can develop independently of obesity (in non-obese people).[52] Obesity is still considered the strongest predictor of diabetes in the general population.[6]

Ethnicity

As described in a previous section, there are dramatic differences in the prevalence of diabetes across various ethnicities. There are virtually no cases of diabetes in some traditional societies like the Bantu of Tanzania and the Mapuche Indians of Chile, while nearly half of the population of Pima Indians and Nauruans are affected.[27, 51] In the United States, diabetes is nearly twice as common among blacks and hispanics as in non-Hispanic whites.[47] This disparity in diabetes prevalence is not explained by underlying differences in the prevalence of obesity, or some other behavioral risk factor, as these values are constant even after adjustment for age, gender, obesity, family history of diabetes, education and income.[6] These same results were found in both the San Luis Valley Diabetes Study of Hispanics in Colorado[32] and the San Antonio Mexican Heart Study.[54] This may indicate that some combination of factors exists which is unknown and perhaps varies only by ethnicity.

C) Gene-Environment Interactions

The relative contribution of genes and environment to the risk of diabetes among individuals in once traditional societies is still unknown. Yet, studies of urban-rural and native-migrant populations indicate substantial differences in the attributable risk due to environmental factors. There is evidence that the risk of obesity and diabetes is increased in communities where exposure to modern conveniences has caused rapid changes to elements of culture such as diet and level of physical activity.[55, 56] There is also evidence of a dose-response relationship from studies of urban, periurban and rural Pacific Islanders.[57] The risk of disease and poorer blood lipid profile increased from least to most modern society and with proximity to the rural centre.[57] Evidence also exists which indicates a higher prevalence of diabetes among landed Mexican and Japanese immigrants to the U.S., and South Asian immigrants to Britain and to South Africa than observed in their home countries.[27, 33] Similarly, Native communities such as the Choctaw of Mississippi, the Ojibway-Cree of Sandy Lake, the Haida of British Columbia, the Natives of Alaska, Algonquins of Quebec and the Pimas of Arizona all have a higher prevalence of diabetes than does the general population of Canada or the United States.[1, 6, 27, 56]

"The Thrifty Genotype Hypothesis"

Neel's thrifty genotype hypothesis has persisted for over thirty-five years as the most widely held explanation for the clustering of diabetes among rapidly developing populations.[58] Neel argued that if there were a 'diabetic genotype' which was highly prevalent in some communities, it must have originally been advantageous for survival.[59] The theoretical rationale rests on the belief that diabetes is largely genetically determined and that the adaptive genotype whose expression was once beneficial is now a detriment to health and survival.[58]

The theory suggests that in those in whom it was expressed, the diabetes genotype was efficient – or 'thrifty' – in storing energy during periods of food surplus and subsequently utilizing energy stores during periods of food scarcity.[58, 59] Therefore, a survival advantage was conferred during periods of food shortages since the food supply was dependent upon the abundance of wild game and a depletion would have otherwise caused great depopulation.[60] Shortly after European explorers began colonizing the 'New World', the low-fat, low-carbohydrate, high-protein traditional diet transformed into an energy-dense, high-fat and high-carbohydrate diet to which the thrifty genotype no longer conferred a survival advantage.[59] Instead, its host was rendered susceptible to obesity and diabetes due to a surplus of high calorie food with no imminent danger of famine.[59, 61] This transition is still apparent today within these communities where both traditional and modern diets are eaten.

Some researchers have felt that little solid evidence exists to substantiate Neel's thrifty genotype hypothesis.[58, 62] Since no single major gene has been isolated and found to account for type 2 diabetes or insulin resistance, those who challenge Neel's theory cite the absence of real genetic evidence for the heredity of the disease.[58] However, recent advances in the fields of genetic research and molecular biology and the identification of candidate genes for insulin resistance and insulin secretory defects may weaken these arguments. Neel himself believed the feast or famine hypothesis presented an overly simplistic view of the physiological adjustments involved in the transition from the traditional to modern lifestyle.[60] However, even Neel's most rigid opponents agree that as a broad explanation, the thrifty genotype hypothesis encapsulated the likely interplay between evolving environmental influences and genetic factors.[58]

1.3 Acculturation: An Environmental Determinant of Health?

1.3.1 Defining and Measuring Acculturation

Epidemiologic evidence suggests that exposure to sociocultural change is a risk factor for non-communicable diseases such as diabetes and obesity in susceptible populations.[56, 57, 65] Studies have observed that within many of these populations life is evolving from a traditional to a modern existence [3, 5, 10, 20, 59, 62, 66] and therefore this cultural phenomenon is better described as a 'transition'. Various terms have been used interchangeably to describe this trend, for example, acculturation, modernization, urbanization, westernization, and globalization, however, they all attempt to describe the same cultural phenomenon: the adoption of a dominant culture's characteristics within a less dominant culture or community.

Within each community studied, different strategies have been used to define and measure associations between acculturation and various health indices. For example, Australian Aborigine communities have undergone a rapid urbanization. During this shift from rural to urban living, the rate of diabetes has steadily increased.[32, 67, 69] Based on a study of Australians with a high diabetes prevalence rate (11%), Wise proposed that the recent shift toward urban living was associated with an increased risk of diabetes.[31] An even higher prevalence of diabetes (17%) was observed among 'fringe dweller' Aborigines of the West Kimberly whose lifestyle was characterized by poverty, high unemployment, high alcohol consumption and a generally poor diet.[67] Recently, the prevalence rate of diabetes among Aborigines in New South Wales was reported in excess of 20%.[32] Further comparison of diabetes rates among newly urbanized Australian communities indicated that Aborigines who have moved to urban centers experienced a greater prevalence of diabetes and other adverse health outcomes than prevalence rates seen among individuals still in rural villages.[66]

Similar findings were evident among Melanesians of Papua New Guinea where a survey revealed that rates of diabetes and serum lipid levels are greater in urban coastal and periurban highland subjects than among their rural counterparts.[30, 57] A seven-question survey was used to gather information on aspects of cultural change in the community and an 'index of relative modernity' was used to differentiate the influence of the changing culture among the Coastal and Highland New Guineans. While acculturation was never explicitly defined, based on the elements of culture which were studied, it is apparent that education and employment were considered central to the definition. Area of origin, years of contact with government, educational level, type of employment, number of years in an urban center, father's employment and type of housing were all items which contributed to a categorization of the degree of 'acculturation'.[68] Equal weighting was applied to each response and an additive scale produced a maximum score of 50 to each individual. Subsequent statistical analyses investigated correlations with known biological risk factors such as blood pressure, 2-h plasma glucose concentrations and measures of body adiposity such as triceps skinfolds, and body mass index.[68] In addition to sex, age, village, BMI, fat distribution, glucose tolerance, and physical activity, the 'index of relative modernity' figured as a contributor to variations in cholesterol and triglyceride concentrations.

Studies of the Choctaw of Mississippi also showed significant associations between lifestyle factors and health. In these analyses, acculturation was defined as the sum total of sociocultural change and was viewed as a representation of physical (body mass), behavioral (physical activity), and sociocultural ('lifestyle incongruity') factors.[56] An acculturation score was constructed using a community-wide questionnaire to assess the association between acculturation and glucose tolerance, body mass index and other anthropometric variables.[56] The Dressler 25-item style of life questionnaire assessed accumulation of consumer goods and exposure to messages in the mass media. Possession of household items were measured and included telephones, color televisions, tape decks, stereos, microwaves, VCRs, cable TVs, air conditioning, cars, travel habits, and exposure to various media outlets.[56] The score from the style of life survey was then compared to a score which described economic status to produce a new variable: 'lifestyle incongruity' (i.e., style of life – economic status = lifestyle incongruity).[56]

The psychosocial stress represented the extent to which individuals lived within their means – seen as traditional behavior– rather than beyond their means – seen as a North American phenomenon. The stress which the authors described was also considered a symptom of acculturation as a result of the Choctaw having an indeterminate status in the region; that is, neither wholly traditional nor wholly North American.[56] In addition to this chronic perceived stress, the authors postulated that changes in physical activity and body mass comprised the aspects of acculturation which influenced rates of diabetes and other health outcomes.[56] The findings indicated that a higher lifestyle incongruity was associated with higher plasma glucose and body mass index (BMI). The investigators concluded that acculturation increased one's risk of diabetes and obesity.[56]

Studies of the Australian Aborigines, Papua New Guineans and Choctaw of Mississippi support the theory that certain cultural and environmental factors play at least some role in the development of diabetes and obesity. However, while a general association between lifestyle factors such as diet and exercise and health is fairly well established,[6] the search for the etiological components of acculturation is not far advanced.[68] Furthermore, while most studies have linked sociocultural change to increased diabetes, at least one investigation could not replicate these findings.[65]

This particular study investigated the association between acculturation and diabetes among Mexican Americans and found a negative association or beneficial effect from acculturation on diabetes and obesity.[65] This study defined acculturation as "a multidimensional process in which individuals whose primary learning has been in one culture (i.e., the Mexican or Mexican American culture) take over characteristic ways of living from another culture (i.e., the mainstream, non-Hispanic white culture)".[65] For Mexican American men and women, diabetes prevalence was inversely related to their socio-economic group.[65] These results were inconsistent with the findings from the Australian Aborigines, Papua New Guineans, Mississippi Choctaw as well as other studies including investigations of Japanese Americans among whom a positive association between diabetes and acculturation has been documented.[56, 65]

The questionnaire given to the Mexican Americans in San Antonio yielded scales on three separate dimensions of acculturation: functional integration with mainstream society, value placed upon preserving traditional culture and attitude toward traditional family structure and sex-role organization.[65] In addition to these scales, a ranking of socioeconomic status was also included: the Duncan Socioeconomic Index which measures SES based on occupational prestige. Like acculturation, the investigators treated socioeconomic status as an indicator of cultural assimilation.[65] This study stressed that these findings were community specific and were not to be extrapolated to other communities.[65] According to them, acculturation should be viewed as the combined influence of each of the singular evolving elements of a given culture, one of which is a culturally appropriate measure of socioeconomic status.

Disparities such as the results from the Mexican American study in San Antonio are not inexplicable. Some believe that the rate of disease among persons undergoing the acculturation process will rise or fall to approach the rate found in the host culture.[65] Support for this hypothesis is found in the prevalence of diabetes among Japanese Americans, Pacific Islanders, and Mexicans in America.[65] However, this does not explain how a community's prevalence rate can surpass that of the host society's as is the case among many aboriginal groups in North America and Australia. Discrepancies in the direction of the association for acculturation-health studies indicate that there is still much to understand regarding the health-effects of acculturation.

It has been found that for studies of acculturation within traditional-minded First-Nations communities with high rates of diabetes, general concepts of socioeconomic status are not applicable.[56, 70] As a result, community based surveys are specifically designed to assess SES in the absence of standard indicators (salary scale, home ownership).[70] Socioeconomic status –as it is known in mainstream society– does not necessarily correspond to the true economic status of traditional individuals nor would it provide a picture of the level of acculturation in these traditional communities. Thus, acculturation and SES ought to be viewed as distinct, though not necessarily uncorrelated social phenomena.

1.4 Sandy Lake, Ontario: Determinants of Health

1.4.1 Historical Perspective of the Sioux Lookout Zone

The native communities that settled within the region now known as the Sioux Lookout Zone are the descendents of the plains Ojibwa and Cree people.[71] Historically, these tribes were completely dependent upon the roaming herds.[72] They were a nomadic and migratory people and the hunt yielded flesh for food, deer hides for summer clothing, buffalo hides for winter garments and bones and horns for tools.[71] During the winter season, meat was accumulated, dried, mixed with fat and pulverized into a high-energy source of food known as pemmican.[71] This could be kept for a year without spoiling and represented the major source of energy during periods of food scarcity. Health risks endemic to the region were primarily a result of exposure to natural elements: trauma, infections, starvation, childbirth, and harsh climate, to name but a few.[3, 72]

The plains Ojibwa and Cree began interacting with European explorers during the early 1600's.[72] Europeans were taught by the natives how to survive the harsh North American climate using hunting, fishing and trapping techniques and in return, European customs, ideologies and conveniences were introduced. New to Canada were iron tools and weapons – necessities for hunting and fighting; alcohol – previously unknown to North America; and the fur trade.[71]

Acute infectious diseases introduced by the Europeans lead to an epidemiologic transition, resulting in increased rates of smallpox, measles, influenza, typhus and diptheria.[4, 5, 72] Acute infectious diseases like malaria, yellow fever and hookworm, believed to have been introduced by African Slaves, also played an important role.[4, 5] Some authorities believe that many of these diseases did exist in pre-contact North America but only emerged as major killers once the stress, disorganization and epidemic-induced panic was felt upon European contact.[5]

Changes in various aspects of culture have profoundly effected the lifestyle of Native North Americans and are evident in Sandy Lake, Ontario, in the Sioux Lookout Zone. Most notably, hunting and fishing are no longer essential activities of survival. The implementation of the reservation system and the distribution of food via local grocery stores have made hunting for survival virtually obsolete and travel outside of the contained reservation unnecessary.[72] A growing trend toward modern conveniences and advanced technology now exists within this native community.

Store-bought groceries and therefore the present native diet tends to be high in saturated fats, cholesterol, and simple sugars, and generally low in fiber.[73] The community's reliance upon food sold at a local grocery store rather than the spoils of hunting and fishing has decreased physical activity and changed the general nutrition of community members.[72] Traditionally, wild game and the gathering of wild plants provided a diet high in protein, moderate in fats and relatively low in carbohydrates and dietary fiber.[48, 72] Exposure to Western-media and education have also contributed to social change in the past few decades in Sandy Lake.[72] Except for community gatherings, traditional activities are minimal and life is relatively sedentary in relation to the once active native lifestyle.

This shift in the way of life, from traditional to modern, is thought to be playing a role in the prevalence of diabetes among native Canadians.[5, 72] The inhabitants of the region have witnessed a corresponding rise in non-communicable diseases such as diabetes and obesity in the past few decades, as has been observed in other native communities acculturating to North American lifestyles.[48, 72]

1.4.2 Diabetes in Sandy Lake, Ontario: Distribution and Determinants

The Sandy Lake Health and Diabetes Project (SLHDP) assessed the prevalence of type 2 diabetes and its risk factors in Sandy Lake. The survey was developed specifically for the Sandy Lake community. Extensive ethnographic fieldwork and local pre-testing exercises were completed to help develop a culturally sensitive data collection instrument.[70] Questionnaires were administered in both English and Oji-Cree, depending on the preference of the respondent.[70] A total of 728 community members > 10 years of age (72% of the eligible population) participated in the survey.[2] Interviews were conducted to complete the questionnaire and consisted of four parts: a) prevalence screening and blood analysis, b) household questionnaire, c) anthropometric measurements, and d) blood pressure and fitness testing[70] (see Chapter 2).

Results from the prevalence study were published in 1995.[2] The community age-adjusted prevalence of diabetes was 26.1%.[2] Of those diagnosed with diabetes, 41% were newly diagnosed at the time of the SLHDP.[2] The age-adjusted prevalence of impaired glucose tolerance (IGT) was almost three times more common in females (19.8%) than in men (7.1%).[2] This finding is inconsistent with studies of other native communities in North America but may be explained by the greater prevalence of obesity among females in this community[2] (SLHDP, mean BMI: male subjects: 25.5, female subjects: 27.3).[74]

Studies of other Canadian native communities have yielded prevalence rates for diabetes in excess of the estimated 5% for the general population.[75] However, the community prevalence of diabetes determined by the SLHDP ranks among the highest in the world; only the Pima Indians of Arizona, the Nauruans of Micronesia, and the Melanesians of coastal Papua New Guinea have reported rates of diabetes which surpass those of the Oji-Cree within this isolated native Canadian community.[27]

Diabetes, obesity and their associated risk factors are major health concerns within Sandy Lake.[2] The rise in the rate of diabetes is generally believed to be the result of a genetic predisposition within North American natives acted upon by influences related to a sudden environmental and cultural transition.[20, 43] Similar ethnomedical models of disease causation are suspected to be acting within the Pimas of Arizona, the Choctaw of Mississippi, the Algonquin of Quebec and other first-nations communities in North America. It is commonly held that the increasing non-native influences in Sandy Lake have resulted in an increased risk of diabetes and obesity.[72] Furthermore, the possession of modern devices in the home, changes to educational practices and nutrition and decreased activity are each suspected to be aspects of culture which have changed as a result of modern influences.[56]

The associations between various indices of acculturation and rates of diabetes have helped establish that ethnomedical approaches to disease causation exist and can contribute to an understanding of disease pathways in traditional communities. Furthermore, although each study employed a different definition and instrument in order to measure cultural change, it is clear that acculturation is believed to have some mediating role in the health of these communities.

For Sandy Lake, where the culture is currently changing and therefore both traditional and modern elements coexist, identifying which elements of lifestyle correlate with the presence of diabetes would have practical implications for future health promotion and diabetes prevention strategies.

1.5 Study Objectives

This study was oriented to investigate both biological and cultural determinants of disease. Where possible, associations between purely cultural risk factors and known biological risk factors were determined. These were performed to better understand the causal pathway by which these non-biological risk factors effected disease incidence.

Data from the SLHDP on frequency of consumption of foods, items owned by households and language proficiency were used to:

1. Determine if clusters exist in the data from the Food Frequency, Household Items and Language sections of the SLHDP and to assess whether these represent traditional and North American lifestyles.
2. Determine whether such clusters were associated with the risk of diabetes, IGT and obesity in the community.

This study will therefore ascertain whether aspects of a traditional Ojibway-Cree lifestyle, as opposed to the more ubiquitous lifestyle of the surrounding North American culture were associated with the risk of acquiring type 2 diabetes, IGT or of becoming clinically obese. This information will help clarify the importance of these elements of acculturation and the consequences of the rapidly changing culture on the health of a traditional community.

Chapter 2

Methods

2.0 Introduction

This chapter outlines the methods which were used to meet the objectives of the present analysis. The data are from the Sandy Lake Health and Diabetes Project (SLHDP), a cross-sectional, descriptive, prevalence survey of an isolated native reservation in Northern Ontario. An understanding of the culture as described in the ethnographic analysis which preceded the SLHDP was helpful in identifying variables which contributed to a measurement of acculturation in the community. Since the 'level of acculturation' was not measured explicitly by the SLHDP survey, factor analysis was performed on the data in order to identify the number and nature of underlying factors from the collected data set. Individual scores on each factor were then used in multiple variable logistic and linear regression analyses in order to identify associations between acculturation and various health outcomes in the community. The design of the study, methods used, and the variable selection procedure for factor analysis as well as other statistical procedures are all described in detail in this chapter.

2.1 The Source of Data: The SLHDP

Sandy Lake, Ontario is a semi-isolated native Canadian reservation with 1611 Ojibwa-Cree inhabitants.[77] With the exception of a winter road which allows passage for about six weeks in January and February, the community is accessible only by air for most of the year. Sandy Lake is approximately 2000 kilometers northwest of Toronto and is located in the north-central portion of the Sioux Lookout Zone. This is an area of subarctic boreal forest with 28 communities including Sandy Lake, all serviced by the Sioux Lookout Zone Hospital, federal nursing stations or satellite clinics. The demographic and social profile, the culture and general health of the people of Sandy Lake are considered representative of the native population of northwest Ontario.[78]

The Sandy Lake Health and Diabetes Project (SLHDP) began with an ethnographic analysis of the community in January 1992. Information was gathered by a medical anthropologist and a local assistant on health beliefs and attitudes, food consumption and physical activity, notions of disease causation and opinions on the determinants of health and disease prevention.[72] Local concepts of disease and diet were found to be dichotomized into "Indian" and "white man's" groupings, with the former believed to be healthy and the latter unhealthy.[72] Much of the community believed that a diet high in traditional food and low in white man's "junk food" (sugar, soda) were protective for diabetes. Diabetes and obesity were believed to result from consumption of white man's food. Increasing one's level of physical activity was not seen as a means of controlling obesity or of decreasing the risk of diabetes, and many of the foods believed to be healthy were in fact high in saturated fat content.[72]

Using the information from the ethnographic investigation, a culturally appropriate questionnaire was constructed in order to conduct a community-wide survey. The primary objectives of the project were the following: a) to determine the prevalence of non-insulin dependent diabetes mellitus (type 2 diabetes) and impaired glucose tolerance (IGT) in Sandy Lake, b) to identify anthropometric, metabolic and lifestyle characteristics associated with type 2 diabetes and IGT in the community, and c) to use ethnographic data collection techniques to aid in the development of culturally appropriate data collection instruments and intervention strategies to modify risk factors for diabetes and its complications.[70]

Inclusion criteria were as follows: individuals who were registered members of the Sandy Lake Band and had lived in Sandy Lake for at least six months of the previous year. Registered members of other bands living in the households of Sandy Lake Band members were also included. Students who met these criteria but were studying out of the community were enrolled when they returned home.[70]

Data were collected from July 1993 to March 1995. A total of 728 individuals participated in the prevalence survey, representing a community participation rate of 72%.[2] Individuals aged 10 years and older were administered the structured survey in either English or Oji-Cree. The survey included the following sections: sociodemographics, health beliefs and knowledge, household items, language proficiency, food frequency (food intake over the past three months), 24-hour dietary recall, substance use, activity recall, concepts of body image, and a family tree for history of diabetes.

Table 2.1.1 Description† of the SLHDP Participants by Sex

	Male (n = 305)	**Female** (n = 423)
Age (years)	30.4 ± 15.9	29.4 ± 15.8
Height (cm)	171.5 ± 10.3	160.7 ± 7.1
Weight (kg)	76.1 ± 19.3	71.2 ± 17.4
Body mass index (kg/m^2)	25.6 ± 5.1	27.4 ± 6.1
Education, % none	3.3	4.4
Education, % secondary or more	7.8	7.5

† Values are means ± standard deviations

Table 2.1.1 displays the anthropometric characteristics of the study sample used for the SLHDP. Physical examination included basic anthropometry and was performed with volunteers in either a cotton examination gown or light athletic clothing. Height was measured to the nearest 0.1 cm using an Accustat wall-mounted stadiometer. Weight was measured to the nearest 0.1 kg using the standard hospital balance beam scale. Each measurement was performed twice and the average was used in the analysis. Body mass index (weight [kg] / height [m^2]) was calculated as an indicator of obesity. (For the purposes of the present analysis, obesity will be defined as a body mass index of 24 kg / m^2 or greater as was used in a recent report of Sandy Lake's residents).[44]

A standard 75-g oral glucose tolerance test (OGTT) was administered to all participants and a blood sample for glucose collected after 120 minutes. Diagnoses of diabetes and IGT used the original World Health Organization criteria:[34] individuals were classified as diabetic if their single fasting glucose concentration was ≥ 7.8 mmol/L[†] or if their blood glucose was ≥ 11.1 mmol/L two hours after a 75-g OGTT. Individuals were classified as IGT if their glucose concentration was ≥ 7.8 mmol/L but less than 11.1 mmol/L two hours after a 75-g OGTT. Volunteers were excluded from the OGTT if they had physician-diagnosed diabetes and were currently being treated with insulin or an oral hypoglycemic agent, or if they had physician-diagnosed diabetes and a fasting blood sugar of > 11.1 mmol/L. Children aged 10 - 18 received 1.75 g of glucose per kg of body weight to a maximum of 75 g. Volunteers who were pregnant at the time of recruitment were temporarily excluded and contacted three months postpartum.[72]

A simple three-stage step test was used to estimate maximal oxygen uptake (VO_{2max}), the most widely accepted measure of fitness level.[80] Volunteers stepped on a high exercise step machine for three minutes per stage to a maximum of three stages. Volunteers with medical conditions that contraindicated vigorous exercise (cardiovascular disease, respiratory disease, arthritis, and amputations) were excluded.

The results of the SLHDP were published in February 1997.[78] The age-adjusted[*] community prevalence of diabetes was 26.1% (28% for females, and 24.2% for males). The age-adjusted prevalence of IGT was 13.6% (19.8% for females and 7.1% for males). In those individuals between the ages of 18 and 49 years of age, all measures of obesity were highly associated with diabetes, and obesity was more common in women than in men.

† Note: recommendations to change these diagnostic criteria were made in 1998. Impaired fasting glucose is now 6.1 – 6.9; impaired glucose tolerance requires a fasting plasma glucose <7.0; diabetes mellitus requires a fasting plasma glucose of ≥ 7.0 or a plasma glucose value in the 2-h sample of the oral glucose tolerance test (OGTT) ≥ 11.1 mmol/L.

* Note: Prevalence rates were calculated by age and sex and adjusted for age using the direct method. Two standard populations were used: the Canadian population from the 1991census[81] and the World Standard Population from King and Rewers.[82]

2.2 Acculturation Study Design

The present study was an analysis of the data collected from the Sandy Lake Health and Diabetes Project (SLHDP). The SLDHP was a cross-sectional prevalence survey of diabetes and its associated risk factors in the native reservation of Sandy Lake, Ontario. Participants from the SLHDP completed questionnaires and were subsequently identified as having diabetes, impaired glucose tolerance or obesity. Subjects were included in the present analysis if they had no prior diagnosis of diagnosis of diabetes or IGT and if they were 20 years of age or older. The resulting sample size was 416 individuals. SLHDP participants without diagnosed diabetes, IGT or obesity represented the control group while newly diagnosed cases (i.e.: at the time of the survey) of diabetes, IGT or obese individuals represented the cases (see Chapter 3, section 3.1.2 for full sample demographics). Therefore, the present study which was designed to investigate the association between acculturation and various health outcomes from data collected during the SLHDP, was a case-control analysis nested within the original SLHDP prevalence survey.

2.3 Partnership with the Community

The SLHDP was performed in conjunction with the Sandy Lake Band Council and community members. Community leaders were involved in all aspects of planning and implementation. Ongoing investigations, including follow-up interventions and health promotion initiatives in the community, continue to operate under the supervision and with the full co-operation of individual members and leaders of the community.

Acculturation was not measured directly in the original baseline SLHDP. However, since the data used to derive this measure was taken directly from the SLHDP, no further ethics approval was deemed necessary. The original ethics approval for the SLHDP and the continued relationship between members of the research team and the Sandy Lake leaders was considered sufficient to proceed with the present analysis of the SLHDP data.

2.4 Acculturation Subscale Development

2.4.1 Item Selection

Initial examination of the data revealed that not all of the data would contribute to a determination of acculturation. Predetermined criteria were used to establish whether various sections of the SLHDP's questionnaire were useful in measuring acculturation in Sandy Lake. Information from the survey was included and subjected to factor analysis (scale development) and subsequent tests of association, if the data met three basic criteria for inclusion: a) ethnographic justification, b) data completeness and interpretability, and c) variability of item responses. Ultimately, only the following sections of data from the questionnaire were used in the development of scales measuring acculturation in the community: Food Frequency, Household Items and Language Proficiency.

Ethnographic justification represented the first of three criteria for data inclusion. All information collected during the SLHDP was scrutinized in conjunction with the ethnographic research from the original study and sections were included / excluded on the basis of this first criterion. This rather subjective criterion was chosen in order to ensure that the data contributed some meaningful cultural features to a comprehensive assessment of acculturation. Furthermore, it was chosen to ensure that the data used to describe acculturation in the community actually contributed some knowledge of the ethnography and culture of the region.

Data completeness and interpretability was the second criterion. This ensured that the data were practical to use, which meant that they were interpretable and that they were completed by a maximum number of respondents. If an individual did not complete a single item, factor analysis would automatically delete the individual from analysis, reducing the power of the analysis and further effecting the reliability of the scale.

The final criterion ensured that questionnaire items measured both traditional and North American characteristics. In order to use this criterion effectively, it was necessarily to understand the difference between North American and Traditional responses. This was done with the help of the SLHDP study coordinator (Anthony Hanley) as well as the anthropologist who wrote the original SLHDP survey (Dr. Joel Gittelsohn). This improved the variability of responses in the data and allowed for the development of a scale of acculturation which measured the extent of one's traditional and North American existence separately.

Sections Included

A. FOOD FREQUENCY: In general, a person's diet is closely linked to their culture. Also, the extent to which a member of the community consumes a traditional or North American diet is easily observed by his/her consumption of certain food groups (white man's or native foods). Two separate instruments measured dietary intake on the SLDHP: a 24-hour dietary recall and a Food

Frequency index. Both of these satisfied the first criterion (ethnographic justification), however, only Food Frequency data were used. Although both of these instruments provided insight into the degree of Oji-Cree culture (and in turn, Western culture) as measured through diet, the Food Frequency data provided a general description of usual dietary habits while the 24-hour recall described only an individual's single day consumption. However, the Food Frequency survey relied on an individual's memory to provide answers (with visual prompts to assist in recalling portion sizes). The Food Frequency survey provided information on both traditional and white-man's foods and frequency of consumption (e.g.: once a day, > once a day, etc.). Last, the Food Frequency section was completed by nearly the entire sample for this study (n = 406, representing 99% of the sample). On the basis of these features of the Food Frequency survey, the items from this section met all 3 of the entry criteria and were therefore subjected to factor analysis and included in this examination of acculturation.

B. HOUSEHOLD ITEMS: An individual's possessions and material goods serve as an objective indicator of the degree to which they adhered to traditional customs or the extent to which they have acculturated to a North American lifestyle. The SLHDP surveyed individuals as to whether they possessed specific traditional or modern household items and participants were asked to report the number of both working and broken items in their possession. Since data were collected on both traditional and North American possessions (i.e.: tikinagens versus televisions), and since this section was completed by 319 respondents (representing 78% of the total sample), the section entitled "Household Items" was subjected to factor analysis and included in the present analysis.

C. PROFICIENCY OF LANGUAGE DATA: The extent to which an individual had absorbed the North American culture was most evident in a semi-isolated community like Sandy Lake by observing how individuals communicate and which language was chosen for interacting with one another. Also, the degree to which an individual was proficient in English and Oji-Cree provided important information on the level of education and exposure to traditional teachings in the home, and to North American education through the public school system. The SLHDP measured the proficiency of both Oji-Cree and English in the section entitled, "Language" and collected information on the ability to read, write and speak both languages. This section met all three criteria for inclusion (this section was also completed in full by every respondent in this analysis) and therefore was submitted for factor analysis in order to contribute to an understanding of acculturation in the community.

Sections Excluded

A. 24-HOUR DIETARY RECALL: Some questions included in the SLHDP met the first criterion and would have contributed some ethnographic information to the study. However, this did not meet the second or third criterion (or both) and were therefore excluded. As detailed above, a 24-hour dietary recall questionnaire was administered to the participants of the SLHDP, and while the Food Frequency data were included in the present analysis, the 24-hour recall was excluded. The 24-hour recall was excluded for the following reasons: A) the data did not provide a general picture of food consumption but rather a single day's food intake. Therefore, without multiple 24-hour recall surveys on the same individual, these data would convey less information than the Food Frequency section. B) Certain foods were not eaten frequency but were considered an important aspect of the measures of acculturation measured in this study. For example, certain traditional bush foods are not eaten daily (or even weekly) and would not contribute to the relative consumption of traditional versus North American dietary risk factors. C) The 24-hour recall was not completed in full by the majority of the respondents, as it required a significant time commitment to complete. Also, these data had been converted by nutritional scientists into estimates of macro and micro nutrient intake rather than by food groupings (for example, total sodium and potassium intake rather than total potato chip intake). Data on micro and macro nutrients does not allow for a clear interpretation of the acculturative effects in the community since these nutrients are present in both traditional and white man's foods. Therefore, factor analysis would not have resulted in factors measuring level of Oji-Cree or North American dietary acculturation.

B. BODY IMAGE: Another area of the survey which was excluded although the first criterion was met was the section entitled, "perception of body image". It was believed that as a traditionally-minded community acculturates to Western society, greater rates of body dissatisfaction and eating disorders would be observed.[76] This was due to a discrepancy in the general body image standard between the traditional and the North American culture.[76] The results of previous analyses of SLHDP data on body image showed that body perception was closely linked to age (and likely degree of modernization). These inherent associations with acculturation could have contributed to the model of acculturation examined here. However, once submitted for factor analysis, there were no interpretable results (i.e.: no underlying factors in the data). The fact that there were no usable results from factor analysis led to exclusion from further analyses.

C. METHOD OF FOOD PREPARATION: A final group of items which was subjected to factor analysis after satisfying the first criterion for inclusion was those assessing the "method of food preparation". This portion of the survey measured the various means by which specific foods were cooked and eaten. For example, 13 different foods were listed and participants were asked to list the most common method used for food preparation. The options were, "smoked, boiled in water, baked, boiled in fat and fried". Based on what was known about the Oji-Cree culture, these variables were re-coded so as to create a continuum of traditional cooking methods. For example, from most traditional technique (smoked) to least traditional or most North American technique (fried). These data were subjected to factor analysis. The results were uninterpretable (did not show simple structure – see section on factor analysis). It was concluded that there were two major problems with these data. First, only those who prepared the food (most commonly the woman of the home) completed this section, therefore the sample size was markedly reduced. Second, factor analysis only yielded factors describing the various foods which clustered as a result of similar food preparation techniques, not a cluster of the methods themselves (i.e.: knowing that moosemeat, bannock and pork are all cooked similarly imparted no useful information on the extent to which the community has become acculturated). If, however, the factors had produced indices of food preparation methods (e.g.: traditional preparation techniques – smoked, boiled, etc – versus North American techniques – deep fried, fried, etc.) this would have been more valuable. Each individual could then have been given a score on "traditional cooking methods" and "North American cooking methods" and then these scores could have been included in a regression model to test for associations with various health outcomes. However, the absence of meaningful factors, together with the small sample size, led to the exclusion of these data from the analysis of acculturation in the community.

2.4.2 EXPLORATORY FACTOR ANALYSIS

Exploratory factor analysis (EFA) is different from confirmatory factor analysis (CFA) and also from principle components analysis (PCA). Results of PCA produce a linear combination of the *observed variables*. Factor analysis also produce linear combinations of observed variables, but the results are viewed as linear combinations of the *underlying factors*. However, these two data reduction techniques are commonly used together to identify underlying factors in data sets. In fact, when using SAS's PROC FACTOR, data are first subjected to PCA and then to EFA. PCA presents a number of components that account for most of the variance in a set of observed variables while factor analysis confirm this number (as determined from PCA) as well as determine the nature of the underlying factors.

Specific criteria are used in order to determine the number of underlying factors from the results of PCA. However, during PCA, the number of factors extracted equals the number of variables. In PCA, factors that are extracted have two characteristics. First, each factor is uncorrelated with previously extracted factors (orthogonal). Factor analysis extends PCA by rotating the factors to produce factors which represent both the number and nature of the underlying factors of a data set.[77] Second, each successive factor extracted will account for a maximum amount of variance that has not already been accounted for by other, previously extracted factors.[77] It is up to the researcher to determine the number of factors which contribute the majority of the cumulative proportion of the variance within the data (by convention, this number is usually set at 80% of the cumulative variance – see " The SAS Procedure – *the data step* " below).

This is where *exploratory* factor analysis (EFA) differs from *confirmatory* factor analysis (CFA). In EFA, the researcher uses conventional data step procedures during PCA to first determine the number of underlying factors in the data. Then, the nature of each factor is subjectively determined using factor analysis. The final lists of items that load on each factor are then used as a representation of the nature of each factor. In CFA, there is less investigation into potential factors since the focus is on confirming that the questionnaire which was designed to form clusters within the data did, in fact, cluster on each of the pre-set factors. While technically the same as EFA, CFA is generally used when surveys are specifically constructed with a number of underlying factors built in and the researchers wish to investigate the success of the survey's intent and design. By contrast, EFA is appropriate when a survey has obtained measures on a number of variables and one intends to determine if there are underlying factors that are responsible for covariation in the data.

In the SLHDP, there were no prearranged objectives to study acculturation or the extent to which individuals lead a North American or traditional existence, so all analyses used EFA. In other words, even though the hypothesis for the present analysis is that items will cluster along cultural lines (Traditional and North American) this is still considered EFA since the questionnaire was not designed with these in mind and all statistically significant factors found from EFA will be included in the analysis.

This study attempted to first identify if any underlying factors existed in the data, and if so, whether they could be interpreted as reflecting traditional and North American lifestyle questions. Next, EFA served as the data reduction technique, allowing for a determination of the number and nature of underlying constructs in the data set. It was hypothesized that in this study, at least two factors would be identified in each section: one defined by traditional responses and one by North American responses. For example, for the Food Frequency section, items would cluster which measured intake of traditional foods and white man's foods separately; within the Household Items questionnaire, variables which measured the possession of traditional goods would cluster separately from modern goods; and for the Language data, a division was hypothesized between the items that measured proficiency in Oji-Cree and English.

A) THE SAS PROCEDURE: PROC FACTOR

The same technical procedure was used for each of the three sections. The data steps and procedures used for each of the Food Frequency questionnaire data, the Household Items data and the language proficiency data are included in detail in the Appendix. SAS Version 6.12 was used for all of the analyses. This statistical software package was used in a VMS and Windows environment.

B) INTERPRETING THE FACTORS

To determine the accurate number of factors to retain once PCA has been completed, Hatcher suggests the use of three criteria: the scree test, the proportion of variance, and the interpretability test. Once the number of factors is chosen using these criteria, further assurance that the correct number was chosen is derived from the use of "simple structure". As detailed in the next section, this confirms that the number of factors was chosen correctly and allows for the next step in factor analysis: determining the nature of the underlying factors.

Hatcher's Criteria

a. The Scree Test: Specifying the SCREE option in the PROC FACTOR statement will print a graph which plots the factors along the y-axis and their eigenvalues along the x-axis. Hatcher suggests that the number of meaningful factors is represented by a break in the graph between eigenvalues of higher values and those of lower values. Those appearing after the break are assumed to be unimportant.

b. Proportion of Variance Accounted For: A factor is retained provided that a certain cumulative proportion of the variance is accounted for by that the combined variance of that factor and any previous factors extracted in the data set. Since the proportion of the total variance accounted for by each subsequently extracted factor decreases, the first few factors usually represent the majority of the cumulative variance. For the present analysis a cumulative variance of 70 – 80% was used as a potential cut point in determining the number of factors. This is a criterion that can only be used in FA (not PCA) since the total eigenvalues of all factors is equal to the sum of the communality estimates. For PCA, the total eigenvalues equals the total number of variables being analyzed since each contributes one unit of variance.

c. Interpretability Criteria: Determining the nature and number of factors to retain requires an understanding of the constructs under investigation. An analysis of the contribution of each of the items loading on a given factor should result in a sensible factor which is highly interpretable, or else, the number of factors specified in the NFACT option may not have been acceptable. A trial-and-error method of testing different NFACT values will ultimately lead to an appropriate number of factors retained.

Simple Structure

Once the number of factors is determined and the factors are rotated, an oblique solution is presented. Items which load on each factor are listed and the FLAG option will highlight any items which load > 0.40 on a given factor. This is considered important and will then be used to determine the nature of the underlying factor. However, before this is done, the factors must follow 'simple structure'. There are two conditions which need to be met in order to attain simple structure: a) Most of the variables which load on a given factor must have high loadings on one factor and near zero loadings on other factors. b) Correspondingly, each factor should be loaded heavily by some items and very minimally by others. Ultimately, the standard rule of simple structure is to produce an acceptable coefficient-α value which implies that the items are a reliable representation of the underlying factor. It is important to ensure that the items which load on a given factor share some conceptual meaning and that items on different factors seem to be measuring a different construct.

Other standards which are also used include the requirement that at least three items load on a given factor and also that if one item should load greater than 0.40 on more than one factor, then it is excluded from the interpretation of any factor's underlying nature (see section on coefficient-α reliability testing). Exceptions in which there are relatively strong loadings on more than one factor from a given item are acceptable provided that the coefficient-α determinations are still at an acceptable level. For example, if a factor should be well represented by only two items, and also satisfy scale reliability test requirements (coefficient-α > 0.60) then there is substantial evidence to use this factor in further statistical testing.

Oblique (VARIMAX) Rotation

It is then necessary to determine which items will be included when determining the nature of the factor. Recall that orthogonal solutions (those used in PCA) are linear transformations of the observed variables. The factors that result from the PCA procedure – also known as a VARIMAX rotation – are uncorrelated with each other. Oblique solutions (those used in FA) are also linear transformations; however, they are linear transformations of the underlying factor and not of the observed variables (as in PCA).

For the SLHDP data set, it is hypothesized that similar underlying factors will be seen throughout separate sections; 'Traditional' and 'North American' responses in Diet, Household Items and Language. Thus, uncorrelated factors (oblique solutions) are better suited for this analysis since it is likely that the factors are not completely independent of each other, but rather that a lot of overlap exists with regard to individual responses across the community.

Factor Matrices and their Interpretations

Two factor matrices will be used in the final determination of the number and nature of factors and which items, taken together, represent that factor. First, the 'pattern matrix' will be reviewed. The loadings found in a pattern matrix table are analogous to the standardized regression coefficients obtained in a multiple regression analysis. The loadings within this matrix reveal the unique contribution that each factor makes to the variance of a single variable. Second, reviewing the 'structure matrix' will reveal the correlation between a given factor and a single variable. In summary, the pattern matrix and structure matrix will both be used to determine the meaning of a given factor.

c) FACTOR SCORES

Factor scores for each individual will be estimated by creating linear composites of the observed variables within a given factor. These scores are optimally weighted estimates of an individual's actual score. For example, the measurement of an individual's 'traditional dietary intake' is only estimated from responses on the questionnaire. There may be other variables which would contribute to this score which were not measured, however this analysis is limited to the questionnaire items only and is therefore only an 'estimate' of individuals' absolute intake. An individual's actual score is the absolute quantified traditional intake measured without any limitations.

The factor scores are computed by adding together the individual's (optimally weighted) score on observed variables which were unique as well as those which were common among factors:

Estimated Factor Scores:
PROC FACTOR automatically computes estimated factor scores by requesting the creation of a new data set within the SAS program using the OUT statement.

A subject's estimated factor score uses the following equation:

$F'_1 = b_{11}V_1 + b_{12}V_{12} + b_{13}V_{13} + \ldots b_{1p}V_p$

where
F'_1 = the estimated factor score for factor 1
B_{11} = the scoring coefficient for survey question 1, used in creating estimated factor score 1
V_1 = the subject's score on survey question 1
B_{12} = the scoring coefficient for survey question 2, used in creating estimated factor score 1
B_{1p} = the scoring coefficient for survey question p (the last question), used in creating estimated factor score 1
V_p = the subject's score on survey question p

D) RELIABILITY TESTING USING COEFFICIENT-α

Coefficient-α was used to determine the internal consistency (reliability) of each of the subscales of acculturation. Coefficient-α measures the internal consistency of the data by measuring the correlation between actual observed scores from each item which make up a 'factor' and individuals' scores obtained on that factor.[77] The squared value of this correlation is the percent of variance in the observed variable that is accounted for by the score on the underlying factor. This is the reliability coefficient.

Coefficient-α is high when the items which comprise a scale are highly correlated with one another. The items subjected to factor analyses have been re-coded into dichotomous variables, into nominal data, or Likert-type scales. However, the actual data used in the coefficient-α determination were factor scores and were therefore continuous. Factor score distributions including descriptive data (means, standard deviations and score ranges) were also determined in addition to α scores. The datasteps used for the coefficient-α determination of each subscale are included in the Appendix.

2.5 Scale Measures

2.5.1 FOOD FREQUENCY SUBSCALE

a) General Analysis

The Food Frequency questionnaire is reproduced in Table 2.5.1. Data from the Food Frequency survey was used for the purpose of developing a scale that represented the types of food consumed by individual members of the community. This Food Frequency questionnaire was not externally validated using multiple 24-hour recall surveys or by other testing means, however, the extensive ethnographic research which predated the prevalence survey provided for a culturally appropriate method of dietary data collection.[72, 79] For detailed SAS coding, see Appendix.

Volunteers were asked to report the frequency with which 34 separate food items were consumed over a previous three-month period. Data were collected from 406 SLDHP respondents which satisfied the criterion for sample size which suggests the larger of 100 subjects or 5 times the number of variables.[77] Information was collected using an ordinal scale which measured frequency of consumption from 'none or rare' to 'more than once per day'. Responses for this section of the questionnaire were re-coded to allow for a gradient of intake from 'never' to 'more than once per day' (see Table 2.5.2). The hypothesis was that factor analysis of the Food Frequency items would yield two correlated clusters of responses (2 factors): one defined by store-bought food such as canned fruit, chips, hamburger, and candy, and a second diet, defined by Indian medicine and teas, bannock, wild meats, fish and wild berries. A high score on either factor indicated a high degree of consumption of the food which comprised that particular factor and correspondingly, a low degree of consumption of the foods which comprised the second factor.

Table 2.5.1 Food Frequency Questionnaire from the SLHDP

Instructions for interviewers: Ask the volunteer to think back over the last 3 months to remember usual food that have been eaten and how often.

Food	Food Frequency				
Rare/Never	1-3x/month	1-2x/week	3-6x/week	1/day	>1/day
Fish					
Moosemeat					
Beef / steak / hamburger					
Pork chops / bacon					
Duck or goose					
Rabbit					
Klik or spork					
Eggs					
Butter / lard / margarine					
Cold cereals					
Hot cereals					
Beans					
White bread					
Whole wheat bread					
Bannock					
Macaroni or other pasta					
Indian medicine or Indian tea					
Home made soup					
Chips or french fries					
Other potatoes					
Peas					
Corn					
Carrots					
Other vegetables					
Wild berries					
Fresh fruit from store					
Canned fruit					
Milk					
Carnation mild					
Pop/soda					
Tea/coffee					
Cookies / cake / pastries					
Chocolate / candy / bars					

Logistic regression analyses were performed on the factor scores to determine possible associations between disease outcomes (diabetes, IGT or obesity) and scores from the factors as produced by the factor analyses on these data (See Appendix, for full logistic regression SAS code).

Table 2.5.2 Re-Coding of Food Frequencies for Factor Analysis

Code	Frequency
6	> once per day
5	once per day
4	3 – 6 times per day
3	1 – 2 times per day
2	1 – 3 times per day
1	rare or never

b) Blood Lipid Analysis

High-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol and triglyceride levels were all measured from fasting blood samples drawn at the baseline survey. In order to determine whether an association existed between Food Frequency factor scores and blood lipid measures, a secondary statistical analysis was done on these continuous data. Linear regression was performed. This analysis was performed with and without adjustments for age and sex. The blood lipid scores and food subscale scores represented the outcome variables and exposure variables, respectively. The blood lipid measures could have been converted to categorical data using standard cut-offs from large population-based studies which measured the occurrence of dysplipedemia in the Canadian population.[82] However, it was decided that this community was sufficiently different from the general Canadian population and that standard cut-offs would bias the results toward the null hypothesis; no association would be detected using thresholds from a separate population in which dyslipidemia is more prevalent.

2.5.2 Household Items Subscale

a) General Analysis

Table 2.5.3 is a reproduction of the SLHDP questionnaire which collected information on the possession of certain household items. The household questionnaire was designed to assess socio-economic status in the community in the absence of standard SES indicators (e.g. salary scale, home ownership, etc.). The possession of contemporary technological devices (televisions, telephones, satellites) modes of transportation (automobiles, trucks, ATVs, skidoos) and traditional items (tikinagen, tahkobisin) were all measured using this survey. For detailed SAS coding, see Appendix.

Table 2.5.3 Household Items Questionnaire from the SLHDP

Instructions: please tell me whether you, or anyone in your house owns the following, and how many they own.

Sewing machine
 # working:_____ #broken_____

TV
 # working:_____ #broken_____

Telephone
 # working:_____ #broken_____

VCR
 # working:_____ #broken_____

Big Freezer (not in refrigerator)
 # working:_____ #broken_____

Air Conditioner
 # working:_____ #broken_____

Boat
 # working:_____ #broken_____

Motor for Boat
 # working:_____ #broken_____

Cassette Player / Stereo
 # working:_____ #broken_____

ATV (all-terrain-vehicle)
 # working:_____ #broken___

Microwave Oven
 # working:_____ #broken_____

Satellite Dish
 # working:_____ #broken_____

Washing Machine
 # working:_____ #broken_____

Dryer
 # working:_____ #broken_____

Tikinagen
 # working:_____ #broken_____

Tah-ko-bi-sin
 # working:_____ #broken_____

Car or Truck
 # working:_____ #broken_____

SkiDoo
 # working:_____ #broken_____

Outdoor Washroom
 # working:_____ #broken_____

In total, the survey measured the possession of 19 different household items. A total of 319 participants were included for this subscale development. This satisfied the criterion for sample size which suggests the larger of 100 subjects or 5 times the number of variables.[77] Information on both broken and working items was collected, as well as nominal data as to the number of these items in the home, however, for the purpose of these analyses, these data were re-coded into dichotomous variables as described in Table 2.5.4. Therefore, if a respondent had 2 broken automobiles and 1 working automobile, the response was coded as a '1'.

Table 2.5.4 Re-Coding of Food Frequencies for Factor Analysis
If (# of working items) + (# of broken items) is ≥ 1 then Item = 1.
If (# of working items) + (# of broken items) is < 1 then Item = 0.

The data collected from the Household Item questionnaire was subjected to factor analysis to determine the number and nature of underlying factors in these data. Hatcher's criteria was applied to determine the number of meaningful factors to retain. It was hypothesized that factor analysis would produce at least two factors which would be represented by Traditional or North American items, respectively. A high score on a factor indicated degree of modernity in one's lifestyle and activities.

Logistic regression analyses were performed on the factor scores to determine whether an association existed between disease outcomes (diabetes, IGT or obesity) and factor scores yielded from the Household Items data (see Appendix for full SAS code).

b) Physical Fitness Analysis

The original baseline survey also measured the level of physical fitness based on maximal oxygen uptake ($V0_{2max}$). Although this is the most widely accepted measure of fitness level,[70, 80] only those without medical conditions that contraindicated vigorous exercise were included.[70] For those who participated, the associations between factor scores from the household subscale and $V0_{2max}$ scores were investigated in order to determine whether possession of household items in the factors was associated with fitness level. Since both $V0_{2max}$ scores and factor scores were continuous variables, linear regression analyses were performed with adjustment for age and sex.

2.5.3 Language Proficiency Subscale

Table 2.4.4 is a reproduction of the Language data collection instrument from the SLHDP. Information was gathered on the level of proficiency of both languages through a section of the questionnaire which helped determine proficiency of common languages in the community. Nine questions were asked in the section entitled 'Languages' on the SLHDP. Scripted questions gathered information on ability in writing, reading and speaking English, Oji-Cree and any other language spoken (e.g. other native dialects, etc). The data collected from these last questions (proficiency of any other languages) were found to be incomplete; few respondents answered.

Therefore, the items which examined 'other languages' were not examined in this analysis. The questions regarding the proficiency of Oji-Cree and English were completed by 411 (or 100%) of the sample for this analysis. For the purposes of the present analysis, only the first 6 questions (reading, writing and speaking English and Oji-Cree) were subjected to factor analysis and regression modeling. For a full description of the SAS code used for these analyses, see Appendix.

Table 2.5.5 Language Questionnaire from the SLHDP

English:	Oji-Cree:
Do you read English? (Yes/No)	Do you read Oji-Cree (Yes/No)
Do you speak English? (Yes/No)	Do you speak Oji-Cree (Yes/No)
Do you write English? (Yes/No)	Do you write Oji-Cree (Yes/No)

Other:

Do you read another language? (Yes/No) which? ______________

Do you speak another language? (Yes/No) which? ______________

Do you write another language? (Yes/No) which? ______________

It was hypothesized that factor analysis would produce two distinct factors: one that measured proficiency of English and another that measured proficiency of Oji-Cree. It as believed that these would be separated out by factor analysis as a function of their inter-correlation among responses and furthermore, that each of these scales will produce a very high degree of reliability as shown by coefficient-α. This was because of the inherent bi-modal nature of the data; elders speak Oji-Cree and little English and the younger generations speak English and little Oji-Cree. Although only six items entered into factor analysis represented the minimum number of items necessary to produce two simple factors from factor analysis, it was left to coefficient-α determination to determine the reliability of the scale with so few items. Provided α met a conventional level (according to Hatcher, is 0.60)[77], then the factor was considered a 'reliable' scale.

2.6 Multivariate Analyses of Factor Scores

2.6.1 The Distribution of the Factor Scores Among "Normals"

Normals were defined as having no diabetes, IGT or obesity. Once factor scores were produced for each subscale, proc UNIVARIATE in SAS produced a distribution by quartile of the factor scores among normals. This technique was used to derive the cut-points for a categorical quartile analysis. Datasteps and quartile cut-points used in the analyses are included in the Appendix.

2.6.2 Exposure and Outcomes Variables Used for Regression Analyses

Table 2.6.1 describes the outcome variables and criteria used to diagnose each of the disease states used in regression analyses in this study. These were described in Chapter 1 and reflect the WHO criteria[34] used for diagnoses of diabetes and IGT at the time of the SLHDP. The prevalence and incidence of these conditions, as found by the SLHDP, are also reported in the results section of Chapter 3.

Table 2.6.2 describes the exposure variable and confounding variables used in regression analyses in this study. Factor scores were used as the primary predictor variables for all regression analysis and all results will be presented before and after adjustment for the potential confounding variables listed.

Table 2.6.1 Description of the Outcome Variables used in the Regression Analyses

The following is a list of the outcome variables used in the regression analyses for the present analysis:

i. "Abnormal Metobolism": this was a general term used to describe an individual with either newly diagnosed diabetes, or IGT.

ii. "Newly Diagnosed Diabetes": a category of individuals who were diagnosed with diabetes at the time of the SLHDP. This required a single fasting glucose concentration ≥ 7.8 mmol/L or ≥ 11.1 mmol/L after a 75-g oral glucose tolerance test (OGTT).

iii. "IGT": community members that were diagnosed with impaired glucose tolerance according to the criteria. This required a glucose concentration ≥ 7.8 mmol/L but less than 11.1 mmol/L after a 75-g OGTT.

iv. "Obese": This category describes any individuals with a body mass index greater than or equal to 24 kg/m^2.

v. "Normal": Individuals in this group were not obese and did not have diabetes or IGT. This 'control' group was used as comparison group for each of the above 'case' groups.

Table 2.6.2 Description of the Exposure Variables used in the Regression Analyses

The following is a list of the predictor variables used in the regression analyses for the present analysis:

i. Factor scores: for each logistic regression model, a different subscale factor score was used as the exposure variable. Both categorical and continuous data were used for logistic regression modeling.

ii. Age: age at the baseline survey was used a confounding variable and was included in the model in continuous data form.

iii. Sex: gender was considered to be a confounding variable and was also included in the adjusted model as a dichotomous variable.

2.6.2 Logistic Regression

Logistic regression analyses were used in order to assess the association between individual scores on each of the subscales of acculturation and disease outcomes such as diabetes, IGT and obesity. As described above, diabetes and IGT were diagnosed at the time of the baseline survey (all previously diagnosed participants with diabetes were excluded from this analysis).

For all logistic regression analyses in this study, the main outcome variables were the disease variables (diabetes, IGT or obesity) and the main exposure variables were the factor scores for each given subscale from either the Food Frequency data, Household Items data or Language data. Adjustments for both age and sex were included in the regression models and both adjusted and unadjusted results will be presented. For all of these analyses, these outcome variables were treated as dichotomous variables. However, the factor scores were first analyzed as continuous data and then within a categorical quartile analysis.* The logistic regression analyses estimated the relative risk through an estimate of the odds ratio associated with increasing quartiles for factor scores. Quartiles for factor scores were assessed based on the distribution of factor scores among normal, non-obese, healthy respondents. For full SAS code and listing of the cut-points within the factor scores as produced from PROC UNIVARIATE, see Appendix, section 6.1.3 (for the Food Frequency analysis); Appendix, section 6.2.3 (for the Household Items analysis); and Appendix, section 6.3.3 (for the Language data analysis).

The regression models used in this analysis and which are presented in the Appendix are in the basic form:

Continuous Analysis:

- *Unadjusted model:* Outcome variable = factor score
- *Adjusted model:* Outcome variable = factor score age sex

Categorical Analysis:

- *Unadjusted model :* Outcome Variable = Quartile2 Quartile3 Quartile4
- *Adjusted model:* Outcome Variable = Quartile2 Quartile3 Quartile4 age sex

Linear regression analyses were also performed in this study to determine the association between the measures of Diet, Household Items and Language and blood lipid markers as well as the level of physical fitness. As described above, descriptive results are presented for the data distribution as well as both adjusted and unadjusted models for the continuous data. The models used for the regression analyses are in the basic form as described above.

*Note: For the Food Frequency and Household Item data, a quartile analysis was performed for the categorical data but a bivariate analysis was used for the Language factor scores due to a bimodal distribution in the data. This is described in more detail in Chapter 3, section 3.4.

Chapter 3

Results

3.0 Introduction

This chapter describes the study population and sample in detail. In addition, the study from which the data for the present analysis was taken is described – the Sandy Lake Health and Diabetes Project (SLHDP). Results of factor analyses of the Food Frequency, Household Items and Language data are also presented, as are subscales that were developed in order to describe acculturation in the community. Results from regression analyses of factor scores will demonstrate the association between scores on these subscales and diabetes, IGT and obesity. Results from secondary analyses will highlight the association between scores on the Food Frequency subscale and measures of blood lipids, and between scores on the Household Items subscale and measures of physical activity. Together, these results will document the relationship between the level of acculturation and disease risk in Sandy Lake.

3.1 Demographic Analysis

3.1.2 ACCULTURATION: THE STUDY SAMPLE

Characteristics and anthropometry of participants in the present analysis are presented in Table 3.1.1. A total of 411 of 416 eligible participants were included in the analysis (99% of the eligible SLHDP participants from the original baseline survey). Inclusion in the present study was contingent upon the following criteria: the participant was included in the SLHDP and was 20 years of age or older at the time of the baseline survey. Participants were excluded if they had had a previous diagnosis of diabetes due to the potential for reporting biases. It was believed that with a pre-existing knowledge of one's own illness coupled with even a superficial understanding of risk factors of disease it was reasonable to assume that individuals may have had a tendency to over-report positive lifestyle behaviors such as those with known biological associations with health, and similarly, under-report the frequency of negative lifestyle behaviors. This bias was of particular concern for questions with direct effects on health such as dietary habits and physical activities.

Table 3.1.1 Description† of the Adult Study Sample by Sex

	Male (n = 185)	**Female** (n = 226)
Age (years)	35.9 ± 14.0	35.9 ± 13.9
Height (cm)	174.8 ± 6.2	161.7 ± 5.4
Weight (kg)	82.7 ± 15.2	75.8 ± 14.6
Body mass index (kg/m^2)	27.0 ± 4.5	29.0 ± 5.5
Education, % none	7.6	9.3
Education, % secondary or more	26.5	19.0

Sample size (n) was 411 adults out of a total sample size (N) of 723
† Values are means ± standard deviations

There were more adult women included (54.9%) which reflects the predominance of adult women which participated in the original baseline survey (56.6% of participants in the SLHDP were women). However, at the time of the baseline survey, women comprised 53% of the total Sandy Lake population. Therefore, although due to work-related time constraints, men within the 40 – 49 year age range participated at a lower rate, yet there was still a representative sample from the population.[4]

Anthropometric measurements in the present study were similar to those seen in the SLHDP. Of note was the predominance of obesity in the study sample. Both men and women exhibited a mean body mass index (BMI) which exceeded the lower threshold for clinical obesity (BMI = 24 kg/m^2). The average BMI was higher among women than among men (29.0 ± 5.5 for women, 27.0 ± 4.5 for men). On average, there were more females than males reporting 'no education' and less reporting 'advanced education' which reflects a cultural norm in Sandy Lake as well as other native communities in North America.

Tables 3.1.2 and 3.1.3 display the anthropometric characteristics of the study sample by disease status. Information is provided for six separate disease categories (over the two tables) and includes persons with 'normal health' (without diabetes, IGT or obesity), abnormal glucose metabolism including both newly diagnosed diabetes and IGT, obesity, 'previously diagnosed diabetes' (these participants were excluded from this analysis but are described in Table 3.1.3), 'newly diagnosed diabetes', and 'IGT'. On average, the healthy group was the youngest, had the lowest BMI, the lowest average weight and the most education. Men and women with newly diagnosed diabetes had the highest average BMI. Also, males with IGT reported the least education while males in the normal group reported the most secondary education.

Table 3.1.2 Description† of the Adult Study Population by Sex and by Disease Status

	Normal‡		**New diabetes / IGT**		**Obesity***	
	Male (n = 47)	Female (n = 39)	Male (n = 36)	Female (n = 71)	Male (n = 135)	Female (n = 182)
Age (years)	31.4 ± 12.8	30.0 ± 10.9	44.4 ± 16.7	43.6 ± 15.2	37.0 ± 13.9	37.2 ± 13.9
Height (cm)	173.5 ± 7.0	161.9 ± 5.6	172.6 ± 5.9	160.5 ± 5.4	175.4 ± 5.9	161.6 ± 5.2
Weight (kg)	63.6 ± 6.7	56.2 ± 7.5	88.6 ± 13.8	80.7 ± 13.1	89.7 ± 10.7	80.5 ± 11.8
Body mass index (kg/m^2)	21.1 ± 1.9	21.4 ± 2.2	29.6 ± 3.6	31.3 ± 4.8	29.1 ± 3.0	31.0 ± 4.3
Education, % none	4.3	2.6	16.7	19.7	8.1	11.0
Education, % secondary or more	29.8	20.5	22.2	15.5	25.9	19.2

† Values are means ± standard deviations.
‡ Normal = no diabetes AND no obesity
*Obesity = body mass index > 24 kg / m^2

There were 69 participants in the SLHDP that were previously diagnosed with diabetes and 44 participants that were newly diagnosed at the time of the baseline survey. While these two groups had a similar age distribution by gender, there was a marked difference between these two groups on body weight, BMI and education. On average, those previously diagnosed with diabetes were lighter, and both males and females reported more education than either the newly diagnosed group or the IGT group.

Table 3.1.3 Description† of the Adult Study Population by Sex and by Disease Status

	Previous Diabetes		New diabetes		IGT	
	Male (n = 24)	Female (n = 45)	Male (n = 22)	Female (n = 22)	Male (n = 14)	Female (n = 49)
Age (years)	46.8 ± 10.7	45.2 ± 14.6	43.4 ± 16.4	46.0 ± 15.6	46.0 ± 17.8	42.6 ± 15.1
Height (cm)	173.7 ± 5.2	160.6 ± 6.1	175.1 ± 4.9	159.7 ± 4.5	168.7 ± 5.3	160.8 ± 5.8
Weight (kg)	85.0 ± 15.8	77.5 ± 15.1	93.8 ± 11.9	81.5 ± 9.3	80.5 ± 13.0	80.4 ± 14.5
Body mass index (kg/m²)	28.2 ± 5.0	29.9 ± 5.1	30.5 ± 3.3	32.0 ± 3.8	28.2 ± 3.8	31.0 ± 5.3
Education, % none	20.8	22.2	13.6	31.8	21.4	14.3
Education, % secondary or more	16.7	20.0	31.8	13.6	7.1	16.3

† Values are means ± standard deviations.

3.2 Subscale 1: Food Frequency

3.2.1 Results of Factor Analysis

The minimum number of subjects needed for factor analysis was the larger of 100 subjects, or 5 times the number of variables being analyzed.[77] The total number of available participants was 411, however, factor analysis excluded any individuals for whom the 34-item Food Frequency questionnaire was not completed in full. Therefore, 406 participants were included in the factor analysis (FA) with data on all 34 separate food items.

Factor analysis identified the number of underlying factors and which of the 34 food items defined them. (For complete factor analysis output from SAS, refer to the Appendix). Hatcher's criteria[77] for determining the number of "meaningful" factors to retain suggested that three factors were important, so only these factors were retained for rotation. These three factors are described in more detail in this section, however, for a step-by-step description of how the number and nature of factors from these data were determined, refer to Appendix.

Questionnaire items and corresponding factor loadings are presented in Table 3.2.1. Factors were named using a combination of ethnographic knowledge and statistical principles. Factor names were based on the items which loaded greater than 0.40 on either the factor pattern or the factor structure, or both. This was the method which was used for each of the subscales described in this chapter. Accordingly, nine items were found to load on the first factor. This first factor was labeled "Traditional Food" based on the nature of the items which, together, comprised the factor (for e.g. bannock, berries, duck, Indian medicine, moose, etc.). The second factor had six items which were found to load ≥ 0.40. This factor was made up of vegetables and other healthy food, and was therefore called "Healthy Food". Seven items were found to load on the third factor, and the items which loaded on this factor (chips, chocolate, cookies, and pop) were best classified as "Junk Food".

Table 3.2.1 Factor Analysis Results

Questionnaire Items and Corresponding Factor Loadings from the Rotated Factor Pattern Matrix and Factor Structure Matrix, Decimals Omitted

	Factor Pattern				**Factor Structure**		
Factor 1:	**1**	**2**	**3**		**1**	**2**	**3**
Bannock	*45	-12	18		*45	5	23
Berries	*40	-11	-7		36	-1	-2
Duck	*51	0	-1		*51	14	8
Fish	*69	-1	-8		*67	16	4
Hot chocolate	*41	3	-2		*41	13	6
Indian medicine / teas	*43	0	-6		*42	11	1
Moose	*59	6	-5		*59	21	7
Rabbit	*42	15	-7		*45	25	4
Soup	36	6	16		*41	20	24
Factor 2:							
Carrots	10	*58	-10		24	*58	5
Corn	7	*57	5		23	*60	19
Milk	-3	*42	5		9	*43	14
Other vegetables	4	*69	-8		21	*68	9
Peas	0	*54	4		15	*55	17
Whole wheat bread	-5	*40	1		6	38	9
Factor 3:							
Beef	-15	9	*40		-5	15	*40
Chips	-2	2	*50		8	14	*51
Chocolate	-1	5	39		8	15	*41
Cookies	17	7	38		26	21	*42
Klik / spork	7	-5	*42		13	7	*42
Pop	-14	6	*40		-5	11	39
Pork	10	13	35		20	24	*40

The results of the factor analysis on the Food Frequency data was consistent with the community's perception of food groupings. "Healthy", "Traditional", and "Junk" foods, each comprised of separate items which tended to cluster with regard to reported dietary intake. A high score on any given scale by one individual indicated more frequent consumption of these foods, and less of others, than an individual with a lower score on the same scale.

3.2.2 Results of Scale Reliability and Coefficient- α

Scale reliability estimates were assessed by calculating coefficient-α for each subscale within the Food Frequency data.[83] Each scale had an acceptable Cronbach's-α based on the cutoff suggested by Hatcher[77] for studies of this kind; Traditional Foods α = 0.68, Healthy Food α = 0.71, Junk Food α = 0.62. These scores indicated a good level of consistency – or reliability – in the Food Frequency data.

3.2.3 Factor Score Distributions

Table 3.2.2 presents descriptive data for the three Food Frequency subscales. Factor analysis produced individual factor scores for each participant on each factor and these scores were distributed with a mean value equal to zero which places half of the scores on the positive side of zero and the other half on the negative side of zero. Therefore, negative scores on a given index were interpreted as a lower than average score rather than a negative intake of food. Median values are also presented.

Table 3.2.2 Descriptive Statistics for the Food Frequency Data

	n	Median	Skewness	Kurtosis	Q1 – Q3	min/max
Traditional Food	406	0.9	0.8	1.1	1.1	- 1.6 / 2.6
Healthy Food	406	0.3	0.2	0.1	1.2	- 1.9 / 3.0
Junk Food	406	0.6	0.4	0.7	1.1	- 2.6 / 3.1

Units: frequency of consumption (f.o.c.)

The Food Frequency data were re-coded to allow for a gradient of response. The questionnaire item had 5 responses which corresponded to the frequency of consumption and these items were coded in ascending order according to frequency of consumption (once per month, once per week, twice per week, once per day). Therefore, the unit of measure in the descriptive data (Table 3.2.2) was frequency of consumption per month, or 'f.o.c'. (see Appendix, section 6.1.1 for the SAS variable re-coding data steps).

For each of the subscales, the distribution of factor scores was positively skewed. The traditional foods scores were most skewed while the healthy foods were the least skewed data. The kurtosis of the distributions also indicated that the healthy foods scores – while the least positively skewed – also had the flattest distribution in comparison to the other two scales.

Table 3.2.3 Frequency Distribution: Food Frequency Data

Foods	Normal (n = 84)	Diabetes/ IGT (n = 107)	Diabetes (n = 45)	IGT (n = 62)	Obesity (n = 313)
Traditional					
Q1	21	18	7	11	73
Q2	20	28	10	18	102
Q3	21	24	10	14	62
Q4	22	37	18	19	76
Healthy					
Q1	21	29	15	14	78
Q2	22	28	14	14	66
Q3	20	30	13	17	96
Q4	21	20	3	17	73
Junk					
Q1	21	43	19	24	109
Q2	22	18	5	13	58
Q3	19	24	12	12	75
Q4	22	22	9	13	71

Table 3.2.3 presents the frequency distribution of factor scores by subscale before adjustment for age and sex. Most apparent from these data was the decreasing number of participants with diabetes in the Healthy subscale with increasing quartile. This was translated into a statistically significant odds ratio of 0.2 in the 4th quartile as seen in Table 3.2.6. Once adjusted for age and sex, this value increased to 0.4 and was no longer significant. The values in Table 3.2.3 were also used for sample size and power calculations. Further discussion can be found in Chapter 4. For power/sample size calculation tables, refer to Appendix, section 6.7.

Section 6.4.2 in the Appendix describes the full univariate analysis and the analysis that assessed the normality of factor scores for the three Food Frequency subscales. These distributions were tested for normality using the Shapiro-Wilk test within the PROC UNIVARIATE function in SAS. The test for normality was performed to determine whether the factor scores were a random sample from a normal distribution. Since an assumption of normality was used for all regression analyses, the Shapiro-Wilk test, and a stem-and-leaf plot were used to test the null hypothesis. The W statistic for the healthy food and junk food scores were both 0.98 with acceptable p-values ($p > 0.05$).

Table 3.2.4 Food Frequency: Analysis of Average Age and BMI by Factor Quartile

Means (standard deviations) for age and BMI among 183 men and 221 women

		n	Age (years)	BMI(kg / m²)
All participants		404	36.1 (13.9)	28.1 (5.1)
Sex	Males	183	36.0 (14.2)	27.0 (4.4)
	Females	221	36.1 (13.7)	29.0 (5.4)
MALES:				
Score on Traditional Food Scale:				
	1st Quartile	44	27.8 (6.0)	26.5 (4.0)
	2nd Quartile	57	35.3 (11.3)	27.4 (4.3)
	3rd Quartile	38	39.3 (16.3)	27.3 (4.9)
	4th Quartile	44	42.4 (17.4)	26.6 (4.6)
Score on Healthy Food Scale:				
	1st Quartile	50	36.1 (16.0)	26.9 (4.2)
	2nd Quartile	45	36.2 (16.1)	26.5 (4.5)
	3rd Quartile	51	37.2 (13.7)	26.9 (4.7)
	4th Quartile	37	34.0 (9.6)	27.7 (4.4)
Score on Junk Food Scale:				
	1st Quartile	57	37.6 (15.3)	27.4 (4.1)
	2nd Quartile	38	37.5 (16.1)	27.5 (5.1)
	3rd Quartile	43	35.8 (13.9)	27.3 (4.2)
	4th Quartile	45	32.9 (1.0)	25.7 (4.4)
FEMALES:				
Score on Traditional Food Scale:				
	1st Quartile	50	30.2 (7.4)	29.2 (4.7)
	2nd Quartile	69	33.4 (11.2)	29.3 (5.1)
	3rd Quartile	46	39.0 (12.9)	29.4 (6.8)
	4th Quartile	56	42.3 (17.8)	28.3 (5.1)
Score on Healthy Food Scale:				
	1st Quartile	52	42.1 (17.6)	28.8 (5.1)
	2nd Quartile	44	36.1 (14.9)	28.5 (4.4)
	3rd Quartile	67	35.1 (11.2)	30.3 (5.9)
	4th Quartile	58	31.8 (8.8)	28.2 (5.7)
Score on Junk Food Scale:				
	1st Quartile	75	42.4 (15.7)	29.7 (5.0)
	2nd Quartile	43	32.1 (10.6)	28.3 (5.4)
	3rd Quartile	53	34.5 (12.7)	29.1 (6.3)
	4th Quartile	50	31.7 (10.0)	28.7 (4.9)

This supported the null hypothesis: the factor scores were considered a random sample from a normal distribution. However, the distribution of scores from the Traditional items scale had a W statistic of 0.96, with a p-value < 0.01. This indicated that these scores were not representative of a normal distribution. A log transformation was performed on the traditional items scores and subsequently a satisfactory W statistic was obtained ($W = 0.97$, $p > 0.05$). Regression analyses were then performed with the transformed data. For detailed output including histogram, boxplot and normal probability plots, see Appendix.

Further descriptive data in Table 3.2.4 display the age distribution by factor quartiles stratified by sex. As the quartile for traditional food consumption increased, the average age of men and women increased. This was consistent with what was known about the culture. However, there was no monotonic association between traditional food intake and average BMI. For both sexes, people consuming the highest amount of healthy foods had the lowest average age, but there was no consistent pattern for BMI. Junk food was consumed in greatest amounts by the youngest group for both males and females. This general trend (i.e. younger respondents reporting higher junk food and elders reporting higher traditional food scores) was consistent with the information gathered from the ethnographic research which preceded the SLHDP and with what is known about the culture of Native communities in North American in general.[72]

3.2.4 Association between Food and Health Status

Continuous Data Summary

Table 3.2.5 displays the results of the multivariate logistic regression which used individual factor scores as the continuous exposure variables. Disease variables, 'diabetes or IGT', 'diabetes', 'IGT' and 'obesity', made up the outcome variables. Each of these outcomes was compared to a 'normal' group in which respondents were a) not obese and b) without diabetes or IGT. Odds ratios (OR) were calculated and represented the increased risk among the cases (diseased) relative to the control (healthy) group. In general, an OR of 1.0 represented no association between disease risk and the factors under study. A value above 1.0 indicated increased risk, while an OR under 1.0 indicated reduced risk associated with increased exposure to the given predictor variable. OR's were based on the Q1 – Q3 values as seen in Table 3.2.2: an increased risk of disease by an OR = 2.0 represents a doubling of risk associated with having an increase in factor score equal to the Q1 – Q3 range. In Table 3.2.5, adjustments were made for both age and sex. Odds ratios that were significant are bolded in the table, however, few values were statistically significant at $p \leq 0.05$. Thus, many of the measured effects may be attributed to chance variation since the majority of confidence intervals for the odds ratios did not exclude the value of 1.00.

Table 3.2.5 Food Frequency: Continuous Data

Unadjusted and adjusted† odds ratios (95% confidence intervals) displaying the relationship between food consumption patterns with health outcomes

Subscale	**Diabetes / IGT** (n = 92 / 64)**	**Diabetes** (n = 38 / 64)	**IGT** (n = 54 / 64)	**Obesity*** (n = 249 / 64)
Traditional				
Unadjusted	1.2 (0.9 – 1.7)	1.2 (0.8 – 1.8)	1.1 (0.8 – 1.6)	0.9 (0.7 – 1.2)
Adjusted	0.9 (0.7 – 1.3)	0.9 (0.6 – 1.4)	1.0 (0.6 – 1.5)	**0.7 (0.6 – 1.0)²**
Healthy Food				
Unadjusted	0.9 (0.6 – 1.2)	**0.6 (0.4 – 0.9)¹**	1.1 (0.8 – 1.6)	1.0 (0.8 – 1.3)
Adjusted	1.0 (0.7 – 1.5)	0.7 (0.4 – 1.2)	1.4 (0.9 – 2.2)	1.0 (0.8 – 1.4)
Junk Food				
Unadjusted	**0.7 (0.4 – 1.0)²**	0.7 (0.5 – 1.2)	**0.7 (0.5 – 1.0)²**	0.8 (0.6 – 1.1)
Adjusted	1.0 (0.7 – 1.5)	1.0 (0.6 – 1.6)	1.0 (0.6 – 1.6)	0.9 (0.7 – 1.2)

† adjusted for age (years) and sex.
* obesity is defined as a body mass index of ≥ 24 kg m^{-2}
** n = (sample size of cases / controls)
¹ significant at $p < 0.001$
² significant at $p = 0.05$

Before adjustment for age and sex, scores from the traditional food scale demonstrated OR's above or close to 1.0 for all comparisons. After adjustment, all OR's were ≤ 1.0 and the comparison of normal versus obese people suggested that obese people consumed traditional food less frequently than people of normal weight. Patterns of association with healthy food consumption were inconsistent both before and after adjustment, though before adjustment, the OR was < 1.0 for those with diabetes compared with normal people. Junk food displayed unexpected negative associations with each disease before adjustment but all were very close to 1.0 after adjustment.

Categorical Data Summary

In order to identify departures from linearity in the associations between nutritional factor scores and disease, the factor scores were divided into quartiles based on the univariate distribution of the factor scores among healthy, non-diseased, non-obese individuals and then used in a logistic regression (for a description of the quartile values, see Appendix E).

Overall, there was little evidence of monotonic patterns of risk across the quartiles for all three subscales (Table 3.2.6). This suggested that the data were not linear. Within the traditional food subscale, increasing levels of consumption resulted in non-linear patterns from the first to the 4th quartile for each disease state. Once adjusted for age and sex, within the higher quartiles (Q3, Q4) for traditional foods, the OR's were all < 1.0. This indicated some evidence of protection from consumption of the traditional food items which constituted this subscale, however, no p-values were statistically significant at $p<0.05$. Though not attaining a conventional level of statistical significance, the most notable results from this subscale were from the obese group which demonstrated a protective effect from traditional food consumption within the upper two quartiles (OR's for both were equal to 0.6, $p > 0.10$).

The odds of having diabetes were considerably reduced (OR = 0.4, $p = 0.18$) for those within the 4th quartile for healthy food consumption indicating some protective effect after adjustment for age and sex. However, results from the IGT group were unexpected: age and sex adjusted odds ratios above 2.5 in the third and fourth quartiles of healthy food consumption indicated increased risk with greater consumption of healthy food. While neither of these values satisfied a conventional level of statistical significance, the magnitude of the odds ratios was surprising. Also unexpected was the lack of association between healthy food and risk of obesity as all odds ratios were near unity both prior to and after adjustment for age and sex.

The results from the junk food subscale were also unexpected. Before adjustment for age and sex, all odds ratios were below 1.0 and two reached a conventional level of significance. The first of these results indicated that a high intake (4th quartile) of junk food was associated with substantial reduction in risk of diabetes or IGT (OR = 0.5, $p=0.07$). The second result suggested a modest intake of junk food (2nd quartile) was negatively associated with diabetes risk (OR = 0.4, $p = 0.03$). However, neither of these associations remained significant after adjustment for age and sex. The only positive associations between junk food and risk of disease was found among the IGT group within the 2nd and 4th quartiles and the 3rd quartile in the diabetes group. However none of these reached a conventional level of statistical significance. In all, the data from the junk food subscale provides no evidence of positive associations between the intake of junk food and disease as would have been expected.

Table 3.2.6 Food Frequency: Categorical Data

Unadjusted and adjusted† odds ratios (95% confidence intervals) displaying the relationship between food consumption and disease

	Diabetes / IGT (n = 107 / 84)**		Diabetes (n = 45 / 84)		IGT (n = 62 / 84)		Obesity* (n = 313 / 84)	
Subscale	***Unadjusted***	***Adjusted***	***Unadjusted***	***Adjusted***	***Unadjusted***	***Adjusted***	***Unadjusted***	***Adjusted***
Traditional Food								
Q2	1.6 (0.7 – 3.8) *p = 0.26*	1.2 (0.5 – 3.2) *p=0.61*	1.5 (0.5 – 4.7) *p = 0.49*	1.2 (0.4 – 3.9) *p = 0.79*	1.7 (0.7 – 4.5) *p = 0.27*	1.4 (0.5 – 4.1) *p = 0.58*	1.5 (0.7 – 2.9) *p = 0.27*	1.2 (0.6 – 2.4) *p = 0.61*
Q3	1.3 (0.6 – 3.2) *p = 0.51*	0.7 (0.3 – 1.8) *p = 0.40*	1.4 (0.5 – 4.5) *p = 0.54*	0.7 (0.2 – 2.3) *p = 0.51*	1.2 (0.5 – 3.4) *p = 0.63*	0.5 (0.2 – 1.8) *p = 0.29*	0.8 (0.4 – 1.7) *p = 0.64*	0.6 (0.3 – 1.2) *p = 0.12*
Q4	1.2 (0.9 – 4.5) *p = 0.12*	0.9 (0.34 – 2.28) *p = 0.79*	2.5 (0.9 – 7.1) *p = 0.10*	0.9 (0.3 – 3.1) *p = 0.90*	1.7 (0.7 – 4.3) *p = 0.30*	0.9 (0.3 – 2.9) *p = 0.89*	0.9 (0.5 – 2.0) *p = 0.99*	0.6 (0.3 – 1.3) *p = 0.19*
Healthy Food								
Q2	0.9 (0.4 – 2.0) *p = 0.84*	1.3 (0.5 – 3.2) *p = 0.63*	0.9 (0.4 – 2.3) *p = 0.81*	1.1 (0.4 – 3.3) *p = 0.81*	1.0 (0.4 – 2.5) *p = 0.55*	1.8 (0.5 – 6.3) *p = 0.36*	0.8 (0.4 – 1.6) *p = 0.54*	0.9 (0.4 – 1.7) *p = 0.66*
Q3	1.1 (0.5 – 2.4) *p = 0.84*	1.4 (0.6 – 3.0) *p = 0.48*	0.9 (0.4 – 2.4) *p = 0.85*	1.0 (0.3 – 2.9) *p = 0.98*	1.3 (0.5 – 3.3) *p = 0.61*	2.7 (0.8 – 9.0) *p = 0.11*	1.3 (0.7 – 2.6) *p = 0.46*	1.3 (0.7 – 2.6) *p = 0.45*
Q4	0.7 (0.3 – 1.6) *p = 0.38*	1.0 (0.4 – 2.7) *p = 0.92*	**0.2 (0.1 – 0.8)** ***p = 0.02***	0.4 (0.1 – 1.6) *p = 0.18*	1.2 (0.5 – 3.1) *p = 0.68*	2.6 (0.8 – 8.6) *p = 0.13*	0.9 (0.5 – 1.9) *p = 0.85*	1.0 (0.5 – 2.1) *p = 0.96*
Junk Food								
Q2	**0.4 (0.2 – 0.9)** ***p = 0.03***	0.8 (0.3 – 1.9) *p = 0.55*	**0.3 (0.1 – 0.8)** ***p = 0.02***	0.4 (0.1 – 1.4) *p = 0.16*	0.5 (0.2 – 1.3) *p = 0.15*	1.1 (0.4 – 3.2) *p = 0.86*	0.5 (0.3 – 1.0) *p = 0.05*	0.6 (0.29 – 1.2) *p = 0.14*
Q3	0.6 (0.3 – 1.4) *p = 0.23*	1.0 (0.4 – 2.6) *p = 0.95*	0.7 (0.3 – 1.8) *p = 0.46*	1.1 (0.4 – 3.1) *p = 0.92*	0.6 (0.2 – 1.4) *p = 0.21*	0.8 (0.3 – 2.4) *p = 0.68*	0.8 (0.4 – 1.5) *p = 0.43*	0.9 (0.4 – 1.8) *p = 0.70*
Q4	**0.5 (0.2 – 1.1)** ***p = 0.07***	1.0 (0.4 – 2.5) *p = 0.10*	0.5 (0.2 – 1.2) *p = 0.12*	0.8 (0.3 – 2.4) *p = 0.69*	0.5 (0.2 – 1.3) *p = 0.15*	1.2 (0.4 – 3.7) *p = 0.69*	0.6 (0.3 – 1.2) *p = 0.16*	0.8 (0.4 – 1.6) *p = 0.50*

Note: Statistically significant associations are bolded
† Adjusted for age (years) and sex
* Obese is defined as a body mass index ≥ 24 kg/m²
** n = (sample size for cases / control)

3.2.5 Analyses of Blood Lipid Markers

Descriptive and Multivariate Analyses

Table 3.2.7 displays the means (standard deviations) for blood lipids by quartile score for the Food Frequency subscales. Descriptive analyses of each of the food subscale scores showed few observable trends with regard to HDL, LDL, total cholesterol or triglyceride levels. Although this table is purely descriptive, it demonstrates that this community has a relatively good lipid profile. For example, despite the average age difference from Q1 to Q4 for scores on some of the subscales, in some cases, the LDL and HDL mean values were identical.

Table 3.2.7 Food Frequency: Analysis of Blood Lipid Markers

Means (standard deviations) for age, high and low density lipoproteins (HDL, LDL), cholesterol and triglyceride levels

	n	Age	HDL	LDL	Cholesterol	Triglycerides
All participants						
	404	36.1 (13.9)	1.2 (0.3)	2.8 (0.7)	4.7 (0.9)	1.5 (0.7)
Sex						
Males	183	36.0 (14.2)	1.2 (0.3)	3.0 (0.8)	4.9 (1.0)	1.6 (0.8)
Females	221	36.1 (13.7)	1.3 (0.2)	2.6 (0.6)	4.5 (0.7)	1.4 (0.6)
Score on Traditional Food Scale:						
1st Quartile	94	29.0 (6.9)	1.2 (0.3)	2.8 (0.7)	4.6 (0.9)	1.4 (0.6)
2nd Quartile	126	34.3 (11.3)	1.2 (0.3)	2.8 (0.8)	4.7 (0.9)	1.6 (0.7)
3rd Quartile	84	39.1 (14.5)	1.3 (0.3)	2.8 (0.7)	4.7 (0.9)	1.4 (0.8)
4th Quartile	100	42.3 (17.6)	1.3 (0.3)	2.8 (0.7)	4.8 (0.8)	1.5 (0.8)
Score on Healthy Food Scale:						
1st Quartile	102	39.2 (17.0)	1.3 (0.3)	2.9 (0.7)	4.8 (0.9)	1.5 (0.7)
2nd Quartile	89	36.2 (15.4)	1.3 (0.3)	2.8 (0.8)	4.7 (1.0)	1.4 (0.7)
3rd Quartile	118	36.0 (12.3)	1.2 (0.3)	2.7 (0.8)	4.7 (0.8)	1.5 (0.7)
4th Quartile	95	32.6 (9.1)	1.2 (0.3)	2.7 (0.7)	4.7 (0.8)	1.6 (0.9)
Score on Junk Food Scale:						
1st Quartile	132	40.7 (15.6)	1.2 (0.3)	2.9 (0.7)	4.8 (0.8)	1.5 (0.6)
2nd Quartile	81	34.6 (13.7)	1.3 (0.3)	2.7 (0.7)	4.6 (0.8)	1.4 (0.7)
3rd Quartile	96	35.1 (13.2)	1.2 (0.3)	2.8 (0.8)	4.7 (0.9)	1.6 (0.7)
4th Quartile	95	32.3 (10.4)	1.2 (0.3)	2.7 (0.7)	4.6 (0.8)	1.5 (0.8)

Units: mmol / L

Table 3.2.8 displays results of the linear regression analysis in which the food factors were analyzed as continuous measures. After adjustment for age and sex, no obvious associations were noted between any of the food scales and any of HDL, LDL, total cholesterol or triglyceride levels. However, none of the results from this analysis had p-values under 0.20.

Table 3.2.8 Blood Lipids: Continuous Data

Unadjusted and adjusted† beta coefficients (standard errors) displaying the relationship between factor scores and blood lipid markers

Subscale	HDL	LDL	Triglycerides	Total Cholesterol
Traditional				
Unadjusted	0.02 (0.02)	0.05 (0.04)	0.01 (0.02)	0.03 (0.02)
Adjusted	0.01 (0.02)	-0.03 (0.04)	0.02 (0.02)	-0.01 (0.02)
Healthy Food				
Unadjusted	-0.02 (0.02)	-0.05 (0.04)	0.02 (0.02)	-0.02 (0.02)
Adjusted	-0.02 (0.02)	0.01 (0.04)	0.03 (0.02)	0.01 (0.02)
Junk Food				
Unadjusted	-0.01 (0.02)	-0.10 (0.04)	0.01 (0.02)	-0.03 (0.03)
Adjusted	0.00 (0.02)	-0.06 (0.04)	0.01 (0.02)	-0.01 (0.02)

† adjusted for age (years) and sex.
Note: none of these values was significant at the $p < 0.20$ level
Units: mmol / L

3.3 Subscale #2 – Household Items

3.3.1 Results of Factor Analysis

As with the development of previous subscales, the total number of available participants was 411. Factor analysis excluded any individuals for whom the 19-item Household Items questionnaire was not completed in full. Therefore, 319 observations were included in the factor analysis (FA) with data on all 19 separate household items.

Hatcher's criteria[77] were again used in order to determine the number of "meaningful" factors to retain (see chapter 2). The results from the factor analysis are presented in Table 3.3.1. There were two factors retained and the items which loaded on each factor at > 0.40 are listed. For a step-by-step description of how the number and nature of factors from these data were determined, refer to Appendix.

The principle of 'simple structure' was also used again to determine the items which loaded on each factor. A combination of ethnographic knowledge and statistical principles (FA) was used in order to ascertain underlying constructs as represented by each group of items. Seven items loaded on the first factor and five on the second factor. The items which loaded on each of the factors could be distinguished by the following: The activities which corresponded with the items within factor 1 were modern activities (i.e.: microwave for cooking, electric dryer for washing, etc). Items which clustered into factor 2, namely, sewing machine, big freezer, boat, and a motor for the boat were items used to make traditional activities more efficient such as beadwork, and storing the spoils of a big hunt and fishing. The only exception to this interpretation was the item 'washing machine', found within traditional items. However, it was believed that some of the respondents might have interpreted this item as a manual washboard, rather than an electric appliance.

Due to a combination of the statistical criteria and interpretability of the items as they clustered, the factors were therefore named as follows: factor 1 = "modern items for modern activities" or for short, "modern items", and factor 2 = "modern items for traditional activities" or for short, "traditional items" (see Appendix for full factor analysis results).

Table 3.3.1 Household Items Data: Factor Analysis Results

Questionnaire Items and Corresponding Factor Loadings from the Rotated Factor Pattern Matrix and Factor Structure Matrix, Decimals Omitted

	Factor Pattern			Factor Structure		
Factor 1:	1	2		1	2	
Television	*62	-10		*58	20	
VCR	*57	-14		*51	14	
Microwave	*47	2		*48	24	
Telephone	*44	12		*50	33	
Dryer	*50	-11		*44	13	
Car	38	4		*40	22	
Cassette player	35	9		*40	26	
Factor 2:						
Sewing machine	-8	*50		16	*47	
Big freezer	12	*50		36	*56	
Boat	5	*62		35	*65	
Motor for the boat	13	*70		*46	*76	
Washing machine	15	*50		39	*57	

*Note: the items 'car' and 'cassette player' loaded on factor 1 in the factor structure =0.40, however just slightly under 0.40 for the factor pattern. Due to the interpretability of the items as contributors to the underlying meaning of the factors and the borderline factor loadings, they are listed as items loading on factor 1.

3.3.2 Results of Scale Reliability and Coefficient-α

Scale reliability estimates were assessed by calculating coefficient-α.[83] The modern items scale had an $\alpha = 0.64$ and the traditional items scale had an $\alpha = 0.73$, acceptable levels of Cronbach's-α (> 0.60) which indicated a satisfactory level of consistency in the household items responses.

3.3.3 Factor Score Distributions

Table 3.3.2 presents descriptive data for the two Household Item factor scores. The values are reported in the unit of possessions ('r.p.', or reported # of possessions). Factor analysis produced individual scores for each participant on each factor and these scores were set with a mean of zero. The data shown in this table presents the distribution of scores for each factor across the entire community. Similar to the Food Frequency data, a test for normality was performed to determine whether the factor scores for Household Items were a random sample from a normal distribution, and could therefore satisfy the assumption of normality used for regression analyses.

Table 3.3.2 Descriptive Statistics for the Household Items data

	n	Median	Skewness	Kurtosis	Q1 – Q3	min/max
Modern Activities	319	0.4	0.5	0.1	1.2	- 1.6 / 2.6
Traditional Activities	319	0.5	0.7	- 0.3	1.4	- 1.3 / 2.5

Units: reported possessions (r.p.)

The Shapiro-Wilk test was specified in the SAS procedure and the results are presented with a stem-and-leaf plot in Appendix F. The W statistics for the Household Items factor scores were 0.95 ($p< 0.001$) and 0.91 ($p< 0.001$), respectively. Based on these results, the factor scores did not follow a normal distribution. Therefore, the null hypothesis should have been rejected. Even after log transformations were performed, the W statistic's test for normality showed no improvement for either subscale (Appendix: Table 6.5.8 and 6.5.9). However, the W statistic was used in conjunction with the stem and leaf plot and the plot of the normal distribution and in using all three tests, it was determined that the factor score distributions were similar enough to a normal distribution to warrant further analyses which required an assumption of normality (for example, regression analyses).

Table 3.3.3 Frequency Distribution: Household Items

Activity	Normal (n = 64)	Diabetes/IGT (n = 92)	Diabetes (n = 38)	IGT (n = 54)	Obesity (n = 249)
Modern					
Q1	18	10	5	5	44
Q2	14	15	5	10	44
Q3	19	25	13	12	60
Q4	13	42	15	27	101
Traditional					
Q1	19	17	9	8	71
Q2	14	13	5	8	41
Q3	16	19	7	12	52
Q4	15	43	17	26	85

Table 3.3.3 displays the frequency distribution of the scores in each subscale. Most notably from these data was frequency of participants with diabetes and IGT with increasing quartiles in both subscales. Of particular note was the increase in the number of participants within the Modern Activities subscale with increasing quartiles. Table 3.3.6 (see section 3.3.4) displays the categorical analysis and presents odds ratios for these quartiles. The OR corresponding to quartile 4 within the

modern activities subscale was 3.4 and 6.4 for diabetes and IGT, respectively. Further description of the odds ratios can be found in section 3.3.4. The values from Table 3.3.3 were also used for power and sample size calculations (see section 6.7 for descriptive tables or Chapter 4 for the discussion of sample size determination).

Table 3.3.4 Household Items Data: Analysis of Average Age and BMI by Factor Quartile

Means (standard deviations) for age and BMI among sample

		n	Age (years)	BMI (kg/m^2)
All participants		213	31.8 (9.8)	27.5 (5.1)
Sex	Males	110	32.5 (10.5)	26.3 (4.1)
	Females	113	30.9 (9.1)	28.8 (5.7)
Males:				
Score on Modern Items Scale:				
	1st Quartile	29	33.6 (11.6)	25.5 (3.9)
	2nd Quartile	19	32.6 (9.0)	25.5 (4.5)
	3rd Quartile	24	29.6 (8.2)	26.3 (4.0)
	4th Quartile	38	33.6 (11.2)	27.4 (4.1)
Score on Traditional Items Scale:				
	1st Quartile	34	30.2 (8.4)	26.7 (4.4)
	2nd Quartile	25	33.9 (9.4)	25.2 (4.3)
	3rd Quartile	20	31.9 (8.9)	26.4 (3.9)
	4th Quartile	31	34.5 (13.8)	26.7 (3.9)
Females:				
Score on Modern Items Scale:				
	1st Quartile	14	26.5 (5.1)	27.6 (6.5)
	2nd Quartile	22	29.9 (8.3)	27.8 (5.2)
	3rd Quartile	25	30.8 (10.7)	28.8 (6.1)
	4th Quartile	42	33.0 (9.1)	29.9 (5.3)
Score on Traditional Items Scale:				
	1st Quartile	25	28.6 (7.3)	29.5 (5.6)
	2nd Quartile	16	31.0 (5.5)	28.9 (5.4)
	3rd Quartile	26	30.6 (9.0)	28.1 (6.3)
	4th Quartile	27	34.1 (11.9)	28.7 (5.4)

Table 3.3.4 displays results of the descriptive analysis of age and BMI distribution by quartiles of the factor scores. The data were stratified by sex and are presented for both factors. For men, average age showed no observable linear trend from lowest to highest quartile for modern items with no relationship apparent for BMI scores. In contrast for females, increased reported possession of modern items was positively associated with average age and, though less strongly, with increased average BMI. The traditional items factor scores were not associated with average age or BMI among men; among women, older women reported possession of more traditional items, though

there was no association with BMI. These observations are consistent with what is known about the culture of Sandy Lake, however, they do not explain why the association between age and the factors describing possession of modern or traditional items appear to differ for men and women.

3.3.4 MULTIVARIATE ANALYSIS: LOGISTIC REGRESSION

Continuous Data Summary

Logistic regression analysis was performed in order to assess the relationship between the risk of disease and exposure to the items in these subscales, represented as continuous factor scores (Table 3.3.5). Odds ratios refer to change in risk of disease outcome relative to 1.0 unit increase in factor score. This value was chosen based on the Q1-Q3 range determined from univariate analysis of the factor scores (see Appendix F).

The analyses displayed in Table 3.3.5 used individual factor scores as the continuous exposure variables and four categories of disease were used as outcome variables: diabetes or IGT, diabetes, IGT and obesity. Odds ratios displayed are both adjusted and unadjusted for age and sex. All but one of the odds ratios were statistically significant at the $p < 0.05$ level.

Table 3.3.5 Household Data: Continuous Data Summary

Unadjusted and adjusted† odds ratios (95% confidence intervals) displaying the relationship between the possession of either modern or traditional labor-saving devices with health outcomes for individuals 20 years of age and older

Activities	**Diabetes / IGT** (n = 92 / 64)**	**Diabetes** (n = 38 / 64)	**IGT** (n = 54 / 64)	**Obesity*** (n = 249 / 64)
Modern				
Unadjusted	**1.9 (1.3 – 2.9)²**	**1.7 (1.0 – 2.9)²**	**2.0 (1.3 – 3.1)²**	**1.5 (1.1 – 2.2)²**
Adjusted	**2.4 (1.4 – 3.9)¹**	**2.1 (1.2 – 3.7)²**	**2.4 (1.4 – 4.3)²**	**1.5 (1.1 – 2.1)²**
Traditional				
Unadjusted	**2.2 (1.5 – 3.3)¹**	**2.1 (1.3 – 3.4)²**	**2.5 (1.5 – 4.1)¹**	**1.5 (1.1 – 2.1)²**
Adjusted	**1.9 (1.2 – 2.9)²**	**1.7 (1.0 – 3.0)²**	**2.2 (1.3 – 3.9)²**	1.3 (0.9 – 1.9)³

† adjusted for age (years) and sex.
* obesity is defined as a body mass index of ≥ 24 kg m^2
** n = (sample size of cases / controls)
¹ significant at $p < 0.001$
² significant at $p < 0.05$
³ significant at $p < 0.10$

An increased score on the modern activities subscale was significantly ($p < 0.05$) positively associated with risk for all disease states both before and after adjustment for age and sex. Positive associations were also observed between scores on the traditional scale and risk of each disease outcome prior to age and sex adjustment. After adjustment, these positive associations remained, but were less strong (the highest OR was 2.2 comparing participants with IGT to normals and the lowest OR was 1.3 comparing obese people to those of normal body mass). All OR's were positive for all disease states (and both factors) after age and sex adjustment and remained significant at $p < 0.05$ for all conditions except when obese and normal people were compared on the traditional factor scores which was significant at $p < 0.10$.

Categorical data summary

Table 3.3.6 displays results of the categorical data analysis of the Household Items using quartiles of factor scores to inspect any departure from linearity in the data. All significant associations at $p < 0.05$ are bolded in the table. The highest risk for both factors was evident for all disease states in the 4th quartile. The greatest adjusted odds ratios were observed for modern items within the 4th quartile comparing people with IGT to normals (OR= 6.3, p = 0.005) and for traditional items within the 4th quartile comparing these same two groups (OR = 4.2, p = 0.05).

For the factor describing possession of modern items, odds ratios increased with increasing quartiles of the factor in all comparisons, with the strongest association observed in the comparison of people with IGT or diabetes to normal people. Here the odds ratio increased from 1.4 to 3.2 to 4.9 across Q2-Q4 with a p-value under 0.05 for Q3 and Q4. For the entire traditional items factor, adjusted odds ratios in Q2 and Q3 were negative for all disease states and were only significantly elevated in Q4. This indicates that for a modest score on traditional items, there is some protective effect, however, for those in the upper quartile for traditional items there is significant added risk (Table 3.3.6)

Table 3.3.6 Household Data: Categorical Data

Unadjusted and adjusted† odds ratios (95% confidence intervals) displaying the relationship between possession of modern devices for either traditional or modern usage

	Normal vs. Diabetes / IGT (n = 92 / 64)**		Normal vs. Diabetes (n = 38 / 64)		Normal vs. IGT (n = 54 / 64)		**Normal vs. Obese* (n = 249 / 64)**	
Activity	*Unadjusted*	*Adjusted*	*Unadjusted*	*Adjusted*	*Unadjusted*	*Adjusted*	*Unadjusted*	*Adjusted*
Modern								
Q2	1.4 (0.5 – 4.0) *p = 0.52*	1.4 (0.5 – 4.0) *p=0.55*	0.9 (0.2 – 3.9) *p = 0.93*	1.0 (0.3 – 3.7) *p = 0.10*	1.9 (0.5 – 6.7) *p = 0.33*	1.2 (0.3 – 4.3) *p = 0.8*	1.0 (0.5 – 2.3) *p = 0.94*	1.1 (0.5 – 2.2) *p = 0.88*
Q3	2.1 (0.8 – 5.8) *p = 0.51*	**3.2 (1.1 – 9.0) *p = 0.03***	2.4 (0.7 – 8.4) *p = 0.17*	2.6 (0.7 – 9.0) *p = 0.14*	1.8 (0.5 – 6.6) *p = 0.37*	3.0 (0.8 – 10.4) *p = 0.09*	1.2 (0.5 – 2.6) *p = 0.74*	1.6 (0.7 – 3.5) *p = 0.23*
Q4	**4.9 (1.8 – 13.1) *p = 0.002***	**4.9 (1.6 – 14.8) *p = 0.005***	**3.4 (1.0 – 11.6) *p = 0.05***	3.1 (0.8 – 12.5) *p = 0.11*	**6.4 (2.0 – 21.0) *p = 0.002***	**6.3 (1.7 – 22.9) *p = 0.005***	**3.0 (1.4 – 6.7) *p = 0.006***	**2.4 (1.0 – 5.6) *p = 0.04***
Traditional								
Q2	0.9 (0.3 – 2.4) *p = 0.80*	0.5 (0.2 – 1.5) *p = 0.22*	0.5 (0.1 – 1.9) *p = 0.32*	**0.3 (0.1 – 1.2) *p = 0.08***	1.3 (0.40 – 4.4) *p = 0.62*	0.6 (0.2 – 2.3) *p = 0.47*	0.8 (0.4 – 1.7) *p = 0.5*	0.6 (0.3 – 1.3) *p = 0.17*
Q3	1.1 (0.4 – 3.0) *p = 0.80*	0.6 (0.2 – 1.8) *p = 0.36*	0.83 (0.3 – 2.8) *p = 0.76*	**0.3 (0.1 – 1.1) *p = 0.07***	1.5 (0.5 – 5.0) *p = 0.49*	0.8 (0.2 – 3.1) *p = 0.78*	0.8 (0.4 – 1.7) *p = 0.54*	0.6 (0.3 – 1.2) *p = 0.14*
Q4	**2.8 (1.1 – 6.8) *p = 0.03***	**3.0 (0.9 – 9.5) *p = 0.07***	1.9 (0.7 – 5.5) *p = 0.24*	2.2 (0.6 – 8.4) *p = 0.25*	**3.9 (1.3 – 11.4) *p = 0.01***	**4.1 (1.0 – 16.6) *p = 0.05***	1.4 (0.7 – 3.0) *p = 0.38*	1.7 (0.6 – 4.5) *p = 0.30*

Note: Statistically significant associations are bolded
† Adjusted for age (years) and sex
* Obese is defined as a body mass index ≥ 24 kg/m^2
** n = (sample size for cases / controls)

3.3.5 Association between Physical Fitness and Household Items

Descriptive and Multivariate Analyses

Means and standard deviations of $V0_{2max}$ scores are described for each quartile of the modern and traditional Household Item factor scores and stratified by sex for each of the 2 subscales in Table 3.3.7. A total of 213 participants were assessed by the fitness test. The exclusion of almost half of the sample was due to the stringent criteria which excluded volunteers with medical conditions that contraindicated vigorous exercise

Table 3.3.7 Household Data: Descriptive Analysis of Physical Activity

Means (standard deviations) of age, weight, body mass index (BMI) and measured maximal oxygen uptake† (VO_{2max}), by factor score and by sex

		n	Age (years)	BMI(kg / m²)	Weight (kg)	VO_{2max}
All participants		213	31.8 (9.8)	27.5 (5.1)	78.7 (14.7)	55.2 (8.2)
Sex	Males	110	32.6 (10.5)	26.3 (4.1)	80.9 (14.3)	57.6 (8.9)
	Females	103	30.9 (9.1)	28.8 (5.7)	76.3 (14.9)	52.8 (6.7)
MALES:						
Score on modern household items:						
	1st Quartile	29	33.6 (11.6)	25.5 (3.9)	78.6 (14.0)	55.8 (9.9)
	2nd Quartile	19	32.6 (10.0)	25.5 (4.4)	77.5 (12.6)	57.3 (8.8)
	3rd Quartile	24	29.6 (8.2)	26.3 (4.0)	80.6 (14.3)	59.7 (9.0)
	4th Quartile	38	33.6 (11.2)	27.4 (4.1)	84.6 (15.1)	57.7 (7.9)
Score on traditional household items:						
	1st Quartile	34	30.2 (8.4)	26.7 (4.4)	83.4 (14.6)	56.9 (8.7)
	2nd Quartile	25	33.9 (9.4)	25.2 (4.3)	76.8 (14.9)	57.9 (8.1)
	3rd Quartile	20	31.9 (8.9)	26.4 (3.9)	80.5 (13.5)	59.6 (7.3)
	4th Quartile	31	34.5 (13.8)	26.7 (3.9)	81.7 (13.9)	56.7 (10.4)
FEMALES:						
Score on modern household items:						
	1st Quartile	14	26.5 (5.1)	27.6 (6.5)	71.8 (15.6)	54.4 (7.2)
	2nd Quartile	22	29.9 (5.2)	27.8 (5.2)	73.2 (13.3)	52.9 (6.3)
	3rd Quartile	25	30.8 (10.7)	28.8 (6.1)	77.2 (14.9)	53.6 (5.7)
	4th Quartile	42	33.0 (9.1)	30.0 (5.3)	78.8 (15.3)	51.6 (7.2)
Score on traditional household items:						
	1st Quartile	34	28.6 (7.3)	29.5 (5.6)	77.7 (12.7)	53.4 (6.0)
	2nd Quartile	16	31.0 (5.5)	28.9 (5.4)	75.1 (14.2)	52.0 (8.0)
	3rd Quartile	26	30.6 (9.0)	28.1 (6.3)	75.9 (17.1)	52.6 (7.7)
	4th Quartile	27	34.1 (11.9)	28.7 (5.5)	75.5 (16.2)	52.5 (5.9)

† adjusted for % lean body mass

The average age of the fitness test participants was 31.8 years, slightly lower than the study average of 35.9 years for all subjects, and the average BMI for men and women who completed the fit-test was slightly less than the study sample average (men: 26.3 versus 27.0, women: 28.8 versus 29.0). These results reflect the inclusion criteria's tendency to only include healthy individuals.

Table 3.3.7 indicates that average age was highest in Q4 of each subscale for both sexes. Only Q1 within the 'modern items' scale displayed an average age equal to Q4 for that subscale. Also with the exception Q1 for modern items for men, the lowest average age was found in the first quartile scores for both subscales. This may reflect the tendency for older aged individuals to possess and report more items in general.

For both men and women, average BMI and weight values increased from Q1 – Q4 for the modern items subscale. This further confirms that modern labor-saving devices used for modern activities were associated with diabetes, IGT and particularly with obesity (Section 3.3.4). The average BMI and weight for men and women within the traditional items subscale showed no apparent increase across Q1 – Q4. This again confirms the results of section 3.3.4 since no statistically significant association was reported between scores on the traditional items subscale and obesity.

Average $V0_{2max}$ score for all males was 57.6 (sd = 8.9) and for females was 52.8 (sd = 6.7). Among male respondents for both factors, $V0_{2max}$ scores increased from Q1 – Q3 of both factors and then dropped again for Q4. The female respondents showed no clear trends for $V0_{2max}$ scores across the quartiles, though the highest $V0_{2max}$ was observed among women in the first quartile of each factor score. Further analyses including regression models were run in order to investigate these associations.

Table 3.3.8 displays the results from the regression modeling between factor scores and measures of BMI, weight and $V0_{2max}$ when all measures were continuous variables. Factor scores were used as exposure variables with body mass index (BMI), average weight and $V0_{2max}$ scores used as outcome variables – note that beta-coefficients are presented both before and after adjustment for age and sex. These results are somewhat consistent with previous findings reported here on the Household Item factor scores. In particular, scores for both modern and traditional items were positively associated with BMI and average weight for modern items, however only the modern items were significantly associated ($p < 0.05$) after adjustment for age and sex.

Table 3.3.8 Physical Activity: Continuous Data

Unadjusted and adjusted† beta coefficients (standard errors) displaying the relationship between factor scores and BMI, average weight and $V0_{2max}$

Subscale		BMI	Weight	VO_{2max}
Modern	Unadjusted	**0.9 (0.4)**	1.7 (1.1)	- 0.000098 (0.00063)
	Adjusted	**0.8 (0.4)**	**2.0 (1.1)**	0.0006 (0.0005)
Traditional	Unadjusted	0.3 (0.4)	0.7 (1.1)	- 0.0001 (0.00062)
	Adjusted	0.3 (0.4)	0.8 (1.1)	0.0008 (0.0005)

† adjusted for age (years) and sex.

The association between $V0_{2max}$ and factor scores was less strong than the relationship between BMI and average weight and the factor scores. Prior to adjustment for age and sex, the slope for both factors and $V0_{2max}$ were negative. Once age and sex were added to the regression models, the slopes became positive and similar in magnitude for both factors. Since the p-values were not significant for these associations ($p >> 0.10$) and the slopes were close to zero, there was no detectable association between scores on these subscales and VO_{2max}.

3.4 Subscale #3 – Language Proficiency

3.4.1 Results of Factor Analysis

The total number of available participants for this analysis was 411. Factor analysis excluded any individual for whom the 6-item questionnaire was not completed in full. In the original SLHDP, all participants completed the Language questionnaire. Therefore, for this section a total of 411 observations were inputted into the factor analysis with data on the 6 separate questions regarding proficiency of language.

Table 3.4.1 Language Data: Factor Analysis Results

Questionnaire Items and Corresponding Factor Loadings from the Rotated Factor Pattern Matrix and Factor Structure Matrix, Decimals Omitted

		Factor Pattern			Factor Structure		
Factor 1:		1	2		1	2	
	Reading English	*99	4		*98	39	
	Speaking English	*96	4		*88	-36	
	Writing English	*90	1		*96	*-42	
Factor 2:							
	Reading Oji-Cree	-4	*92		*-44	*94	
	Speaking Oji-Cree	5	19		-3	18	
	Writing Oji-Cree	-6	*91		*-46	*94	

Results from the factor analysis from the Language subscale development are presented in Table 3.4.1. There were only two factors retained and the items which loaded on each factor at > 0.40 are listed for both the factor pattern and structure. According to Hatcher, the results from the factor pattern should be relied upon to interpret the meaning of each factor.[77] By convention, the factor structure is also presented in Table 3.4.1, however, Hatcher suggests that these results only be used to help interpret the nature of a factor in the case of strongly correlated factors with loadings in excess of 10.00 on the factor pattern. Therefore, although the factor structure matrix seen in Table 3.4.1 shows items which load on more than one factor at > 0.40 – an apparent violation of the rule of 'simple structure' – the results of the pattern matrix confirm that only two distinct factors satisfy the rule of simple structure and can thus be interpreted. (For complete SAS output from factor analysis, see Appendix G).

The factor pattern which represents the standardized regression coefficients or pattern loadings showed that the survey items which tested English skills clustered together, resulting in high loadings on factor 1. Similarly, items which tested for Oji-Cree skills clustered into a single factor 2. Thus, factor 1 was termed 'English Proficiency'. For factor 2, only 2 of the 3 Oji-Cree items clustered together, 'reading' and 'writing' Oji-Cree loaded > 0.40 while speaking Oji-Cree did not reach this conventional level. While having only two items load on one factor is an apparent breach of Hatcher's criteria, the rationale used to justify further analyses was the acceptable level of scale reliability as determined by the coefficient-α determination (see section 3.4.2 for a summary of these results or for full SAS results and output, see Appendix G).

The construct underlying the items of factor 2 described a level of proficiency not merely attained from a casual aptitude of Oji-Cree; for example, not just the ability to speak, but specifically to read and write as well. Therefore, while factor 2 measures a level of 'proficiency' in Oji-Cree, the quality of that ability is not an informal adeptness. Rather, a high score on factor 2 represents sophisticated knowledge of the language and perhaps an exposure to some form of Oji-Cree education. Factor 2 was therefore named, "Oji-Cree Education".

3.4.2 Results of Scale Reliability and Coefficient-α

The principle of 'simple structure' was used to determine the items which loaded on each factor (see Chapter 2, section 2.4.2 (b)). These criteria typically require a minimum of three items per factor in order to ensure a high degree of reliability in the scale. As previously mentioned, factor 2 produced only 2 items which loaded ≥ 0.40. While this was a breach of Hatcher's criterion in the literal sense, the theoretical rationale was maintained; the coefficient-α for factor 1 was 0.96 and for factor 2 was 0.98. This represents a very high degree of reliability and for factor 2, the two items sufficiently represent the underlying construct of "Oji-Cree Education" (see Chapter 4, section 4.3.3 for further discussion).

3.4.3 Factor Score Distributions

Table 3.4.2 presents descriptive data for the individual level factor scores from the Language data. Factor analysis produced two distinct factors and every participant eligible for the present analysis (n=411) was given a score on each subscale. This table presents the distribution of scores for each factor across the entire sample (see Appendix G for full output).

Table 3.4.2 Descriptive Statistics for the Language Data

	n	Median	Skewness	Kurtosis	Q1 – Q3	min/max
English Proficiency	411	- 0.8	- 2.6	4.7	0.1	-2.9 / 0.37
Oji-Cree Education	411	0.3	0.9	- 1.1	1.4	-0.7 / 1.6

units: level of proficiency

Table 3.4.3 displays the results of the frequency distribution of subscales scores for both language factors. The heavily skewed nature of the data is also evident in this table since there should be a relatively even number of participants in each level of scores within the 'normal' group. These data were also used for power and sample size calculation (see section 6.7 for descriptive table or Chapter 4 for discussion).

Table 3.4.3 Frequency Distribution: Language Data

Language	Normal (n = 85)	Diabetes/IGT (n = 108)	Diabetes (n = 45)	IGT (n = 63)	Obesity (n = 317)
English					
Lower	32	53	22	31	121
Upper	53	55	23	32	196
Oji-Cree					
Lower	61	65	28	37	219
Upper	25	43	17	26	98

Similar to the Food Frequency and Household Item data, an analysis was undertaken to determine whether the factor scores for the Language data were a random sample from a normal distribution, and therefore satisfied the assumption of normality used for regression analyses. The Shapiro-Wilk test was specified in the SAS procedure and the results are presented with a stem-and-leaf plot in Appendix. The test for normality demonstrated that the distributions of these two factor scores showed a marked departure from normal distributions.

The W statistics for the factor 1 and factor 2 distributions were 0.39 ($p < 0.001$) and 0.59 ($p < 0.001$), respectively. This indicated that the two distributions of scores did not follow a normal distribution. Therefore, log transformations were performed and the W statistic was tested again. These methods were unable to improve the W statistic's test for normality. The univariate analysis showed that the data were generally bimodal and therefore the data were re-coded into a dichotomous variable (as opposed to the quartile variables used in the previous subscales). This dichotomous variable was created using the median score for normal people as the cut-off for 'below' and 'above' the median value. The cut point was determined using the results from the univariate analyses of normals to determine the median value of the factor scores. Although distribution problems were evident and were unchanged by transformation techniques, linear and logistic regression were still performed in keeping with the procedures from the other two subscales. However, the interpretation of results focuses on the categorical data analysis. (For full Language analysis results, see Appendix G).

Table 3.4.4 Language Data: Analysis of Average Age and BMI by Factor Bi-Level Score

Means (standard deviations) for age and BMI among sample

		n	Age (years)	BMI (kg / m^2)
All participants		411	35.9 (13.8)	28.2 (5.1)
Sex	Males	185	35.9 (14.0)	27.0 (4.5)
	Females	226	35.9 (13.6)	29.0 (5.5)
MALES:				
Score on English Proficiency Scale:				
	Below median	79	41.7 (17.6)	27.3 (4.6)
	Above median	105	31.8 (8.9)	26.9 (4.3)
Score on Oji-Cree Education Scale:				
	Below median	125	32.3 (10.2)	27.0 (4.3)
	Above median	60	44.1 (17.8)	27.1 (4.7)
FEMALES:				
Score on English Proficiency Scale:				
	Below median	80	43.6 (17.7)	28.9 (5.0)
	Above median	147	31.6 (8.0)	29.1 (5.7)
Score on Oji-Cree Education Scale:				
	Below median	159	31.6 (8.8)	29.1 (5.7)
	Above median	68	45.7 (17.2)	28.8 (5.0)

Table 3.4.4 displays results of the descriptive analysis of age and BMI distributions with stratified bi-level scores from the SLHDP Language data. These data are also stratified by sex and are presented for both factors. As expected, the average ages for men and women with English Proficiency scores above the median were lower than those English Proficiency scores below the median. Similarly, the average ages for men and women with scores above the median on Oji-Cree Education were older that those with lower Oji-Cree Education scores. These observations confirm what is known about the elders of the community and their increased exposure to the Oji-Cree culture.

The average BMI for males with English Proficiency scores below the median was higher than the average BMI for males who scored above the median. This was consistent with the findings for the average age of these groups since the older group displayed an average BMI greater than the younger group. This was expected, given that age is significantly associated with obesity (as measured by BMI) in Sandy Lake and in the Canadian population in general.[13] A similar association was not found within the male groups above and below the median for Oji-Cree Education scores. These scores differed only by 0.1 kg / m^2.

The average BMI score for females above and below the median for scores on English Proficiency and Oji-Cree Education were unexpected. Here, the older groups displayed lower average BMI scores than the younger groups for both factors. This was opposite to the findings from the male groups and also to what is known about the association between BMI and age in the general Canadian population.[84]

3.4.4 MULTIVARIATE ANALYSIS: LOGISTIC REGRESSION

Continuous Data Summary

Results of logistic regression analysis using factor scores as continuous variables are presented in Table 3.4.5. Increased proficiency in English, as reflected by an increased factor score, was positively associated with the risk of each disease condition after age and sex adjustment. These associations were statistically significant in the comparison of normal individuals to those with diabetes / IGT (OR = 1.8), IGT (OR = 2.0) to normal individuals at $p < 0.05$ and obesity at $p < 0.10$.

In contrast, increased Oji-Cree Education, as reflected by an increased score on factor 2, was negatively associated with risk (after age and sex adjustment) though only the association with diabetes reached a conventional level of statistical significance ($p < 0.10$).

Table 3.4.5 Language Data: Continuous Data

Crude and adjusted† odds ratios (95% confidence intervals) displaying the relationship between proficiency in English and Oji-Cree Education and various health outcomes

Factor	**Diabetes / IGT** (n = 108 / 85)**	**Diabetes** (n = 45 / 85)	**IGT** (n = 63 / 85)	**Obesity*** (n = 317 / 85)
English Proficiency				
Crude	**0.6 (0.5 – 0.9)**[1]	**0.6 (0.4 – 0.8)**[1]	**0.7 (0.5 – 0.9)**[1]	0.8 (0.6 – 1.1)
Adjusted	**1.8 (1.0 – 3.0)**[1]	1.4 (0.8 – 2.4)	**2.0 (1.1 – 3.6)**[1]	**1.5 (1.0 – 2.3)**[2]
Oji-Cree Education				
Crude	1.2 (0.9 – 1.6)	1.1 (0.8 – 1.7)	1.3 (0.9 – 1.8)	1.0 (0.8 – 1.3)
Adjusted	0.7 (0.5 – 1.1)	**0.6 (0.3 – 1.0)**[2]	0.8 (0.5 – 1.3)	0.8 (0.6 – 1.1)

† adjusted for age (years) and sex
* body mass index ≥ 24 kg / m²
** sample size: (cases / controls)
[1] significant at $p < 0.05$
[2] significant at $p < 0.10$

Categorical Data Summary

Individual factor scores for English Proficiency and for Oji-Cree Education were categorized into a two-level categorical analysis using scores above and below the median value for normals.

Table 3.4.6 Language Data: Categorical Data

Unadjusted and adjusted† odds ratios (95% confidence intervals) displaying the relationship between Language factor scores and various health outcomes

Factor	**Diabetes / IGT** (n = 108 / 85)**	**Diabetes** (n = 45 / 85)	**IGT** (n = 63 / 85)	**Obesity*** (n = 317 / 85)
English Proficiency				
Crude	0.6 (0.4 – 1.1)	0.6 (0.3 – 1.3)	0.6 (0.3 – 1.2)	1.0 (0.6 – 1.6)
Adjusted	1.1 (0.5 – 2.3)	1.6 (0.6 – 4.5)	0.8 (0.3 – 2.0)	1.2 (0.7 – 2.0)
Oji-Cree Education				
Crude	1.6 (0.9 – 2.9)	1.5 (0.7 – 3.1)	1.7 (0.9 – 3.4)	1.1 (0.6 – 1.8)
Adjusted	0.6 (0.3 – 1.3)	**0.4 (0.1 – 0.9)**	0.7 (0.3 – 1.8)	0.7 (0.4 – 1.3)

† adjusted for age (years) and sex
* body mass index ≥ 24 kg / m²
** sample size: (cases / controls)

Results of the logistic regression analysis with the Language factors categorized above and below the median are presented in Table 3.4.6. Odds ratios refer to change in risk of disease among diseased compared to controls at different levels of exposure. The exposure in this analysis was the category score for factor scores (above or below the median).

For this analysis, only one odds ratio value was significant at $p < 0.05$ and it is bolded in the table. The risk of each condition increased with English Proficiency and decreased with Education in Oji-Cree, though only the association between diabetes and Oji-Cree Education reached statistical significance. The risk of diabetes was reduced by 64% for those with higher Oji-Cree Education once adjusted for age and sex (OR = 0.36, $p < 0.05$).

While the results of the categorical data analyses (Table 3.4.6) were similar to those of the continuous analysis (Table 3.4.5), the treatment of the data as categorical instead of continuous produced less results which were significant at the conventional level of $p < 0.05$. While the categorization of the continuous data was done in order to improve the lack of normality in the distribution of the data, it caused a loss of power among the associations. This can be seen in the reduced number of statistically significant OR's in Table 3.4.6 as opposed to Table 3.4.5.

Chapter 4

Conclusions and Discussion

4.0 Introduction

The ultimate objective of this study was to identify which elements of acculturation were associated with diabetes, IGT or obesity. The data from the Sandy Lake Health and Diabetes Project were used to define and measure acculturation in the community of Sandy Lake, Ontario and were chosen based on a biocultural model of disease. The first two variables used in the analysis, Food Frequency and Household Items have a biologic component – the former, a direct link with disease and the latter an indirect link via level of physical activity. In previous studies, these biologically active variables had exhibited associations with the disease processes under investigation. However, in a community undergoing such profound and rapid social changes, cultural factors were also thought to be potential sources of disease. Therefore, although data from the Language section had no known biologic connection with diabetes, IGT or obesity and was therefore a purely cultural factor, it was included in the model of disease risk as an indicator of individual cultural affiliation.

The discussion in this chapter will highlight the results and measurement considerations of each subscale as well as delineate the importance of integrating the anthropologic and epidemiologic perspective into a cohesive, biocultural model of disease risk.

4.1 Food Frequency Subscale

4.1.1 Does a Modern Diet Lead to Diabetes?

The ethnographic analysis which preceded the Sandy Lake Health and Diabetes Project determined that the community of Sandy Lake clearly distinguished between two categories of foods. "Indian foods" which consisted of bush meats and wild plants were considered healthy while "white man's foods" which encompassed 'junk foods' such as pop, candy and chocolate were perceived as unhealthy.[72] In part, the present study was designed to determine whether consumption of certain groups of foods were associated with diabetes, IGT or obesity and thereby ascertain whether a 'modern diet' may be linked to the high prevalence of these diseases in the community.

Previous work within communities rapidly acculturating to a host-society has focused primarily on physical, behavioral and socio-cultural risk factors for diabetes. Researchers in this area have suggested that diabetes risk is likely associated with a particular dietary pattern and that only a culturally appropriate measure could uncover this hypthosis.[56] However, despite this claim, only a small number of studies have investigated the risk of diabetes, IGT or obesity within the context of acculturation,[56, 65, 88] and fewer yet have demonstrated that consumption of particular food groups may be related (positively or negatively) to the risk of diabetes, IGT or obesity within a rapidly acculturating First Nations community.[48, 87, 89]

Despite these hypotheses and the belief that diabetes risk is associated with a particular dietary pattern,[56] the present study observed virtually no statistically significant association between the consumption of a modern diet and diabetes, IGT or obesity in Sandy Lake, Ontario. With the exception of a few associations which did not reach a conventional level of statistical significance, the risk of diabetes remained unchanged with increased consumption of three food groups (traditional, healthy, or junk foods) which were found to characterize the local diet. These findings were unexpected given the close relationship between diet and obesity, in particular since obesity is a known risk factor for diabetes.[6, 56]

The lack of significant findings was unexpected in light of the - albeit limited - evidence from past studies and what is known about diet and its effect on health. For example, the finding that consuming traditional food had no association with any of the disease outcomes was a surprise given the study undertaken by O'Dea in Australia of a group of Aborigines with diabetes who were exposed to a traditional diet for a 3 month period.[85] In contrast to the typical urban diets, the composition of the diet during this temporary 'reversion to a traditional lifestyle' was made up of < 10% carbohydrates. This dramatic change in the composition of the diet demonstrated a positive effect on blood glucose control. In addition, it was believed that the primary mediating factor was the positive influence that the traditional diet had on the participants' weight (participants lost an average of 8kg over the 3 month study period).[85]

Critics of O'Dea's study could point out that any change in general activity could have resulted in improved insulin sensitivity due to increased physical endeavors (hunting and fishing), however, the authors found that the level of physical activity in the urban setting was not dramatically different from that of the pre-study period. Therefore, the authors reported that "the level of physical activity during the study period was not particularly high...there was no correlation between level of physical activity and improvement in any of the metabolic parameters."[85] The findings from the present analysis with regard to traditional food consumption are therefore unable to confirm the findings from Australia. Consumption of high amounts of traditional foods in Sandy Lake, Ontario displayed no protection from diabetes or IGT and only a statistically insignificant association with obesity. The effects of physical activity will be considered under a later section.

Any inconsistency between the Sandy Lake and Australian studies may have also been the result of differences between communities in the extent to which traditional foods were used. Gittelsohn,[48] who conducted the ethnographic analysis which preceded the SLHDP, found that traditional foods were eaten infrequently in conjunction with traditional activities and that a high level of traditional food consumption did not suggest that other traditional customs were commonly practiced.[48, 72] In Sandy Lake, game-hunting and fishing have become uncommon primary methods of gathering food, and bush meats and wild foods are most commonly consumed during frequent big catered community feasts.[48] Therefore, highly active lifestyles may be more uncommon in Sandy Lake and the consumption of traditional foods may not necessarily be associated with a highly active traditional lifestyle as is the case in the Australian study. For the purposes of this study, a high consumption of traditional food does not represent an active lifestyle or suggest a large extent of traditional customs in one's lifestyle.

Another unexpected observation from the present analysis was that no association was found between the risk of diabetes, IGT or obesity and a high consumption of junk food. This result was surprising for the following reasons: a) the diet of this native community has changed dramatically over the past 50 years and this change has closely followed the rise in the prevalence of diabetes. As the ethnographic analysis observed, community members believed that the incorporation of the 'white man's diet' into native culture was linked to this rise in diabetes and obesity;[72] b) The junk food subscale was made up of items high in simple sugars, high in fat and low in dietary fiber, which according to Trowell's hypothesis, is associated with diabetes in susceptible populations.[86]

The association between diabetes and dietary-fiber depleted foods was confirmed by Wolever and colleagues in an analysis of the Sandy Lake data. They determined that a 1 standard deviation increase in fiber intake reduced the risk of having diabetes by 39%.[87] The main difference between the dietary analysis from the present study and that of Wolever and colleagues was the choice of instrument used as well as the inclusion / exclusion criteria. Wolever's analysis used data from the 24-hour dietary recall while the present study used Food Frequency data exclusively. This difference in data collection may account for the difference in findings: a 24-hour recall survey does not reflect an individual's long-term intake since food intake may vary widely from day-to-day; in contrast, a Food Frequency questionnaire is designed to detect such long-term dietary patterns.[90] In this case, it was advantageous to sacrifice the precise measurement of food intake from a single day for the more crude measurement of diet over an extended period of time in order to gauge ordinary dietary habits. Furthermore, if significant associations were to be observed, the real effects were likely to be even stronger than those observed due to the difficulty in obtaining accurate dietary intake measurements and therefore a dilution of the actual associations.[87]

Gittelsohn and colleagues also used the Food Frequency questionnaire data in order to analyze dietary patterns and create a subscale of scores on various food groups. However, unlike the present analysis, Gittelsohn and colleagues' measure of junk food consumption demonstrated an association with both diabetes and obesity.[48] These findings provided evidence that the Food Frequency instrument was a valid measure since it was able to detect statistically significant associations with diabetes and various food groupings.

The lack of similarity in the findings between Gittelsohn's study and the present analysis can be summarized as follows: First, Gittelsohn used data from the entire community for factor analysis of the dietary data in order to identify food groupings (N = 728 participants) and the results produced seven dietary scales.[48] The present analysis used only adults (n = 411) and factor analysis of the Food Frequency data resulted in three dietary scales. Second, although Gittelsohn and colleagues acknowledged that there was a difference in the diet reported between those newly diagnosed and those previously diagnosed with diabetes, both groups were included in their analysis.[48] The present analysis excluded people with previously diagnosed diabetes due to the potential for over-reporting positive lifestyle behaviors associated with good eating habits. The exclusion of previously diagnosed participants from the present analysis accounted for a great deal of the difference reported between these two studies. This was confirmed by Gittelsohn and colleagues who suggested that when previously diagnosed individuals were excluded from the 'case group', no significant findings were observed.[48] This was contrary to what was expected by these researchers since they acknowledged that those aware of their condition (those who had been previously diagnosed with diabetes) were more likely to report altered diets and therefore have less strong associations with diabetes, IGT or obesity.[48]

The lack of consistency between previous analyses of diet in Sandy Lake and the results of the present factor analyses of the dietary data, together with the inability to confirm the findings of previous studies of diabetes and dietary intake, suggest that the factor analyses of the dietary data have not added to the understanding of the association between diet and the prevalence of diabetes in other communities. This indicates that using dietary factors may not be the most effective way to look at associations between diet and disease. It appears preferable to directly study associations between disease and food nutrients for which there is a prior biologic hypothesis predicting an association between diet and disease. Factor analysis was helpful as a method to better understand which foods were eaten together in this cultural setting. Further investigation into food nutrients with known biological associations with disease may use this knowledge of food groupings to help assess the risk of consuming foods commonly eaten with those with known biological importance to the disease process.

4.1.2 Serum Lipids and Diet in Sandy Lake

Prior to the study by Hodge and colleagues in Papua New Guinea, research showed uniformly low cholesterol levels among the Melanesian population.[56] However, after investigating three communities with varying levels of urbanization, Hodge reported that low blood cholesterol levels were no longer protecting the Melanesian population from coronary heart disease. It was found that the prevalence of dyslipidemia was increased, as was the rate of heart disease.[56] This group also found that higher fat intake was associated with increased blood lipid levels and that the highest fat intake was found among the most acculturated community.[57, 88] These findings were of interest because like the communities in Papua New Guinea, the Oji-Cree community of Sandy Lake had historically been thought to have a good community lipid profile.
Hodge also found that the rural participants were leaner and had a lower prevalence of diabetes than their urban counterparts, a finding which indicated that some protection was associated with the traditional lifestyle which was characteristic of the highland natives of New Guinea.[88] It was suspected that within the Sandy Lake sample, a similar finding would emerge; those who consumed a diet high in traditional food would demonstrate protection from dyslipidemia due to the hypothesis that a traditional diet represents one aspect of a traditional lifestyle which includes a high level of physical activity. Therefore, it was hypothesized that a diet high in traditional food would be associated with low levels of dyslipedemia while a diet high in junk food would be associated with high levels of dyslipidemia.

The current investigation found that there was virtually no statistically significant relationship between any of the food groupings and blood lipid levels; the consumption of a diet high in junk food or a traditional diet displayed no association with any of LDL, HDL, total cholesterol or triglyceride levels even once adjusted for age and sex. Therefore, while it was possible that the Sandy Lake community had not been exposed to the high fat diet which characterizes the North American diet for a sufficient amount of time to have had a noticeable effect on blood lipid levels, the current finding indicated that there was no detectable association between scores on the Junk food, Healthy Food or Traditional Food subscales and blood lipid levels. However, until a follow up study of the Sandy Lake community is performed (one is currently being designed), there will be no way to detect whether the lipid profile has remained unchanged with continued exposure to these acculturative processes.

There were other considerations which may account for the discrepancy between the present results and Hodge's findings in Papua New Guinea. First, Hodge and colleagues found correlations between their measures of modernity and diet when dietary intake was measured by total caloric consumption.[89] In the current study, there was no way to estimate the average daily caloric intake and furthermore, no adjustment for total energy intake was made since there was evidence that adjusting for energy intake within the dietary data in Sandy Lake would not have improved the significance of the findings; Wolever used two methods to adjust the SLHDP 24-hour dietary recall data by energy intake and found that neither resulted in a significant difference from the unadjusted analysis.[87] In theory, knowledge of total calorie intake would have had its advantages for determining the possibility of a genetic susceptibility of a community if consumption of a high calorie diet increased the risk of diabetes, IGT or obesity regardless of the foods being eaten. As proposed by Neel, this genetic characteristic would have once offered a survival advantage, but is now associated with diseases of disordered metabolism such as diabetes, IGT and obesity.[61] However, without an estimate of long term energy intake and with the evidence from Wolever's analysis, no adjustment was made for energy intake within the Food Frequency data.

Thus, as was considered in relation to the lack of association between the three dietary factors and diabetes, IGT and obesity, it may be that the three dietary factors identified by factor analysis are not those which are biologically important determinants of blood lipid levels, such as energy, fat or fiber intake.

4.1.3 Diet as Risk Factor

It is widely believed that the high rate of diabetes in First Nations communities around the world are due to recent lifestyle changes associated with diet and physical activity acting on a susceptible genotype.[43, 87] Dietary intake may be considered a direct biologically active risk factor for obesity due to its causative, physiological association with weight gain. A directly bio-active risk factor like food intake is typically correlated with the prevalence of obesity and obesity is in time related to the risk of diabetes and IGT. Therefore, for this study, obesity was not considered a confounding variable, but rather a mediating factor on the biological pathway between diet and diabetes. Obesity was therefore treated as a health outcome and not as a confounding variable and dietary data were not stratified by the presence of obesity.

Non-biologically active risk factors are aspects of the social environment which may be associated with biologically active factors, but without any direct biological activity themselves. For example, within developing countries, obesity has been associated with factors related to modernity such as education, sophistication of housing, type of employment, and duration of residence in an urban centre.[88] While there is a considerable body of literature on the relationship between non-biologically active risk factors and the overall prevalence of diabetes or obesity, there are few studies which measured the relationship between biologically active risk factors for diabetes or obesity within the context of a First Nations community. While there may be advantages to studying these 'non-biologically active' cultural determinants of health, ultimately, directly biologically active factors are important to analyze in order to plan intervention strategies and wellness programs.

However, there are relatively few studies which have looked at diet – arguably the most directly 'biologically active' of the three elements of acculturation used in this analysis – among First Nations communities and how changes to a community's diet have effected the risk of diabetes. The available research has found some evidence of associations between diet and diabetes, however, the findings have been inconsistent and difficult to generalize across populations. For example, dietary fiber and protein intake was inversely associated with the risk of newly diagnosed diabetes in Sandy Lake.[87] In Australian Aborigines it was found that decreased total energy intake was associated with improved blood glucose control.[85] In comparison, the SLHDP determined that there were no significant effects on risk from energy, fat, starch or simple sugars.[87]

While no biologically active or non-biologically active risk factor has been universally linked to diabetes or obesity in the First Nations communities studied, there are similarities among these communities which allow for certain generalizations. For example, as described in Chapter 1, the Pima Indians have the highest recorded community rate of diabetes in the world yet they consume a diet similar to that of the general population of the U.S.[47] Similarly, the people of Sandy Lake no longer rely on game hunting or fishing for their primary source of food and with the availability of store-bought foods, the diet available to this isolated community is similar to that of the average Canadian in an urban setting. However, like the Pimas, the Oji-Cree of Sandy Lake display a rate of diabetes which surpasses the national average rate of diabetes in the Canadian population. This implies that aspects of the diet may be less important than the change in the culture with regard to the food gathering processes or perhaps that genetic susceptibility, as described by Neel, increases the risk of diabetes among Native North Americans when they are exposed to the Western diet.[61]

The hypotheses put forth in the present study claimed that a 'modern diet' was contributing to the rise in the risk of diabetes in Sandy Lake, Ontario. This was either due to the actual consumption of various 'high risk' foods, or a low consumption of certain 'protective' foods. However, the risk of diabetes, IGT or obesity was not significantly increased with increasing consumption of a modern diet; similarly, no food consistently demonstrated a protective effect. It has been hypothesized by Dressler that perhaps particular aspects of the diet may have been less important than the rapid change in diet and the secondary consequences as a result of that shift in the food supply.[56] In other words, since hunting and actively gathering food has become obsolete, the lack of any necessary physical activity for survival, acting on a susceptible genotype, may be contributing to the increase in the prevalence of diabetes and obesity.

4.2 Labor-Saving Devices (Household Economic Items)

4.2.1 WAS THE POSSESSION OF MODERN HOUSEHOLD ITEMS A RISK FACTOR FOR DIABETES?

The Household Items questionnaire from the Sandy Lake Health and Diabetes Project was designed to assess socio-economic status in the absence of standard indicators (e.g.: salary scale, home ownership).[70] For the present study, the Household Items questionnaire provided information on the extent to which individuals accumulated labor-saving devices for either traditional or modern activities. It was hypothesized that the possession of modern items was associated with an increased risk of diabetes and obesity due to the likelihood that such items reduced physical activity and promoted sedentism. In addition, it was thought that the possession of traditional items indicated an adherence to a traditional way of life and subsequently that individuals reporting a high degree of these items had a greater amount of physical activity in their lives due to the heightened level of activity which characterized the traditional Oji-Cree existence relative to the typical North American existence.

It was hypothesized that a greater risk of diabetes and obesity would be associated with the modern items subscale and that a decreased risk of diabetes and obesity would be associated with reporting greater amounts of traditional possessions in the home.

Some results from this section were unexpected but not unexplainable. First, as hypothesized, possession of modern items was associated with an increased risk of diabetes, IGT and obesity, a finding which supports the theory that these items may reduce physical activity and ultimately, that they may lead to diseases of disordered metabolism such as diabetes and obesity. However, the possession of items for traditional activities was also found to increase the risk of diabetes, IGT and obesity. The evidence strongly suggested that a high score on either the modern or traditional scale increased the risk of diabetes, IGT and obesity.

This led to a reinterpretation of the underlying factors which made up each of the scales. Originally, it was hypothesized that the items on each scale represented the possession of either traditional or modern items. However, on second evaluation, it was realized that the items from each scale were all modern items; they were not handcrafted, passed down from generation to generation or the result of a cultural or historical transition. In fact, the majority of the items on both the modern and traditional subscales were labor-saving devices designed to streamline the activity for which they were designed. These labor-saving devices likely reduced the amount of physical activity needed to perform each task, whether the task was traditional or modern. Viewed in this light, it was therefore reasonable that scores from both the traditional and modern items subscales were associated with an increased risk of health outcomes like diabetes, IGT and obesity.

The possession of labor-saving devices (for any purpose) could potentially have resulted in a reduction in daily physical activity. Furthermore, within populations believed to be susceptible to diabetes and obesity any variable that may be associated with a reduction in physical activity may have led to an increased risk of obesity. Therefore, a decrease in physical activity could lead to an increase in the community-wide prevalence of diabetes and IGT.

Results from three separate studies of First Nations communities undergoing rapid acculturation have documented the influence of obesity on the prevalence of diabetes. First, Hodge's study of Melanesians of Papua New Guinea showed that modernity was characterized by reduced physical activity and that modernization was a significant predictor of obesity and diabetes.[88] Like the SLHDP, Hodge's analysis was a cross sectional study which attempted to identify risk factors for diabetes through a community wide questionnaire. Biological measures such as blood glucose and serum lipids were taken during the individual level interviews. However, unlike the current analysis, the level of modernity was defined by community affiliation; four separate communities were surveyed and the degree of modernization varied among the villages. The degree of modernization was based on each village's proximity to the coastline.[88]

O'Dea tested the concept of implementing a traditional lifestyle as a means to improve diabetes control in a study of Australian Aborigines. In this study, a prospective design was used to show that the implementation of a physically active lifestyle improved the measures of diabetes control.[85] O'Dea's study concluded that a health strategy which promoted an active lifestyle could protect against obesity as well as new cases of diabetes.[85] However, O'Dea's study could be criticized for its assertions that reversing the urbanization process should improve all aspects of diabetes.[85] While it was important to promote the benefit of leading a traditional lifestyle, O'Dea's analysis failed to show how risk of disease may vary based on cultural affiliation. The actual improvement in blood glucose control was likely mediated by the increase in physical activity which resulted in weight loss.

O'Dea succeeded in showing that by increasing the level of physical activity, an individual's blood glucose can be improved and sometimes kept within normal limits. However, this study did not measure the actual activity in the community nor did it determine whether those who engaged in regular physical activity had a lower prevalence of diabetes or improved blood glucose control. This was an important study because it demonstrated that regular physical activity could improve the symptoms of diabetes within a community believed to be susceptible to this disease. However, the study was unable to demonstrate how variables in the community can be useful in determining an individual's level of activity or how real lifestyle activities in the community can protect against diabetes.

In contrast, the present study attempted to draw parallels between personal possessions and the risk of diabetes, IGT and obesity. Most importantly, this analysis demonstrated that possessions with separate cultural affiliations predicted risks for diabetes, and obesity. In combination with O'Dea's findings, these results suggest that there may have been a difference in level of activity between those with different cultural affiliations and more importantly, that health outcomes vary based on those affiliations. This is relevant for designing culturally specific health promotion strategies in First Nations communities.

Dressler's study of the Mississippi Choctaw was similar to the present analysis as it used a lifestyle questionnaire to determine the level of acculturation and results were related to health measures made concurrently. In this community-wide survey, a 25-item 'style of life' scale assessed (among other variables) the accumulation of consumer goods. A positive association was found between average glucose levels and scores on the style of life scale and a negative association was observed with physical activity.[56] These results confirmed Dressler's hypothesis that acculturation, a concept which included such sociocultural factors as the accumulation of consumer goods, was associated with an increased risk of diabetes and obesity within the Mississippi Choctaw community.[56]

Dressler's study demonstrated that an important association might exist between modern items within a once traditional community and diseases of disordered metabolism such as diabetes, IGT and obesity. The present study took this one step further by demonstrating that the possession of labor-saving devices increases the risk for these diseases, but also, that labor-saving devices of any kind, regardless of their cultural affiliation, may increase the risk for diabetes, IGT and obesity.

4.2.2 WAS THE HOUSEHOLD ITEMS VARIABLE A PROXY MEASURE FOR PHYSICAL ACTIVITY?

The association between labor-saving devices and diseases of disordered metabolism was likely mediated via the influence of labor-saving devices on reducing the average physical activity required among the community members. For example, the traditional scale included such items as a motor for a boat and skidoos. Each of these items represents the modernization of the process for fishing and for hunting. While the cardiovascular benefit of exercise in general is well documented, it can be assumed that those who have come to rely on labor-saving devices for simple daily chores do not benefit from any physical endeavors usually associated with these activities. Moreover, within a community susceptible to obesity and diabetes, it can be assumed that any physical activity could be beneficial on weight and diabetes management.

Since the SLHDP gathered data on individual fitness level through the sub-maximal step test, determining the association between this variable and the subscale scores for labor-saving devices demonstrated how well the Household Items scale scores estimated an individual's level of physical fitness. In fact, no association was observed between Household Item scores and the VO_{2max} scores from the submaximal step test.

Initially, this may suggest that an association between possessions of labor-saving devices and fitness level did not exist since those who rely heavily on these items should also demonstrate a low level of physical activity in their lifestyle. However, the lack of association observed between these two variables may be explained in part by methodological considerations. First, the small sample size and strict inclusion criteria from the step test resulted in a group which was younger and healthier than the sample that participated in the household survey; only 66% of those who completed the household survey also completed the step test. A further complication stems from the fact that only one individual completed the household questionnaire for an entire household. This implies that an individual's level of exposure was defined by the person (usually a parent) who completed the survey. The predominance of young adults still living in their parents homes complicates this issue since this group would be most likely to have traditional or modern affiliations which were different from those of their parents. It had been suspected that a strong cohort effect was evident in the community in which the older generations were more closely tied to the traditional customs while younger adults held less affiliation with the traditional way of life. Therefore, any reported possessions that had a traditional affiliation may not have actually been used by the younger generations, but merely been in the home and reported by the elders. This measurement error would have biased the results towards the null so that those reporting a high amount of traditional possessions would have showed little difference from those reporting high modern possessions.

Further research would be needed to determine whether an individual's use, rather than the household possession of, labor-saving devices either reduces physical activity, or is related to risk of diabetes, IGT or obesity.

4.2.3 BIO-ACTIVITY AMONG THE HOUSEHOLD ITEMS VARIABLE

The Household Items variable was not considered a purely biologically active variable since it was not proven to be a proxy measure of level of physical fitness. In addition, there is no evidence to suggest that variables measuring household possessions are part of the pathophysiology of diabetes. Despite this, the Household Items subscale score was a significant risk factor all three diseases studied; a clear association was found between the accumulation of modern items for either traditional or modern activities and diabetes, IGT and obesity. Although this variable showed no statistical evidence of approximating a real biological variable (VO_{2max}), and thus, may not be considered a completely bio-active variable, the sampling biases may have made the detection of any relationship between household items and level of physical fitness (VO_{2max}) unlikely.

In part, the possessions variable may be considered an indirectly bio-active variable due to its likely – yet, unproven – association with reduced physical activity. However, more likely, it may be an indicator of culture in the home, which does predict risk, independent of the 'bio-active' component of physical activity.

4.3 Language Subscale

4.3.1 LANGUAGE: A PURELY CULTURAL VARIABLE

The studies of diet demonstrated that pre-existing evidence of a direct biologic link to diabetes was not sufficient to demonstrate strong associations with the prevalence of diabetes, IGT or obesity in this community. The results from the Language section suggest that a purely cultural variable, despite being entirely non-biologic in nature, may be important in assessing the risk of diabetes in the community.

Information obtained from the Language section was quite different from that derived from either of the previous sections on Food Frequency or Household Items. First, every member of the community who took part in the SLHDP completed the Language questionnaire, and therefore, data from 100% of the respondents was included in this section of the analysis.

Second, 'Language Proficiency' was a purely cultural variable with no known biologic association with the prevalence of diabetes or obesity. In contrast, both food consumption and labor-saving devices may impact health directly (or indirectly, at least). Previous studies have shown that a specific diet as well as the extent of activity in an individual's lifestyle may each impact long term health. This includes an increased risk of diabetes and obesity, particularly within a community believed to be susceptible to diseases of disordered metabolism like Sandy Lake, Ontario.[3, 6, 8, 15]

Third, the Language subscale was composed of purely North American items on factor 1 and purely Traditional items on factor 2. This separation of items along cultural lines was not demonstrated in either of the previous two sections. For example, the dietary subscale included a third factor called "Healthy Foods" which were not classifiable as one of either 'modern' or 'traditional' items; the Household Items factors were all considered 'modern' items and only the activity for which they were designed were distinguished as either traditional or modern. Only the Language items were categorized as 'Oji-Cree' or 'English' – corresponding directly to traditional and modern culture. These two distinct factors helped to demonstrate the risk of diabetes, IGT or obesity as they related to cultural affiliation in the community.

Cultural affiliation has been defined in previous studies using sociocultural parameters such as household possessions, exposure to mass media as well as other proxy measures which represent assimilation into the mainstream – or host – culture.[2, 9, 15] However, no study was found in the literature which described the association between language proficiency and diseases such as diabetes, IGT or obesity in a First Nations community. In fact, within the existing literature, little emphasis has been placed on language ability and no study has been found in the medical literature which measured language proficiency as an indicator of cultural affiliation. Therefore, while the present study provided an estimate of the risk associated with an individual's proficiency in either the English or Oji-Cree language, it suggests that language is an important indicator of cultural affiliation and may therefore be used in future studies to estimate the association between disease risk and culture in First Nations communities.

The findings from the Language section also helped to confirm the initial hypothesis that exposure to Oji-Cree culture (as measured by the Oji-Cree Education subscale) was associated with a reduced risk of diabetes. Although not reaching a conventional level of statistical significance, it is important to note that both IGT and obesity demonstrated negative associations with the Oji-Cree subscale as well. If, in fact, some element of the Oji-Cree culture provides protection from diabetes and obesity and this protection is related to the Oji-Cree Education factor, then it is reasonable to assume that the utilization of traditional teachings and the promotion of an Oji-Cree Education could serve as a focal point to health promotion strategies in the community.

Correspondingly, the positive association between diabetes and obesity and the English Proficiency scores provided evidence that an individual's cultural affiliation with the North American culture, in some way, increased the risk of diabetes, IGT and obesity. While the results from this section were not all statistically significant, the association between diabetes and English Proficiency scores were as strong as the association between Oji-Cree Education and diabetes, yet, in the opposite direction. This finding would further support strong Oji-Cree content in the school curricula, as well as community-wide promotion of traditional activities. Although the language variable has no known biologic associations with disease, increasing children's exposure to traditional language and customs would hopefully lead to the promotion of other, biologically active aspects of life which may promote health and well-being.

4.3.2 EVIDENCE OF A STRONG COHORT EFFECT

The lack of statistically significant findings within the language results might be explained by reasons other than true null effects. The associations demonstrated between scores on the Language subscales and diabetes, IGT and obesity may have been 'diluted' by the confounding effect of age and its relationship with diabetes prevalence in the community. In particular, a strong cohort effect was evident in relation to three age groups in the community: For example, those over 50 years of age were not educated in residential schools and maintained strong traditional connections. This group was most likely to be affected by diabetes and IGT because of their age.[2] Those 30 – 50 years old had little Oji-Cree education due to the implementation of residential schooling during the middle part of the 20^{th} century. However, this group had higher rates of diabetes and obesity during the SLHDP relative to the younger groups in the community because of their advanced age – the rates of diabetes were found to increase steadily with age in both men and women in all age groups.[2] Finally, those under age 30 were the most exposed to and influenced by Western media. This group had received an education most similar to that of a typical North American curriculum – with only minimal Oji-Cree content. However, despite having the strongest socio-cultural similarities with their 'North American' counterparts, the prevalence of diabetes and IGT was more than double the national rate for this same group (10.0% in Sandy Lake compared to 4% in the general Canadian population in this same cohort).[91] Therefore, the presence of high rates of diabetes, IGT and obesity and the lack of a normal distribution of scores from the Language subscale within young (<30) and old age groups (> 50), might have resulted in a dilution of the measurable effect.

Evidence of a strong age-related effect can be seen in the dramatic difference in the risk of disease before and after the Language data were adjusted for age and sex. This is demonstrated from the table of adjusted odds ratios found in Chapter 3. Generally, the adjusted odds ratios for the Language subscale scores were less strong and opposite in direction from the unadjusted rates. This might be indicative of a very strong effect from age (and from sex). This strong confounding effect might be improved with a stratified analysis of each age group, yet the small sample size would preclude separate factor analyses on each of these subgroups.

4.3.3 Hatcher's Criteria: An Exception to the Rules

The reliability of a factor describes the consistency of the scores that are obtained on the observed variables. An instrument is said to be reliable if it is shown to provide consistent scores upon repeated administration. Coefficient-α will be high when the items that constitute a particular scale are highly correlated with one another and the general rule of thumb is that a scale requires at least 3 items in order to attain a standard level of reliability. This is a component of 'Hatcher's Criteria' for simple structure.[77]

In the case of the Language subscale, the criterion that at least 3 items be included in a given factor was not satisfied. Only 2 items were included in the Oji-Cree Education subscale. However, the primary purpose of including at least 3 items on a given factor is to ensure that a sufficient level of scale reliability is attained.[77] In fact, this factor scored higher on the coefficient-α test than any other subscale in the analysis.

Thus, despite its apparent breach of Hatcher's criteria for simple structure, the Oji-Cree Education subscale was considered a reliable factor for the following reasons: first, 'Language' was easily measured from the questionnaire and data was collected on 100% of the SLHDP participants. Therefore, the raw data was complete and likely to display little error. Second, the finding through reliability testing showed that the subscale maintained an extremely high degree of accuracy (coefficient-α determination). Last, random error within the Language data would have reduced the measured association between the Language scales and disease toward the null; therefore, the true association would likely be even stronger than that observed here. In fact, the actual association between Oji-Cree Education and diabetes might be even stronger than what was reported in Chapter 3.

Therefore, using theoretical rationale and intuitive sense, the Language subscale that measured Oji-Cree Education was considered reliable, justifying its use in further analyses.

4.4 Sample Size and Power Calculation

The sample size and power calculations for each subscale and factor are included in Table 6.7.1 in the Appendix and frequency values may be found in the frequency distribution tables from Chapter 3. In each case, a power of 80% and a significance level of 0.05 were used to determine the minimum number of cases needed to detect a significant difference between the cases and non-cases. The Schlesselman[95] formula for a case-control design was used to detect the number of cases needed. More information on the sample size determination and power calculation may be found in the Appendix, section 6.7.

The sample size estimates for the number of cases needed to detect as significant differences between cases and non-cases on the magnitude actually observed in the study explained why there were mostly null results for the food frequency subscales. There were insufficient cases to detect a significant difference between all of the associations.

The number of cases included in the Household Items section was sufficient to detect a real association with a power of 80% and a significance level of 0.05 for most of the associations. The modern items subscale demonstrated a significant effect in each unadjusted association and in all but the IGT group, there was a sufficient number of cases to detect a real association. For the traditional subscale, the null findings are explained by an insufficient number of cases to satisfy a power of 80% and significance level of 0.05 for the magnitude of effect involved. For the diabetes or IGT group, 144 cases were needed and only 92 were available; for the diabetes group, 294 cases were needed and only 38 were available; for the IGT group, 98 were required and only 54 participated and for the obesity group there were only 249 available for a group which required 8,508 participants.

For the Language Data, there were insufficient cases to detect a real association for all categorical analyses. This helped to explain why only one odds ratios was significant at the 0.05 level from this analysis. The significant value demonstrated a strong protective effect from Oji-Cree language (OR = 0.40). The remaining findings from this section, though statistically non significant and though not reaching a sufficient level of power did still suggest a protective effect from the other disease outcomes as a result of reporting a high Oji-Cree Education score.

4.5 Future Research and Lessons Learned

4.5.1 Viability of the Biocultural Model of Disease in First Nations

Within a First Nations community, culture and tradition pervade the practice of medicine and the promotion of health as well as the idealistic vision of a healthy lifestyle. Only an integrative approach which synthesizes biomedicine with traditional beliefs of disease prevention and management can produce effective wellness strategies.

The approach used in this study combined the biology of disease with cultural risk factors. This is not a new strategy; this was the foundation of the thrifty genotype hypothesis proposed by Neel in 1962. His belief was that gene(s) once advantageous to survival now render its host susceptible to obesity and diabetes due to the surplus of food characteristic of present day North America.[59, 61] Neel's model provided a framework within which cultural risk factors acted on a genetic susceptibility to disease.

In this study, several findings contributed to the framework described by Neel. First, despite a biologic link to diabetes and obesity, the finding that food intake was a poor predictor of disease outcome suggested that the prevalence of diabetes and obesity in the community is affected by more than diet alone. This may point to a genetic susceptibility to disease or to some other risk factor. Second, the existence of an a purely non-biological risk factor for diabetes – language proficiency – lends evidence to the theory that cultural risk factors can independently or via a mediating biologic risk factor, affect disease risk.

4.5.2 The Biocultural Approach: Where Epidemiology and Anthropology Meet

Interest in integrating the biological and cultural perspective has developed due to the belief that both biology and culture may each contribute to disease causation.[37, 63] Integrating these two perspectives requires an in depth knowledge of history and local context as well as an understanding that the consequences of illness are not limited to physical symptoms but may also shape the cultural response to disease within a community.[63] In addition, a biocultural synthesis requires that basic epidemiologic principles be adhered to while considering social, cultural, political and economic factors which may contribute to a population's susceptibility or the cause of disease.

Fundamental differences exist between the basic principles of these two disciplines. For example, the anthropologic literature suggests that any factor that influences the ability of a population to adapt successfully to their environment is a potential source of disease.[63] According to the epidemiologic perspective, a causal inference can only be made when a gradient of response is preceded by an exposure which produces a strong and confirmed association which is biologically plausible.[94] In the anthropologic literature, this approach to disease causation has been criticized for being overly conservative and for failing to fully or accurately realize that culture is part of the environment which interacts with the disease process.[63] Furthermore, an association which does not appear to be biologically credible at one time may eventually prove to be so.

Indeed the epidemiologic community recognizes the potential for a seemingly implausible association representing an expansion of scientific knowledge.[92] Important associations once considered implausible include the cigarette smoking–lung cancer association as well as the circumcision–cervical cancer relationship, to name but a few. Identification of these associations has helped to attain a greater understanding of these diseases as well as the risk factors which precede them. However, not all biologically implausible associations ought to be considered as potentially significant findings in hiding. Lilienfeld suggested that only if statistical associations are found in the absence of a proven, physiological mechanisms, should further investigations be carried out [92]

This study is an example of a successful biocultural synthesis; the analysis was oriented to the ethnography of the community while respecting the scientific methods of epidemiology and biostatistics. Moreover, the significant associations found among the Language scores and the Household Items scores represents an expansion of knowledge in the area of cultural determinants of health. With future research, groups of biological risk factors may demonstrate strong associations with these cultural risk factors and help to identify a model for disease risk in this community.

2.5.3 FUTURE RESEARCH

The majority of the literature which has reported acculturation has done so as a 'broad concept', describing it as a social phenomenon. In some studies, physical, economic and sociocultural aspects of culture have been combined in various ways to quantify a single risk factor, called 'acculturation'.

This study has described elements of culture which are undergoing the acculturation process. Each was viewed as separate risk factors and subsequently tested for an association with diabetes, IGT and obesity.

In this study, an attempt was made to determine if the Household Items scores were actually a proxy measure for physical activity. The possession of labor-saving devices likely reduced the amount of physical activity, however, an accurate measure of physical activity was not available so this assertion could not be tested. It would be important to identify how non-biological risk factors like Household Items and Language relate to risk factors with known biological associations with diabetes and obesity. This research helped to identify that associations may exist between non-biological risk factors associated with culture and the prevalence of diabetes. Future research should be directed at identifying through what biological channels these cultural risk factors are expressed.

In particular, through follow up studies from the original SLHDP, surveys should include sections which measure the extent to which individuals take part in traditional activities like hunting, sweat lodges, as well as other Oji-Cree behaviors.

Future research should also be directed at combining the results of this study with investigations of genetic mutations in Sandy Lake. In particular, it was identified that a rare variant in a gene encoding HNF-1α was strongly associated with type 2 diabetes in Sandy Lake.[93] It would be extremely interesting to investigate a possible association between scores from these subscales and the presence of this mutation.[93] This would provide a rare opportunity to examine associations between risk factors associated with the changing environment and genes identifying individuals' susceptibility to disease, as well as to explore the possibility of interactions between cultural and genetic factors.

5.0 Works Cited

1. Bennett PH, Burch TA, Miller M. Diabetes mellitus in American (Pima) Indians. *Lancet* 1971; ii: 125 – 128.

2. Harris SB, Gittelsohn J, Hanley A, Barnie A, Wolever TMS, Gao J, Logan A, Zinman B. The prevalence of NIDDM and associated risk factors in native Canadians. *Diabetes Care* 1997; 20(1): 185 – 187.

3. Newman, M. Aboriginal new world epidemiology and medical care, and the impact of old world disease imports. *Am J Physical Anthropol* 1976: (45) 667 – 72.

4. Snow DR, Lanphear KM. European contact and Indian depopulation in the Northeast: The timing of the first epidemics. *Ethnohistory* 1988: (35)15 – 33.

5. Kunitz SJ. Disease and Social Diversity: the European impact on the health of non-Europeans. New York, Oxford University Press, 1994.

6. Gohdes D, Diabetes in North American Indians and Alaska natives, in National Diabetes Data Group: Diabetes in America (2nd ed.) Washington, DC, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 1995, pp 683 – 702. (NIH Pub. No. 95 – 1468).

7. Dagogo-Jack S, Santiago JV. Pathophysiology of type 2 diabetes and modes of action of therapeutic interventions. *Arch Int Med* 1997; 157 (16): 1802 – 1817.

8. The Diabetes Control and Complications Trial Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993; 329: 977 – 986.

9. Lawson ML, Gerstein HC, Tsui E, Zinman B. Effect of intensive therapy on early macrovascular disease in young individuals with type 1 diabetes: a systematic review and meta-analysis. *Diabetes Care* 1999; 22(2S): 35B – 39B.

10. Zimmet P, Kelly West Lecture 1991 Challenges in diabetes epidemiology – from west to the rest. *Diabetes Care* 1992; (15) 232 – 252.

11. Atkinson MA, Maclaren NK. Mechanism of disease: the pathogenesis of insulin-dependent diabetes mellitus. *N Engl J Med* 1994; 331(21): 1428 – 1436.

12. Orchard TJ. From diagnosis and classification to complications and therapy. *Diabetes Care* 1994; 17: 326 – 338.

13. National Diabetes Data Group: Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* 1979; 28: 1039 – 1057.

14. DeFronzo RA, Ferrannini E. Insulin resistance: a multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipedemia and atherosclerotic cardiovascular disease. *Diabetes Care* 1991; 14(3): 173 – 194.

15. Pontiroli AE, Galli L. Duration of obesity is a risk factor for non-insulin-dependent diabetes mellitus, not for arterial hypertension or for hyperlipidaemia. *Acta Diabetologica.* 1998; 35(3): 130-6.

16. Maggio CA. Pi-Sunyer FX. The prevention and treatment of obesity. Application to type 2 diabetes. *Diabetes Care* 1997; 20(11): 1744-66.

17. Robert HR. Glucose control and insulin resistance in non-insulin-dependent diabetes mellitus. *Ann Intern Med* 1996; 124(1 pt 2): 97 – 103.

18. Karam JH. Reversible insulin resistance in non-insulin-dependent diabetes mellitus. [Review] [24 refs] *Hormone & Metabolic Research.* 1996; 28(9): 440-4..

19. Grundy SM. Dietary therapy in diabetes mellitus: is there a single best diet? *Diabetes Care* 1991; 14: 796 – 801.

20. Zimmet PZ, McCarty DJ, de Courten MP. The global epidemiology of non-insulin dependent diabetes mellitus and the metabolic syndrome. *Journal of Diabetes and its Complications* 1997; 11(2): 60 – 8.

21. Reaven GM. Role of insulin resistance in human disease. *Diabetes* 1988; 37: 1595 – 1607.

22. Haffner SM. The insulin resistance syndrome revisited. *Diabetes Care* 1996; 19(3): 275 – 277.

23. Amos AF, McCarty DJ, Zimmet P. The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabetic Medicine* 1997; 14 Suppl 5: S1 – 85.

24. Tan H, Maclean DR. Epidemiology of diabetes mellitus in Canada. *Clin Invest Med* 1995; 18: 240 – 246.

25. Zargar AH, Wani AI, Masoodi SR, Laway BA, Bashir MI. Mortality in diabetes mellitus--data from a developing region of the world. *Diabetes Research & Clinical Practice* 1999; 43(1): 67-74.

26. O'Rahilly S. Science, medicine, and the future: Non-insulin dependent diabetes mellitus: the gathering storm. *BMJ* 1997; 314(7085): 955 – 959.

27. King H, Rewers M, WHO Ad Hoc Diabetes Reporting Group. Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. *Diabetes Care* 1993; 16: 157 – 177.

28. Vinicor F. Is diabetes a public health disorder? *Diabetes Care* 1994; 17: 22 – 7.

29. Collins VR, Dowse GK, Ram P, Cabealawa S, Zimmet PZ. Non-insulin dependent diabetes and 11-year mortality in Asian Indian and Melanesian Fijians. *Diabetic Medicine* 1995: 13: 125 - 132

30. Martin FIR, Wyatt GB, Griew AR, Haurahelia M, Higginbotham L. Diabetes mellitus in urban and rural communites in Papua New Guinea. Studies of prevalence and plasma insulin. *Diabetologia* 1980; 18: 369 – 374.

31. Wise PH, Edwards FM, Craig RF, Evans B, Murchland JB, Sutherland B, Thomas DW. Diabetes and associated variables in the South Australian Aborigines. *Aust NZ J Med* 1976; 6: 191 – 196.

32. Cameron WI, Moffitt PS, Williams DRR. Diabetes mellitus in the Australian Aborigines of Bourke, New South Wales. *Diabetes Res Clin Pract* 1986; 2: 307 – 314.

33. McKeigue PM, Marmot MG, Syndercombe-Court YD, Cottier DE, Rahman S, Riemersma RA, et al. Diabetes hyperinsulinemia and coronary risk factors in Bangladeshis in East London. *Br Heart J* 1988; 60: 390 – 6.

34. World Health Organization (WHO): Prevention of diabetes mellitus, Geneva, Technical Report Series no. 844, 1994.

35. Jun HS, Bae HY, Lee BR, Koh KS, Kim YS, Lee KW, Kim H, Yoon JW. Pathogenesis of non-insulin-dependent (type II) diabetes mellitus (NIDDM) – genetic predisposition and metabolic abnormalities. *Advanced Drug Delivery Reviews* 1999; 35: 157 – 177.

36. King H, Dowd JE. Primary prevention of type 2 (non-insulin dependent) diabetes mellitus. *Diabetologia* 1990; 33: 3 – 8.

37. Leatherman TL. A biocultural perspective on health and household economy in southern Peru. *Medical Anthropology Quarterly* 1996; 10(4): 476 – 95.

38. Zimmet P, Dowse G, Finch C, Serjeantson S, King H. The epidemiology and natural history of NIDDM – lessons from the South Pacific. *Diab Metab Rev* 1990; 6: 91 – 124.

39. Frisch RE, Wyshak G, Albright TE, Albright NL, Schiff I. Lower prevalence of diabetes in female former college athletes compared with non-athletes. *Diabetes* 1986; 35: 1101 – 1105.

40. Helmrich SP, Ragland DR, Leung RW, Paffenbarger RJ. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med* 1991; 325: 147 – 152.

41. Manson JE, Rimm EB, Stampfer MJ, Colditz GA, Willett WC, Krolewski AS Rosner B, Hennekens CH, Speizer FE. Physical activity and incidence of non-insulin dependent diabetes mellitus in women. *Lancet* 1991; 338: 774 – 778.

42. Daniel M, Gamble D. Diabetes and Canada's aboriginal peoples: the need for primary prevention. *Int J Nurs Stud* 1995; 32(3): 243 – 259.

43. Zimmet, PZ. The pathogenesis and prevention of diabetes in adults. *Diabetes Care* 1995; (18)7: 1050 – 1064.

44. Zinman B, Hanley AJG, Harris SB, Kwan J, Fantus IG. Circulating tumor necrosis factor – α concentrations in a native Canadian population with high rates of type 2 diabetes mellitus. *Jourrnal of Clinical Endocrinology and Metabolism* 1999; 84(1): 272 – 278.

45. Barrett-Connor E. Epidemiology, obesity, and non-insulin dependent diabetes mellitus. *Epidemiol Rev* 1989; 11: 172 – 181.

46. UK Prospective Diabetes Study Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes. *Lancet* 1998; 352: 854 – 865.

47. Carter JS, Pugh JA, Monterrosa A. Non-insulin dependent diabetes mellitus in minorities in the United States. *Ann Intern Med* 1996; 125(3): 221 – 232.

48. Gittelsohn J, Wolever TMS, Harris SB, Harris-Giraldo R, Hanley JG, Zinman B. Specific patterns of food consumption and preparation are associated with diabetes and obesity in a native Canadian community. *Journal of Nutrition* 1997;541 – 547.

49. Knowler WC, Pettitt Savage PJ, Bennett PH. Diabetes incidence in Pima Indians: contributions of obesity and parental diabetes. *Am J Epidemiol* 1981; 113: 144 – 156.

50. Ravussin E, Valencia ME, Esparza J, Bennett PH, Shulz LO. Effects of a traditional lifestyle on obesity in Pima Indians. *Diabetes Care* 1994; 17(9): 1067 – 1074.

51. Knowler WC, Pettitt DJ, Saad MF, Bennett PH. Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis. *Diab Metab Rev* 1990; 6: 1 – 27.

52. Barnett AG, Eff C, Leslie RDG, Pyke D. Diabetes in identical male twins. A Study of 200 pairs. *Diabetologia* 1981; 20: 87 – 93.

53. Marshall JA, Hamman RF, Baxter J, Mayer EJ, Fulton DL, Orleans M, Rewers M, Jones RH. Ethnic differences in risk factors with the prevalence of non-insulin dependent diabetes mellitus. The San Luis Valley Diabetes Study. *Am J Epidemiol* 1993; 137: 707 – 718.

54. Haffner SM, Hazuda HP, Mitchell BD, Patterson JK, Stern MP. Increased incidence of type II diabetes mellitus in Mexican Americans. *Diabetes Care* 1991; 14: 102 – 108.

55. Bennett, PH. Type 2 diabetes among the Pima Indians of Arizona: an epidemic attributable to environmental change? / discussion. *Nutrition Reviews* 1999; 57(5): part 2. S51 – S54.

56. Dressler WW, Bindon JR, Gilliland. Sociocultural and behavioral influences of health status among the Mississippi Choctaw. *Med Anthrop* 1996; 17: 165 – 180.

57. Hodge AM, Dowse GK, Erasmus RT, Spark RA, Nathaniel K, Zimmet PZ, Alpers MP. Serum lipids and modernization in coastal and highland Papua New Guinea. *Am J Epidemiol* 1996; 144(12): 1129 – 1142.

58. Swinburn BA. The thrifty genotype hypothesis: How does it look after 30 years. *Diabetic Medicine* 1996; 13(8): 695 – 699.

59. Neel JV. Diabetes Mellitus: A 'thrifty' genotype rendered detrimental by 'progress'? *Am J Human Genetics* 1962; 14: 353 – 362.

60. Neel JV. The 'thrifty genotype' in 1998. *Nutrition Reviews* 1999; 57(5): part 2. S2 – S9.

61. Neel JV. The thrifty genotype revisited. In : Kobberling J. Tattersall R, eds. The genetics of Diabetes Mellitus. London: Academic Press, 1982: 282 – 293.

62. McDermott R. Ethics, epidemiology and the thrifty gene: biological determinism as a health hazard. *Soc Sci Med* 1998; 47(9): 1189 – 1195.

63. Armelagos GJ, Leatherman T, Ryan M, Sibley L. Biocultural synthesis in medical anthropology. *Medical Anthropology* 1992; 14(1): 35 – 52.

64. Haffner SM. Epidemiology of type 2 diabetes: risk factors. *Diabetes Care* 1998: 21 (3): 3-6.

65. Hazuda HP, Haffner SM, Stern MP, Eifler CW. Effects of acculturation and socioeconomic status on obesity and diabetes in Mexican Americans: The San Antonio Heart Study. *Am J Epidemiol* 1988; 128(6): 1289 – 1301.

66. McDermott R, O'Dea K, Rowley K, Knight S, Burgess P. Beneficial impact of the homelands movement on health outcomes in central Australian aborigines. *Australian & New Zealand Journal of Public Health* 1998. 22(6): 653-8.

67. Bastian P. Coronary heart disease in tribal Aborigines – the West Kimberly Survey. *Aust NZ J Med* 1979; 9: 284 – 292.

68. King H, Collins AM. A modernity score for individuals in Melanesian Society. *Papua New Guinea Med J* 1989; 32: 11 – 22.

69. O'Dea K, Spargo RM, Nestel PJ. Impact of Westernization on carbohydrate and lipid metabolism in Australian Aborigines. *Diabetologia* 1982; 22: 148 – 153.

70. Hanley AJG, Harris SB, Barnie A, Gittelsohn J, Wolever TMS Logan A, Zinman B. The Sandy Lake Health and Diabetes Project: design, methods and lessons learned. *Chronic Dis Canada* 1995; 16: 149 – 156.

71. Herstein HH, Hughes LJ, Kirbyson RC. Challenge and Survival: The History of Canada. Prentice Hall of Canada Ltd. Scarborough, Canada: 1970.

72. Gittelsohn J, Harris SB, Burris KL, Kakegamic L, Landman LT, Sharma A, Wolever TMS, Logan A, Barnie A, Zinman B. The use of ethnographic methods for applied research on diabetes among the Ojibway-Cree in Northern Ontario. *Health Education Quarterly* 1996; 23(3): 365 – 382.

73. Wolever TMS, Hamad S, Gittelsohn J, Hanley AJG, Logan A, Harris SB, Zinman B. Nutrient intake and food use in an Ojibway-Cree community in northern Ontario assessed by 24h dietary recall. *Nutr Res* 1997; (17) 603 – 618.

74. Hanley AJG, Harris SB, Gao JX, Kwan J, Zinman B. Serum immunoreactive leptin concentrations in a Canadian aboriginal population with high rates of NIDDM. *Diabetes Care* 1997; 20(9): 1408 – 1415.

75. Delisle HF, Ekoe JM. Prevalence of non-insulin-dependent diabetes mellitus and impaired glucose tolerance in two Algonquin communities in Quebec. *Can Med Assoc J* 1993; 148: 41 – 47.

76. Gittelsohn J, Harris SB, Thorne-Lyman AL, Hanley AJG, Barnie A, Zinman B. Body image concepts differ by age and sex in an Ojibway-Cree community in Canada. *J Nutr* 1996: 126; 2990 – 3000.

77. Hatcher L. A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. SAS Institute, Inc: Cary, North Carolina. 1998.

78. Stevens, J. Applied multivariate statistics for the social sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.

79. Gittelsohn J, Harris SB, Whitehead S, Wolever TMS, Hanley AJG, Barnie A, Kakegamic L, Logan A, Zinman B. Developing diabetes interventions in an Ojibwa-Cree community in Northern Ontario: Linking qualitative and quantitative data. *Chronic Diseases in Canada* 1995;16 (4): 157 – 164.

80. Siconolfi SF, Garber LE, Lasater TM, Carleton RA. A simple step test for estimating maximal oxygen uptake in epidemiologic studies. *Am J Epidemiol* 1985; 121: 382 – 90.

81. Health and Welfare Canada. Condensed Canadian Nutrient File, revised 1991. Ottawa: Health Protection Branch, 1991.

82. Connelly PW. MacLean DR. Horlick L. O'Connor B. Petrasovits A. Little JA. Plasma lipids and lipoproteins and the prevalence of risk for coronary heart disease in Canadian adults. Canadian Heart Health Surveys Research Group. *CMAJ* 1992; 146(11): 1977-87.

83. Cronbach, LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297 – 334.

84. Cairney J, Wade TJ. Correlates of body weight in 1994 National Population Health Survey. *International Journal of Obesity and Related Metabolic Disorders* 1998; 22(6): 584 – 591.

85. O'Dea K. Marked improvement in carbohydrate and lipid metabolism in diabetic Australian aborigines after temporary reversion to traditional lifestyle. *Diabetes* 1984; 33: 596 – 603.

86. Trowell HC, Dietary-fiber hypothesis of the etiology of diabetes mellitus. Diabetes 1975; 24: 762 - 5.

87. Wolever TMS, Hamad S, Gittelsohn J, Gao J, Hanley AJG, Harris SB, Zinman B. Low dietary fiber and high protein intakes associated with newly diagnosed diabetes in a remote aboriginal community.

88. Hodge AM, Dowse GK, Koki G. Modernity and obesity in coastal and highland Papua New Guinea. Int J Obes Relat Metab Disord 1995; 19: 154 - 161.

89. Hodge AM, Montgomery J, Dowse GK. Diet in an urban Papua New Guinea population with high levels of cardiovascular risk factors. Ecol Food Nutr (in press).

90. Willet, W. Food Frequency Methods in Nutritional Epidemiology. Oxford University Press, New York: 1990.

91. Health Canada: Diabetes in Canada, National Statistics and Opportunities for Improved Surveillance, Prevention, and Control. Minister of Public Works and Government Services Canada, 1999.

92. Lilienfeld DE, Stolley PD. Foundations of Epidemiology, 3rd Edition. Oxford University Press, London, 1994.

93. Hegele RA, Henian C, Harris SB, Hanley AJG, Zinman B. The hepatic nuclear factor-1α G319S variant is associated with early-onset Type 2 Diabetes in Canadian Oji-Cree. *J Clin Endocrinol Metab* 1999; 84: 1077 – 1082.

94. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology, 2nd Edition. Oxford University Press, New York, 1996.

95. Schlesselman JJ. Sample size requirements in cohort and case-control studies of disease. *American Journal of Epidemiology* 1974; 99(6): 381 – 384.

Appendices

6.0 Appendix A: SAS Data Steps

Summary

The following data steps were used in the development of each of the subscales within each subscale. Each subscale was coded differently, however, the general options were kept the same throughout the analyses.

The Data Steps

The PROC FACTOR statement was used within SAS. Since there were two separate data reduction techniques run, the data step options for each are listed separately below. First, the general options are listed which applied to both PCA and FA. Then, the options used for PCA are described and finally, the options from FA are detailed. For full data step code, see "datastep" in Appendices B, C and D).

GENERAL PROC OPTIONS:

a. FLAG = 0.40

The square of any factor loading tells the proportion of the variance in the data explained in a particular questionnaire item by a factor. The FLAG option highlights factor loadings whose absolute value is greater than 0.40 (in both PCA and EFA), and therefore reveals items for which 16% of the variance is explained by the factor on which it loads. The value of 0.40 is chosen as the minimum value that is allowed when determining which items will be used to determine the nature of the underlying factor as suggested by Hatcher[77] and Stevens[78] for analyses of this kind.

b. SCREE:

A scree plot is produced when this option is used. This creates a plot which graphically displays the eigenvalues and allows for a visual determination of the number of factors to retain. The scree test is one of Hatcher's three criteria used in determining the number of factors to retain when analyzing the results of PCA (see Hatcher's criteria below). This test looks for a break between the components with large eigenvalues and those with small eigenvalues. Those components which appear before a substantial break are assumed to be important and are retained for rotation.

c. OUT = x::

This specifies that a new data set of individual scores on each factor be made. Further examinations, such as regression analyses, may be completed using this new data set of values.

d. ROUND:

Eliminates all decimal places in the factor loadings.

e. SIMPLE:

Requests simple descriptive statistics.

PRINCIPLE COMPONENTS ANALYSIS OPTIONS:

a. METHOD = PRIN:

This specifies that the initial factor extraction will use the uncorrelated (orthogonal) principal axis method. Since this procedure is the first data reduction technique which the data are exposed to, there is no empirical reason to suggest a specific number of factors to extract. Therefore, PROC FACTOR extracts a number of factors equal to the number of variables in the analysis. However, only the items on the first few factors contribute a substantial amount of variance. The first factor will account for the majority of the variance seen in the data, and each successive factor will gradually account for smaller amounts of variance. The results of PCA are then used to specify the actual number of factors to be submitted to factor analysis (see NFACT option below).

b. MINEIGEN = 1.00:

This option is commonly used in principal axis methods (PCA) and is used to help solve the 'number of components problem'. This option allows for the specification of a minimum eigenvalue necessary in order to retain and rotate a factor (the 'eigenvalue' refers to the amount of variance accounted for by a given factor). Most commonly, the critical eigenvalue that a factor must display is a value of 1.00. This value is chosen since in PCA it is assumed that each variable is contributing one unit of variance. Therefore, any component which displays greater than this value is considered to contribute a "meaningful" amount of variance. However, when performing factor analyses, it is assumed that each variable contributes its prior communality estimate to the variance, which is calculated as its squared multiple correlation (the communality refers to the common variance of a single variable). This value is always less than 1.00 and so using the eigenvalue-one criterion in factor analysis would produce spurious results. During PCA, the MINIEGEN option automatically discards any factors with eigenvalues < 1.0 and only those that are ≥ 1.00 are held for rotation. In order to use the MINEIGEN = 1.00 option correctly, the NFACT option should be deleted from the data steps. The NFACT option (described next) usurps the MINEIGEN option since it explicitly directs factor analysis to include a specified number of factors to be retained and rotated. NFACT is not used in PCA, but is only used in factor analysis. In this study, the MINEIGEN option was first used with PCA in order to determine the number of potential factors and only then was the NFACT option used.

FACTORY ANALYSIS OPTIONS:

a. ROTATE = PROMAX:

This option directs SAS to perform factor analysis. As described previously, the method performs an oblique rotation which uses correlated factors.

b. NFACT = x:

Since using PCA yields a number of factors equal to the number of variables, it is necessary to decide how many are "meaningful" and should be held for rotation. As described above, during PCA, only the first few factors account for sufficient amounts of variance in the data and therefore the remaining factors account for relatively less. However, PCA produces a solution to the 'number of components' question within an uncorrelated, orthogonal solution. If factor analysis is being performed (as it is in this analysis), then the number of components chosen during PCA will have to be verified once they are rotated and the oblique solution is produced. PCA allows the researcher the opportunity to decipher the number of components and then, in order to run factor analysis, this number is specified in the NFACT option. Once the factor analysis is performed and the factors retained are rotated (oblique solution), separate criteria are used to ensure the number chosen from PCA for the NFACT option was indeed correct. In an orthogonal (PCA) method, each variable is considered to contribute one unit of variance and therefore this is a valid method of determining "meaningful" factors. However, in factor analysis, each variable instead contributes its prior communality estimate (SMC) which is always less than 1.00. Prior communality estimates are squared multiple correlations between a given variable and the other observed variables. The eigenvalue-one criterion is therefore not a valid method of confirming that the number of factors retained for factor analysis was correct. Therefore, Hatcher's criteria[77] are used.

c. PRIORS = SMC:

This requests that the squared multiple correlations between a given variable and the other observed variables are used as that variable's prior communality estimate. The communality of a variable refers to the proportion of 'common variance' in an observed variable. The common variance corresponds to the proportion of total variance in the variable that is accounted for by the common factors. Since it is not possible to know a variable's communality estimate prior to factor analysis, for this first procedure, they are estimated using squared multiple correlations (SMC). The SMC's are automatically determined by PROC FACTOR by using multiple regression in SAS. A variable with a large communality will load heavily on any of the retained factors.

6.1 Appendix B: Development of Subscale #1:

Fx: represents one of the three factors. For analytic purposes, each factor number was inputted separately as 1, 2, or 3.

6.1.1 Questionnaire Item Re-Coding

Each item was originally coded as:

1 = > than once per day
2 = once per day
3 = 3 – 6 times per week
4 = 1 – 2 times per week
5 = 1 – 3 times per month
6 = rare or never

Using the following re-coding Language in SAS V6.12, a gradient of food intake was created from lowest (0) to highest (5):

```
if FOODa = 1 then aFOOD = 5;
else if FOODa = 2 then aFOOD = 4;
else if FOODa = 3 then aFOOD = 3;
else if FOODa = 4 then aFOOD = 2;
else if FOODa = 5 then aFOOD = 1;
else if FOODa = 6 then aFOOD = 0;
else delete;
```

6.1.2 Other Variable Re-Coding

a. Body Mass Index cut-off was set at 24 or greater for "obesity" classification:

```
if bmi ge 24 then obe = 1;else
if bmi  < 24 and diab = 3 then obe = 0;
```

b. Diabetes classifications:

- original code:
  diab = 1 = newly diagnosed diabetes
  diab = 2 = newly diagnosed IGT
  diab = 3 = normal
  diab = 4 = previously diagnosed diabetes

- new code (which excludes anyone obese from normal groups):

```
if diab = 1 then newdiab = 1;
else if diab = 3 and obe = 0 then newdiab = 0;
else if diab = 2 | diab = 4 then newdiab = .;

if diab = 4 then prevdiab = 1;
else if diab = 3 and obe = 0 then prevdiab = 0;
else if diab = 1 | diab = 2 then prevdiab = .;

(abmet = one of either IGT or newly diagnosed diabetes)
if diab = 1 | diab = 2 then abmet = 1;
else if diab = 3 and obe = 0 then abmet = 0;
```

c. Education classification:

- original code:
  a6educ = 1 = none
  a6educ = 2 = some elementary school
  a6educ = 3 = elementary school completed
  a6educ = 4 = some secondary school
  a6educ = 5 = secondary school completed
  a6educ = 6 = some college of trade school
  a6educ = 7 = college or trade school graduate
  a6educ = 8 = some university
  a6educ = 9 = university graduate
  a6educ = 10 = graduate school
  a6educ = 11 = other

- new code:

```
if a6educ = 11 then educ = . ;
else if a6educ = 1 then educ = 1;
else if a6educ = 2 then educ = 2;
else if a6educ = 3 then educ = 3;
else if a6educ = 4 then educ = 4;
else if a6educ = 5 then educ = 5;
else if a6educ = 6 then educ = 6;
else if a6educ = 7 then educ = 7;
else if a6educ = 8 then educ = 8;
else if a6educ = 9 then educ = 9;
else if a6educ = 10 then educ = 10;
```

6.1.3 FACTOR ANALYSIS

- Data Step:

```
proc factor
        data=a1
        simple
        method=prin
        priors=smc
        nfact=3
        scree
        rotate=promax
        round
        flag=.40
        out=a2;
var bannock beans beefs wildberries carm carrots
cannedfruit chips chocolate coke candy corn
duck eggs freshfruit fish hotchocolate
indianmedicine/teas klik lard macaroni/pasta margarine milk
moose potatoes othervegetables peas pop pork
rabbit soup tea wholewheatbread whitebread;
run;
```

- Exclusion of any individual without a factor score:

```
        if fac1 = . or fac2 = . or fac3 = . then missing = 1;
```

- Univariate distribution of the factors:

```
proc univariate data =a2;
        var factor1 factor2 factor3;
```

- Determine the quartiles of scales excluding obese and diseased;

```
data univ; set a2;
if igt = 1 or newdiab = 1 or obe = 1 then delete;
proc univariate data = univ;
var factor1 factor2 factor3;
run;
```

- Transformations of the factor scores;

```
data a3; set a2;
fac1=factor1;fac2=factor2;fac3=factor3;

proc univariate plot normal data = a3;
var fac1 fac2 fac3;
run;
```

- Creation of dummy variables (as determined by univariate analysis)

FACTOR1

```
if fac1 <  -0.61206 then  Q1 = 0; else
if fac1 ge -0.61206 and fac1 <  -0.01342  then  Q1 = 1; else
if fac1 ge -0.01342 and fac1 <  0.497519 then  Q1 = 2; else
if fac1 ge 0.497519 then  Q1 = 3;

Q2_fx=0; Q3_fx=0; Q4_fx=0;
if Q1 = . then do;
Q2_fx = . ; Q3_fx = . ; Q4_fx = . ;
        end;
if Q1 = 1 then Q2_fx = 1; else
if Q1 = 2 then Q3_fx = 1; else
if Q1 = 3 then Q4_fx = 1;
run;
```

FACTOR2

```
if fac2 <  -0.62744 then  Q1 = 0; else
if fac2 ge -0.62744 and fac2 < -0.04808  then  Q1 = 1; else
if fac2 ge -0.04808 and fac2 < 0.63817  then  Q1 = 2; else
if fac2 ge  0.63817 then  Q1 = 3;

Q2_f2=0; Q3_f2=0; Q4_f2=0;
if Q1 = . then do;
Q2_f2 = . ; Q3_f2 = . ; Q4_f2 = . ;
        end;
if Q1 = 1 then Q2_f2 = 1; else
if Q1 = 2 then Q3_f2 = 1; else
if Q1 = 3 then Q4_f2 = 1;
run;
```

FACTOR3

```
if fac3 < -0.37074  then  Q1 = 0; else
if fac3 ge -0.37074 and fac3 < 0.006454  then  Q1 = 1; else
if fac3 ge 0.006454 and fac3 < 0.578681  then  Q1 = 2; else
if fac3 ge 0.578681 then  Q1 = 3;

Q2_f3=0; Q3_f3=0; Q4_f3=0;
if Q1 = . then do;
Q2_f3 = . ; Q3_f3 = . ; Q4_f3 = . ;
        end;
```

```
if Q1 = 1 then Q2_f3 = 1; else
if Q1 = 2 then Q3_f3 = 1; else
if Q1 = 3 then Q4_f3 = 1;
run;
```

6.1.4 COEFFICIENT-α SCALE RELIABILITY TESTING

FACTOR1:

```
proc corr data = a3 alpha nomiss;
var bann1 berr1 duck1 fish1 hotc1 indn1 rabt1 soup1;
run;
```

FACTOR2:

```
proc corr data = a3 alpha nomiss;
var carr1 corn1 milk1 othv1 peas1 wtbd1;
run;
```

FACTOR3:

```
proc corr data = a3 alpha nomiss;
var beef1 chip1 choc1 coky1 klik1 pop1;
run;
```

6.1.5 LOGISTIC REGRESSION ANALYSES

Completed for continuous and categorical data including quartile analysis from dummy variables:

(facx = each factor was inputted separately here as 1, 2 or 3)

Outcome: Abnormal Metabolism (abmet)

```
proc logistic des data = a3;
model abmet = facx;
proc logistic des data=a3;
model abmet = Q2_fx Q3_fx Q4_fx;
proc logistic des data=a3;
model abmet = facx agescm sex;
proc logistic des data=a3;
model abmet = Q2_fx Q3_fx Q4_fx agescm sex ;
```

Outcome: Newly diagnosed diabetes (newdiab)

```
proc logistic des data = a3;
model newdiab = facx;
proc logistic des data=a3;
model newdiab = Q2_FX Q3_fx Q4_fx;
proc logistic des data=a3;
model newdiab = facx agescm sex ;
proc logistic des data=a3;
model newdiab = Q2_FX Q3_fx Q4_fx agescm sex ;
```

Outcome: IGT

```
proc logistic des data=a3;
model igt = facx;
proc logistic des data=a3;
model igt = Q2_fx Q3_fx Q4_fx;
proc logistic des data=a3;
model igt = facx agescm sex ;
```

```
proc logistic des data=a3;
model igt = Q2_fx Q3_fx Q4_fx agescrn sex ;
```

Outcome: Obestiy (obe)

```
proc logistic des data=a3;
model obe = facx;
proc logistic des data=a3;
model obe = Q2_fx Q3_fx Q4_fx;
proc logistic des data=a3;
model obe = facx agescrn sex;
proc logistic des data=a3;
model obe = Q2_fx Q3_fx Q4_fx agescrn sex;
```

6.1.6 ANALYSIS OF LIPID VARIABLES

(sorted by factor quartile Q and by disease status D)

```
proc sort data = a3;
by Q D;
run;

proc means data = a3;
var eldl ehdl echolest etrigly;
by Q D;
run;
```

6.2 Appendix C: Development Subscale #2:

Fx: represents one of the two factors. For analytic purposes, each factor number was inputted separately as a 1, or 2.

6.2.1 Questionnaire Item Re-Coding

Individual Household Items were re-coded in the following manner to reflect an additive total for # of items possessed, working or broken. This was completed for each of the Household Item variables:

```
if workingHHOLD >= 1 or brokenHHOLD >= 1 then HHOLD = 1;
if workingHHOLD  = 0 or brokenHHOLD  = 0 then HHOLD = 0;
```

6.2.2 Other Variable Re-Coding

Please see Appendix A as these variables were re-coded the same for each section.

- Inclusion/exclusion:
```
if agescrn < 20 then delete;
if prevdiab = 1 then delete;
```

6.2.3 Factor Analysis

- Data Step:
```
proc factor
     data=a1
     simple
     method=prin
     priors=smc
     nfact=2
     scree
     rotate=promax
     round
     flag=.40
     out=a2;
var sewing machine TV telephone VCR BigFreezer
AirConditioner Boat Motor4Boat CassttePlayer
WashMachine Microwave SatelliteDish WashMachine
ClotherDryer Tikinagen Tahkobisin Automobile SkiDoo
ATV;
     run;
```

- determining quartiles while excluding obe and diseased;
```
data univ; set a2;
if obe = 1 or newdiab = 1 or igt = 1 then delete;
proc univariate data = univ;
var factor1 factor2;
run;
```

- TRANSFORMATION OF FACTOR SCORES BY;
```
data a3;set a2;
*transformation of factor1;
fac1 = factor1;
*transformation of factor2;
fac2 = factor2;
*exclude anyone who does not have a factor score;
```

```
if fac1 =. or fac2 = . then delete;
```

- CREATION OF DUMMY VARIABLES FOR QUARTILE ANALYSIS;

FACTOR1

```
if fac1 <  -0.78287  then  Q1 = 0; else
if fac1 ge -0.78287 and fac1 < -0.29415  then  Q1 = 1; else
if fac1 ge -0.29415 and fac1 <  0.205505 then  Q1 = 2; else
if fac1 ge 0.205505 then  Q1 = 3;

Q2_fx=0; Q3_fx=0; Q4_fx=0;
if Q1 = . then do;
Q2_fx = . ; Q3_fx = . ; Q4_fx = . ;
end;
If Q1 = 1 then Q2_fx = 1; else
if Q1 = 2 then Q3_fx = 1; else
if Q1 = 3 then Q4_fx = 1;
run;
```

FACTOR2

```
if fac2 <  -0.6628   then  Q1 = 0; else
if fac2 ge -0.6628   and fac2 < -0.32047  then  Q1 = 1; else
if fac2 ge -0.32047   and fac2 < 0.250364  then  Q1 = 2; else
if fac2 ge  0.250364 then  Q1 = 3;

Q2_f2=0; Q3_f2=0; Q4_f2=0;
if Q1 = . then do;
Q2_f2 = . ; Q3_f2 = . ; Q4_f2 = . ;
        end;
if Q1 = 1 then Q2_f2 = 1; else
if Q1 = 2 then Q3_f2 = 1; else
if Q1 = 3 then Q4_f2 = 1;
        run;
```

- ANALYSIS OF FITNESS DATA;

```
data mean;
set a3;
if vo2unit = . then delete;

proc sort data = mean;
by Q1 a4sex;
run;

proc means data = mean;
var agescrn BMI avewgt vo2unit;
by Q1 a4sex;
run;

proc means data = mean;
var agescrn BMI avewgt vo2unit;
run;

proc sort data = mean;
by a4sex;
run;

proc means data = mean;
var agescrn BMI avewgt vo2unit;
```

```
by a4sex;
run;
```

6.2.4 Coefficient-α Scale Reliability Testing

FACTOR1:

```
proc corr data = a3 alpha nomiss;
var drvw drelpw dvcrw dmicrow ddryerw dcarw dcasetw;
run;
```

FACTOR2:

```
proc corr data = a3 alpha nomiss;
var dseww dfrezw dboatw dmboatw dwasmw;
run;
```

6.2.5 Logistic Regression Analyses

Completed for continuous and categorical data including quartile analysis from dummy variables:

(facx = each factor was inputted separately here as 1, 2 or 3)

Outcome: Abnormal metabolism (IGT or diabetes)

```
proc logistic des data = a3;
model abmet = facx;
proc logistic des data=a3;
model abmet = Q2_fx Q3_fx Q4_fx;
proc logistic des data=a3;
model abmet = facx agescm sex ;
proc logistic des data=a3;
model abmet = Q2_fx Q3_fx Q4_fx agescm sex;
```

Outcome: Newly diagnosed diabetes

```
proc logistic des data = a3;
model newdiab = facx;
proc logistic des data=a3;
model newdiab = Q2_fx Q3_fx Q4_fx;
proc logistic des data=a3;
model newdiab = facx agescm sex ;
proc logistic des data=a3;
model newdiab = Q2_fx Q3_fx Q4_fx agescm sex ;
```

Outcome: IGT

```
proc logistic des data=a3;
model igt = facx;
proc logistic des data=a3;
model igt = Q2_fx Q3_fx Q4_fx;
proc logistic des data=a3;
model igt = facx agescm sex;
proc logistic des data=a3;
model igt = Q2_fx Q3_fx Q4_fx agescm sex ;
```

Outcome: Obesity

```
proc logistic des data=a3;
model obe = facx;
```

```
proc logistic des data=a3;
model obe = Q2_fx Q3_fx Q4_fx;
proc logistic des data=a3;
model obe = facx agescm sex ;
proc logistic des data=a3;
model obe = Q2_fx Q3_fx Q4_fx agescm sex ;
```

6.3 Appendix D: Development of Subscale #3:

Fx: represents one of the two factors. For analytic purposes, each factor number was inputted separately as a 1, or 2.

6.3.1 QUESTIONNAIRE ITEM RE-CODING

Individual Language items were coded in the following manner to reflect total for # of items possessed, working or broken. This was completed for each of the Household Item variables:

```
ENGLISH:
        if a8engspk = 2 then engspk = 0;else
        if a8engspk = 1 then engspk = 1;
        if a8engrd = 2 then engrd = 0;else
        if a8engrd = 1 then engrd = 1;
        if a8engwrt = 2 then engwrt = 0;else
        if a8engwrt = 1 then engwrt = 1;

OJI-CREE
        if a8ojispk = 2 then ojispk = 0;else
        if a8ojispk = 1 then ojispk = 1;
        if a8ojird = 2 then ojird = 0;else
        if a8ojird = 1 then ojird = 1;
        if a8ojiwrt = 2 then ojiwrt = 0;else
        if a8ojiwrt = 1 then ojiwrt = 1;
```

6.3.2 OTHER VARIABLE RE-CODING

Please see Appendix A as these variables were re-coded the same for each section.

6.3.3 FACTOR ANALYSIS

- Data Step:

```
proc factor
     data=josh
     simple
     method=prin
     priors=smc
     nfact=2
     scree
     rotate=promax
     round
     flag=.40
     out=t1;
Var engrd engspk engwrt ojird ojispk ojiwrt;
run;
```

- determining quartiles while excluding obe and diseased;

```
if obe = 1 or newdiab = 1 or igt = 1 then delete;
proc univariate data=a1;
var factor1 factor2;
run;
```

- TRANSFORMATION OF FACTOR SCORES;

```
data a2;set a1;
*log transformation of factor1;
f1 =factor1;
*log transformation of factor2;
f2 =factor2;

*exclude anyone who does not have a factor score;
if fac1 =. or fac2 = . then delete;
```

- CREATION OF DUMMY VARIABLES FOR QUARTILE ANALYSIS

```
FACTOR1
        if f1 < 0.3701433820 then Q1 = 0; else
        if f1 ge 0.3701433820 then Q1 = 1;

        Q2_f1=0;
        if Q1 = . then do;
        Q2_f1 = . ;
        end;
        if Q1 = 1 then Q2_f1 = 1;

FACTOR2
        if f2 < -0.615286805 then Q1 = 0; else
        if f2 ge -0.615286805 then Q1 = 1;

        Q2_f2=0;
        if Q1 = . then do;
        Q2_f2 = . ;
        end;
        if Q1 = 1 then Q2_f2 = 1;
        run;
```

6.3.4 COEFFICIENT-α SCALE RELIABILITY TESTING

```
FACTOR1:
    proc corr data = a2 alpha nomiss;
    var engrd engwrt engspk;
    run;

FACTOR2:
    proc corr data = a2 alpha nomiss;
    var ojird ojiwrt;
    run;
```

6.3.5 LOGISTIC REGRESSION ANALYSIS

Completed for continuous and categorical data including bivariate analysis from dummy variables:

(facx = each factor was inputted separately here as 1 or 2)

```
Outcome: Abnormal metabolism (IGT or diabetes)
    proc logistic des data = a2 ;
    model abmet = fx;
    proc logistic des data = a2 ;
```

```
    model abmet = Q2_fx;
    proc logistic des data = a2 ;
    model abmet = fx agescm sex ;
    proc logistic des data = a2 ;
    model abmet = Q2_fx agescm sex;
```

Outcome: Newly diagnosed diabetes

```
    proc logistic des data = a2 ;
    model newdiab = fx;
    proc logistic des data = a2 ;
    model newdiab = Q2_fx;
    proc logistic des data = a2 ;
    model newdiab = fx agescm sex;
    proc logistic des data = a2 ;
    model newdiab = Q2_fx agescm sex;
```

Outcome: IGT

```
    proc logistic des data = a2 ;
    model igt = fx;
    proc logistic des data = a2 ;
    model igt = Q2_fx;
    proc logistic des data = a2 ;
    model igt = fx agescm sex ;
    proc logistic des data = a2 ;
    model igt = Q2_fx agescm sex ;
```

Outcome: Obesity

```
    proc logistic des data = a2;
    model obe = fx;
    proc logistic des data = a2 ;
    model obe = Q2_fx;
    proc logistic des data = a2 ;
    model obe = fx agescm sex ;
    proc logistic des data = a2 ;
    model obe = Q2_fx agescm sex;
    run;
```

6.4 Appendix E: Food Frequency SAS Factor Analysis

Summary

Appendix A takes a step-by-step approach to describing factor analysis and the specific process used in selecting the number and nature of factors within the Food Frequency data set. Results of factor analysis and the univariate distribution of factor scores from SAS are included in full. The next two Appendices begin with a summary of the nature and number of factors within the household and Language data, respectively. These are followed by the complete SAS output. The step-by-step procedure is only described in Appendix A for the Food Frequency data.

Table 6.4.1 Food Frequency Data: Eigenvalues - SAS Output

Initial Factor Method: Principal Factors

Prior Communality Estimates: SMC

BANN1	BEAN1	BEEF1	BERR1	CARM1	CARR1	CFRT1	CHIP1	CHOC1	COKY1	COLC1	CORN1
0.264841	0.176092	0.234862	0.199254	0.226791	0.332735	0.248405	0.311666	0.255354	0.254649	0.278123	0.376626

DUCK1	EGGS1	FFRT1	FISH1	HOTC1	INDN1	KLIK1	LARD1	MACA1	MGRE1	MILK1	MOOS1
0.262673	0.188065	0.213217	0.432155	0.238014	0.175488	0.206635	0.220512	0.203016	0.135549	0.265684	0.362067

OTHP1	OTHV1	PEAS1	POP1	PORK1	RABT1	SOUP1	TEA1	WHBD1	WTBD1
0.229209	0.400562	0.332480	0.227058	0.245850	0.276814	0.277618	0.194362	0.229774	0.198211

Eigenvalues of the Reduced Correlation Matrix: Total = 8.67460805 Average = 0.25513553

	1	2	3	4	5	6	7	8	9
Eigenvalue	**3.6727**	**1.8495**	**1.5120**	0.9971	0.6923	0.5479	0.5074	0.3833	0.3321
Difference	1.8231	0.3375	0.5150	0.3048	0.1444	0.0405	0.1241	0.0512	0.0531
Proportion	0.4234	0.2132	0.1743	0.0949	0.0798	0.0632	0.0585	0.0442	0.0383
Cumulative	0.4234	0.6366	0.8109	0.9258	1.0056	1.0688	1.1273	1.1715	1.2098

	10	11	12	13	14	15	16	17	18
Eigenvalue	0.2790	0.2030	0.1796	0.1417	0.1213	0.0698	0.0367	0.0293	-0.0132
Difference	0.0759	0.0234	0.0380	0.0203	0.0516	0.0331	0.0073	0.0425	0.0162
Proportion	0.0322	0.0234	0.0207	0.0163	0.0140	0.0080	0.0042	0.0034	-0.0015
Cumulative	1.2419	1.2653	1.2860	1.3024	1.3164	1.3244	1.3286	1.3320	1.3305

	19	20	21	22	23	24	25	26	27
Eigenvalue	-0.0294	-0.0397	-0.0496	-0.0619	-0.1130	-0.1311	-0.1415	-0.1510	-0.2029
Difference	0.0104	0.0099	0.0122	0.0511	0.0181	0.0104	0.0095	0.0519	0.0179
Proportion	-0.0034	-0.0046	-0.0057	-0.0071	-0.0130	-0.0151	-0.0163	-0.0174	-0.0234
Cumulative	1.3271	1.3225	1.3168	1.3097	1.2966	1.2815	1.2652	1.2478	1.2244

	28	29	30	31	32	33	34
Eigenvalue	-0.2208	-0.2359	-0.2588	-0.2697	-0.2993	-0.3093	-0.3528
Difference	0.0150	0.0230	0.0109	0.0296	0.0100	0.0434	
Proportion	-0.0255	-0.0272	-0.0298	-0.0311	-0.0345	-0.0357	-0.0407
Cumulative	1.1989	1.1718	1.1419	1.1108	1.0763	1.0407	1.0000

3 factors will be retained by the NFACTOR criterion.

6.4.1 STEP-BY-STEP APPROACH TO FACTOR ANALYSIS

a) Determining the Number of Factors to Retain

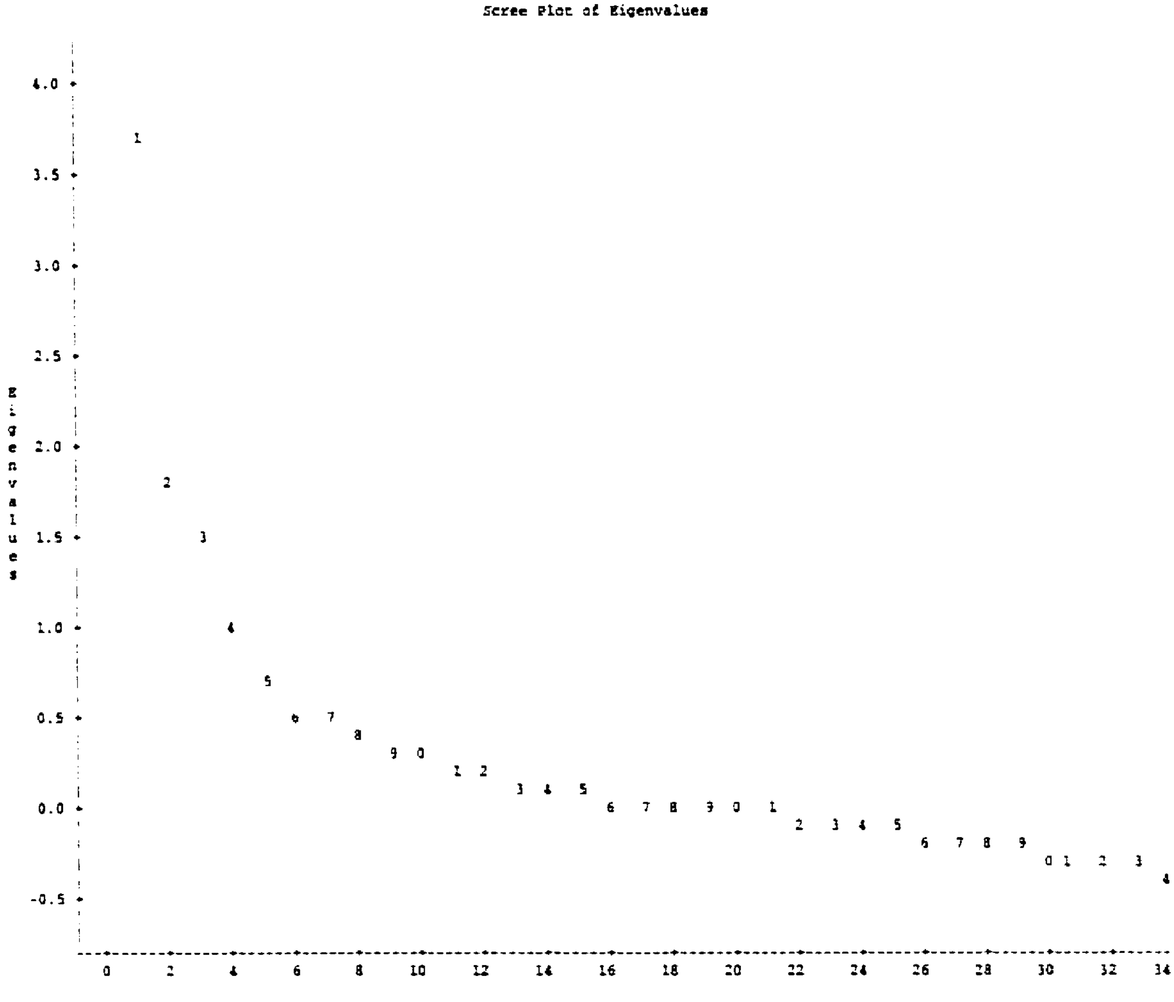
Table 3.5.1 displays the results of the initial output from PROC FACTOR. As explained in Chapter 2, the eigenvalue-one criterion applies less in a factor analysis than in a principle components analysis when determining the number of factors to retain for rotation. The initial output in Table 3.5.1 is the same data as produced from principle components analysis and all factor analyses begin with this initial data output. This is done in order to determine the number of factors to direct

PROC FACTOR to retain in the next step of the analysis (through the NFACT selection). In the case of the Food Frequency data, the eigenvalue-one criterion suggested retaining three factors (note in Table 6.4.1 boldface that the eigenvalues for factors 1 – 3 were all > 1.0).

The second criterion used to determine the number of meaningful factors also suggested that three factors should be retained. The cumulative percent of common variance accounted for by the first three factors was 81% which (Table 6.4.1). This is greater than the 80% cut-off specified in Chapter 2. Also, after the third factor, no single factor accounts for greater than 10% of the variance.

Table 6.4.2 Food Frequency Data: Scree Plot – SAS Output

Initial Factor Method: Principal Factors



The third criterion used was the scree plot (Table 6.4.2). The major break in the eigenvalues appears after the third point. In addition, all of the eigenvalues from #4 onward are below 1.0.

As a result of the proportion of variance accounted for, the scree plot and eigenvalue one criterion, factor analysis was performed and the NFACT criterion in the data step was set to extract 3 factors.

b) Determining which Items Load on each Given Factor

Hatcher suggests that the factor pattern be used as the primary output to determine the items which load on each given factor.[10] Any loadings of at least 0.40 were flagged and these were considered meaningful. Items which loaded on more than one factor were excluded in the interpretation of the factor, however there was no such example in the Food Frequency data (Table 6.4.3) since each of the three factors displayed items which only loaded (≥ 0.40) on that factor.

Table 6.4.3 Food Frequency Data: Promax Rotated Factor Pattern – SAS Output

Rotation Method: Promax

Rotated Factor Pattern (Std Reg Coefs)

	FACTOR1	FACTOR2	FACTOR3
BANNOCK	45*	-12	18
BEAN	4	8	31
BEEF	-15	9	40*
BERRIES	40*	-11	-7
CARM	14	-33	33
CARROTS	10	58*	-10
CANNEDFRUIT1	30	-2	28
CHIP	-2	2	50*
CHOCOLATE	-1	5	39
COOKIES/CAKE	17	7	38
COLDCEREALS	-1	18	29
CORN	7	57 *	5
DUCK	51 *	0	-1
EGGS	3	-2	30
FRESHFRUIT	15	28	12
FISH	69*	-1	-8
HOTCHOCOLATE	41*	3	-2
INDIANMEDICINE	43*	0	-6
KLIK	7	-5	42*
LARD	4	-12	38
MACARONI/PASTA	-8	22	25
MARGARINE	-11	23	12
MILK	-3	42*	5
MOOSEMEAT	59*	6	-5
POTATOES	19	22	19
OTHERVEGETABLES1	6	69*	-8
PEAS	0	54*	4
POP	-14	6	40*
PORK	10	13	35
RABBIT	42*	15	-7
SOUP	36	6	16
TEA	-2	-11	21
WHITEBREAD	-19	11	36
WHOLEWHEATBREAD	-5	40*	1

NOTE: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 have been flagged by an '*'.

c) Rotated Factor Pattern Matrix

The first rotated factor pattern matrix is listed in Table 6.4.3. All variables with loadings ≥ 0.40 were flagged with asterisks (*). Recall that the results from this table are regression coefficients on the variables in the factors. According to Hatcher, this matrix is always more likely to display simple structure than the factor structure matrix.[10] It is for this reason that he suggests that this matrix strongly influence identifying the nature of the underlying factors.

Table 6.4.4 summarizes the results from the promax rotated factor pattern. The items which loaded on each of the three factors formed cohesive, interpretable factors and were therefore named according to the meaning of the underlying construct which the items were likely measuring. Namely, factor 1 = traditional food; factor 2 = healthy food; factor 3 = junk food.

Table 6.4.4 Summary of the Variables which Loaded ≥ 0.40 on each Factor

Factor 1:	**Factor 2:**	**Factor 3:**
Bannock	Carrots	Beef
Berries	Corn	Chips
Duck	Milk	Chocolate (marginal at 0.39)
Fish	Other vegetables	Klik
Hot Chocolate	Peas	Pop
Indian Medicine	Whole wheat bread	
Moosemeat		
Rabbit		

The final determinations of the underlying constructs and the items which load on each was not officially completed until the final pattern matrix was reviewed (see next section). This data output only contributes to an understanding of the nature of the factors when there is a great deal of intercorrelation between the factors. The loadings seen in the factor pattern are not constrained to range between –1.00 and +1.00 therefore in some rare cases where there is a great deal of correlation between the factors, the loadings can be > 10.00. The final structure matrix is then used to help understand the grouping of the items in a factor since it shows the correlation between a single item and a single factor. This matrix is not generally used on its own to determine the nature of the underlying factors since it rarely displays 'simple structure'.

d) Final Determination of the Relationship between the Factors

The structure matrix is presented in Table 3.5.5. Here, the loadings represent the correlation between the variable and the factor. The variables that loaded on each of the factors were similar, as was the apparent meaning of the factors.

Table 6.4.5 Food Frequency Data: Factor Structures – SAS Output

Rotation Method: Promax

Factor Structure (Correlations)

	FACTOR1	FACTOR2	FACTOR3
BANN1	45 *	5	23
BEAN1	11	16	14
BEEF1	-5	15	40 *
BERR1	36	-1	-2
CARM1	11	-21	28
CARR1	24	58 *	5
CPRT1	35	13	33
CHIP1	8	14	51 *
CHOC1	8	15	41 *
COKY1	26	21	42 *
COLC1	9	25	33
CORN1	23	60 *	19
DUCK1	51 *	14	8
EGGS1	7	6	30
FPRT1	25	35	21
FISH1	67 *	16	4
HOTC1	41 *	13	6
INDN1	42 *	11	1
KLIK1	13	7	42 *
LARD1	8	-2	36
MACA1	2	25	29
MGRE1	-1	23	15
MILK1	9	43 *	14
MOOS1	59 *	21	7
OTHP1	28	32	28
OTHV1	21	68 *	9
PEAS1	15	55 *	17
POP1	-5	11	39
PORK1	20	24	40 *
RABT1	45 *	25	4
SOUP1	41 *	20	24
TEA1	-1	-6	18
WHBD1	-10	14	35
WTBD1	6	38	9

NOTE: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 have been flagged by an '*'.

6.4.2 FACTOR SCORE DISTRIBUTIONS

Tables 3.5.6 – 3.5.10 display the results of the analyses to assess the normality of the data for the Food Frequency subscales. Transformations were performed with the data and are as indicated. These results are reported in the following sections: Food Frequency subscales – section 3.2.3; Household Items – section 3.3.3; Language items – 3.4.3.

Traditional Food Scores

Table 6.4.6 Factor Scores Distribution: Traditional Food Items

Moments

N	406	Sum Wgts	406
Mean	0	Sum	0
Std Dev	0.884929	Variance	0.783099
Skewness	0.765233	Kurtosis	1.086894
USS	317.1551	CSS	317.1551
CV	.	Std Mean	0.043918
T:Mean=0	0	Pr>|T|	1.0000
Num ^= 0	406	Num > 0	182
M(Sign)	-21	Pr>=|M|	0.0417
Sgn Rank	-2864.5	Pr>=|S|	0.2264
W:Normal	0.95858	Pr<W	0.0001

Quantiles(Def=5)

100% Max	3.747133	99%	2.605635
75% Q3	0.491083	95%	1.498912
50% Med	-0.08363	90%	1.185817
25% Q1	-0.59683	10%	-1.08313
0% Min	-1.78119	5%	-1.27182
		1%	-1.5675
Range	5.528327		
Q3-Q1	1.087913		
Mode	0.459451		

Extremes

Lowest	Obs	Highest	Obs
-1.78119(	364)	2.605635(	141)
-1.59522(	8)	3.098084(	238)
-1.59184(	323)	3.157071(	349)
-1.57685(	146)	3.16686(	140)
-1.5675(	9)	3.747133(	373)

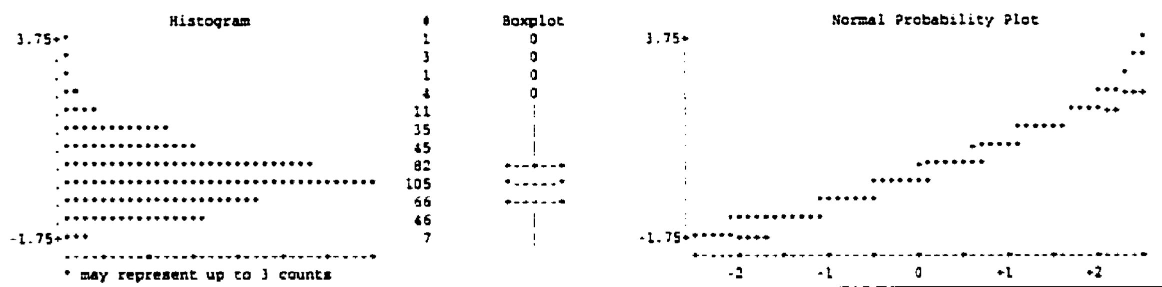


Table 6.4.7 Factor Scores Distribution: Traditional Food Items (Transformed)

Variable=FACTOR1 (LOG TRANSFORMED)

Moments

N	406	Sum Wgts	406
Mean	2.298801	Sum	933.313
Std Dev	0.086514	Variance	0.007485
Skewness	0.471561	Kurtosis	0.370464
USS	2148.532	CSS	3.031322
CV	3.763459	Std Mean	0.004294
T:Mean=0	535.3968	Pr>|T|	0.0001
Num ^= 0	406	Num > 0	406
M(Sign)	203	Pr>=|M|	0.0001
Sgn Rank	41310.5	Pr>=|S|	0.0001
W:Normal	0.974472	Pr<W	0.0540

Quantiles(Def=5)

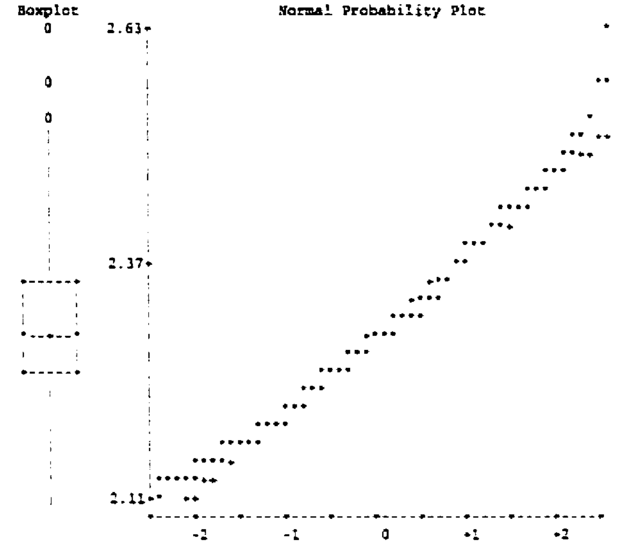
100% Max	2.62083	99%	2.534144
75% Q3	2.350526	95%	2.442252
50% Med	2.294187	90%	2.414647
25% Q1	2.241047	10%	2.187945
0% Min	2.106425	5%	2.166557
		1%	2.132094
Range	0.514405		
Q3-Q1	0.109479		
Mode	2.347506		

Extremes

Lowest	Obs	Highest	Obs
2.106425(	364)	2.534144(	141)
2.128801(	8)	2.572466(	238)
2.129203(	323)	2.576959(	349)
2.130984(	146)	2.577703(	140)
2.132094(	9)	2.62083(	373)

Stem	Leaf	#
262	1	1
260		
258		
256	278	3
254		
252	4	1
250	23	2
248	12	2
246	02599	5
244	22483899	8
242	01222557822345668	17
240	023550224557	12
238	11344558899900223378889	23
236	0123445588001123448	19
234	11125577888899112566779	23
232	00022233445690111244445566666789	32
230	0011111123334456758899900111233334555678899	43
228	011122366779001112223334566666778899	36
226	012223444555777888900022223444555567788899	42
224	000011112224567899901111122315555567777799	42
222	03668900112345566999	20
220	0023367701233344889	19
218	133455668000112233336889	24
216	12466799222356999	17
214	12483478	8
212	991235	6
210	6	1

Multiply Stem.Leaf by 10**-2



Healthy Food Scores

Table 6.4.8 Factor Scores Distribution: Healthy Food

Moments

N	406	Sum Wgts	406
Mean	0	Sum	0
Std Dev	0.881127	Variance	0.776384
Skewness	0.232999	Kurtosis	0.076285
USS	314.4355	CSS	314.4355
CV	.	Std Mean	0.04373
T:Mean=0	0	Pr>|T|	1.0000
Num ^= 0	406	Num > 0	208
M(Sign)	5	Pr>=|M|	0.6552
Sgn Rank	-575.5	Pr>=|S|	0.8082
W:Normal	0.977858	Pr<W	0.0512

Quantiles(Def=5)

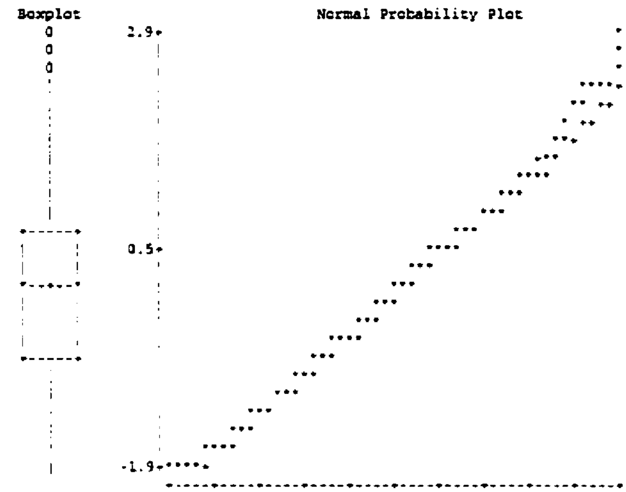
100% Max	2.960085	99%	2.231815
75% Q3	0.601133	95%	1.326803
50% Med	0.046365	90%	1.11923
25% Q1	-0.64435	10%	-1.13825
0% Min	-1.92237	5%	-1.44537
		1%	-1.83231
Range	4.882453		
Q3-Q1	1.24548		
Mode	-0.57238		

Extremes

Lowest	Obs	Highest	Obs
-1.92237(	245)	2.231815(	185)
-1.91941(	372)	2.316989(	265)
-1.89165(	150)	2.565498(	373)
-1.85846(	55)	2.605938(	140)
-1.83231(	30)	2.960085(	99)

Stem	Leaf	#
28	6	1
26	1	1
24	7	1
22	0332	4
20	249	3
18	25	2
16	78	2
14	22189	5
12	144455701333	12
10	09912245667799	14
8	233456788999991134467788	24
6	0013333445566777778891334444778999	34
4	0011112233334678901333366778899	31
2	133344566778880233345566668899	30
0	00145555555667778999000222334555666777777789	44
-0	6555543332188776555421110	25
-2	9887776653322222100009887766555432210	37
-4	97777766554433221077776543210	29
-6	8888755543322109988766555431000	31
-8	88544442210099988755410	23
-10	98644998875543210	17
-12	865431086433330	15
-14	7543105	7
-16	74219931	8
-18	229630	6

Multiply Stem.Leaf by 10**-1



Junk Food Scores

Table 6.4.9 Factor Scores Distribution: Junk Food

Moments

N	406	Sum Wgts	406
Mean	0	Sum	0
Std Dev	0.864105	Variance	0.746677
Skewness	0.368288	Kurtosis	0.692869
USS	302.404	CSS	302.404
CV	.	Std Mean	0.042885
T:Mean=0	0	Pr>|T|	1.0000
Num ^= 0	406	Num > 0	193
M(Sign)	-10	Pr>=|M|	0.3457
Sgn Rank	-1506.5	Pr>=|S|	0.5250
W:Normal	0.981673	Pr<W	0.2240

Quantiles(Def=5)

100% Max	3.071112	99%	2.391854
75% Q3	0.493985	95%	1.52062
50% Med	-0.04445	90%	1.055348
25% Q1	-0.58972	10%	-1.01815
0% Min	-2.6462	5%	-1.29077
		1%	-1.95851
Range	5.717312		
Q3-Q1	1.083708		
Mode	-0.79179		

Extremes

Lowest	Obs	Highest	Obs
-2.6462(	318)	2.391854(	174)
-2.40035(	141)	2.528639(	400)
-2.08135(	24)	2.557278(	246)
-2.01381(	103)	2.69121(	185)
-1.95851(	48)	3.071112(	99)

```
                 Histogram                       #             Boxplot
   3.25+*                                        1                0
       .*                                        3                0
       .**                                       6                0
       .****                                    12                |
       .********                                23                |
       .*******************                     56                |
   0.25+*******************************         92             +--+--+
       .*********************************       98             *-----*
       .************************                70             +-----+
       .************                            36                |
       .**                                       5                |
       .*                                        3                0
  -2.75+*                                        1                0
        ----+----+----+----+----+----+---
        * may reresent up to 3 counts

                       Normal Probability Plot
   3.25+                                                     *
       |                                                    **
       |                                                *****++
       |                                            ****+++++
       |                                        *****+
       |                                  +*******
   0.25+                             +*******
       |                        ********
       |                  *******
       |            *********
       |   *****++
       |**
  -2.75+*
        +----+----+----+----+----+----+----+----+----+----+
             -2        -1         0        +1        +2
```

6.5 Appendix F: Household Items SAS Factor Analysis

Table 6.5.1 Household Data: Eigenvalues – SAS Output

Initial Factor Method: Principal Factors

Prior Communality Estimates: SMC

DSEWW	DTVW	DTELPW	DVCRW	DFREZW	DAIRCW	DBOATW	DMBOATW	DCASETW	DWASRMW
0.318678	0.386457	0.335790	0.325177	0.373123	0.244385	0.550634	0.650264	0.201456	0.257183

DMICROW	DSATLW	DWASMW	DDRYERW	DTIKIW	DTAHW	DCARW	DSKIW	DATVW
0.266619	0.213306	0.398718	0.274866	0.312037	0.258880	0.187816	0.353120	0.258177

Eigenvalues of the Reduced Correlation Matrix: Total = 6.16668527 Average = 0.32456238

	1	2	3	4	5	6	7	8	9	10
Eigenvalue	3.4168	1.0928	0.9763	0.7028	0.5848	0.4440	0.2923	0.1632	0.1025	0.0235
Difference	2.3240	0.1165	0.2735	0.1180	0.1407	0.1518	0.1291	0.0607	0.0790	0.0408
Proportion	0.5541	0.1772	0.1583	0.1140	0.0948	0.0720	0.0474	0.0265	0.0166	0.0038
Cumulative	0.5541	0.7313	0.8896	1.0036	1.0984	1.1704	1.2178	1.2443	1.2609	1.2647

	11	12	13	14	15	16	17	18	19
Eigenvalue	-0.0173	-0.0368	-0.1208	-0.1730	-0.1840	-0.2039	-0.2691	-0.2987	-0.3286
Difference	0.0196	0.0840	0.0522	0.0110	0.0199	0.0651	0.0297	0.0298	
Proportion	-0.0028	-0.0060	-0.0196	-0.0281	-0.0298	-0.0331	-0.0436	-0.0484	-0.0533
Cumulative	1.2619	1.2559	1.2363	1.2083	1.1784	1.1454	1.1017	1.0533	1.0000

2 factors will be retained by the NFACTOR criterion.

Table 6.5.2 Household Data: Scree Plot – SAS Output

Initial Factor Method: Principal Factors

Scree Plot of Eigenvalues

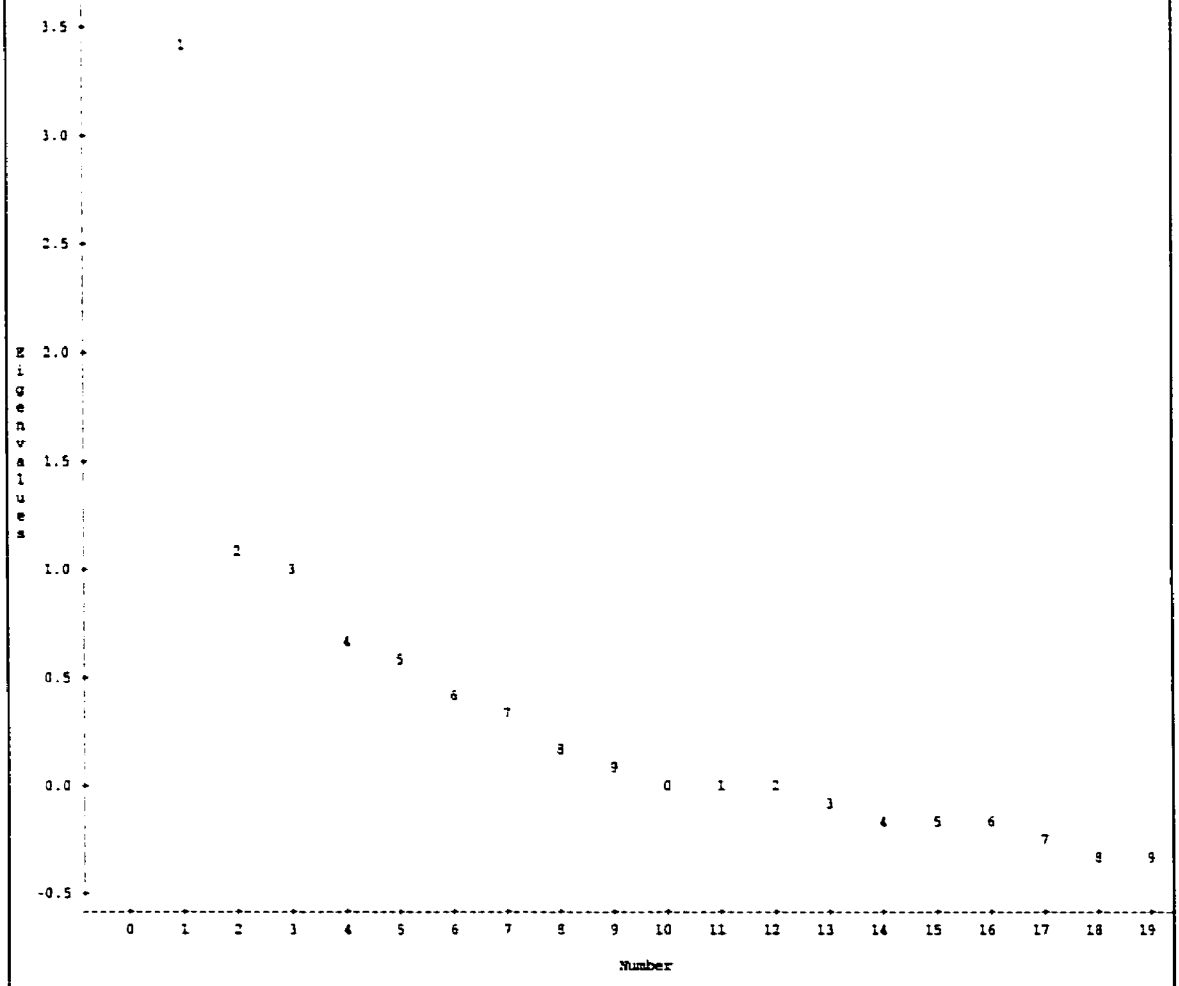


Table 6.5.3 Household Data: Promax Rotated Factor Pattern – SAS Output

Rotation Method: Promax

Rotated Factor Pattern (Std Reg Coefs)

	FACTOR1	FACTOR2	
DSEWW	-1	45 *	Sewing Machine: # Working
DTVW	64 *	-11	TV: # Working
DTELPW	42 *	14	Telephone: # Working
DVCRW	57 *	-14	VCR: # Working
DFREZW	15	48 *	Big Freezer: # Working
DAIRCW	27	5	Air cond.: # Working
DBOATW	2	66 *	Boat: # Working
DMBOATW	7	75 *	Motor for Boat: # Working
DCASETW	39	8	Cassette: # Working
DWASRMW	-19	49 *	Washroom: # Working
DMICROW	45 *	4	Microwave: # Working
DSATLW	32	6	Satellite: # Working
DWASMW	20	47 *	Wash Machine: # Working
DDRYERW	54 *	-14	Dryer: # Working
DTIKIW	25	8	Tikin: # Working
DTAHW	-7	9	Tah: # Working
DCARW	36	6	Car: # Working
DSKIW	20	45 *	Ski: # Working
DATVW	-4	30	ATV: # Working

NOTE: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 have been flagged by an '*'.

Table 6.5.4 Household Data: Inter-Factor Correlations – SAS Output

Inter-factor Correlations

	FACTOR1	FACTOR2
FACTOR1	100 *	46 *
FACTOR2	46 *	100 *

Table 6.5.5 Household Data: Factor Structure – SAS Output

Factor Structure (Correlations)

	FACTOR1	FACTOR2	
DSEWW	20	44 *	Sewing Machine: # Working
DTVW	59 *	18	TV: # Working
DTELPW	49 *	33	Telephone: # Working
DVCRW	51 *	12	VCR: # Working
DFREZW	37	54 *	Big Freezer: # Working
DAIRCW	29	17	Air cond.: # Working
DBOATW	32	66 *	Boat: # Working
DMBOATW	42 *	78 *	Motor for Boat: # Working
DCASETW	43 *	26	Cassette: # Working
DWASRMW	3	40 *	Washroom: # Working
DMICROW	47 *	24	Microwave: # Working
DSATLW	35	21	Satellite: # Working
DWASMW	41 *	56 *	Wash Machine: # Working
DDRYERW	47 *	11	Dryer: # Working
DTIKIW	29	19	Tikin: # Working
DTAHW	-3	5	Tah: # Working
DCARW	39	23	Car: # Working
DSKIW	41 *	55 *	Ski: # Working
DATVW	9	28	ATV: # Working

NOTE: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 have been flagged by an '*'.

Table 6.5.6 Factor Scores - Modern Activities

Moments

N	319	Sum Wgts	319
Mean	0	Sum	0
Std Dev	0.875244	Variance	0.766052
Skewness	0.540646	Kurtosis	0.044544
USS	243.6047	CSS	243.6047
CV	.	Std Mean	0.049004
T:Mean=0	0	Pr>|T|	1.0000
Num ^= 0	319	Num > 0	154
M(Sign)	-5.5	Pr>=|M|	0.5756
Sgn Rank	-1396	Pr>=|S|	0.3980
W:Normal	0.955179	Pr<W	0.0001

Quantiles(Def=5)

100% Max	2.567095	99%	2.325274
75% Q3	0.515582	95%	1.743325
50% Med	-0.03062	90%	1.182689
25% Q1	-0.57753	10%	-1.09985
0% Min	-1.62577	5%	-1.33737
		1%	-1.55469
Range	4.192865		
Q3-Q1	1.093107		
Mode	0.611346		

Extremes

Lowest	Obs	Highest	Obs
-1.62577(	219)	2.325274(	11)
-1.62577(	196)	2.325274(	145)
-1.55469(	399)	2.325274(	146)
-1.55469(	391)	2.567095(	374)
-1.46132(	359)	2.567095(	375)

```
Missing Value          .
Count                 94
% Count/Nobs       22.76
```

```
Stem Leaf                                  #
  24 77                                    2
  22 3333                                  4
  20
  18 355777                                6
  16 2223889944499                        13
  14
  12 133499                                6
  10 245588278                             9
   8 1112588889145888                     16
   6 111112222677779666                   18
   4 144566770266688                      15
   2 1222233355555589990111122677         28
   0 1134455566677999999900002233344478888 37
  -0 9999988775555542221100998443333211   34
  -2 73332100553330000                    17
  -4 99755555533320075555221              23
  -6 997444333211199888877553322200       30
  -8 7666666400844332210000               22
 -10 9966211107654431                     16
 -12 774443398853                         12
 -14 556643330                             9
 -16 33                                    2
     ----+----+----+----+----+----+----+--
Multiply Stem.Leaf by 10**-1
```

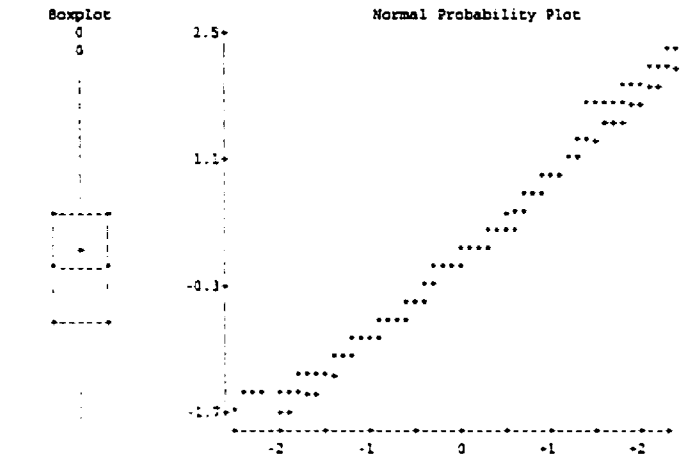


Table 6.5.7 Factor Scores Distributions - Traditional Activities

Moments

N	319	Sum Wgts	319
Mean	0	Sum	0
Std Dev	0.905481	Variance	0.819896
Skewness	0.717507	Kurtosis	-0.3431
USS	260.727	CSS	260.727
CV	.	Std Mean	0.050697
T:Mean=0	0	Pr>|T|	1.0000
Num ^= 0	319	Num > 0	136
M(Sign)	-23.5	Pr>=|M|	0.0099
Sgn Rank	-1559	Pr>=|S|	0.3451
W:Normal	0.913483	Pr<W	0.0001

Quantiles(Def=5)

100% Max	2.529557	99%	2.449112
75% Q3	0.693955	95%	1.729997
50% Med	-0.18675	90%	1.310507
25% Q1	-0.73771	10%	-0.99025
0% Min	-1.2866	5%	-1.14683
		1%	-1.26488
Range	3.816156		
Q3-Q1	1.431668		
Mode	1.079648		

Extremes

Lowest	Obs	Highest	Obs
-1.2866(	219)	2.137618(	5)
-1.2866(	196)	2.449112(	260)
-1.27938(	137)	2.529557(	366)
-1.26488(	399)	2.529557(	380)
-1.26488(	391)	2.529557(	381)

```
Missing Value          .
Count                 94
% Count/Nobs       22.76
```

```
Stem Leaf                                      #
  24 5333                                      4
  22
  20 044444                                    6
  18 00                                        2
  16 8003999                                   7
  14 223333339955                             12
  12 33555911                                  8
  10 588888866889999                          15
   8 11277778899999988                        17
   6 1159992236677                            13
   4 0048811666668                            13
   2 0055555559911444                         16
   0 11334444448902223555557                  23
  -0 977666555443210066422111                 24
  -2 997552211086644100                       18
  -4 999777764332000988776665522222210        33
  -6 99999988777644399988876666665444443332221 41
  -8 988887777554422218888775433222200000     36
 -10 998555221004444422220                    21
 -12 9986655333                               10
     ----+----+----+----+----+----+----+----+-
Multiply Stem.Leaf by 10**-1
```

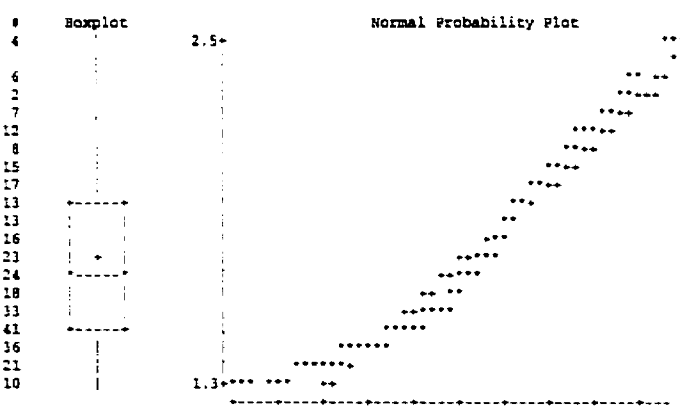


Table 6.5.8 Factor Scores - Modern Activities (Transformed)

Moments

N	319	Sum Wgts	319
Mean	2.298846	Sum	733.3318
Std Dev	0.086226	Variance	0.007435
Skewness	0.322335	Kurtosis	-0.23008
USS	1688.181	CSS	2.364305
CV	3.750838	Std Mean	0.004828
T:Mean=0	476.1755	Pr>|T|	0.0001
Num ^= 0	319	Num > 0	319
M(Sign)	159.5	Pr>=|M|	0.0001
Sgn Rank	25520	Pr>=|S|	0.0001
W:Normal	0.965417	Pr<W	0.0001

Quantiles(Def=5)

100% Max	2.531082	99%	2.511652
75% Q3	2.352858	95%	2.463285
50% Med	2.299519	90%	2.414367
25% Q1	2.232428	10%	2.186068
0% Min	2.125159	5%	2.159019
		1%	2.133611
Range	0.405923		
Q3-Q1	0.12043		
Mode	2.361924		

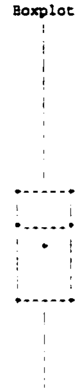
Extremes

Lowest	Obs	Highest	Obs
2.125159(	219)	2.511652(	11)
2.125159(	196)	2.511652(	145)
2.133611(	399)	2.511652(	146)
2.133611(	391)	2.531082(	374)
2.144606(	359)	2.531082(	375)

Missing Value	.
Count	94
% Count/Nobs	22.76

```
Stem Leaf                                      #
 252 11                                        2
 250 2222                                      4
 248
 246 33377022444                              11
 244 33348899                                  8
 242 44                                        2
 240 122559347999                             12
 238 00025777789336669                        17
 236 222223333688880666                       18
 234 266788991377799                          15
 232 000114445555677777701222333344888        33
 230 000011233666788999991222222333344556777  39
 228 3333333335555777789001222233599          31
 226 0015999112277999                         16
 224 1122466666668889124777                   22
 222 0026667778999112222445588999             28
 220 1222222488055667789999                   22
 218 355560122246                             12
 216 0466915599                               10
 214 5578881559999                            13
 212 5544                                      4
     ----+----+----+----+----+----+----+----
Multiply Stem.Leaf by 10**-2
```

Boxplot



Normal Probability Plot

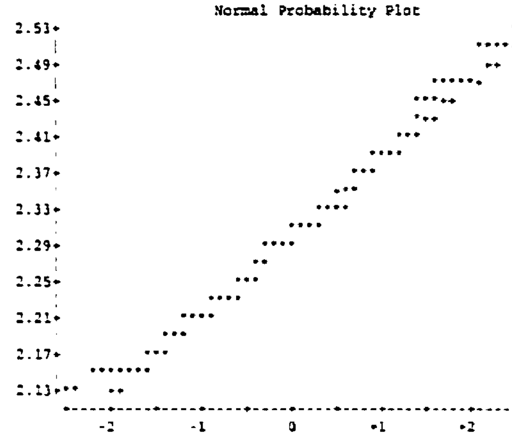


Table 6.5.9 Factor Scores Distributions - Traditional Activities (Transformed)

Moments

N	319	Sum Wgts	319
Mean	2.298635	Sum	733.2644
Std Dev	0.088309	Variance	0.007799
Skewness	0.573046	Kurtosis	-0.60565
USS	1687.987	CSS	2.479938
CV	3.841819	Std Mean	0.004944
T:Mean=0	464.8988	Pr>|T|	0.0001
Num ^= 0	319	Num > 0	319
M(Sign)	159.5	Pr>=|M|	0.0001
Sgn Rank	25520	Pr>=|S|	0.0001
W:Normal	0.925133	Pr<W	0.0001

Quantiles(Def=5)

100% Max	2.52809	99%	2.521649
75% Q3	2.369679	95%	2.462149
50% Med	2.283733	90%	2.425732
25% Q1	2.225951	10%	2.198307
0% Min	2.164862	5%	2.180776
		1%	2.167351
Range	0.363228		
Q3-Q1	0.143728		
Mode	2.40511		

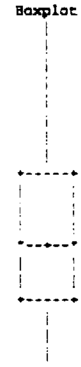
Extremes

Lowest	Obs	Highest	Obs
2.164862(	219)	2.49631(	5)
2.164862(	196)	2.521649(	260)
2.16569(	137)	2.52809(	366)
2.167351(	399)	2.52809(	380)
2.167351(	391)	2.52809(	381)

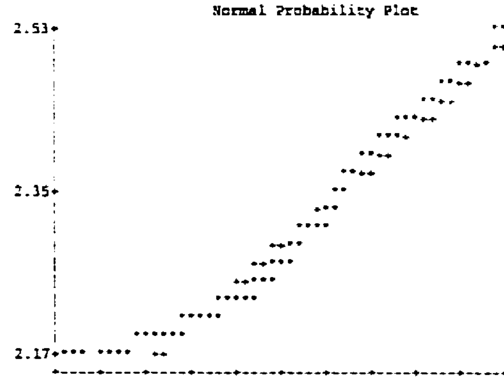
Missing Value	.
Count	94
% Count/Nobs	22.76

```
Stem Leaf                                        #
 252 2888                                        4
 250
 248 588666                                      6
 246 00277788                                    8
 244 11778                                       5
 242 11146655666666                             14
 240 25555552244555588                          17
 238 11266667788888866                          17
 236 2260002236677                              13
 234 2260022777779                              13
 232 3377777771133666                           16
 230 1111234567777770225555888889               28
 228 1334557777778891233778                     22
 226 0012335770011346689                        19
 224 022244445788911133345566669999             30
 222 00000111223346661112234444445566666777999  41
 220 001133446668000111446677889999             30
 218 11144566333335555789999                    23
 216 5567799222557                              13
     ----+----+----+----+----+----+----+----+-
Multiply Stem.Leaf by 10**-2
```

Boxplot



Normal Probability Plot



6.6 Appendix G: Language Proficiency SAS Factor Analysis

Table 6.6.1 Language Data: Eigenvalues – SAS Output

Initial Factor Method: Principal Factors

Prior Communality Estimates: SMC

ENGRD	ENGSPK	ENGWRT	OJIRD	OJISPK	OJIWRT
0.928445	0.771547	0.907312	0.840827	0.040301	0.843636

Eigenvalues of the Reduced Correlation Matrix: Total = 4.33206749 Average = 0.72201125

	1	2	3	4	5	6
Eigenvalue	3.3039	1.1426	0.0134	-0.0020	-0.0513	-0.0746
Difference	2.1613	1.1292	0.0154	0.0493	0.0233	
Proportion	0.7627	0.2638	0.0031	-0.0005	-0.0118	-0.0172
Cumulative	0.7627	1.0264	1.0295	1.0291	1.0172	1.0000

2 factors will be retained by the NFACTOR criterion.

Table 6.6.2 Language Data: Scree Plot – SAS Output

Initial Factor Method: Principal Factors

Scree Plot of Eigenvalues

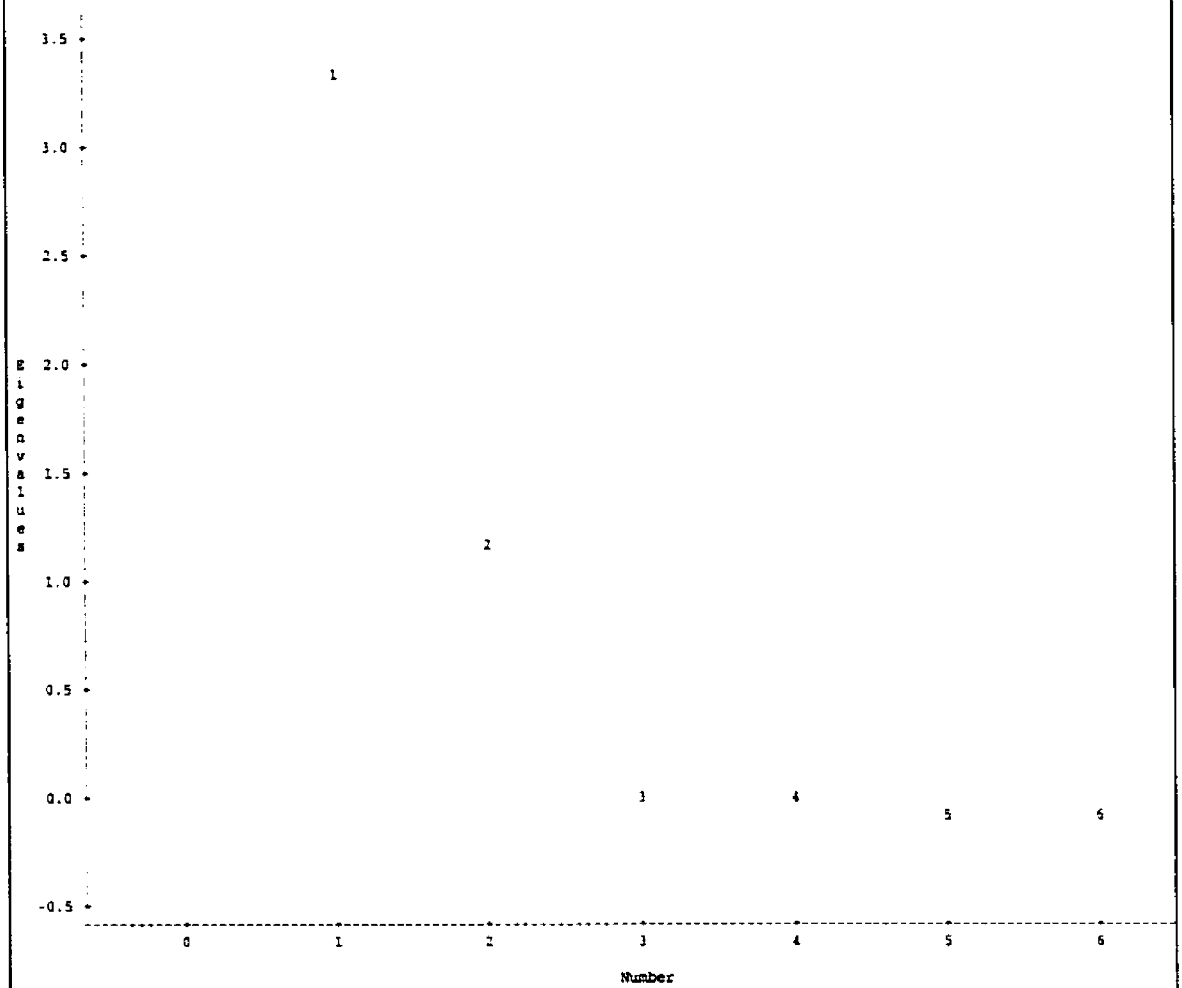


Table 6.6.3 Language Data: Rotated Factor Pattern – SAS Output

```
Rotated Factor Pattern (Std Reg Coefs)

                     FACTOR1   FACTOR2

        ENGRD          99 *        4
        ENGSPK         90 *        4
        ENGWRT         96 *        1
        OJIRD          -4         92 *
        OJISPK         10         21
        OJIWRT         -6         91 *

NOTE: Printed values are multiplied by 100 and rounded to the nearest integer.  Values greater than 0.4 have been flagged by an
'*'.
```

Table 6.6.4 Language Data: Factor Structure Correlations – SAS Output

```
                 Factor Structure (Correlations)

                     FACTOR1   FACTOR2

        ENGRD          98 *      -40 *
        ENGSPK         86 *      -36
        ENGWRT         96 *      -42 *
        OJIRD         -45 *       94 *
        OJISPK          1         17
        OJIWRT        -47 *       94 *

NOTE: Printed values are multiplied by 100 and rounded to the nearest integer.  Values greater than 0.4 have been flagged by an
'*'.
```

Table 6.6.5 Factor Scores Distributions: English Proficiency

```
                 Moments                                    Quantiles(Def=5)                                     Extremes

N              411  Sum Wgts         411      100% Max  0.370144       99%  0.370144       Lowest    Obs     Highest    Obs
Mean             0  Sum                0       75% Q3   0.370144       95%  0.370144      -2.9416(   355) 0.370144(   406)
Std Dev   0.983489  Variance     0.96725       50% Med  0.370144       90%  0.370144      -2.9416(   354) 0.370144(   407)
Skewness  -2.56588  Kurtosis    4.694815       25% Q1   0.261135       10%  -2.47193      -2.9416(   310) 0.370144(   408)
USS       397.5397  CSS         397.5397        0% Min   -2.9416        5%   -2.9416      -2.9416(   180) 0.370144(   409)
CV               .  Std Mean    0.048453                                1%   -2.9416      -2.9416(   174) 0.370144(   410)
T:Mean=0         0  Pr>|T|        1.0000      Range     3.311747
Num ^= 0       411  Num > 0          365      Q3-Q1     0.109009
M(Sign)        159  Pr>=|M|       0.0001      Mode      0.370144
Sgn Rank     24256  Pr>=|S|       0.0001
W:Normal  0.389766  Pr<W          0.0001

                                             Missing Value          .
                                             Count                  1
                                             % Count/Nobs        0.24
```

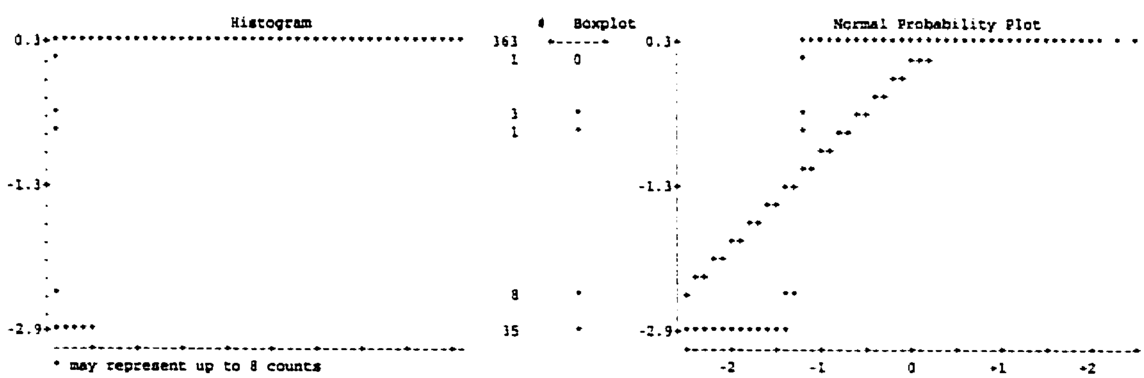


Table 6.6.6 Factor Scores Distributions – Oji-Cree Education

Moments

N	411	Sum Wgts	411
Mean	0	Sum	0
Std Dev	0.9578	Variance	0.91738
Skewness	0.933586	Kurtosis	-1.07873
USS	377.0433	CSS	377.0433
CV	.	Std Mean	0.047187
T:Mean=0	0	Pr>|T|	1.0000
Num ^= 0	411	Num > 0	126
M(Sign)	-80	Pr>=|M|	0.0001
Sgn Rank	-2506	Pr>=|S|	0.2861
W:Normal	0.58557	Pr<W	0.0001

Quantiles(Def=5)

100% Max	1.555791	99%	1.555791
75% Q3	1.522121	95%	1.536953
50% Med	-0.61529	90%	1.536953
25% Q1	-0.61529	10%	-0.61529
0% Min	-0.70094	5%	-0.68611
		1%	-0.68611
Range	2.256734		
Q3-Q1	2.137408		
Mode	-0.61529		

Extremes

Lowest	Obs	Highest	Obs
-0.70094(	172)	1.555791(	53)
-0.70094(	26)	1.555791(	55)
-0.70094(	12)	1.555791(	136)
-0.68611(	403)	1.555791(	192)
-0.68611(	402)	1.555791(	349)

Missing Value	.
Count	1
% Count/Nobs	0.24

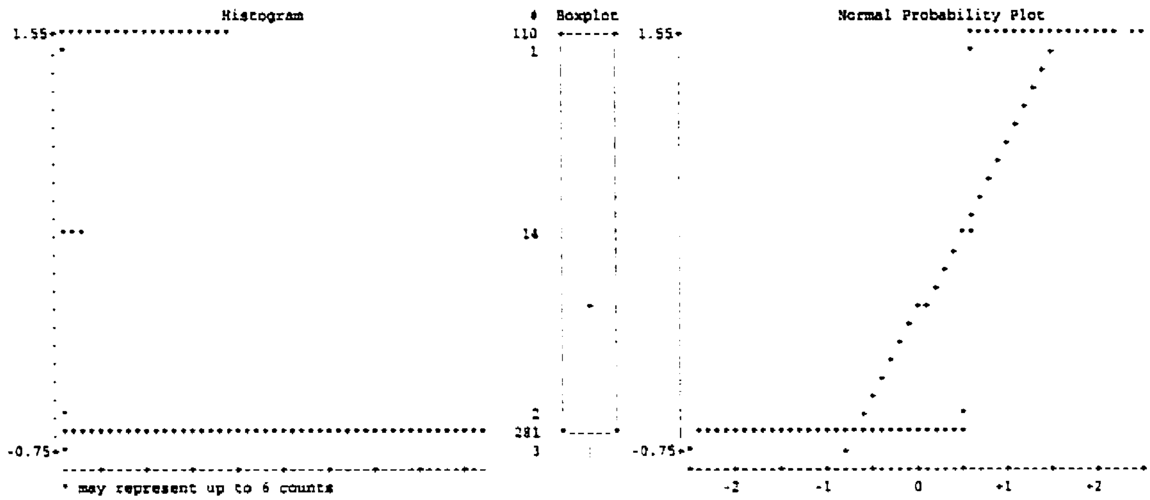


Table 6.6.7 Factor Scores Distributions: English Proficiency (Transformed)

Variable=F1 log(factor 1 + 10)

Moments

N	411	Sum Wgts	411
Mean	2.296721	Sum	946.2492
Std Dev	0.113805	Variance	0.012952
Skewness	-2.58037	Kurtosis	4.765402
USS	2178.594	CSS	5.323106
CV	4.95511	Std Mean	0.005607
T:Mean=0	409.6333	Pr>|T|	0.0001
Num ^= 0	411	Num > 0	411
M(Sign)	206	Pr>=|M|	0.0001
Sgn Rank	42539	Pr>=|S|	0.0001
W:Normal	0.384449	Pr<W	0.0001

Quantiles(Def=5)

100% Max	2.338931	99%	2.338931
75% Q3	2.338931	95%	2.338931
50% Med	2.338931	90%	2.338931
25% Q1	2.328363	10%	2.018639
0% Min	1.954218	5%	1.954218
		1%	1.954218
Range	0.384713		
Q3-Q1	0.010567		
Mode	2.338931		

Extremes

Lowest	Obs	Highest	Obs
1.954218(	355)	2.338931(	406)
1.954218(	354)	2.338931(	407)
1.954218(	310)	2.338931(	408)
1.954218(	180)	2.338931(	409)
1.954218(	174)	2.338931(	410)

Missing Value	.
Count	1
% Count/Nobs	0.24

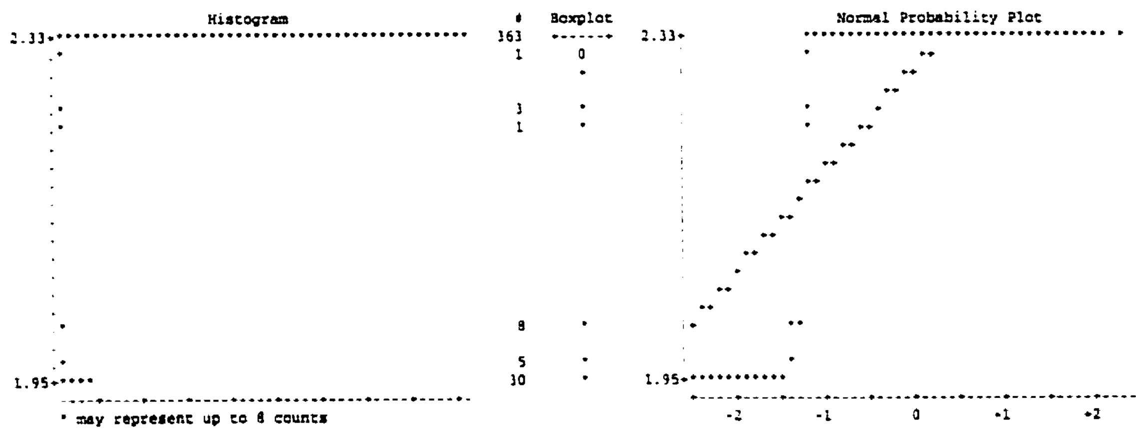


Table 6.6.8 Factor Scores Distributions: Oji-Cree (Transformed)

Variable=F2 log(factor 2 + 10)

Moments

N	411	Sum Wgts	411
Mean	2.298245	Sum	946.8767
Std Dev	0.092033	Variance	0.00847
Skewness	0.926049	Kurtosis	-1.09299
USS	2179.635	CSS	3.481187
CV	4.004484	Std Mean	0.004534
T:Mean=0	506.8764	Pr>|T|	0.0001
Num ^= 0	411	Num > 0	411
M(Sign)	206	Pr>=|M|	0.0001
Sgn Rank	42539	Pr>=|S|	0.0001
W:Normal	0.587194	Pr<W	0.0001

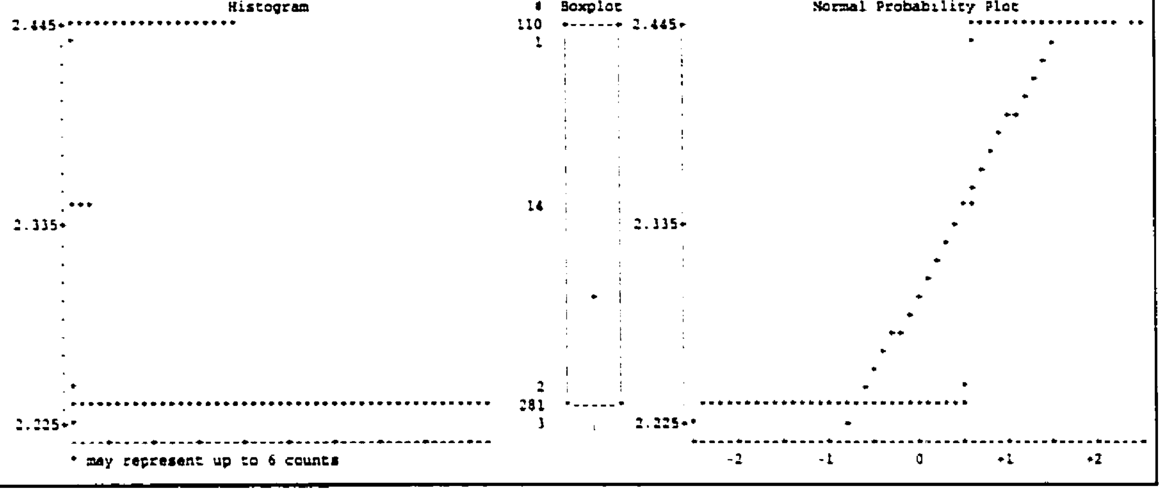
Quantiles(Def=5)

100% Max	2.447187	99%	2.447187
75% Q3	2.444269	95%	2.445555
50% Med	2.239082	90%	2.445555
25% Q1	2.239082	10%	2.239082
0% Min	2.229913	5%	2.231507
		1%	2.231507
Range	0.217274		
Q3-Q1	0.205187		
Mode	2.239082		

Extremes

Lowest	Obs	Highest	Obs
2.229913(	172)	2.447187(	53)
2.229913(	26)	2.447187(	55)
2.229913(	12)	2.447187(	136)
2.231507(	403)	2.447187(	192)
2.231507(	402)	2.447187(	349)

Missing Value	.
Count	1
% Count/Nobs	0.24



6.7 Power / Sample Size Calculation

These calculations were based on frequency distributions detailed in Chapter 3.

Table 6.7.1 Sample size (SS) required to detect a significant association for effects of the magnitude observed in these analyses

	Diabetes/IGT	Diabetes	IGT	Obesity
Food				
Traditional	897	247	7417	602
Healthy	7088	161	841	1140
Junk	888	6672	409	6544
Number observed:	(107)	(45)	(62)	(313)
Activity				
Modern	**56**	**37**	60	**138**
Traditional	144	294	98	8508
Number observed:	(92)	(38)	(54)	(249)
Language				
English	233	241	228	106091
Oji-Cree	269	410	208	11669
Number observed:	(108)	(45)	(63)	(317)

Bolded values are those which surpassed the minimum SS needed for that group

Using the formula for sample size determination in a case-control design:[95]

$$n = \{Z_{\alpha}\sqrt{2\mu(1-\mu)} + Z_{\beta}\sqrt{f(1-f) + p_1q_1}\}^2 / (f - p_1)^2$$

where n = the required sample size of cases for *each* group.

f = p (exposure / diseased)

where exposure is defined as having a score within the 3rd or 4th quartile of subscale scores.

$\mu = \frac{1}{2}f\{1 + R / [1 + f(R-1)]\}$

where R is the OR for the 'exposed' group.

$q_1 = 1 - p_1$

$Z_{\alpha/2} = 1.96$

Z_{β} = 0.842 (with significance level (α) at 0.05 and Power ($1-\beta$) = 0.80).