Physician Characteristics and Compliance With Thyroid Function Testing Practice Guidelines

Joanne McKechnie
Dept of Epidemiology & Biostatistics
McGill University, Montreal

November 1993

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of M A.

Optional Short Title

Physician Compliance With Thyroid Function Testing Guidelines

PHYSICIAN CHARACTERISTICS AND COMPLIANCE WITH THYROID FUNCTION TESTING PRACTICE GUIDELINES

1.0 BACKGROUND:

1.1 Factors Associated With Variation in the Utilization of Medical Services

- 1.1.1 Factors Associated With Patients
- 1.1.2 Factors Associated With Organizations
- 1.1.3 Factors Specific to Physicians and Their Practices

1.2 Physician Compliance With Practice Guidelines

- 1.2.1 Definition of Practice Guidelines
- 1.2.2 Factors Inhibiting the Adoption of Practice Guidelines
- 1.2.3 Physicians' Awareness of and Compliance With Practice Guidelines

1.3 Thyroid Function Testing

- 1.3.1 The Most Commonly Used Thyroid Function Tests
- 1.3.2 Thyroid Function Testing Guidelines
- 1.3.3 Thyroid Function Testing in Alberta

2.0 METHODS

2.1 Research Question and Hypotheses
2.2 Description of Variables

- 2.2.1 Dependent Variable
- 2.2.2 Independent Variables

2.3 Study Design
2.4 Data Collection and Target Population
2.5 Data Analysis

3.0 RESULTS:

3.1 Univariate and Bivariate Analysis
3.2 Multivariate Analysis

4.0 DISCUSSION:

4.1 What Results Suggest
4.2 Further Research Needs

- 4.2.1 Further Research Needs
- 4.2.2 Recommendations

ABSTRAIT

Cette étude utilize les données d'un régime provincial d'assurance-santé pour examiner certains facteurs liées au médecin et à la pratique afin de prédire l'observance aux lignes directrices pour l'épreuve du fonctionnement du thryoïde. Les facteurs en question sont: le sexe du médecin, l'année de naissance, l'université ou pays de graduation, type de pratique (seul ou en group), le milieu de la pratique (métropolitain, rural ou urbain), la région géographique (division de recensement), la certification du Collège des Médecins de famille du Canada (CMFC), la volume des patients du médecin, et la proportion des patients qui sont sujet à l'épreuve du fonctionnement du thryoïde.

On a trouvé des variations considérables en l'observance aux lignes directrices par les médecins. En général, les médecins se sont confirmés aux lignes directrices en 73% des cas. Très peu (approximativement 7%) de la variation est expliqué par les facteurs examinés. Malgré le degré d'explication du modèle, certains facteurs semble avoir un effet positive sur l'observance aux lignes directrices. Ces facteurs sont: le sexe masculin, la pratique localisée dans le sud de la province, la pratique dans un milieu urbain, la certification du CMFC, et d'avoir met a l'épreuve du fonctionnement du thryoïde proportionellement plus de patients.

Les résultats de cette étude suggèrent que les médecins ont besoin de plus d'enseignment au sujet des lignes directrices de l'épreuve du fonctionnement du thryoïde et aussi que de la recherche additionelle est encore besoin pour mieux comprendre les variations en l'observance liées au sexe et la region géographique.

ABSTRACT

This study uses claims data from a provincial health care insurance plan to examine the ability of certain physician and practice characteristics to predict compliance with thyroid function testing (TFT) guidelines. The characteristics examined are the practitioner's gender, year of birth, university or country of graduation, type of practice (solo or clinic), location of practice (metropolitan, urban, or rural), geographical area of practice (census division), CCFP certification, patient caseload, and proportion of patient caseload tested for thyroid functioning.

This study found significant variation in the degree of compliance with TFT guidelines. On average, physician-ordered thyroid function tests complied with guidelines only 37% of the time. Very little of the variation (approximately 7%) was explained by the characteristics examined. Although the explanatory power of this model was low, certain characteristics appear to contribute to better compliance. They are: being male, practicing in the southern third of the province, practicing in an urban location, having CCFP certification, and testing proportionately more of one's patient caseload for thyroid functioning.

The results of this study suggest that physicians require better training regarding TFT and that further research should be done to understand the marked differences in compliance by gender and geographic location.

PHYSICIAN CHARACTERISTICS AND COMPLIANCE WITH THYROID FUNCTION TESTING PRACTICE GUIDELINES

1.0 BACKGROUND:

Economic restraint is forcing all those involved in health care delivery to examine and evaluate the allocation of limited resources. The Watanabe Report[1], released in 1989, indicated that one area that had been allowed to grow with little monitoring or control was the utilization of laboratory services. The report identified the utilization of thyroid function tests as an area of particular concern.

Many physicians appear to be unaware of the impact of the number, type, and cost of the laboratory tests that they generate[2]. This, in addition to the emergence of numerous new technologies, increasing demands from consumers, static or decreasing levels of funding, and evidence of unexplained variations in utilization between seemingly similar populations, has sparked interest in the development of practice guidelines. Several protocols or guidelines for thyroid function testing have been published in the last decade. A preliminary review by Alberta Health of test ordering patterns suggested that physicians tend not to order thyroid function tests in a manner consistent with these guidelines. This raises the following questions: Is there variation across physicians regarding the proportion of times they order thyroid function tests in a manner consistent with existing guidelines? If so, are there characteristics of physicians and/or their practices which are associated with this variation in the rate of compliance with the guidelines?

A small literature examines the relationship between practice patterns and characteristics specific to physicians or their practices. More often, however, explanations given for causes of variation are made informally and not substantiated. For example, part time practitioners are sometimes said to make up for a lack of clinical confidence by an over-reliance on diagnostic testing. Similar allegations have been made regarding female practitioners, those trained in other countries, new learners versus old, and more.

1.1 Factors Associated With Variation in the Utilization of Medical Services

Research on small area variations (SAV) of medical services has shown that variations in many surgical and medical procedures do exist and are often large[1,4,5,6,7] These studies have noted that while medical need is an important factor, it does not account for the majority of the variation At best, studies examining variation in utilization have been able to explain only as much as 40% of variation, of which two-thirds represents health status or need[8] Thus, the search continues for the causes of variation in utilization.

Other factors considered as potential contributors to the utilization of medical services relate to patient, organizational and physician specific characteristics[6,8,9,10,11] Following is a discussion of the findings to date regarding these factors:

1.1.1 Factors associated with patients

Patient factors identified as contributors to utilization include health status or need for medical care; demographic factors such as, age, sex, race, socio-economic status and education, and consumer awareness.

Mortality and morbidity statistics clearly show that age and sex predispose individuals to differential utilization of health care services. Females have lower mortality but higher morbidity and utilization than males, and the young and the elderly are most at risk of disease and utilize more medical services than the middle aged group Work by Roos and Shapiro[7] in Manitoba suggest, however, that use among the elderly is not evenly distributed They show that relatively few elderly persons become heavy users and thus weight the aggregate

Variations due to race and socio-economic status should not be notable in Canada given our nationalized health care system. However, there is still some concern that segments of our society, such as aboriginal people and those in remote rural areas, are not receiving equitable health care. In some cases, access is the problem; in others it may be cultural and/or educational

differences in expectations and acceptance

1.1.2 Factors associated with organizations

Organizational factors are associated with the availability and supply of resources and method of payment

Availability and supply variables include human resources, availability of hospital beds, funding availability, and waiting lists Studies suggest that higher ratios of physicians to the population and specialists to generalists increase utilization, as does having a regular source of medical care. In addition, rates of utilization vary between solo and group practice, comprehensive health clinics, out-patient and emergency care, and walk-in clinics It is unclear how much these differences reflect real differences in need.

While method of payment is not a critical component in Canadian studies of utilization, it does have potential implications, such as requests to extra bill[a], uninsured services, limitations set on services, the use of non-physician practitioners, user fees, and employment benefit packages.

In addition, the level of information available to consumers may affect their attitudes about risks, benefits, and need.

1.1.3 Factors associated with physicians and their practices

Why do physicians practice the way they do? Eisenberg[10] has proposed three roles the physician plays, each with several factors that may contribute to variation in utilization of medical services. These three roles are: physician as his/her own agent; physician as patients' agent; and, physician as guarantor of the social good.

---

[a] The term 'extra-bill' refers to charges to a patient by their physician, in addition to the amount the physician claims from the health care insurance plan for an insured service.

Under 'physician as own agent', Eisenberg includes the following factors income seeker, style of practice; personal characteristics of physician; practice setting, role of clinical leadership, and clinical factors, such as clinical judgement, perceptions of cultural/clinical attitudes, and prevailing customs

As for the 'physician as the patients agent' he asserts that the literature provides support that physicians' practice patterns are driven largely by their patients' best interests.

Add to this the role of 'physician as guarantor of social good' and medical decision making becomes even more complex as physicians weigh the general social good with the factors identified above, especially in times of economic restraint Eisenberg asserts that the collision of the collective social interest with the interests of an individual doctor or patient has become more prominent as concerns about health care costs mount and as the cost of care for individuals becomes increasingly shared.

In summary, given the large number of variables that could affect utilization, it is not surprising that variations exist.

For the purposes of this paper, only factors associated with the practitioner's personal characteristics and practice setting will be examined in more detail The particular factors selected were chosen because: (1) they represent physician characteristics frequently discussed informally within the profession as potential contributing factors and (2) data were available for these characteristics through the provincial health care data base and the National Physician database.

(a) Factors associated with the physician

Eisenberg[10] says that a physician's specialty, age, sex, experience, and type of training come together to represent a sort of physician profile. Following is a description of the limited research that has been done examining the effect of these specific physician characteristics on the

utilization of health care services:

Specialty - Studies by Eisenberg and Nicklin[12] and Childs[13] suggest that internists and family physicians order more diagnostic tests than general practitioners, and that specialists order more than non-specialists.

Age and/or length of practice - A few studies have examined the effect of length of practice on the appropriateness of hospital admissions. An AHCPR study[14], based on the RAND Health Insurance Experiment, found that 27% of admissions by physicians practicing for more than 15 years were inappropriate, compared to 20% for those physicians who had been practicing fewer than 15 years. Other studies[12,15,16] have found that younger doctors tended to provide more ancillary services but shorter lengths of stay than older colleagues. Two studies[12,13] examining variations in diagnostic test utilization noted older physicians used fewer tests.

Gender - While several studies have looked for significant differences by gender, few have found important effects. Williams et al.[17] suggest that there are grounds for believing that female practitioners bring with them distinctive values and interests, which may affect the way they conduct their professional practices. Their study found that women earned significantly less than men with similar levels of experience, practice types, and patient loads. They did not establish reasons why this was the case, but speculated that women may provide fewer services per patient and/or less expensive services than their male counterparts. Similar US studies support these speculations[18,19], noting that female physicians earned less even after the hours worked and patient loads were controlled. Williams et al. interpreted these differences to be partially a result of influences by other roles played by women. For example, they found female practitioners identified their responsibilities as child bearers and family care givers as a primary influence on the organization and conduct of their practices. This interpretation is corroborated by Woodward[20] who found that across career types, women worked shorter hours per week than men and the presence of children further reduced hours of work for women only. Moreover, Maheux[21] suggests that female practitioners have to find ways to attempt to reconcile family and professional responsibilities

The Williams' study also found that only one-third of women, compared with one half of the males in their study, were in specialties, and that female practitioners were more likely to work in urban/metro areas than their male counterparts (69.4% versus 60.3%).

Experience/awareness - A study by Nagurney et al[22] found that overall physicians' awareness of the cost of diagnostic tests was generally very low. However, Epstein and McNeil[23] noted that high-use physicians tended to be more aware of their high usage and Hoey and Cummings[24,25] found that the provision of price information appeared to decrease the use of diagnostic testing in studies.

Type of training - Conflicting evidence exists as to the effect the degree of research orientation of a medical school has on a physicians' test ordering habits[26,27] In some cases physicians from a school with a stronger emphasis on research were found to use more diagnostic tests, other studies have shown them to be more conservative in this respect Gillespie et al.[28] found little difference in test utilization by graduates of two different US medical schools, both working out of the same large Veterans Adminstration Hospital

Eisenberg and Nicklin[12] found that the amount of medical knowledge affected physicians' rates of utilization. For example, internists and family physicians ordered more diagnostic tests than general practitioners. This may reflect a trend whereby general practitioners refer complicated cases to specialists for assessment and lab work up.

(b) Factors associated with the physicians' practice

Eisenberg[10] contends that the place where physicians practice and the influence of their colleagues has an important effect on the level of utilization, and that the effects of group style and peer pressure are probably stronger within more formally organized practices. For example, physicians at one teaching hospital reported that perceived pressure from others was a major influence on the use of diagnostic tests[29].

Location of Practice - Location of practice in this paper refers to whether one practices in a rural, urban or metropolitan area. No reference to this type of breakdown was noted in the literature. One could speculate that urban and metro physicians might have better access to ongoing education such as seminars, rounds and collegial communication and easier access to medical libraries However, the advent of tele-conferences, the increased availability of computerized literature searches and on line text services, and CME course availability in rural and urban centres, reduce the credibility of this argument.

Type of Practice - Type of practice in this paper refers to whether one practices in a solo or a group practice. As pointed out by Williams et al.[30,31] a simple dichotomization of practice type into solo or group is an over simplification and fails to address the importance of a growing number of practices that are neither group nor solo. However, the solo/group differentiation persists and is all that is recorded in many health care data bases. The Williams' paper gave no information regarding differences in utilization of health care services between practice types. It cited other work (in press) suggesting that the organizational efficiencies of group practice may be correlated with the number of patient visits per physician, and hence with professional incomes, so that increased efficiency can in practice generate higher system costs. Williams also found that younger physicians and female practitioners were less likely than older, male practitioners to be in a solo practice. Paulick and Roos[32] concurred that more recent medical graduates are more likely to be in group practices. The implication of these findings is that forms of group practices will likely become more prevalent due to the new graduates and the increasing number of women entering the profession.

1.2 Physicians' Compliance With Practice Policies

1.2.1 Definition of practice policies

Practice policies are standardized specifications for care developed by a formal process that incorporates the best scientific evidence of effectiveness with expert opinion.

In a recent article in the *Journal of the American Medical Association*, Eddy[11,14,15] described three types of practice policies according to their intended flexibility: standards, guidelines, and options. The author uses 'standards' to refer to policies that are intended to be applied rigidly. Exceptions will be rare and difficult to justify. Thus, violation of standards could trigger consideration of malpractice. 'Guidelines' are, in his terminology, more flexible. They are meant to be followed in most cases but can and should be tailored to fit individual needs depending on the patient, the setting, and other factors. Deviation by itself does not imply malpractice. 'Options' are neutral with respect to recommending the use of an intervention. They merely note that different interventions are available and do not attempt to rank them.

Eddy suggests that placement of an intervention policy in one of these categories depends upon 1) the extent to which the outcomes of the intervention are known, 2) the extent to which the preferences of the patients for the outcomes are known; and 3) the spectrum of preferences among patients. If patients are split on their choice of preference, then the onus is on the physician to adequately describe the options and their potential outcomes to each patient and elicit their individual response.

1.2.2 Factors inhibiting the adoption of practice policies

Factors suggested in the literature which may inhibit adoption of practice policies include personal characteristics of the physician and/or the patient, peer opinion, tradition, organization of practice, financial incentives, and consumer expectations.

1.2.3 Awareness of and compliance with practice policies.

The US National Institutes of Health recently evaluated its Consensus Conference program[16]. Part of the evaluation was to look at its effectiveness in disseminating the conference guidelines. It also examined physician and practice characteristics that might predict adoption and utilization of these guidelines.

The NIH suggested that changes in practice are a result of: 1) development of new knowledge, 2) transfer of that knowledge; and 3) adoption and utilization of the knowledge. Practitioners usually require more than knowledge to change behaviour. Disseminating information does not automatically assure its acceptance and application. Features of the health care system, characteristics of practitioners, and the innovations themselves all exert powerful influences on the probability of innovation adoption.

a) Awareness of practice policies

The NIH found that few programs will capture the attention of all practising physicians and that an appropriate bench mark for gauging success of dissemination efforts may not be 100 percent awareness. It found that approximately one third of those practising physicians for whom the conference was relevant were sure that they had heard about the recommendations. NIH researchers found that dissemination efforts had been more successful in reaching specialists than general or family practitioners. This was partially accounted for by differences in information habits. The primary method of dissemination of conference findings has been through published reports and articles in general and specialized medical journals. NIH results showed that specialists and academic physicians made greater use of professional journals, while less specialized journals had a broader-based readership. The *New England Journal of Medicine* attracted non-surgical specialists; *the Journal of the American Medical Association* attracted specialists and general practitioners.

While respondents clearly regarded professional journals as important sources for keeping up to

date and for deciding whether to use new medical procedures, collegial discussion was also an important avenue of communication. Conferences and Continuing Medical Education (CME) courses were most important for hearing about and deciding to use new procedures. In particular, CME courses may be a valuable method of targeting otherwise hard-to-reach physicians, such as general and family practitioners and those in private practice.

Physicians identified several characteristics that made information useful to them. The overwhelming majority preferred new information that highlighted clinical rather than research implications. The NIH results failed to support the idea that physicians do not want to be told what to do. On the contrary, physicians said that clinically-oriented information that is summarized and updated is most useful. Seventy percent preferred summary information rather than complete information accompanied by evidence. There was a broad demand for practice-relevant information, especially among general and family practitioners.

Thus the route to specialists and academic practitioners is fairly direct, through scientific journals. The route to hard-to-reach practitioners is more indirect, through general interest journals, the popular press, written reports or bulletins, and CME programs. In general, dissemination was most effective when the information was didactic and concise.

The NIH study also found a positive association between a physician's age and his/her awareness of the recommendations. The researchers commented that while increasing age may reflect the influence of additional experience in practicing medicine, one might also expect that age would be a negative predictor of awareness, in that younger physicians who were more recently trained may be more up-to-date on sources of information. They stated that previous studies examining this relationship have established no consistent effects.

In general, the NIH study found that primary care physicians and practitioners in private practice tended to be the most difficult to reach through these methods of dissemination. Out of sixty conferences presented over the past decade, approximately half of the physicians surveyed said they had heard of the recommendations from at least one conference, but were less aware of

specific conferences Only 15% to 30% of the physicians who should have heard specific recommendations did so. The program was more successful at reaching specialists than generalists Frequent attendance at CME courses was the most consistent single predictor of consensus conference recommendation awareness.

In summary, the NIH report concluded that two factors were most commonly associated with awareness of conference recommendations: 1) information habits, ie. reading the *New England Journal of Medicine* (guidelines were not printed in *the Journal of the American Medical Association* at the time of the study), specialty journals, and even science and popular magazines; and 2) participation in CME programs.

b) Compliance with practice policies

Literature discussing the model of diffusion of information, ideas or knowledge suggests that the adoption of new ideas is a multi-stage process and that application of the information or new ideas does not necessarily follow immediately from awareness of relevant information. Reasons for various time lags in adoption of new practices have been explained in terms of variations in attitudes, personality or other characteristics, such as early versus late adopters. Using the model of diffusion one would assume that. 1) physicians want to keep abreast of new medical developments; 2) they devote time and effort to do so; and 3) when they encounter information that suggests they should change some aspect of the way they treat patients, they are generally willing to do so. It also assumes a two step process, whereby educationally influential leaders will adopt the changes first, followed by their less influential counterparts. The model implies that these early adopters differ from other physicians in some respects. Perhaps they are better informed, or more closely affiliated with academic medicine. It also assumes that physicians who are up-to-date in one aspect of practice will tend to be up-to-date in other areas as well.

The NIH findings do not support these assumptions. Even a large number of measured background characteristics and information habits could not explain much of the variance in specific measures of physicians' conformity to appropriate practices as defined by NIH consensus

recommendations. Moreover, conformity in one area did not predict conformity in another The NIH authors concluded that the diffusion model fails because it does not specify conditions under which physicians are motivated to change their behaviour. Exposure to new information is neither necessary nor sufficient to produce change It appears that patient treatment is routine, habitual, and automatic and is unlikely to change without a motivating trigger, such as failure to achieve a satisfactory outcome, or the physician's conclusion that a change in treatment will result in more efficient or convenient way of doing things for either the physician or the patient.

The NIH findings imply that: 1) It is more difficult to change practice behaviour than one might think; and 2) successful interventions are apt to either supply a motivating trigger along with information, or capitalize on existing motivations. If physicians have no compelling reason to change their practice, and the new information supplies none, physicians are unlikely to change Thus, informational interventions will work best when they include some way of motivating physicians to change.

The NIH study found that, overall, few physician characteristics or information habits strongly and consistently predicted physician compliance with conference guidelines Two exceptions were relevant specialty and age. Generally, specialists had showed higher than average conformity, but cardiac and thoracic surgeons, internists and family practitioners did not. Academic activities (teaching and publishing) occasionally predicted conformity but this was not consistent. Contrary to other findings, this study found that younger physicians were more likely than older ones to endorse the guidelines.

Lomas et al.[37] surveyed hospitals and obstetricians in Ontario before and after the release of a widely distributed and nationally endorsed consensus statement recommending decreases in the use of caesarian sections. They compared reported versus actual behaviour change They found that most obstetricians (87% to 94%) were aware of the guidelines and that most (82 5% to 85%) agreed with them. One third of both obstetricians and hospitals reported changing their practice as a result of the guidelines. Specifically, they reported that rates of caesarian sections in women with previous sections dropped significantly. However, the survey showed that knowledge of the

content of the guidelines was poor. Furthermore, a review of actual practice after publication of guidelines showed that caesarian section rates were 15 to 49% higher than the rates reported by obstetricians. In fact, the rates were similar to pre-guideline implementation. The researchers concluded that guidelines for practice may predispose physicians to consider changing their behaviour, but that unless there are other incentives or the removal of disincentives, guidelines are unlikely to effect practice in any significant way.

Cohen et al.[38] in Manitoba looked at physicians' compliance with guidelines for Papinicolaou testing in relation to the following physician-specific variables: age of physician, place of undergraduate training, specialty, and number of different patients seen over a one year period (an indicator of practice size)

They found that only 5% of physicians complied with guidelines more than 75% of the time. Overall, only 56% of women were screened appropriately (a rate only somewhat better than chance alone, as they note). They concluded that physician-specific characteristics did not appear to affect the overall results. A characteristic of a physician's' practice that was significantly related to compliance with the guidelines were having a high proportion of patients visiting for gynaecological, obstetrical, or general medical examinations. Variables that were negatively associated with compliance were being a gynaecologist and having a high proportion of patients who lived in inner city or rural areas.

These authors claim that it is unlikely that physicians were unaware of the several available guidelines regarding Papinicolaou testing. Rather, the low compliance rate may have resulted from confusion relating to which set of guidelines to use or simple disagreement with existing guidelines.

1.3 Thyroid Function Testing (TFT)

The following section describes the most commonly used thyroid function tests and reviews the literature discussing guidelines for their use

Thyroid hormones (thyroxine and triiodothyronine) are released through a complicated, interactive process among the hypothalamus and the pituitary and thyroid glands Blood levels are regulated by a glycoprotein called thyroid stimulating hormone (TSH), which is secreted by the anterior pituitary gland. These hormones play a vital role in influencing metabolic processes.

When thyroxine (T4) and triiodothyronine (T3) are released from the thyroid gland they bind almost completely to plasma proteins (approximately 99.97% of T4 and 99.7% of T3). This means that approximately 0.03% of T4 and 0.3% of T3 are 'free' in the plasma. It is these free portions that are active[50]. Thus, many tests are aimed at measuring these free quantities.

Many factors can influence the release and regulation processes of thyroid hormones. Their delicate balance can be upset as a result of congenital or inborn errors of metabolism, or from diseases such as Hashimoto's thyroiditis, Graves' disease, or tumours. Thyroid dysfunction can also result from non-thyroidal illnesses (NTI) or conditions, such as pregnancy or estrogen therapy, acute infections, major systemic illnesses, and certain drug therapies. Thus, the assessment of thyroid function is often appropriate and is not always a simple process.

1.3.1 Description of most commonly used thyroid function tests

T4T - The T4 total or T4T measures the amount of total thyroxine, bound and unbound, circulating in plasma. This level usually increases in hyperthyroidism and always decreases in hypothyroidism. It is the test most commonly performed to assess thyroid function. Its drawback is that the concentration of free or unbound circulating thyroxine in the plasma is commonly affected by numerous non-thyroidal illnesses (NTI) and by various conditions that affect protein binding.

FT4 or FT4I (also referred to as **T7** in some literature) - The free, or unbound, portion of thyroxine in plasma can be measured by direct or indirect methods. The indirect method calculates the free T4 (FT4I) from the results of the total T4 and a T3U test (see description further on). Newer direct methods of measurement called FT4 are being introduced. Measurement of free T4 is more reliable than total T4 because it normalizes the T4 in cases where protein binding is abnormal.

T3T or T3RIA - T3, total by radio-immunoassay, measures total triiodothyronine (T3), bound and unbound, circulating in plasma. Elevations of T3 generally parallel those of T4. It is rarely measured as a test for hypothyroidism.

T3U - T3 uptake test (T3U) is the most common test for measuring protein binding capacity. It is also sometimes called the thyroid hormone-binding ratio (THBR). Generally, the T3U test is used only as part of the calculation with T4T or T3T to establish the free T4 or T3, and it has no other generally established use.

TSH - Measures the concentration of thyroid stimulating hormone (TSH) in plasma. Elevated TSH levels reflect a decreased feedback inhibition of TSH secretion by the pituitary and indicate the presence of hypothyroidism. Older methods of TSH detection were used primarily to test for hypothyroidism; they were not deemed reliable predictors for low-normal levels in hyperthyroidism. Newer methods of measurement, called 'sensitive' or 'ultra sensitive' TSH measures, do have the degree of sensitivity required and are being used to assess both hyper and hypothyroidism. In addition to their good discriminatory power for low-normal results, TSH tests are not affected by protein binding problems and are much more reliable in cases of NTI.

Thyroid antibodies - Two types of circulating thyroid antibodies can be measured, anti-thyroglobulin and anti-microsomal antibodies. High titres may be indicative of auto-immune disease, while negative or low titres are usual in thyroid disfunction. Low positive titres are sometimes found in apparently normal individuals, thus producing 'false positive'. However,

these tests are generally not used as part of thyroid function assessment, are not discussed as part of thyroid function guidelines, and are used relatively infrequently by Alberta physicians The two tests accounted for only 2 5% of all testing. Thus they are excluded from this study

1.3.2 Thyroid function testing guidelines

Thyroid function tests are used to assess patients for the presence and type of thyroid disease and to monitor the progress and effects of therapy (if any) of patients known to have thyroid disease. It has been estimated that approximately one third of all requests for thyroid function tests are made to monitor patients receiving therapy[10]

At the time of this study there appeared to be confusion within the medical community regarding the use of thyroid function tests This confusion was a result of rapid developments within the field of clinical chemistry that make it difficult for the average practitioner to stay informed, and the lack of "established guidelines"[b]. While numerous groups and individuals had proposed various versions of the 'progressive profile' approach, none had been adopted as the definitive set of practice guidelines. A 'progressive approach' to laboratory testing suggests that specific general tests should be done and results reviewed prior to ordering more specialized or specific tests. For example, at the time of the study the University of Alberta Hospital laboratory would not conduct further thyroid function tests unless an abnormal result was found in a preliminary TSH(sensitive) test.

What should practicing general practitioners in Alberta have been expected to know about ordering of thyroid function tests? What information was available to them prior to the study?

General practitioners could have received this information in several ways, including journal publications, protocols published or distributed by professional groups, clinical chemistry texts,

---

[b]Subsequent to the collection of data for this study the Alberta Medical Association, at the request of the Utilization Monitoring Committee - Medical Services in Alberta, developed and published thyroid testing guidelines in April 1992.

Continuing Medical Education presentations, conferences, information from the Alberta Medical Association, Alberta College of Physicians and Surgeons, or Alberta Health, training at medical school, and/or as advice or information from colleagues. Assessing the latter form of dissemination was beyond the scope of this study.

I have attempted to determine what information was provided by any of these entities in the past few years by reviewing the literature back to 1980, and the other sources back to 1985. Following is a review of information available:

a) Published guidelines

All guidelines for the use of thyroid function testing published during the past decade support the use of a progressive profile approach.

In 1982, the British Columbia Association of Laboratory Physicians[40] adopted a protocol for progressive thyroid testing. The protocol required a T4 as the primary test with further testing to be determined by the laboratory physician. For example, a typical protocol might call for a TSH test when the T4 was found to be less than some critical value. If both the T4 and the TSH are low, a T3U or other test of thyroid hormone binding should be carried out (they suggest that a FT4 could replace this step).

In 1985, Nolan[41] proposed a protocol starting with a T4, followed by a FT4I if the T4 was abnormal. If in turn the FT4I was high, a T3 should be done; if low, a TSH. Nolan's primary purpose was to detect unsuspected thyroid disease. However, he says this diagnostic strategy is also applicable to thyroid function testing in general.

In 1985, Caldwell et al.[42] tested a progressive profile on patients with suspected thyroid disease. They cautioned that this protocol had not been tested on patients with NTI or patients receiving thyroid replacement therapy. They advocated the use of a sensitive TSH as the primary test. If the TSH was low, a FT4 should follow. If the FT4 is normal, a FT3 should be done to confirm

diagnosis. If the TSH is high, a FT4 should follow

In 1986, Beck[43] applied a decision analysis model and sensitivity analysis to determine cost effectiveness of both the Nolan and Caldwell protocols. Given the current cost of specific tests within his institution, Beck concluded that the Nolan protocol was most cost-effective However, if the cost of the newer sensitive TSH test should drop the cost would be similar

In 1987, Kalra et al.[44,45] concluded that neither the T4 nor the FT4 can effectively discriminate between sufficient and excess thyroxine replacement They support the use of the sensitive TSH as the sole initial test in evaluating optimum thyroxine therapy They also support the use of the sensitive TSH as the primary test for detection of either hyper or hypothyroidism If additional testing is required they advocate the use of a FT4

In 1987, John, Gow, and Klee[30,46,47] supported the use of the sensitive TSH as the initial screening test for thyroid disorders

In 1987, Ericsson et al.[48] evaluated the use of the sensitive TSH as a first line test in patients in whom thyroid disease was suspected This study was conducted in four clinics within a Swedish hospital. They concluded that the TSHs' predictive value was too low Thus, they advocated combining a T3T (or FT3) and a sensitive TSH as the first line test

In 1988, Young[49] and Watanabe[1] promoted the use of the T4 as the primary test, followed by a T3U (used only to calculate the FT4I) and then a TSH if subsequent tests were abnormal

In 1988, Massey[50] supported Beck's stepped approach starting with a sensitive TSH, followed by a FT4, and then a T3 or FT3 only if previous steps were abnormal. He cautions that this approach is adequate only for uncomplicated patients in whom the main question is one of thyroid status, not major problems of binding.

In 1990, the Canadian Association of Pathologists[40] proposed that the primary test be either a T4 or sensitive TSH. If this test was beyond critical limits, further testing would be conducted.

In 1990, Schlossberg[51] suggested that the first test should be a FT4 combined with a sensitive TSH He did not feel the TSH alone was adequate.

In 1990, Helfand and Crapo[52,53] suggested that either a T4, a FT4, or a sensitive TSH could be suitable as a primary test for screening or case-finding. They recommend the sensitive TSH for monitoring therapy.

In 1990, the American Thyroid Association published a discussion of the merits of various tests but did not offer any testing protocols.

Only information published prior to July 1, 1989, the beginning of the study period, was reviewed. The criteria established to determine categories representing those physicians who ordered thyroid function tests in a manner "consistent" versus "non-consistent" with guidelines will be based on information available prior to the study period only. The data used reflect claims to the Alberta Health Care Insurance Plan (AHCIP) for services provided between July 1, 1989 to June 30, 1990.

Following is a summary of practice guidelines published prior to July 1, 1989.

Figure 1 Published Journal Articles and Protocols

Year	Author & Journal	Primary Test		Secondary Test		Tertiary Test
1982	Br Columbia Association of Lab Pathologists	T4T[c]	if low	TSH	if low	T3U[f]
			if high	T3U[f]		
1985	Nolan Am J Clin Path	T4T	if abnormal	FT4I[d]	if low	TSH
					if high	T3T
1985	Caldwell Lancet	TSH(sens)	if low	FT4[e]	if normal	FT3[g]
			if high	FT4		
1986	Beck supported Nolans' protocol but would rather support TSH(sens) if cost decreased.					
1987	Kalra/John, Gow/Klee Clin Chem, Clin Biochem, J Clin Endo & Metab	TSH(sens)		---		---
1987	Ericsson Scand J Clin	T3T(orFT3) plus TSH(sens)		---		---
1988	Young/Watanabe Alberta Doctors' Digest & Watanabe Report	T4T		T3U[f]		TSH(or T3T
1988	Massey/Can Fam Med	TSH(sens)		FT4		T3T(or FT3)[g]

[c] As an alternate to T4T they suggested a free T4, if available, noting that this would eliminate the need for T3U

[d] A T3U would have been necessary to calculate the FT4I.

[e] They suggest the use of the T3U and T4T (to calculate a FT4I) if the FT4 is not available

[f] To calculate a FT4I.

[g] At the time of this study no labs in Alberta were performing this test This information was gathered through telephone discussions with either the lab director or the lab supervisor for all labs included in the study.

b) Clinical chemistry texts

Several texts of endocrinology and laboratory diagnosis also advocate the use of the progressive profile, starting with a T4 (or FT4I) as the first line test; these include Watts and Keffler (1982)[54], Fischbach (1988)[55], and Teitz (1989)[56].

c) Continuing Medical Education (CME) presentations

A review of the University of Alberta's CME presentations since 1985 indicated that information regarding thyroid function testing was provided on two occasions: a seminar in 1988 and a tele-conference in March 1989. Both were attended mainly or exclusively by general practitioners. Both were presented by the same endocrinologist, who encouraged use of a profile. The progression advised was the same as that described in the next section 'medical school training' because this same endocrinologist was part of the team that prepared the material for teaching.

d) Medical school training at University of Alberta

The University of Alberta publishes an Endocrinology Handbook to be used by medical students. The following progressive approach to thyroid function testing has been advocated in this handbook at least two years prior to the onset of this study:

For diagnosing hypothyroidism:

Start with a TSH (sens) if available, and follow up with a FT4I only if the TSH is high. If TSH(sens) is not available, start with FT4I, followed by TSH if FT4I is low or normal.

For diagnosing hyperthyroidism:

Start with TSH(sens) if available, follow up with a T3RIA if the TSH is suppressed. If TSH(sens) is not available, start with FT4I, followed by T3RIA if FT4I is high or normal.

e) Summary

All of these authors and educators advocate a progressive approach to thyroid function testing and monitoring. All except Ericsson[48] proposed using either the THS (sens), a T4T (or a calculated FT4I), or a FT4 (if available) as the primary test. All of these protocols were available to Alberta practitioners. Albertan physicians should, therefore, be familiar with the guidelines.

1.3.3 Utilization of thyroid function testing in Alberta

Thyroid function testing can be carried out in hospital or private laboratories. Because hospitals operate under global funding, data regarding utilization of specific laboratory tests is difficult or impossible to collect. In addition, at the time of this study, hospital laboratories were not necessarily using a common coding system or terminology for recording laboratory tests. On the contrary, laboratory tests done in private labs were billed under the fee for service mechanism using standardized fee codes and are recorded in detail in the database held by the Alberta Health Care Insurance Plan (AHCIP) of the Alberta government. The data used for this study include only the data held in the AHCIP database. This represents all thyroid function tests ordered by Alberta physicians for Alberta patients through private laboratories between Jul 1, 1989 and June 30, 1990 and excludes all such tests provided in hospital laboratories.

For the year ending June 30, 1990, AHCIP paid almost thirteen million dollars for thyroid function tests to private laboratories. As noted above, this does not include similar tests done in hospital laboratories. Discussions with some hospital laboratory directors suggest that inclusion of hospital data might double these figures.

Thyroid function tests accounted for approximately 13% of the total amount paid for all services to the laboratory and pathology section of the AHCIP Schedule of Medical Benefits for the year ending June 30, 1990. The following table gives the number and cost of thyroid function tests ordered by all Alberta physicians and then by general practitioners only during this period:

Table 1 Number of Thyroid Function Tests and Costs for the Year End June 30, 1990.

Fee Code[h]	Unit Cost	# Services - all physicians	Yearly Cost - all physicians	#Services(%)[i] - GP's only	Yearly Cost - GP's only
	$		$		$
E650T(TSH)	34.50	192,782	$6,613,253	167,104 (87)	$5,734,971
E550U(T4)	12.60	229,125	$2,868,897	199,524 (87)	$2,499,436
E350(T3U)	12.30	113,172	$1,391,979	93,765 (83)	$1,153,309
E550W(T3T)	34.50	40,282	$1,381,762	33,710 (84)	$1,156,824
E353(T4C)	17.70	24,160	$425,767	21,345 (88)	$ 376,321
Total		**599,521**	**$12,681,656**	**515,448 (86)**	**$10,920,860**

Table 1 shows that general practitioners accounted for approximately eighty-six percent (86%) of all thyroid function tests ordered through private laboratories. The remaining 14% attributed to specialists may appear low as they might be expected to order significantly larger numbers of thyroid function tests. However, endocrinologists and other specialists are usually associated with large hospitals. As such, their testing would normally be done through hospital laboratories, and the majority of their lab tests (including thyroid function tests) will not be reflected in the AHCIP database.

Three areas of concern were identified[57] as potential contributors to the high cost of thyroid function testing in Alberta. They are:

---

[h] TSH (thyroid stimulating hormone), T4T (total thyroxine), T3U (tri-iodothyronine uptake), T3T (total Tri-iodothyronine), T4C (thyroxine corrected for abnormal binding protein).

[i] Figures in brackets ( ) represent the proportion of tests claimed by general practitioners compared to all practitioners in the province.

a) Number of Albertans being tested.

Several studies suggest that the prevalence of overt hyperthyroidism in the general population is less than 1.0%[55,58,59] and that among adults, prevalence rates range from 1.9% to 2.7% for females and approximately 1.6% for males[60,61]. Studies of hospitalized geriatric patients set the prevalence as high as 2.3%[62,63] The prevalence rate for overt hypothyroidism (including iatrogenic cases) is approximately 0.5% to 1% in the general population[53,55] Again, if only the adult population is considered the prevalence rate was approximately 1.4% to 1.9% for females, less than 0.1% for males[61], and 0.2% to 5% for the geriatric population[53,62,64,65]

For the year ending June 30, 1990, the data extracted for this study show that 223,978 individuals (approximately 9% of the total Alberta population) received one or more thyroid function tests.

There appears to be confusion within the medical community regarding the type of patient to be tested. As pointed out by Platt[65], the Watanabe Report[1] suggested using the progressive profile for 'thyroid disease screening', an article in Doctor's Digest[49] recommended its' use 'to detect unsuspected thyroid disease', and an article by Nolan[41] suggested it be used for 'case-finding for unsuspected thyroid disease'.

Neither the United States Preventive Services Task Force[66] nor the Canadian Task Force on the Periodic Health Examination[67] recommend screening for thyroid disease in any age or risk group other than newborn screening for cretinism.

b) Confusion regarding the type and sequence of tests to be utilized.

The claims data extracted for this study show great variation regarding the type and sequence of tests. This variation may be a result of rapid developments within the field of clinical chemistry and the lack of officially adopted guidelines that make it difficult for the average practitioner to keep informed.

The large number of occasions (approximately 13,500 during the past year) when five or more thyroid function tests were ordered at one time for one patient by a general practitioner is indeed cause for concern One of the arguments against progressive testing, put forward by some lab directors and physicians, is that there would be a cost incurred by additional physician office visits for subsequent testing This can be overcome if the onus is put on the labs to carry out the progressive profile (if required) on the original sample.

An example of this approach is the successful implementation of such a strategy at the University of Alberta Hospital Laboratory[1]. Early in the fall of 1990, the lab began utilizing a progressive profile approach to thyroid function testing. Each request for testing was subject to this protocol. A physician could write additional orders on the requisition form to override this approach. Between October 1990 and February 1991 approximately 6000 requests for thyroid function testing were received by the lab. Of these, only 217 requested the progressive approach be overridden. Analysis indicated that the majority of tests specially requested would have been done anyway as part of the progressive profile approach. Sixty-four special requests would not have been covered by the progressive approach, but when the special requests were carried out it was noted that only one special request was really warranted and that the progressive profile would have been more appropriate for the rest. Thus, the progressive approach seemed to be appropriate in all but one of the roughly 6000 cases.

c) Current fees paid for tests.

Many of the fees for thyroid function testing in the current Alberta Health Care Insurance Plan (AHCIP) - Schedule of Medical Benefits were instituted several years ago and may not reflect the current state of technology. While one might argue that the cost per test for private laboratories will differ from that for hospital laboratories because of such factors as variations

---

[1] Information on this approach was gathered through interviews with Dr. Keith Walker, Director of the Department of Laboratory Medicine, University of Alberta and Dr. David Fawcett, Medical Biochemist - Endocrinology in the Division of Medical Biochemistry, Department of Laboratory Medicine, University of Alberta .

in volume, type of equipment utilized, or levels of technical skills, the results of a recent costing exercise conducted by Drs. Walker and Fawcett at the University of Alberta Laboratory compared with the fees being paid by the AHCIP suggest the AHCIP fees may be out of line with current actual costs. Table 2 compares the fees for thyroid function tests in the AHCIP Schedule of Medical Benefits with those established by the University of Alberta Hospital for 1991

Table 2 Thyroid Function Test Fees from the AHCIP Fee Schedule and the University of Alberta Hospital Laboratory (1991)

Fee Code	Fee Schedule Unit Cost $	University of Alberta Lab Unit Cost[k] $ (approx)
E550T (TSH) old or sens	34.50	5 50
E550U (T4)	12.60	5.85
E350 (T3U)	12.30	5 75
E550W (T3T)	34.50	5.85
E353 (T4C)	17.70	1.40[l]
no fee code (FT4) claimed as E550U	12.60	5 50

The fees charged by private laboratories are at least twice as high as the University of Alberta laboratory. More importantly, the test for TSH (old or new) is charged at a rate over six times higher. Given the current economic restraints on health care budgets and the rationalization of services it seems imperative that fees for these tests be reassessed.

---

[k] The University of Alberta rates include the cost of kits and reagents, technical equipment and depreciation and a professional component (staff salaries, support, etc) It does not include costs for rent or leases. Some would argue that the University laboratory is able to keep prices low due to high volume but several of the private labs also benefit from large volumes.

[l] The University of Alberta Hospital laboratory charges this amount for the calculation of the T4C in addition to the fees for the T4 and the T3U. The AHCIP fee schedule rate assumes that either the T4C is calculated free of charge when the T4 and the T3U are claimed or only the T4C is claimed, but not both. A review of the claims in the study data base shows that rarely were all three claimed at one time.

During the year ending June 30, 1991, 1766 general practitioners ordered 515,448 thyroid function tests for 200,579 different registrants through 12 private laboratories, excluding two laboratories that had only two claims each. They were dropped from the analysis as it was assumed to be a claims transaction error. Four large laboratories accounted for 85% of all thyroid function tests claimed through the AHCIP.

I contacted each laboratory director or senior billing clerk by telephone and asked the following questions: 1) What options or check boxes were printed on their labs' requisition forms and thus available choices for the physician?; and 2) How were these options translated into claims to Alberta Health?

Of the 12 laboratories contacted, some fell under the same director or manager and thus used the same requisitions and protocol for claims transcription and submission. There were seven distinct lab groupings.

Several issues regarding thyroid function testing were identified during this survey. First, the AHCIP fee schedule had not kept pace with current technology. For example, some laboratories were using the new sensitive TSH, others the older version. The fee schedule has only one fee coded for TSH testing; thus, the same code and same payment was used for both. Second, some laboratories used the newly developed free T4, or the 'direct' method of assessing thyroxine levels corrected for protein binding. There is no fee code for this. The two laboratories using this technique used different fee codes to claim it. There was also confusion regarding claims for the 'indirect' method of calculating thyroxine levels corrected for protein binding. Although there is a fee code in the schedule, E353 - thyroxine corrected for protein binding, few labs used it. Instead, they claimed a E550U - thyroxine (or total T4) and E350 - T3 uptake using the results of these two tests to calculate a free T4 index or thyroxine level corrected for protein.

Requisitions from Lab C have a check box for 'thyroid profile', in addition to check boxes for the individual thyroid function tests. Discussion with the director revealed that if a physician checked both the "thyroid profile" box (this automatically gives all five major thyroid function

tests) and one or more individual thyroid function tests that his lab would do all tests requested regardless of any duplication.

While one could argue that physicians ordering tests through Lab C should know how their requisitions are interpreted, and that they did, in fact, have the option of ordering tests on an individual basis, it is quite probable that many physicians did not consider the consequences of checking off the thyroid profile box. Lab C accounted for 13.1% of all thyroid function tests ordered by general practitioners. A review of data from all labs showed that there were 13,408 occasions when 5 or more tests were claimed at one visit. Lab C accounted for 7242 or 54% of all such requests. The remaining 46% was evenly divided among three other major laboratories. Following is a summary of the information gathered:

Table 3 Fee Codes Utilized by Lab Groups When Transcribing Orders Into Claims

Lab Group	T4T	T3U	FT4	T7 (or FT4I)[m]	T3T	TSH	Thyroid Profile
Lab A	E550U	E350	-	E350/E550U	E550W	E550T	-
Lab B[n]	E550U	E350	E353	E350/E550U	E550W	E550T	-
Lab C[n,o]	E353	E350	E550U	E350/E353	E550W	E550T	T4T,T3U, FT4 or FT4I, T3T,TSH
Lab D	E550U	E350	-	E350/E550U	E550W	E550T	-
Lab E	E550U	E350	-	E350/E550U	-	E550T	-
Lab F	E550U	E350	-	E350/E550U	-	E550T	-
Lab G	E550U	E350	-	E350/E353	E550W	E550T	-

[m] All labs offered a T7 or FT4I calculation (indirect method) Some automatically provided this, others on request only. All but labs C and G claimed T4T and T3U for this calculation

[n] Only these two lab groups indicated they performed the newer FT4 (direct method); they used different fee codes.

[o] In addition to the above options, Lab C offers a 'Thyroid Profile' which if selected gives a practitioner all 5 thyroid function tests (T4T, T3U, FT4 or FT4I, T3T and TSH). Lab C represents two labs run by the same director.

2.0 METHODS:

2.1 Research Question and Hypotheses

This study attempts to answer two questions: 1) whether general practitioners in Alberta vary in their degree of compliance with thyroid function testing guidelines; and 2) if there is variation, whether physician or practice characteristics predict better or worse compliance.

The data lends itself to determining if guidelines for thyroid function testing were followed. It does not determine if the patients tested met the criteria for thyroid function testing.

The hypotheses are:

H_{ALT1}: Compliance with thyroid function testing guidelines varies among general practitioners in Alberta.

H_{ALT2}: One or more of the following factors, associated with general practitioners or their practice, is correlated with the degree of compliance: gender, age, year of graduation, place of graduation, type of practice (group or solo), location of practice (metropolitan, urban, rural), geographic area of practice (census division), practice size, proportion of patients tested for thyroid functioning, and/or whether or not they have certification by the College of Family Physicians of Canada (CCFP).

2.2 Description of Variables

2.2.1 Dependent variable

The dependent variable is a classification system that assigns a score to each physician in the study indicating the proportion of times he/she orders TFT's in a manner consistent with TFT guidelines existing at the time of the study.

To do this, each testing occasion, for each physician, had to be assessed as to its appropriateness A 'testing occasion' is the ordering of one or more thyroid function tests for one patient on one date of service by a physician. A cumulative score, indicating the proportion of times the practitioner ordered in an appropriate manner, was developed for each physician This was accomplished in the following way

1. The data were reviewed to identify all the different test ordering combinations used by the study participants. With five thyroid function tests, $2^5-1=31$ combinations of at least one test are possible. All thirty-one combinations were ordered at least once on a single testing occasion

2. Criteria and a rationale, based on thyroid function testing guidelines and protocols outlined in the first part of this paper, were developed to determine whether or not each of the thirty one combinations was consistent with one or more of the thyroid function testing guidelines

Each testing occasion was then classified as 'consistent' or 'non-consistent' with the criteria A practitioner may order in a manner consistent with guidelines on some occasions and non-consistently on others; thus, a proportion of the number of times the practitioner ordered in a consistent manner will be calculated This proportion is the dependent variable in this report Each practitioner will have a value on a scale from 0 0 to 1.0, with a higher score indicating a higher proportion of times when testing was carried out on a manner consistent with existing guidelines. For the remainder of this paper the dependent variable will be referred to as score

These criteria were developed with assistance from a medical biochemist - endocrinology and a laboratory director associated with a large laboratory at the University of Alberta Hospital The endocrinologist was, at the time of the study, a lecturer on thyroid disease for the University of Alberta and co-author of the thyroid disease training manual used by interns, residents, and CME course participants.

Following are the criteria (Yes=compliant, No=non-compliant):

1. **T3U** (E350) - this test was claimed by all labs. Most laboratories agreed that on its own it is rarely a useful test. However, most labs hold blood specimens for 2 to 8 days, and if a T4T is found to be abnormal, the physician can request a T3U on the same sample. Therefore, I assumed that if the same physician had claimed a T4T for that patient in the preceding 10 days, then this category would be Yes Otherwise No.

2. **T3U, T4C** (E350, E353) - Yes, because some labs claim these two fee codes when they provide a FT4I or T7.

3. **T3U, T4C, TSH** (E350, E353, E550T) - No, all three at one time is not a progressive approach.

4. **T3U, T4C, TSH, T4** (E350, E353, E550T, E550U) - No, not a progressive approach. In addition, should not claim E353 (corrected T4) and E550U (uncorrected T4) together.

5. **T3U, T4C, TSH, T3T** (E350, E353, E550T, E550W) - No, not a progressive approach.

6. **T3U, T4C, TSH, T4, T3T** (E350, E353, E550T, E550U, E550W) - No, not a progressive approach.

7 **T3U, T4C, T4** (E350, E353, E550U) - No, this duplicates testing. A T4C would give the same information as a T4T plus a T3U.

8. **T3U, T4C, T4, T3T** (E350, E353, E550U, E550W) - No, not a progressive approach.

9. **T3U, T4C, T3T** (E350, E353, E550W) - No, this duplicates information. A T3U is part of the T4C.

10.**T3U, TSH** (E350, E550T) - Generally No, T3U not useful except with a T4T. In addition, when ordered with a TSH is not a progressive approach. However, this scenario may be appropriate for a secondary line of testing if a prior T4T was abnormal. Therefore, it is classified as yes if a T4T was claimed in the previous 10 days

11.**T3U, TSH, T4** (E350, E550T, E550U) - No, not a progressive approach.

12.**T3U, TSH, T3T** (E350, E550T, E550W) - No, not a progressive approach.

13.**T3U, TSH, T4, T3T** (E350, E550T, E550U, E550W) - No, not a progressive approach.

14.**T3U, T4** (E350, E550U) - Yes, this can be used to calculate a T7 or FT4I.

15.**T3U, T4, T3T** (E350, E550U, E550W) - No, the T3T should be secondary test.

16.**T3U, T3T** (E350, E550W) - Yes, can be used to calculate a FT3 Index similar to the FT4 Index.

17.**T4C** (E353) - Yes, but only if claimed by labs 6205 or 6861 because they use this to claim a FT4. Otherwise, No

18.**T4C, TSH** (E353, E550T) - No, these should be In a sequential order unless done as a simultaneous assay, ie. a packaged kit that provides results for both tests No lab in the study group indicated that they were performing this simultaneous assay

19.**T4C, TSH, T4** (E353, E550T, E550U) - No, not a progressive approach

20.**T4C, TSH, T4, T3T** (E353, E550T, E550U, E550W) - No, not a progressive approach

21.**T4C, TSH, T3T** (E353, E550T, E550W) - No, not a progressive approach.

22.**T4C, T4** (E353, E550U) - No, they are essentially the same test.

23.**T4C, T4, T3T** (E353, E550U, E550W) - No, not a progressive approach.

24.**T4C, T3T** (E353, E550W) - No, not a progressive approach.

25.**TSH** (E550T) - Yes, a TSH is appropriate as a primary, secondary or tertiary test

26.**TSH, T4** (E550T, E550U) - No, not a progressive approach

27.**TSH, T4, T3T** (E550T, E550U, E550W) - No, not a progressive approach.

28.**TSH, T3T** (E550T, E550W) - No, not a progressive approach

29.**T4** (E550U) - Yes, a T4T is appropriate.

30.**T4, T3T** (E550U, E550W) -No, T3T should be a secondary or tertiary test

31.**T3T** (E550W) - Yes, if a T4T (E550U) by any lab but the same practitioner, same patient or the FT4 (E353) by labs 6205 or 6861 with the same practitioner, same patient, has been ordered in the preceding 45 days Otherwise, No.

These criteria had certain limitations when they were applied to the claims database Most notable was the inability to know: 1) whether a specific episode was for a first assessment or for a follow up visit; 2) whether the patient was being screened, assessed for symptoms, or monitored (with or without therapy); and 3) the results of any previous testing These three scenarios were reflected in my decisions about whether each category was consistent with the guidelines. A further limitation is that it was generally not feasible to review each testing occasion in relation to other testing occasions for the same patient For example, a physician could see a patient this week and order a series of tests that are deemed appropriate because they fit the criteria. The physician could order the same series of tests for each of the next three

weeks, and again the orders at each testing occasion would be considered appropriate. However, if these testing occasions were taken as a whole, one might consider that the repeated testing was inappropriate Thus, the criteria err on the side of leniency regarding the repetition of testing. The magnitude of the data prohibited manual review of all the data. However, the first 10,000 records were printed. A number of test ordering occasions were examined to ensure that the computer logic developed did interpret each occasion according to the established criteria and the logic was as complete as possible.

Given the magnitude of the problem of utilization of laboratory testing (thyroid function testing in particular) it is worthwhile to assess the factors affecting utilization of thyroid function testing in spite of the limitations of the data. While there will always be the argument for the individual or 'odd case' scenario, these criteria should hold for nearly all situations encountered by a general practitioner. Even if a testing scenario is occasionally inappropriate, the physician's score would not be greatly affected. These criteria were deemed to be appropriate for the majority of situations and as such should be reflected in the routine ordering patterns of practitioners.

The following table lists the thirty-one combinations of test ordering that I found in the review of the claims data, with their frequencies and percentages:

Table 4 Number of Times Each Test or Combination of Tests Was Ordered on the Same Day, for One Patient, by the Same General Practitioner, during the Year Ending June 30, 1990[p]

FEE CODE[q]	NUMBER OF TIMES ORDERED (%)	CONSISTENT YES/NO
1 T3U	386	yes * [r]
2. T3U, T4C	1,375	yes
3 T3U, T4C, TSH	894	no
4 T3U, T4C, TSH, T4	216	no
5 T3U, T4C, TSH, T3T	6	no
6 T3U, T4C, TSH, T4, T3T	711	no
7 T3U, T4C, T4,	9	no
8 T3U, T4C, T4, T3T	12	no
9 T3U, T4C, T3T	4	no
10 T3U, TSH	436	yes *
11 T3U, TSH, T4	44,720 (22 2)	no
12 T3U, TSH, T3T	93	no
13 T3U, TSH, T4, T3T	11,327 (5 6)	no
14 T3U, T4	10,523 (5 2)	yes
15 T3U, T4, T3T	1,703 (0 9)	no
16 T3U, T3T	76	yes
17 T4C	447	yes *
18 T4C, TSH	1,356	no
19 T4C, TSH, T4	277	no
20 T4C, TSH, T4, T3T	19	no

[p] In-province data only, date-of-service data from July 1, 1989 to June 30 1990 Because physicians have 180 days to submit claims, payment data was reviewed through to December 31, 1990.

[q] TSH (thyroid stimulating hormone), T4T (total thyroxine), T3U (tri-iodothyronine uptake), T3T (total Tri-iodothyronine), T4C (thyroxine corrected for abnormal binding protein).

[r] An asterisk indicates Yes - with conditions (see criteria)

21 T4C, TSH, T3T	28	no
22 T4C, T4	26	no
23 T4C, T4, T3T	2	no
24 T4C, T4T	3	no
25 TSH	21,679 (10 8)	yes
26 TSH, T4	49,275 (24 5)	no
27 TSH,T4, T3T	6,335	no
28 TSH, T3T	729	no
29 T4	48,801 (24 3)	yes
30 T4, T3T	1,309 (0 7)	no
31 T3T	343	yes *
TOTAL	**201,201 (100%)**	

Of the 201,201 test ordering occasions noted in the table above approximately 42% could be considered 'consistent with' guidelines if all conditions were present.

2.2.2 Independent variables

The independent variables used in this study are: 1) sex of practitioner; 2) practitioners' year of birth; 3) practitioners' year of graduation; 4) university/country of graduation; 5) type of practice (clinic/solo); 6) location of practice (metropolitan/urban/rural); 7) geographical area of practice (census division); 8) total patient count (this is a discrete count of all patients seen by the practitioner during the study period and includes patients seen for TFT's as well as other patients); 9) proportion of the total patient count tested for thyroid functioning; and 10) whether the practitioner is a certificant of the College of Family Physicians of Canada.

2.3 Data Collection and Target Population

Data used for this study are taken from Alberta Health Care Insurance Plan (AHCIP) claims

records, which include all fee for service claims submitted by private practitioners through private laboratories. This database does not include services provided in hospitals, hospital laboratories, or any institutions that are globally funded. Some rural physicians, particularly those in the northern region, rely heavily on the public laboratories for testing These tests will not be captured in the database and as such may affect the results noted by geographic region or for the 'rural' category. The records extracted from this database were matched with additional physician specific information taken from the National Physician Database The extraction of these data and matching of the information from the National Physician Database was done by the Information Technology Division of Alberta Health

The target population of this study is general practitioners in the province of Alberta who were paid by the AHCIP for services provided between July 1, 1989 and June 30, 1991 Specialists were excluded from the study because: 1) as noted earlier only a portion of their lab testing is reflected in the AHCIP database; 2) 86% of thyroid function tests in the database were attributable to general practitioners; and 3) given the nature of this study, I wished to narrow the scope of investigation.

This initial database contained data for all thyroid function tests claimed by 1,766 general practitioners for services provided between July 1, 1989 and June 30, 1990. I also had physician and practice specific information available for these practitioners. Because physicians have 180 days to submit claims and because AHCIP claims are stored on a date of payment basis, it was necessary to review all claims for these tests submitted through December 31, 1990 (that is, six months beyond the end of the study period). From this review, any tests that were performed during the period under review were extracted to form the initial database

As with any large administrative database, errors and inconsistencies were noted. Some of the errors encountered were: 1) lab tests attributed to non-existent or terminated physician billing numbers; 2) lab tests assigned to active billing numbers which fell within the range of in-province billing numbers but were found to be either a generic number used for out of province claims or a "slough" number used for tests submitted with an invalid practitioner number; and 3) lab tests

attributed to physicians for whom there was no record of any patient visits These latter cases were reviewed individually. One practitioner was found to be on salary at a community health clinic. As such, this practitioner could not submit claims for patient visits, and lab tests ordered were submitted for payment to the lab. These tests were left in the data base. The remainder of physician billing numbers with lab tests attributed to them but no patient visits appear to have been errors. All lab tests and practitioner billing numbers that could not be directly linked to an active Alberta physician were dropped from the study. This amounted to approximately 2.3% of the total number of physicians in the original database.

There were also problems in the way in which some labs prepared their requisitions and transcribed the data. As a result, claims submitted from Lab C (noted in Table 3) were not included in the database. As discussed earlier, Lab C represents two laboratories in Calgary under the direction of the same lab pathologist. Use of these submissions posed two problems. First, the requisition form offered practitioners a check box for 'thyroid profile' which automatically gives all five major TFT whether required or not. This lab accounted for 13% of all TFT claimed and 54% of all test ordering occasions when 5 or more tests were ordered at one time. While it can be argued that it is not appropriate to order all five tests simultaneously and that the requisition did offer the opportunity to order the tests individually, it is likely that many physicians did not consider the implications of requesting the profile option. This lab director indicated that if a physician ordered both the profile and checked off individual tests, the lab would provide all tests requested even if it meant duplication of services. A second reason this lab would be a problem for the study is that the tests ordered by the physician were transcribed into AHCIP fee codes in a manner quite different from practice in other labs and it was not feasible to devise a simple criterion for classifying test ordering behaviour if Lab C were included.

Dropping this laboratory meant a loss of approximately 13% of the data in terms of number of services and amount paid for lab tests but less than 5% of physicians. Exclusion of the data for this laboratory group may affect the results for the Calgary census division and/or the 'urban' category.

Thus, following data cleaning and evaluation of outliers, the final data base contained claims data and physician and practice information for 1636 general practitioners or 92 7% of the original database. The data were sorted and initial reports produced using SAS on a mainframe computer. The later steps of data cleaning and the subsequent data analyses were done on a personal computer using SPSS for Windows

3.0 RESULTS:

The study population was limited to all general practitioners (including family practitioners with CCFP accreditation) who practiced in Alberta, who were paid (in part or in full) by the Alberta Health Care Insurance Plan for services provided between July 1, 1989 to June 30, 1990, and who had at least one patient tested for thyroid functioning at a provincially funded laboratory during that period.

3.1 Univariate and Bivariate Analysis

Univariate and bivariate analysis of the effect of several physician and practice-specific variables on a physician's score will be examined. Potential interaction variables will be examined during the multivariate analysis phase. The probability level for all analyses in this paper will be set at $p < .05$.

Lower than desirable statistical power may affect some the subsequent analyses The probability of making a Type II error (failing to detect a real difference) is called beta The quantity 1 - beta is called power, the probability of observing an effect in the sample. Power can best be managed by ensuring the sample size is adequate The 'sample' for this study is actually the 'population', ie. all general practitioners in the province who fit the entrance criteria. Overall, the sample size is adequate but it is a predominantly male sample (1197 males and 439 females). When the data are broken down by more than one variable such as gender and place of graduation or region of the province, the number of females in each cell sometimes becomes lower than desirable. Some

of the non-significant results may be a reflection of low statistical power.

3.1.1 Dependent variable

The dependent variable for each practitioner is a score representing the proportion of times he/she ordered TFT in a manner consistent with one or more of the existing TFT guidelines. This is a continuous variable It will be referred to as Score for the remainder of the paper

Scores ranged from zero (0.0) to one (1.0); the mean score being 0.37. Approximately one third of all practitioners had a score less than 0.1 (or 10%). The remainder of the scores were equally distributed across the rest of the scoring range.

The distribution of physician scores is reported in Figure 2 and Table 6 on the following table:

Figure 2 Histogram of Physician Scores

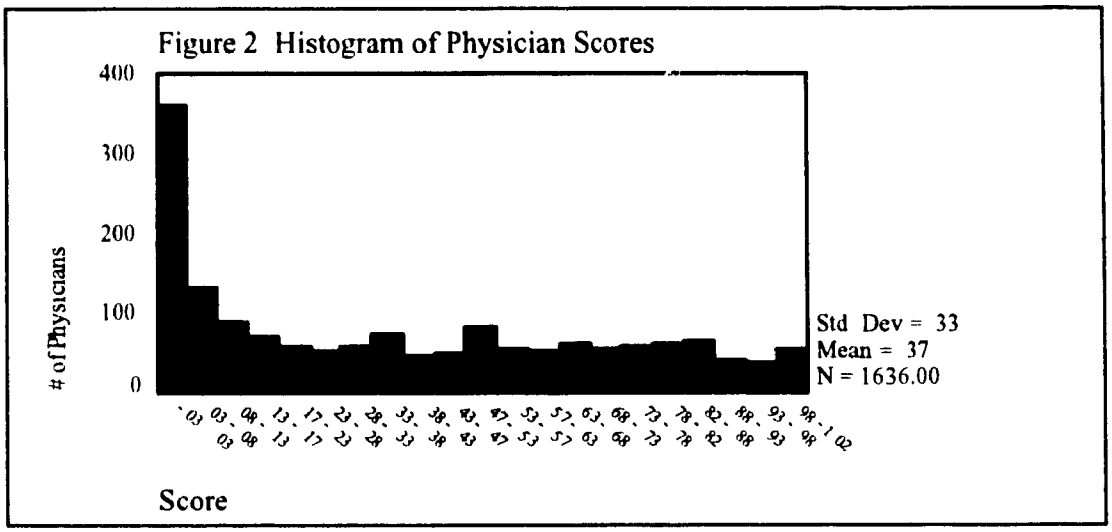


Table 6 Frequency Distribution of Physician Scores

Score	Frequency	Percent	Cum Percent
<.03	362	22.1	22.1
.03-<.08	132	8.1	30.2
.08-<.13	91	5.6	35.8
.13-<.17	72	4.4	40.2
.17-<.23	58	3.6	43.8
.23-<.28	54	3.3	47 1
.28-<.33	61	3.7	50.8
.33-<.38	75	4.6	55.4
.38-<.43	48	2.9	58.3
.43-<.47	50	3.1	61.4
.47-<.53	84	5.1	66.5
.53-<.57	56	3.4	69.9
.57-<.63	52	3.2	73.1
.63-<.68	64	3.9	77.0
.68-<.73	55	3.4	80.4
.73-<.78	61	3.7	84.1
.78-<.82	62	3 8	87.9
.82-<.88	65	4 0	91.9
.88-<.93	40	2.4	94.3
.93-<.98	39	2.4	96.7
.98-1.00	55	3.4	100.0
Total	1636	100.0	100.0

Mean .3678 SD .326 Min .0000 Max 1.0000

These results support the first hypothesis, which states "There will be variation in the degree to which general practitioners in Alberta comply with existing thyroid function testing guidelines."

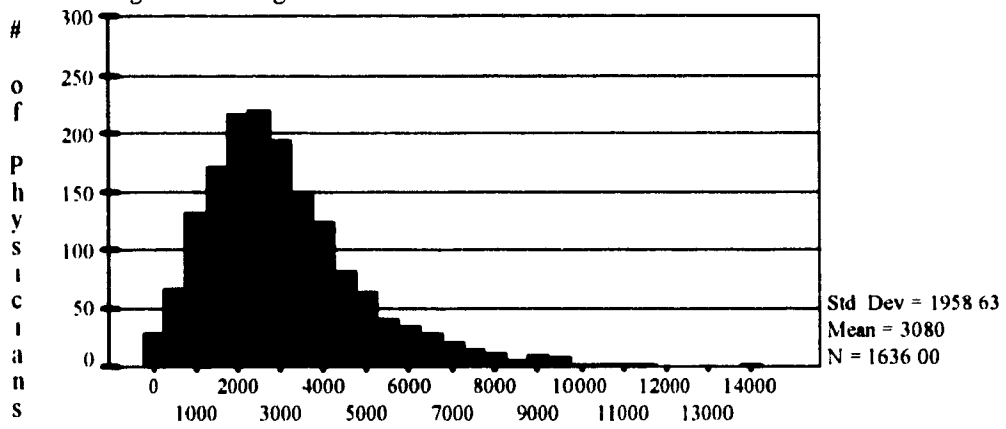
3.1.2 Independent variables

(1) *Gender of practitioner*. There are 1197 males and 439 females in the study, a ratio of 2.7·1 The average score for males is .3899; for females 3078. The standard deviations (SD) are similar. A T-test for the difference between means is significant at $p < .0001$ These findings illustrate two major points. First, female scores are significantly lower than those for males. Second, because males form the majority of the study population their results will tend to influence the outcome for the group as a whole. Because of these differences and because gender is generally considered to be a factor in various practice parameters, scores for male and female physicians are examined separately where appropriate.

(2) *Total Patient Caseload*

The range for total patient caseload (total number of patients seen by a physician for all causes) is extremely wide (6 to 14,414 patients seen during the study period). The average caseload is 3,080 patients, with 95% of physicians seeing between 6 and 6,998 patients. Eighteen physicians saw in excess of 9,500 patients. Figure 3 below shows the number of physicians by the total patient caseload.

Figure 3 Histogram of Total Patient Caseload



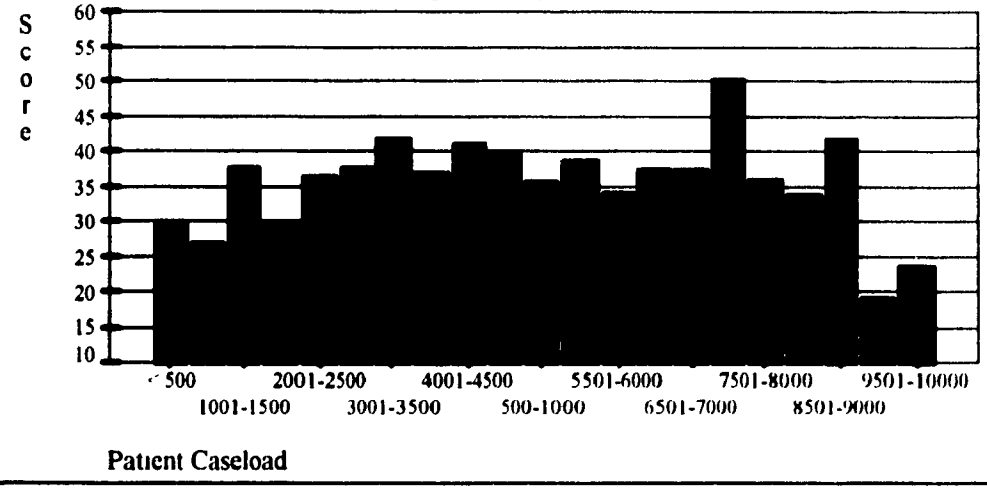
The effect of total patient caseload (as a continuous variable) on Score is not significant at $p < .05$. Table 7 and Figure 4 below categorize total patient caseload into groups, showing the number of physicians and average score per group.

Table 7 Frequency Distribution and Average Score per Caseload Category

Patient Caseload	Number of Physicians	Percent	Average Score per Caseload Category
< 500	59	3.6	.30
500- 1000	97	5.9	.27
1001- 1500	152	9.3	.38
1501- 2000	203	12.4	.30
2001- 2500	215	13.1	.37
2501- 3000	211	12.9	.38
3001- 3500	174	10.6	.42
3501- 4000	140	8.6	.37
4001- 4500	91	5.6	.41
4501- 5000	76	4.6	.40
5001- 5500	55	3.4	.36
5501- 6000	33	2.0	.39
6001- 6500	30	1.8	.34
6501- 7000	25	1.5	.38
7001- 7500	16	1.0	.38
7501- 8000	12	.7	.50
8001- 8500	13	.8	.36
8501- 9000	8	.5	34
9001- 9500	8	.5	.42
9501+	18	1.1	.23
Total	1636	100.0	.37

Mean .37 SD .33 Variance .11 T-test .944 Sig T .3455

Figure 4 Bar Chart of Average Score by Patient Caseload



The differences between the mean scores of the various patient caseload categories, as shown in Table 7 and Figure 4, are not significant. However, it should be noted that the eighteen physicians who saw more than 9,500 patients during the study period had comparatively low scores.

A profile of these 18 physicians shows that 16 are males; the average year of birth is 1953; 17 practice in a metropolitan area; 14 in a clinical setting; 10 in the Edmonton census division and 8 in the Calgary census division; 9 graduated from the University of Alberta, 2 from the University of Calgary; 4 are CCFP certificants; and, on average, they tested only 82 patients for thyroid function testing during the study period. In summary, this group is comprised largely of male physicians working in large volume clinics in Edmonton and Calgary One could speculate that these are large walk-in clinics that generally provide a very basic level of care. Although they see many more patients than average they are testing very few for thyroid functioning and as such will not unduly bias the findings of the testing scenarios.

Total Patient Caseload (controlling for gender)

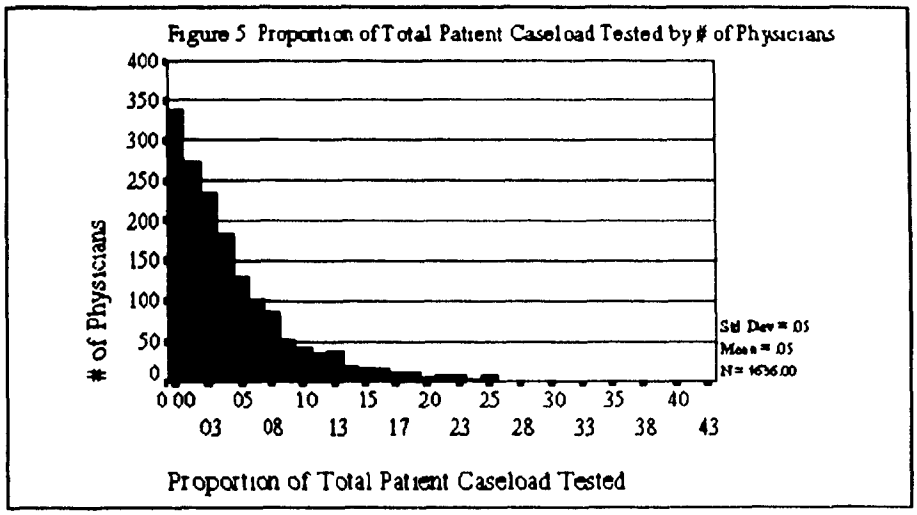
On average, male physicians had a patient caseload of 3,306 (SD 2,027), compared to 2,464 (SD 1,608) for females. This difference in practice size is significant at $p < .01$. However, when total patient caseload is regressed on Score, both in total and controlling for gender, the relationship remains non-significant.

In summary, there appears to be little direct relationship for either males or females between their score and their total patient caseload. Because total patient caseload is significantly different between males and females, and because this practice parameter may interact with other independent variables, it will be examined as a possible interaction term in the multivariate phase of analysis.

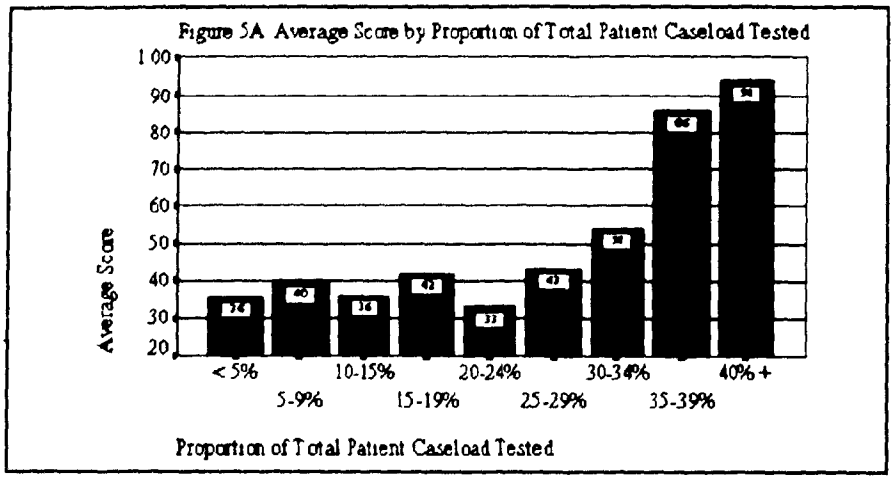
(3) *Percentage of Total Patient Caseload tested for Thyroid Functioning*

The percentage of total patients tested for thyroid functioning ranges from .01% to 42.5%, the average being 4.7%. This distribution is skewed to the right, with over half the physicians testing less than 3% of their patients. The range within $\pm$ 2 SD was quite narrow, 0 to 15.5% The average percentage of patients tested by female practitioners was 6.4%, compared to 4.1% for males. The T-test for the difference between the mean proportions tested for males and females is significant at $p < .01$.

Figure 5 shows the number of physicians by the proportion of their total patient caseload tested for thyroid functioning:



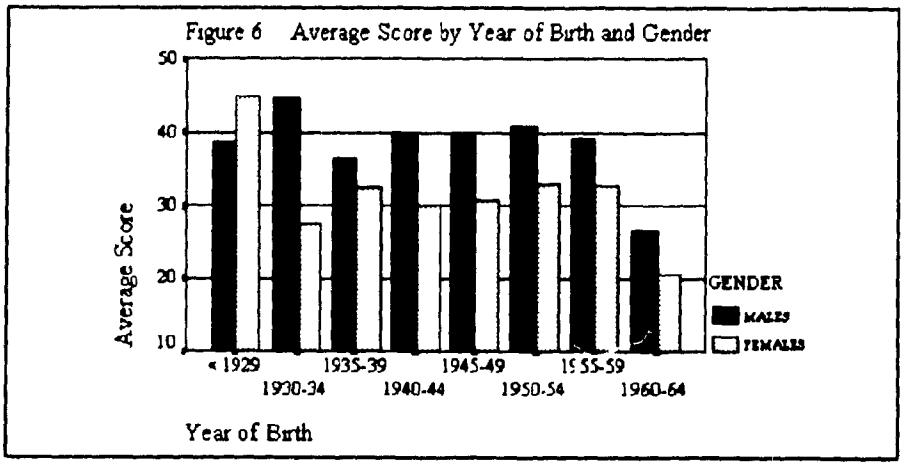
As Figure 5A on the following page illustrates, the relationship between Score and the proportion of total patient caseload tested is slightly positive. Score increased 0.4% for every increase of 1% in the number of patients tested for thyroid functioning This relationship is significant at the $p < .01$ value. It remained significant at this level when gender was controlled.



In summary, a physician's score, regardless of gender, increases approximately 0.4% for every 1% increase in the proportion of patients tested and is significant at $p < .01$.

(4) *Physician's year of birth and year of graduation.*

These two variables are understandably highly correlated. The distribution of year of birth appears fairly normal, skewed slightly to the left. The average year of birth is 1946 with 95% of all physicians falling within the range 1923 to 1964. The average year of birth is 1945 (SD 11.7) for males; 1952 (SD 8.5) for females. Figure 6 reports Score by year of birth and gender:



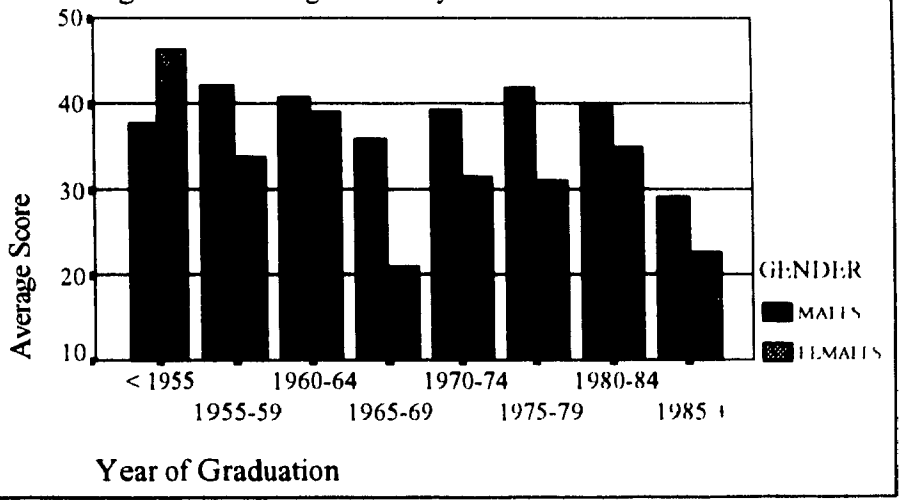
Note: The smallest n per cell is 9 for females with year of birth in the 1935-39 category.

Overall, there is a slightly negative relationship between year of birth and Score. Younger physicians have lower scores than older physicians. For every increasing year of birth, Score decreases by 0.2%. This relationship is significant at $p < .01$ (Sig T= .0085). However, this relationship is not significant when gender is controlled.

Similarly, year of graduation is skewed to the left. The average year of graduation was 1973, with 95% of all physicians graduating since 1952. The average year of graduation is 1971 (SD 11.3) for males and 1977 (SD 8.0) for females. This relationship was also significant, but not quite at the level of year of birth (Sig T = .0110). As above, this variable is not significant when gender is controlled.

As Figure 7 shows, more recent graduates appear to have lower Scores.

Figure 7 Average Score by Year of Graduation and Gender



Note: The smallest n per cell is 8 for females in the < 1955 category.

In summary, there is little difference in the ability of year of birth or year of graduation to predict Score. Because age is such as important demographic factor and because there is a significant difference in average age between male and female physicians (a factor not picked up by year of graduation) I chose to retain year of birth for the remainder of the analysis. Year of

graduation will be dropped.

(5) Practitioner's university or country of graduation.

Forty-eight different universities or countries of graduation were represented in the study population. This information was missing for only thirteen physicians. These counties and universities were aggregated along geographical and political lines in Table 8 if their average scores were similar, otherwise they were left as a separate group.

Table 8 Average Score by University or Country of Graduation

University/Country of Graduation	Average Score	SD	(m,f)
USA/Mex/other	477	.368	(15,2)
Saskatchewan	.437	.317	(40,8)
Univ of Calgary	420	.336	(102,76)
Ontario/Quebec	.379	.331	(70,46)
Univ of Br Columbia	375	.287	(18,9)
Univ of Alberta	.370	.314	(469,171)
UK/Ire/Aust/NZ/Africa	364	.343	(328,63)
Manitoba	.349	.344	(35,7)
Europe/USSR	.307	.328	(36,25)
India/Pakistan/Asia	.258	.292	(69,26)
Maritimes	.206	.268	(7,6)
missing = 13	F = 2.190	Sig F = .0102	

There is considerable variation among groupings. Of particular interest to this study is the notable difference between the University of Calgary (.420) and University of Alberta (.370). The difference among the average scores of all groups is significant at $p < .05$.

When gender is controlled, the relationship is significant at $p < .01$ for males but not significant at $p < .05$ for females.

(6) Census division of practice

There were 19 different census divisions identified in this data set. Scores ranged from .1636 to .7797, the average being .3678. Some very small census divisions were aggregated, if they were geographically adjacent to each other and their average scores were similar, to maintain the confidentiality of individual practitioners. This resulted in the 15 groups shown in Table 9 on the following page

Table 9 Average Score for Each Census Division Grouping

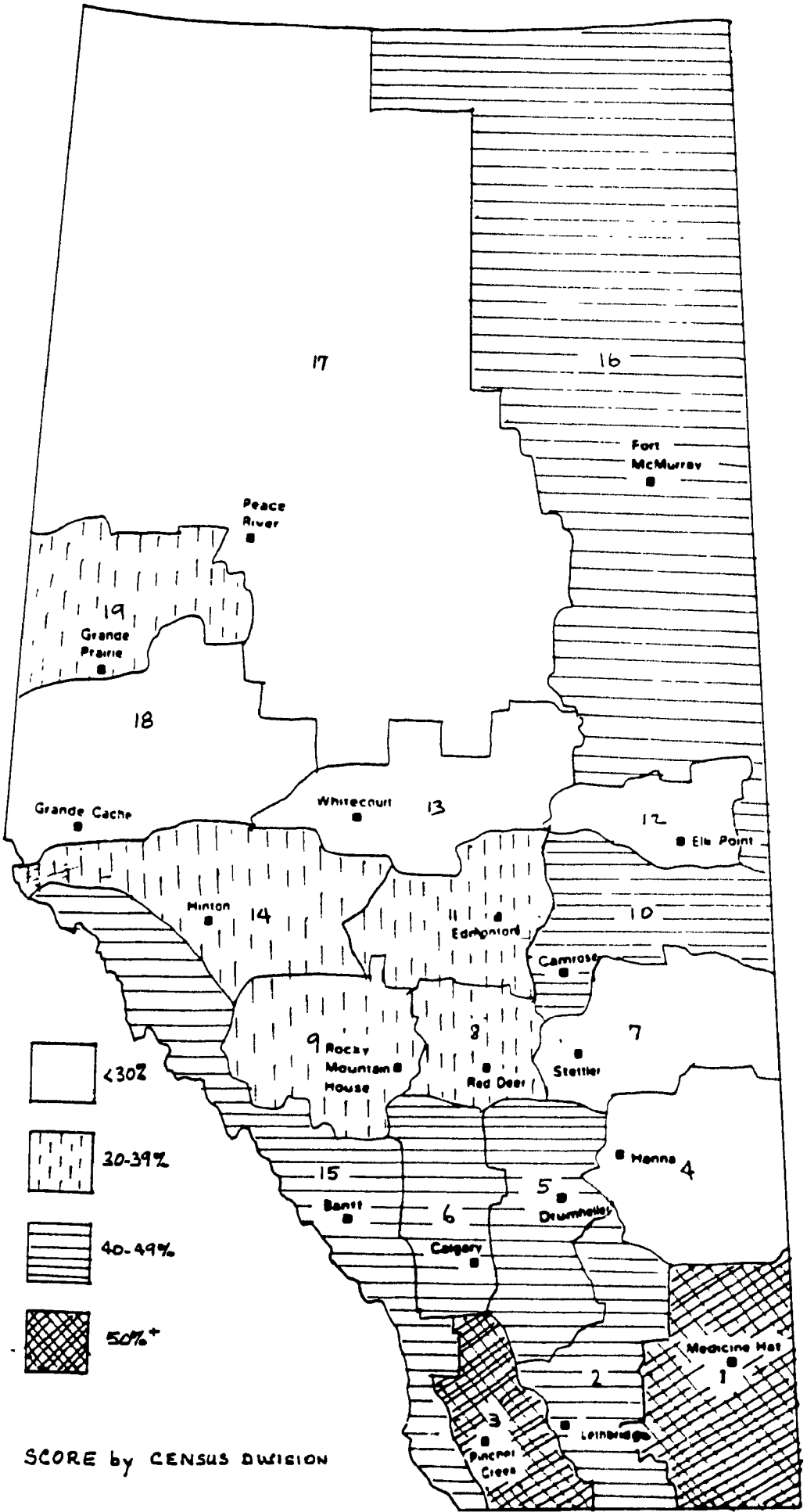
Census Division Grouping	Avg Score	SD	n	(m,f)
Entire Population	.3678	3263	1636	(1197,439)
Medicine Hat	.7795	.1459	37	(33, 4)
Pincher Creek	.5831	3278	20	(18, 2)
Fort McMurray	.4346	.3220	24	(22, 2)
Lethbridge	.4287	.3036	62	(50, 12)
Camrose	.4257	.3230	41	(36, 5)
Drumheller	.4167	.3673	22	(20, 2)
Calgary	.4002	.3273	493	(323,170)
Banff	.3992	.3945	10	(3, 7)
Grande Prairie	.3741	.3442	32	(27, 5)
Red Deer	.3622	.3249	67	(55, 12)
Edmonton	.3255	.3151	713	(512,201)
Hinton/Rocky Mnt House	.3014	.2884	30	(25, 5)
White Crt/Elk Pt	.2570	.2985	36	(32, 4)
Peace River/Grande Cache	.2137	.2990	33	(27, 6)
Stettler/Hannah	.1636	.2417	16	(14, 2)

The difference among the average scores was significant at $p < .0001$.

This table and the map on the following page illustrate that there is similarity in many adjacent geographic areas. While it does not hold true for all census divisions, there appears to be a general trend towards higher scores in the southern third of the province (Calgary and south), average scores in the central region which includes Edmonton, and lower scores in the northern region. Factors such as centre of influence, (eg., University of Calgary or perhaps individuals of influence) may contribute to this trend. It may also reflect the characteristics of physicians

(eg. age, gender, university or country where training occurred) who practice in urban or rural settings. The large difference in average Score between Calgary and Edmonton is interesting. So too is the extreme variation in scores across some urban and rural areas such as Medicine Hat and Pincher Creek, where scores are very high compared to Stettler with its very low average score

Because there appears to be a geographical trend, and because some census division groups still had a very small n when controlling for gender, these groups were further aggregated into north,



SCORE by CENSUS DIVISION

central, and southern geographical areas. Table 10 reports the average score by these three major census division groupings. The differences in average score remain significant at $p < .0001$ when aggregated

Table 10 Average Score for each Major Census Division Grouping

Census Division Grouping	Average Score	SD	n (m,f)
Entire Population	.3678	.3263	1636
South (Medicine Hat, Pincher Creek, Calgary, Banff, Drumheller, Lethbridge)	.4310	.3319	644 (447,197)
Central (Hinton, Edmonton, Red Deer, Camrose, Stettler)	.3293	.3152	867 (642,225)
North (Grande Prairie, Peace River, White Court, Fort McMurray)	.3097	.3231	125 (108,17)
	F = 20.58	Sig F = .0000	

The South region includes the six census divisions in the southern third of the province, including one of the two major cities (Calgary) in the province and one of the two large universities (University of Calgary). The other census divisions in this groping are: Medicine Hat, Pincher Creek, Banff, Drumheller, and Lethbridge. The majority of divisions within this region have higher than average scores.

The Central region includes the five census divisions in the central region of the province; Red Deer, Camrose, Stettler, Hinton/Rocky Mountain House, and Edmonton. This region contains the other large metropolitan area (Edmonton) and the other major university (University of Alberta) in the province. The average score for most of the divisions in this region is close to the overall average. The exceptions are Camrose, whose score is higher and similar to those scores noted in the South region, and Stettler which has the lowest average score in the province.

The North region includes the remaining areas, which are all in the northern part of the province and are either rural or urban centres. This area includes: Grande Prairie, Peace River, White

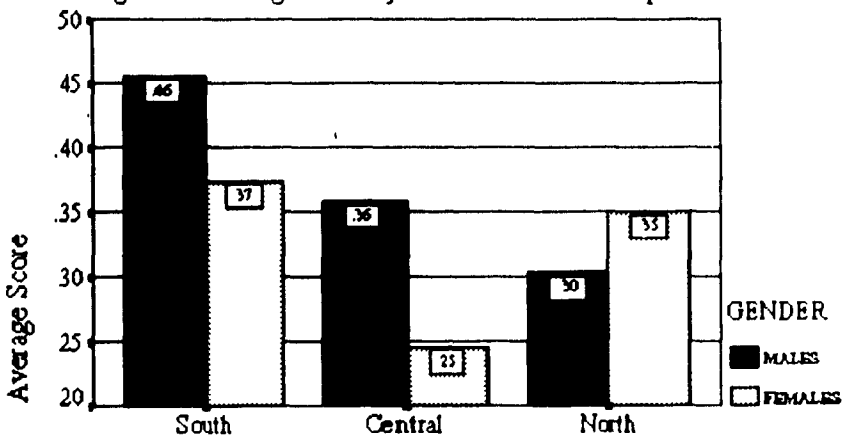
Court, and Fort McMurray. The scores range from higher than average in Fort McMurray to lower than average in Peace River and White Court. Grande Prairie is close to the overall average.

This grouping attempts to retain the important differences noted above, such as the wide range in variation between census divisions, the trend to geographic similarities and the notable difference between the two major cities and universities in the province.

Census Division Grouping Controlled for Gender

The difference in the average scores between regions remains significant at $p < .0001$ when gender is controlled. The trends previously noted hold for males but not for females. As Figure 8 below shows, females in the south and north regions score quite closely but those in the central region are significantly lower. This difference will be examined more closely during the multivariate phase of analysis.

Figure 8 Average Score by Census Division Group and Gender



(7) Location of practice defined as metropolitan, urban, or rural

This variable offers a slightly different perspective from that of census division, because a specific census division may have a mix of metropolitan, urban and rural practitioners. For example, Grande Prairie census division has urban and rural practitioners, while Edmonton and Calgary census divisions have both metropolitan and urban practitioners.

Metropolitan includes the cities of Edmonton and Calgary only Surrounding and adjacent cities such as St Albert and Spruce Grove are considered urban. Urban also includes all other centres in the province designated as cities by Statistics Canada. Rural includes all remaining areas. The following table reports the average Score by location of practice:

Table 11 Location of Practice by Average Score

	Average Score	SD	n (m,f)	
Total	.3678	.3263	1636	
Metropolitan	.3536	.3234	1102	(749,353)
Urban	.4623	.3127	259	(215,44)
Rural	.3359	.3356	275	(233,42)

F 13.45 Sig F .0000

It is clear from the table above that urban practitioners have higher scores than their metropolitan or rural counterparts. The relationship of this variable to Score is significant at $p < .0001$. When gender was controlled males remained significant at $p < .0001$, females at $p < .05$.

(8) Type of practice, clinic or solo.

Type of practice (either clinic or solo) is designated by the AHCIP. In general, a clinic practice is one where physicians share common office space and patient files and submit billings as a group using a group billing number. Solo physicians work independently of other physicians and submit billings under a solo billing number. The following table illustrates that there is little

difference in Score between types of practice:

Table 12 Average Score by Type of Practice

	Average Score	SD	n (m,f)
Total	.3678	.3263	1636
Clinic	.3773	.3269	761 (557,204)
Solo	.3596	.3258	875 (640,235)

F 1.1964 Sig F .2742

In summary, this variable is not significant at $p < .05$ either in total or when gender is controlled.

(9) Physicians who hold a certificate from the College of Family Physicians of Canada

The first exams set for CCFP certification were in 1969. As would be expected, this variable is positively correlated with more recent years of birth and with graduation from Canadian universities. The following table shows Score by CCFP certification:

Table 13 Average Score by CCFP certification

	Average	SD	n (m,f)
Total	.3678	.3263	1636
no	.3477	.3284	1095 (857,238)
yes	.4085	.3185	541 (340,201)

F 12.632 Sig F .0004

It is clear that having CCFP certification is positively correlated with a higher average score. The score for those with CCFP certification was .41 compared to .35 for those without. This overall difference is significant at $p < .001$.

This difference is particularly strong for males. The average score for males with CCFP

certification was .45, compared to .36 for those without CCFP, and remained significant at $p < .0001$. For females the difference was .33, compared to .29 for those without CCFP but was not significant. The findings for males and females remain the same when age is controlled.

This concludes the section on univariate and bivariate analysis. Table 14 summarizes the level of significance for each variable in total and by gender. When the group as a whole was examined, all independent variables except total patient caseload and type of practice (clinic or solo) were significantly related to the dependent variable Score at $p < 0.05$. However, when gender was controlled, only percentage of patients tested, census division groupings, and location of practice remained significant for both genders.

Table 14 Level of Significance for Each Independent Variable Regressed on Score by Gender

Independent Variable	Total Group	Males	Females
Gender	.0000		
Total Patient Caseload	.3455	.7638	.6359
% of Patients Tested	.0070	.0094	.0056
Year of Birth	.0085	.2610	.2609
Country/Univ of Origin	.0102	.0046	.4502
Census Division Group	.0000	.0000	.0000
Location of Practice	.0000	.0002	.0194
Type of Practice	.2742	.3677	.5231
CCFP Certification	.0004	.0000	.1437

In summary, the following are the highlights noted during this phase of the analysis:

(a) The average score achieved was .37, with a full one-third falling below .10. The remaining two thirds of the physicians were evenly spread across the full range of the dependent variable Score.

(b) This is a predominately male sample. The ratio of males to females was 2.7:1. Males scored significantly higher than females (39% versus 31%). However, because gender is correlated with several of the other independent variables, its direct effect on the dependent variable Score is

uncertain at this time. The multivariate analyses in the next stage will be run separately for males and females.

(c) While the difference in average patient caseload between males and females is significant, the effect of this variable on Score is not significant in total or by gender. Because males and females differ so strongly on this variable it will be examined during the multivariate phase of analysis.

(d) The average percentage of patient caseload tested for thyroid functioning was 4.7% but over half tested less than 3% of their caseload. Females tested a significantly higher proportion of patients than male practitioners (6.4% for females versus 4.1% for males). The data also indicate that, in general, the higher the percentage of patients tested the higher the Score. This relationship was significant in total and by gender. Despite the facts that females tested proportionately more patients and that testing proportionately more of one's patient caseload is associated with an increased score, the female average score remains significantly lower than the average male score. There is also an inverse relationship between total patient caseload and proportion of patients tested for thyroid functioning. This trend is influenced, in part, by the eighteen practitioners who saw in excess of 9,500 patients but tested very few for thyroid functioning.

(e) The average year of birth was 1945 for males and 1952 for females. Overall, younger physicians tend to score lower than older physicians. Year of birth is significantly related to Score in total, but not when gender is controlled. This is presumably related to the fact that females are significantly younger and score significantly lower than their male counterparts.

(f) Physician's scores varied greatly depending on university or country of graduation. Approximately 65% of the study group graduated from Canadian universities, 24% from the United Kingdom, Ireland, Australia, New zealand, or Africa, and 6% (69 males, 26 females) from India, Pakistan, or Asia. Of particular interest is the notable difference in scores between graduates from the University of Alberta (in Edmonton) and the University of Calgary (in

Calgary). The average score for U of A graduates was 37%, compared with 42% for U of C graduates. The relationship between this variable and Score was not significant for females when gender was controlled.

(g) The same trend is noted when score is examined by census division. Physicians practicing in the Calgary area scored significantly higher than those in the Edmonton census division. When census divisions were aggregated by geographic area of the province the differences in scores remained significant in total and by gender. Practitioners in the South scored significantly higher than those in the central or northern regions for both males and females. However, males scored lowest in the northern regions while females scored lowest in the central region.

(h) Both male and female physicians practicing in an urban location had significantly higher scores than those in metropolitan or rural areas. Metropolitan scores were only slightly higher than those in rural areas. Metropolitan scores may have been lower, in part, because 80% of the female practitioners, who generally scored lower, practice in this location compared to only 63% of males. In fact, in this study group there are only 44 females in an urban setting and 42 in a rural setting. This relationship will be explored during the multivariate phase of analysis.

(i) No significant difference in average score was noted between clinic and solo practitioners.

(j) Having CCFP certification is positively correlated with Score for males (.45 versus .36). This trend is the same for females but not significantly so (.33 versus .29). Of interest is the difference in the number of practitioners with this certification when gender is controlled. Approximately 29% of males had CCFP certification compared to 46% of females.

3.2 Multivariate Analyses

The analysis in the previous section of the paper identified variables that appear to affect a physician's score. Of the nine independent variables examined, seven will be included in the

multivariate phase of the analysis. Two continuous variables (proportion of patients tested for thyroid functioning and year of birth) and five categorical variables (gender, census division grouping of practice, location of practice - metropolitan, urban or rural, university or country where training occurred, and having CCFP certification) will be retained because they were significant either in total or by gender in the previous analysis. Total patient caseload and type of practice (clinic or solo) will be excluded.

The categorical variables were set up as 'dummy variables'. Each reference group was chosen on the basis of being either the category with the largest number of subjects or the category most relevant to the study. The following table details the variables used in this phase of the analysis and identifies the reference groups:

Table 15 Description of 'Dummy Variables' Showing Reference Groups

Variable	Reference Group	Dummy Variables
Gender	males	Gender F (females)
Census Division Group	Central (includes Edmonton)	North South(includes Calgary)
Practice location	metropolitan	urban rural
Place of graduation	University of Alberta	USA/Mexico/other Saskatchewan Univ of Calgary Ontario/Quebec Univ of Br Columbia UK/Ire/Aust/NZ/Africa Manitoba Europe/USSR India/Pakistan/Asia Maritimes
CCFP certification	no	yes

Note: All dummy variables were coded with the category of interest = 2 and all other categories = 1. For example, the dummy variable 'urban' was coded as urban=2, while rural and metropolitan were coded = 1. Similarly, the dummy variable 'rural' was coded as rural=2, while urban and metropolitan were coded = 1, and so on.

This phase of the data analysis was done to test the second hypothesis that "One or more of the following factors, associated with general practitioners or their practice, is correlated with the degree of compliance: gender, age (measured by year of birth), place of graduation, location of practice (metropolitan, urban, rural), geographic area of practice (census division group), proportion of patients tested for thyroid functioning, and whether or not they have certification by the College of Family Physicians of Canada (CCFP)."

The following describes the steps that will be taken and the statistics used for the remainder of the analysis. As discussed earlier, all regressions were run separately for males and females because they are significantly different on a number of variables.

Step 1: The variables will be entered into the model individually or in blocks, using the SPSS stepwise method. The p level for inclusion in the model will be set at .05. The p level for exclusion will be set initially at .20 to ensure that variables are not prematurely rejected.

Step 2: A trimmed regression model will be run retaining only the variables that were significant at $p < .05$ in Step 1. In addition, any possible interaction terms will be added.

Possible interactions will be determined by examining; (a) any correlation greater than .200 on a correlation matrix of the continuous and "dummy" variables, (b) regression models run with each of the independent variables as the dependent variable observing any significant relationships, and (c) possible interactions based on theory or experience.

These variables and interaction terms will be added as a group using the stepwise method. As before, the analysis will be run separately for males and females. The p values for inclusion and exclusion will remain the same.

Step 3: Variables and interaction terms that are significant at $p < .05$ will be examined to determine the degree, slope, and significance of the relationship of the variables and interaction terms to the dependent variable. The adjusted R-square will be interpreted as the amount of

variation in Score that can be explained by the set of independent variables and interaction terms left in the model.

The statistics used are.

(a) Rsq is used as a measure of the goodness of fit of the model or in simpler terms, it tells me how much of the variation in the dependent variable can be explained by the independent variable being examined. The assumption is that the model is linear The adjusted Rsq, used in this table, corrects Rsq to more closely reflect the goodness of fit of the model in the population. The cumulative Rsq (Cum Rsq) is simply the sum of the adjusted Rsq already in the model

(b) Beta tells us the direction and the degree of the linear relationship between the dependent and the independent variables. For example, it tells me if there is a positive or a negative relationship between the dependent and independent variables and it also tells me to what degree the value of the dependent variable changes for every unit change in the x axis or the independent scale. A standardized beta, used in this table, adjusts the betas to a standard score so they are comparable

(c) The significance test tells me the probability that the differences noted could have occurred by chance. For this study I set the probability level at .05. This means that if there was more than a five percent probability that these results could have occurred by chance I will not consider the results significant.

Step 1: The following table presents the results from Step 1 of the multivariate regression analysis:

Table 16 Regression Results from Step 1 (block entry, stepwise)

Males:

Block #	Variable*	Adj Rsq	Cum Rsq	Standardized Beta	Sig T
1	Year of birth				not sig
2	Proportion tested	.0045	.0045	.0777	.0062
3	Census division group				
	South	.0237	.0282	.1381	.0000
	North				not sig
4	Location of practice				
	Urban	.0148	.0430	.1255	.0000
	Rural				not sig
5	Country/univ trained				
	India/Pakistan	.0069	.0499	-.0801	.0050
	Maritimes	.0047	.0546	-.0741	.0086
	all others				not sig
6	CCFP certification	.0053	.0599	0803	.0057

Females:

Block #	Variable*	Adj Rsq	Cum Rsq	Standardized Beta	Sig T
1	Year of birth				not sig
2	Proportion tested	.0156	.0156	1335	.0041
3	Census division group				
	South	.0404	.0560	.2120	.0000
	North				not sig
4	Location of practice				
	Urban	0166	.0726	.1366	0034
	Rural				not sig
5	Country/univ trained				
	all categories				not sig
6	CCFP certification				not sig

* Only those subcategories of the variable which were significant are listed in the table. The others did not meet the $p < .05$ inclusion criteria.

Table 16 indicates that for males, testing proportionately more patients for thyroid functioning, practicing in the southern third of the province, practicing in an urban location, and having CCFP certification, are all predictive of having a higher than average score On the contrary, having graduated from a university in India, Pakistan, Asia or the Maritimes is related to having a lower then average score.

The variation in Score for males explained in this model (as measured by adjusted Rsq) is only 5.99% Testing proportionately more patients (0.45%), practicing in a southern third of the province (2.37%), practicing in an urban setting (1.48%), or having CCFP certification (53%) are all predictive of having a higher than average score. Having graduated from a university in India or Pakistan (.69%), or a Maritime university (.47%) suggest a lower than average score.

For females, three variables explained approximately 7.26% of the variation noted in Score. These factors are practicing in the southern third of the province (4 04%), practicing in an urban location (1.66%), and testing proportionately more patient (1 56%). All were positive predictors of a higher than average score.

Step 2:

The significant variables identified in step one will form the trimmed regression model used for step 2. In addition, many possible interaction terms were examined.

Some of the correlations noted between independent variables were obvious such as, an association between the Calgary census division and having trained at the University of Calgary, or the relationship between being in a clinic practice and having a larger total patient caseload, or showing that those with a more recent year of birth were more likely to be female or to have CCFP certification.

Other correlations were not so obvious but make sense such as, a positive correlation between physicians who graduated from a medical school in the UK, Ireland, Africa, Australia or New

Zealand and practicing in a rural area. This is reasonable because Alberta has historically recruited practitioners from these countries to fill rural practices. These foreign trained physicians also showed negative correlations between year of birth and having CCFP certification indicating they are generally older than the average study participant and are less likely to have CCFP certification. Given that CCFP certification is generally acquired by Canadian graduates this is expected. The reason these practitioners are older than average is unknown but it may be a result of an influx of UK physicians during the transition to our national health care plan in the nineteen sixties. Alternately, these physicians may have immigrated after several years of practice elsewhere.

After reviewing the various possibilities, the following interaction terms were examined in more detail:

Males

(1) Score by location of practice (metropolitan, urban, or rural) and area of the province (north, central, or south). These results are displayed in the following table:

Table 17 **Average Score (number of physicians) by Type of Practice and Area of the Province - Males**

	North	Central	South	TOTAL
Metro	none	.35 (436)	.42 (305)	.38 (741)
Urban	.42 (36)	.43 (107)	.55 (52)	.47 (215)
Rural	.35 (72)	.31 (96)	.52 (65)	.35 (233)
TOTAL	.30 (108))	36 (639)	.46 (442)	.39 (1187)

missing 8

Points of interest in this table are:

a) There are no metropolitan areas in the northern area of the province;

b) Practicing in any type of practice in the southern third of the province or in any urban region in the province related to a higher score. A combination of these two factors leads to the highest score in the table - 55%, and

c) No other general trends were noted, but three individual cells show average scores below the mean They are north rural, central rural and central metro

(2) Average score of Indian, Pakistani, or Asian graduates by type of practice and area of the province.

There are 69 male graduates from an Indian, Pakistani or Asian university in this study group Their average score is .25 compared to .39 for males overall. Of the 69 physicians, 7 practice in the northern region, 45 in the central region and 17 in the south The scores in all cells were less than .29 except for the cental urban area where these practitioners scored 42, similar to other male physicians in a central urban area whose average score was .43.

Graduates from a Maritime university were not examined further because there are only 7 male graduates and any results may be spurious.

(3) Percentage of practitioners with CCFP certification by the southern area of the province

Forty one percent of the male physicians in the southern region of the province held CCFP certification compared to only 21% of the males in the rest of the province. As noted earlier, having CCFP certification was significantly ($p < .00001$) and positively related to higher scores Thus, a combination of having CCFP and practicing in the southern third of the province predicts a higher than average score.

(4) Proportion of patients tested for thyroid functioning by location of practice (urban) and by graduates from India, Pakistan or Asia.

Urban male physicians tested approximately 3.4% of their patients for thyroid functioning compared to 4.2% by the remainder of the males in the province. This relationship is significant at $p < .05$.

Male practitioners who graduated form India, Pakistan or Asia tested approximately 2.5% compared to 4.2% by the rest as noted above. This relationship was significant at $p < 01$.

Females

(1) Score by location of practice (metropolitan, urban, or rural) and area of the province (north, central, or south). These results are displayed in the following table:

Table 18 **Average Score (number of physicians) by Type of Practice and Area of the Province - Females**

	North	Central	South	TOTAL
Metro	none	.24 (181)	.36 (167)	.30 (348)
Urban	.52 (7)	.33 (20)	.49 (17)	.42 (44)
Rural	.23 (10)	.20 (21)	.37 (11)	.25 (42)
TOTAL	.35 (17)	.25 (225)	.38 (197)	.31 (434)

missing 5

Points of interest in this table are:

a) There are no metropolitan areas in the northern area of the province;

b) Similar to the trends noted with males, working in an urban practice or in the southern third of the province is a predictor for higher than average scores. As noted earlier, it is unclear how exclusion of the data from this laboratory may effect the results seen for the Calgary census division and/or the 'urban' category. Working in a central metropolitan or rural area or in a northern rural

area are predictors of low scores for females On the contrary, females practicing in an urban setting in the south or the north scored particulary high at .52 and .49 respectively.

(2) Average scores for Indian, Pakistani, or Asian graduates by type of practice and area of the province.

There are 26 female graduates from an Indian, Pakistani or Asian university in this study group Their average score is .27 compared to .31 for the females as a whole. Of these 26 physicians, 13 practice in Edmonton and 10 in Calgary. There is little difference in the average score between these centres for this group.

Graduates from a Maritime university were not examined further because there are only 6 female graduates and any results may be spurious

(3) Percentage of practitioners with CCFP certification by the southern area of the province.

Sixty one percent of the female physicians in the southern region of the province held CCFP certification compared to only 32% of the females in the rest of the province. As noted earlier, having CCFP certification was significantly ($p < 001$) and positively related to higher scores Thus, a combination of having CCFP and practicing in the southern third of the province predicts a higher than average score.

As would be expected very few practitioners who were not Canadian trained held CCFP certification. For example, only three out 26 female physicians trained in India, Pakistan, or Asia had this credential.

As a result of the examination noted above, the following interaction terms will be included in the next regression run. Again, the model will be run separately for males and females

Males	Proportion of patients tested * India/Pakistani/Asia graduate
	Proportion of patients tested * Urban location of practice
	Southern Region * India/Pakistani/Asia graduate
	Southern Region * CCFP certification
Females	
	Southern Region * CCFP certification
	Southern Region * India/Pakistani/Asia graduate

The following table details the results of the second regression run, which included all variables that were significant in the first run and the interaction terms noted above. All variable and interaction terms were entered in one block and run in a stepwise method. Following are the results of this regression run:

Table 19 Regression Results from Step 1 (block entry, stepwise)

Males:

Variable[1]	Adj Rsq	Cum Rsq	Standardized Beta	Sig T
Southern Region * CCFP certification	.0316	.0316	.1719	.0000
Location of practice				
Urban	.0131	.0447	.1044	.0003
Prop of patients tested * Urban practice	.0067	.0514	.0833	.0038
Country/univ trained				
India/Pakistan	.0051	.0565	-.0785	.0058
Maritimes	.0047	.0612	-.0742	.0085
Proportion of patients tested				not sig
Census Division Group				
South				not sig
CCFP certification				not sig
Proportion of patients tested * India/Pakistani/Asia graduate				not sig
Southern Region * India/Pakistani/Asia graduate				not sig

[1] Only those subcategories of the variable which were significant are listed in the table. The others did not meet the $p < .05$ inclusion criteria.

Females:

Variable[t]	Adj Rsq	Cum Rsq	Standardized Beta	Sig T
Census division group				
South	.0398	.0398	.2120	.0000
Location of practice				
Urban	.0170	0568	.1366	.0034
Proportion of patients tested	.0158	.0726	.1335	0041
Country/univ trained				
India/Pakistan				not sig
CCFP certification				not sig
Southern Region * CCFP certification				not sig
Southern Region * India/Pakistani/Asia graduate				not sig

The results from Table 19 will be discussed first for males, then for females.

Males

Overall, only 6.12% of the variation noted in males scores was attributable to the factors included in this model.

The combination of practicing in the southern third of the province and having CCFP certification is the most predictive factor identified for males. It explained 3.16% of the variation in Score and was significant at $p < .0001$. On their own, both variables are significantly related to the dependent variable at $p < .0001$. The Chi-Square for these two variables is also significant at $p < .0001$. However, when these two characteristics are entered into the model as an interaction term, they are no longer significant as individual variables Only the interaction term remains

In total, 28.4% of males have CCFP certification; this figure is 41% for males practicing in the southern third of the province compared to 22% in the north and 18% in the central region. The average score for males in the south is .4559 compared to .3584 in the central region and .3032 in the north region. The following table illustrates the average scores by region of the province

and by CCFP status:

Table 20 Average Score by Census Division Group and CCFP Status - Males

Region by CCFP	Average Score	(n)
North	.3032	108
CCFP-yes	.4401	19
CCFP-no	.2740	89
Central	.3584	642
CCFP-yes	.3832	140
CCFP-no	3515	502
South	4559	447
CCFP-yes	.5110	181
CCFP-no	.4183	266

The next largest contributing factor for males is practicing in an urban location. This variable contributed to 1.31% of the variation noted in Score for males and was significant at $p < .001$. As noted earlier, the average score for males in an urban practice is .4707 compared to 3787 in a metropolitan area and 3508 in a rural area and the correlation between location of practice (metropolitan, urban or rural) and census division grouping is low. Approximately 16.1% of practitioners in the south practiced in an urban location compared to 18% overall in the province.

The interaction term of an urban location with the proportion of patients tested for thyroid functioning was also significant ($p < .01$). Once this interaction term is introduced into the model, the independent variable 'proportion of patients tested' is no longer significant. This indicates that these two variables are closely related.

The final contributing variable was university or country from which the physician graduated. The contribution of this variable to variation in Score was low, approximately 1%. Having graduated from either a university in India, Pakistan, or Asia or from a Canadian Maritime

university was predictive of a lower than average score These two categories were significant at $p < .01$.

In summary, for males, the combination of practicing in the southern third of the province and having CCFP certification or practicing in an urban location are predictive of a higher than average score. On the contrary, having graduated from a university in India, Pakistan, Asia, or a Canadian Maritime university was predictive of a lower than average score

Females

The variables included in the model explained 7.26% of the variation noted in the dependent variable Score.

Practicing in the southern region of the province was the most predictive variable for females. Given that the average Score of female graduates from the University of Calgary was 3902 compared to an overall average for females of 3078 and given that 64 out of 76 of these graduates worked in the southern region of the province, the possibility of interactions was examined. None were significant. While the average score for females in the south was 4062, considerably higher than other regions, female graduates from most other Canadian universities who were practicing in the South also had high scores in comparison to their counterparts in other areas of the province. For example, female graduates from the University of Alberta (in Edmonton) scored almost the same as female graduates from the University of Calgary (4062 versus .4063). In addition, while having CCFP certification significantly increased scores for males and given that 62% of females in the south have CCFP certification, one could speculate that there was a relationship present for females This was not the case In fact, the average Score for females practicing in the South was almost identical for those with and without of CCFP certification.

Both practicing in an urban location and the proportion of patients tested contributed a little more than 1.5% each to the explanation of variation in the dependent variable.

In summary, for females, practicing in the southern region of the province, practicing in an urban location and testing proportionately more patients for thyroid functioning are all predictive of a higher than average Score. No negative predictors or significant interaction terms were identified.

4.0 DISCUSSION

4.1 What Results Suggest

4.1.1 Overview

The first result to note is the distribution and mean of the dependent variable Score. While the values spanned the full range of 0.0 to 1.0, the average score was low at .37. In fact, a full one-third of practitioners scored less than .10. The remainder were evenly distributed across the remainder of the range.

It might be argued that the criteria established to classify testing scenarios as appropriate or inappropriate may have been too rigid so that some scores were lower than appropriate. This is not the case. The scoring range and mean accurately reflect the way general practitioners in Alberta were utilizing thyroid function tests at the time of the study. First, the criterion summarized in Table 4 were quite liberal. In fact, any ordering that complied with any of the guidelines listed were accepted. Second, the criteria were developed with the assistance of a medical biochemist - endocrinology and a laboratory director associated with the University of Alberta, Department of Laboratory Medicine. Third, while many physicians did score very low, others scored very well, covering the entire range of the variable. Fourth, there were significant and consistent differences noted across geographic regions and educational qualifications. Fourth, no thyroid function testing guidelines had been broadly distributed or endorsed by any provincial

body prior to or during the study[u]. Thus, there was no one source for reference

A second major point is that little of the variation noted in the dependent variable could be explained by independent variables included in the model. The cumulative adjusted R-square was only 6.12% for males and 7.26% for females. While this is disappointing, it is consistent with the findings of other studies that have attempted to explain variation in practice by physician or practice specific factors. As noted above, studies examining variation in utilization have been able to explain only as much as 40% of the variation, of which two-thirds represents health status or need. This leaves, at best, approximately 13% of the explanation attributable to physician, patient or organizational factors. This study supports these findings in that only six to seven percent of the variation was explained by examining key physician and practice characteristics

4.1.2 Physician Specific Characteristics

This section of the paper discusses the findings for physician specific characteristics

Gender

Several studies[17,18,19,20,21] have examined income distribution, hours of work, caseloads, and the distribution of physicians by specialty and geographic area based upon gender. No studies examining variation in utilization of diagnostic testing by gender were identified

This study's population is predominately male, the ratio of male to female being 2.7:1. Female practitioners scored significantly ($p < .0001$) lower that their male counterparts. The average score for females was .3078 compared to .3899 for males

Several variables differed significantly between male and female practitioners. Some of the major

---

[u] Subsequent to this study, thyroid function testing guidelines were developed by the Alberta Medical Association and endorsed by all major health care organizations and medical schools in the province. These guidelines were distributed in April 1992

differences were: females had smaller average patient caseloads than males (2,363 versus 3,306, $p < .01$), they tested proportionately more patients for thyroid functioning (6.4% versus 4.1%, $p < .01$), they were, on average, younger than males average year of birth (1952 versus 1945, $p < .01$), they were more likely than males to have graduated from a Canadian university (74% versus 62%, $p < .01$), they were more apt to practice in the southern third of the province than males (45% versus 37% of males, $p < .01$), they were more likely to practice in a metropolitan area (80% versus 63% of males, $p < .001$), and they were more likely to have CCFP certification (46% compared to only 28% of males, $p < .01$).

Interestingly, most of these findings should have predicted higher scores for females because these characteristics for the group as a whole are associated with higher than average scores. For example, testing proportionately more patients, graduating from a Canadian university, practicing in the southern third of the province, and having CCFP certification were all predictive of having higher scores. Only the fact that females were younger than males and tended to practice in metropolitan areas would suggest lower than average scores. The latter were not found to be significant during the multivariate analysis. Thus, the lower scores for females are not explained by the variables included in this model. This significant difference in Score between males and females is either a direct result of gender or it is being influenced by other variables not considered in this model, such as patient specific characteristics.

Year of Birth

Some studies suggest that the age of a physician may be associated with the way he or she practices. An AHCRP study[14] found that physicians in practice less than 15 years had 27% inappropriate hospital admissions compared to only 20% inappropriate admissions for those in practice longer than 15 years. Eisenberg[12], Campbell[15] and Goldfarb[16] found that younger doctors tend to provide more ancillary services but shorter lengths of stay than older colleagues. Eisenberg[12] and Childs[13] noted that older physicians ordered fewer tests.

Data from the present study show a very weak negative relationship ($p=.0085$) between year of

birth and score. This trend is similar to the findings discussed above

The average year of birth of physicians in the study was 1946, 95% of all general practitioners in the study were born between 1923 and 1964. The average year of birth for females was 1952, for males 1945. This difference would be expected as the proportion of females graduating from medicine has increased substantially over the past twenty years

When this relationship was examined controlling for gender the results were not significant. This initial significance for the group as a whole probably results from the fact that females are younger and tended to score lower.

Year of birth was not retained as a significant predictor during the multiple regression analysis phase.

Country or University of Graduation

The variation in scores controlling for university or country of graduation ranged from 21% for those trained in a Maritime university to 48% for study participants trained in the United States, Mexico or the Caribbean. Generally, practitioners who trained in Canada had scores slightly higher than average. There was a significant difference in the average score between those trained at the University of Alberta (37%) compared to the University of Calgary (42%) The reasons for this difference are not apparent.

Multiple regression results showed that, for males, having graduated from a university in India, Pakistan, Asia, or the Maritimes was predictive of a lower than average score Various interaction terms with this variable were examined but none were significant.

No predictive values on this variable were identified for females.

CCFP Certification

Eisenberg and Nicklin[12] and Childs[13] suggest that family practitioners order more diagnostic tests than general practitioners. The authors do not speculate on whether this pattern is favourable or unfavourable. Data from this study show that overall, study participants with CCFP certification have significantly higher scores that those without the certification; 41% compared to 35%. This difference is even more pronounced for males where the difference in score is 45 5% compared to 36.5%. The trend for females is similar, 33% compared to 28.9% but this difference is not significant for females.

During the multiple regression phase, this variable was dropped but the interaction term with practicing in the southern third of the province was significant and therefore retained in the model.

As Table 20 illustrated, having CCFP certification clearly increases the chances of having a higher than average score, as does practicing in the southern third of the province. The combination of these factors is particularly strong.

Male practitioners in the south with CCFP certification scored .5110 compared to only .4183 for those in the south without CCFP certification. In addition, males practicing in the south were significantly more likely to have CCFP certification. The rates overall for males with CCFP certification were 28 4% but this figure rose to 41% for males in the south compared to 22% in the north and 18% in the central region of the province.

Given that 62% of the females in the southern third of the province had CCFP certification, one might speculate that this too would contribute to higher than average scores for them. It did not. The average score for females in the south was almost the same for those with and without CCFP certification.

4.1.3 Practice Specific Characteristics

This section of the paper discusses the findings relating to practice specific characteristics.

Studies by Eisenberg et al[10 29] suggest that where physicians practice and the influence of their peers affect the level of utilization of diagnostic testing, and that the effects of group style and peer pressure are probably stronger within more formally organized practices. For example, they noted that physicians at one teaching hospital felt pressure from others to be a major factor in influencing their ordering of diagnostic tests

Geographic location by census division, location of practice (metropolitan, urban and rural) and type of practice (solo or clinic) were examined in this study.

Census Division Grouping

Scores by census division ranged from a low of 16.4% in the Stettler/Hannah census division to a high of 78% in the Medicine Hat census division. As shown earlier on the census map, the scores of some adjacent areas tend to be similar This lends support to the idea that centres of influence or individuals of influence may play a role in setting standards or norms of practice The census divisions were grouped into three major geographical areas of the province based on major trade and transportation boundaries. Most areas within one of the major three regions had similar average scores.

Clearly, practicing in the southern region of the province was highly predictive of a better than average score. The score for the south was .4320 compared to 3293 in the central region and .3097 in the north. This trend held when examined by gender but the difference for females was not as marked between the central and north regions.

Practicing in the south was a significant variable for females ($p < .0001$) There average score was .4062, considerably higher than any other region. Having CCFP certification or having graduated from a Canadian university does not appear to interact with these findings as the results for females in all the south in any of these categories in similar. It appears there is something unique about practicing in this area that leads to higher than average scores

As noted above, practicing in the south is significant for males only as part of an interaction term with CCFP certification. Both variables appear to contribute to this phenomenon but the combination of the two is particularly strong.

Location of Practice (metropolitan, urban or rural)

Location of practice (metropolitan, urban or rural) is a significant predictor of compliance with thyroid function testing guidelines. More specifically, the data show that study participants practicing in an urban setting have the highest scores, 46.2% compared to 35.4% in metropolitan areas and 33.6% in rural areas. It is unclear whether or not the absence of data for tests done by rural physicians in public laboratories has any affect on the results

This variable was also retained in the regression model for both males and females as a positive predictor of higher than average scores. In addition , it was retained, for males, as part of an interaction term with the proportion of patients tested. As noted above, physicians in an urban setting tended to test proportionately fewer patient for thyroid functioning then was done in other practice locations.

Type of Practice (solo or clinic)

Williams et al.[30,31] point out that the dichotomization of practice as solo or group is insufficient and an oversimplification. They found that younger practitioners and female practitioners were less likely than older male physicians to be in solo practice. Paulick and Roos[32] concur that more recent graduates are more likely to be in group practice. These factors confound any variation between group versus solo practices Data from this study support these findings. The effect on score of solo versus clinic practice was not significant, and remained non significant when examined separately for males and females. And, as noted earlier, the percentage of physicians who were practicing as clinic versus solo practitioners was roughly equal for both males and females at approximately 47%.

Total Patient Caseload

Total patient caseload was not a significant predictor of a physician compliance with thyroid function guidelines. The average caseload was 3,080 Generally, the average scores were similar across all levels of caseloads. Reasons for the significant difference in average patient caseload between males and females is not clear from this data. It is not a result of differences in type of practice. Approximately 47% of both males and females are registered as working in a clinic practice. This variable did not appear to be highly correlated with any of the other independent variables.

Proportion of Total Patient Caseload Tested for Thyroid Functioning

Another potential predictor of compliance with thyroid function testing was the proportion of patients in the physician's caseload for whom any testing for thyroid functioning was ordered It is unclear whether those physicians who tested a larger proportion of patients for thyroid functioning actually had patient caseloads who were more at risk of thyroid disease, or whether they were inappropriately testing too many patients, or perhaps their counterparts were testing too few. On average, general practitioners ordered thyroid function tests for 4 7% of the patients they saw during the study period. Females ordered tests for 6.4% of their patients, males for 4.1%.

The data also indicated a mildly positive relationship between proportion of patients tested and Score. This was noted for both males and females. One interpretation of this result could be that inappropriate test ordering may be in part a function of infrequent ordering

During the multivariate analysis this variable was dropped from the model for males but was retained as part of an interaction term with an urban location The significance of this interaction term appears to be a result more of practicing in an urban location than of the proportion of patients tested. Urban males tested approximately 3.4% of their patients and have an average score of .4707, while the remainder of the males tested 4.2% of their patients and had an average

score of .3720.

The proportion of patients tested was significant for females at $p < .01$.

Of special note were six physicians who ordered thyroid function tests on 31% to 42% of their total patient caseload. When they were examined more closely, the characteristics they had in common were that five practiced in an urban setting, all saw had lower than average caseloads (ranging from 682 to 2100), and they all tested appropriately more than 65% of the time. In fact, three had scores in excess of 86% They differed on other key variables For example, three were males and three females, they were a variety of ages, and they were trained at universities around the world

One of the interpretations that could be made of this result is that the more patients a physicians has who have thyroid function disease, the more familiar that physician is with management of thyroid illnesses and thus, the better at ordering appropriately. Another, of course, is that there was mild over-ordering that was masked in part by 'appropriate' ordering. A third, is that these six had established what were in effect specialty practices despite their designation as general practitioners. It is possible in Alberta for general practitioners to refer patients to other general practitioners.

In summary, the most important finding of this study is that it confirms the fact that physician and practice characteristics explain little of the variation noted in utilization. In this case it is the appropriate utilization of thyroid function tests. Second, at the time of this study, general practitioners in Alberta, at the time of the study, were generally not ordering thyroid function tests in a manner consistent with existing thyroid function testing guidelines.

Despite the low explanatory power of this model, certain characteristics do appear to influence a physician's test ordering patterns. Male physicians are ordering these tests in a more appropriate manner than are females. The reasons for this are not clear. This difference between

gender remained significant when other characteristics, such as total patient caseload, age of practitioner, type and location of practice, year of birth, and CCFP certification, were controlled I conclude that either the difference is a direct result of gender or it is a result of other variables not considered in this model.

More specifically, this study suggests that males who practice in the southern third of the province, who have CCFP certification, and who practice in an urban location order thyroid function tests in a significantly better way. On the contrary, male physicians who have graduated from a university in India, Pakistan, Asia, or the Maritimes have significantly lower than average scores.

Female physicians who practice in the southern third of the province, in an urban location, and test proportionately more patients for thyroid function testing, scored significantly higher than other females

Overall, there appears to be something unique about physicians who work in the southern third of the province, in an urban practice, who have CCFP certification, and who test proportionately more of their patient caseload for thyroid functioning. These findings raise the issue of the importance of individuals or centres of influence.

4.2 Further Research Needs and Recommendations

4.2.1 Further Research Needs

Like other studies examining reasons for variation in utilization, this study explains only a small proportion (8.9%) of the variation in compliance with thyroid function testing guidelines Many questions are left unanswered. Some of the specific questions arising from this study are;

1. What other factors are related to compliance and how might they interact with the factors identified in this study? For example, what other factors might explain more of the variation in

appropriate thyroid function testing?

2. Further research should examine the differences in caseload, case mix, and style of practice between male and female practitioners. What factors affect the way in which they practice medicine?

3. Physicians who test proportionately more patients tend to have smaller total caseloads and better compliance scores. Why is this, and what does it tell us? Is it that 'practice makes perfect'? Is their patient case mix different, ie. higher risk?

4. Why the variation among census divisions? What is there about physicians in the census divisions of Medicine Hat and Pincher Creek that causes them to achieve such high scores? Is this a reflection of centres or individuals of influence?

5. What is there about working in an urban practice that relates to having higher scores?

6. What prepares CCFP certificants to order thyroid function tests in a manner more consistent with guidelines? Self selection, type of practice, other?

7. How are general practitioners in Alberta ordering thyroid function tests today? Has their pattern of practice improved since April 1992 when the Alberta Medical Association introduced thyroid function testing guidelines.

8. How do general practitioners order other laboratory and radiology tests? Are they generally being ordered as poorly as thyroid function tests? Do the same practitioner profiles emerge for those who score particularly high or low?

4.2.2 Recommendations

Epstein's[23] findings show that high users of diagnostic testing tended to know that they were high

users, but Nagurney et al[22] suggest that physician awareness of the cost of diagnostic testing was generally low. Studies by Hoey[24] and Cummings[25] show that the provision of price information appeared to decrease the use of diagnostic testing

While physicians are major players, there is consensus that many factors beyond the physician's control, such as patient and organizational characteristics, influence the utilization of health care services. Eisenberg[10,12] states that "although physicians' fees represent only one fifth of health care expenditures (in the USA), they are responsible for decisions that govern the way in which as much as 90% of each health care dollar is used". It is therefore important to understand the factors that influence a physician's level of knowledge and decision making process regarding utilization of health care services.

Some recommendations that come forward from this study are,

1. At minimum, information and training courses regarding thyroid function testing (and perhaps other laboratory and radiology tests) should be provided to various subgroups of general practitioners, such as practitioners new to Alberta who have been trained outside of Canada Given the rather low average score overall, training sessions for the majority of physicians in the province may be in order and cost-effective.

2. Develop a provincial body that coordinates and disseminates practice guidelines across the province. This body should have the support and endorsement of the major medical professional and health care organizations in the province. It should involve some type of evaluation process to ensure disseminated guidelines are being followed

3. The findings from this study, especially the variations among geographic regions, should be published in a forum such as the Alberta Medical Association Doctor's Digest This may generate thought as to what the cause of this variation is, i.e whether it reflects the role of individuals or centres of influence.

1 Watanabe M An Agenda for Action: Report of the Advisory Committee on the Utilization of Medical Services. Government of Alberta, 1989

2.Messenger College of Physicians and Surgeons of Alberta. Number 15, September 1991.

3.Wennberg JE, Gittlesohn A Small area variation in health care delivery.Science.1973;1102·182-98.

4.Sheps S, Anderson G, Cardiff K Utilization management: A literature review and case study. Final Report for NHRDP Project No. 6610-1745, 1990

5.McPherson K. International differences in medical care practices. Health Care Financing Review/Annual supplement 1989;S17-28.

6.Roos NP, Roos LL. Surgical rate variations: do they reflect the health or socio-economic characteristics of the population? Med Care 1982;20:945-58.

7.Roos NP, Shapiro E. The Manitoba Longitudinal Study on Ageing: preliminary findings on health utilization by the elderly Med Care. 1981;9·644.

8.Hulka BS, Wheat JR. Patterns of Utilization·the Patient Perspective. Medical Care 1985;23:438-60.

9.Quality Assurance and Effectiveness in Health Care. Report to the Conference of Deputy Ministers of Health Health and Welfare, Canada, November, 1990.

10.Eisenberg JM. Physician Utilization: The State of Research About Physicians' Practice Patterns Medical Care. 1985;23:461-83.

11.Sheps S, Scrivnes S, Gait J. Perceptions and Realities:Medical and surgical procedure variations - a literature review. Final Report for NHRDP Project No. 6610-1745-HT, 1990.

12.Eisenberg JM, Nicklin D. Use of diagnostic services by physicians in community practice. Med Care 1981;19·297.

13.Childs AW, Hunter ED. Non-medical factors influencing use of diagnostic x-ray by physicians. Med Care, 1972;10:323.

14.Siu A, Manning WG, Benjamin B. Patient, Provider, and Hospital Characteristics Associated With Inappropriate Hospitalization. Am J Pub Health 1990;80:1253.

15.Campbell DM. Why do physicians in neonatal care units differ in their admission thresholds?Social Med 1984;18:365.

16.Goldfarb MG, Hornbrook MC, Higgins G. Determinants of hospital use: a cross diagnostic analysis. Med Care 1983,21:48.

17.Williams AP, Domnick-Pierre K, Vayda E, et al. Women in Medicine: practice patterns and attitudes Can Med Assoc J 1990;143:194.

18.Silberger AB, Marder WD, Wilkie RJ. Practice characteristics of male and female physicians Health Affairs, 1987;6:104

19.Ohsfeldt RJ, Culler SD. Differences in income between male and female physicians J Health Econ, 1986;5:335.

20.Woodward CA, Cohen ML, Ferrier BM. Career Interruptions and Hours Practices Comparison Between Young Men and Women Physicians. Can J Pub Health 1990;81 16

21.Maheux B, Dufort F, Beland F, Jacques A, Levesque A. Female Medical Practitioners' More Preventive and Patient Oriented? Med Care 1990;28:87.

22.Nagurney JT, Brahyam RL, Reader GC. Physician awareness of economic factors in clinical decision making. Med Care, 1979;17:727

23.Epstein AM, McNeil BJ. Relationship of beliefs and behaviour in test ordering.

24.Hoey J, Eisenberg JM, Spitzer WO, Thomas D. Physician sensitivity to the price of diagnostic tests: a U.S. - Canadian analysis. Med Care, 1982;20:302.

25.Cummings KM, Frisof KB, Long MJ, Hrynkiewich G. The effects of price information on physicians testing behaviour. Med Care, 1982;20:293.

26.Pinault R. The effect of medical training factors on physician utilization behaviour Med Care, 1977;15:51.

27.Weil H. Comment on "The effect of medical training factors on physician utilization behaviour." Med Care, 1977;15:869.

28.Gillespie KN, Romeis JC, Virgo KS, Fletcher JW, Elixhauser A. Practice Pattern Variation Between Two Medical Scholls. Med Care, 1989;27:537

29.Williams SV, Eisenberg JM, Pascal LA, et al. Physicians' perceptions about unnecessary diagnostic testing. Inquiry 1982;19:363.

30.Williams AP, Vayda E,Stevenson HM, Burke M, Dominick-Pierre K A Typology of Medical Practice Organization in Canada. Med Care 1990;28:995

31.Williams AP, Dominick-Pierre K, Vayda E, Stevenson HM, Burke M. Women in medicine: practice patterns and attitudes. Can Med Assoc j 1990;143:194.

32.Paulick JM, Roos NP. The young physician: types of practice Can Med Assoc J 1978;118 276

33 Eddy DM. Designing a Practice Policy. JAMA 1990;263:3077.

34 Eddy DM. Practice Policies' Guidelines for Development. JAMA 1990;263:1839.

35.Eddy DM. Guidelines for Policy Statements. the Explicit Approach. JAMA 1990;263:2239.

36.Kanouse DE et al. Changing Medical Practice Through Technology Assessment: An Evaluation of the NIH Consensus Development Program. The Association of Health Services Research and Health Administration Press, 1989. A RAND Corporation Study.

37 Lomas et al. Do Practice Guidelines Guide Practice? NEJM 1989;321:1306.

38.Cohen MM, Roos NP, MacWilliam L, Wadja A. Assessing Physicians' Compliance With the Guidelines for Papanicolaou Testing. Unpublished work presented at the Canadian Conference on Health Services Research, Toronto, June 1990

39.John R, Evans PE, Scanlon MF, Hall R. Clinical value of an immunoradiometric assay of thyrotropin for patients with non-thyroidal illness and taking various drugs. Clin Chem. 1987;33:566-9.

40.McNeely M, Rasaiah B. Thyroid Testing Protocol: An Opinion From the Canadian Association of Pathologists CAP Newsletter. 1990;Sept/Oct:8-10.

41.Nolan JP, Tarsa NJ, Dibenedetto G. Case-finding for unsuspected thyroid disease: Costs and benefits. AJCP.1985,83 346-55.

42.Caldwell G, Kellett HA, Gow SM, et al. New strategy for thyroid function testing. Lancet. 1985;May 18:1117-9.

43.Beck JR. Laboratory decision science applied to chemometrics:strategy testing for thyroid function. Clin Chem. 1986;32:1707-13.

44.Kalra J, Hart IR. Value of Free Thyroxine (FT4), Free Triiodothyronine (FT3), and Sensitive Thyrotropin (TSH) Assay in the Assessment of Optimal Thyroxine Therapy. Clin Biochem 1987;20:265-7.

45 Kalra J, Laxdal VA, Walker SD, Massey KL. Evaluation and Assessment of High Sensitivity Thyrotropin Methods as an Index for Thyroid Function. Clin Biochem 1987;20:307-12.

46.Gow SM, Nicol K, Seth J, et al. Laboratory and diagnostic performance of a coated well immunoradiometric assay for serum thyrotropin. Clin Biochem. 1987;24:273-78.

47.Klee GG, Hay ID. Assessment of Sensitive Thyrotropin Assays for an Expanded Role in Thyroid Function Testing: Proposed Criteria for Analytic Performance and Clinical Utility. J Clin Endocrinol Metab. 1987,64:461-71.

48.Ericsson UB, Fernlund P, Thorell JI. Evaluation of the usefulness of a sensitive immunoradiometric assay for thyroid stimulating hormone as a first-line thyroid function test in a unselected patient population. Scand J Clin Lab Invest. 1987;47 2155-221.

49.Young DG. Physicians can reduce costs. Alberta Doctors' Digest 1988,Jan/Feb 11

50.Massey KL. The Laboratory Approach to Thyroid Disease A Stepped Approach. Can Fam Physician. 1988;34:365-68.

51.Shlossberg AH. Thyroid Tests: Guidelines for Primary Care Physicians. Can Fam Physicians. 1990;36:1807-10.

52.Helfand M, Crapo LM. Monitoring Therapy in Patients Taking Levothyroxine. Annals of Internal Medicine. 1990;13:450-4.

53.Helfand M, Crapo LM. Screening for Thyroid Disease. Annals of Internal Medicine. 1990;112 840-49.

54.Practical Endocrine Diagnosis, (3rd ed). Watts NB, Ketfler JH Publ Lea & Febiger, Philadelphia, 1982.

55.A Manual of Laboratory Diagnostic Tests, (3rd Ed). F Fischbach, 1988.

56.Textbook of Clinical Chemistry Teitz NW., WB Saunders Co

57.McKechnie JL. Review of Thyroid Function Testing in Alberta Report prepared for the Utilization Monitoring Committee - Medical Services, Government of Alberta, 1991.

58.Gordin A, Maatela J, Miettenen A et al. Serum thyrotropin and circulating thyroglobulin and microsomal antibodies in a Finnish population. Acta Endocrinol. 1979;90 33-42

59.Evered DC, Ormston BJ, Smith PA et al Grades of hypothyroidism. Br Med J. 1973,1 657-662

60.Turnbridge WGM, Evered DC, Hall R et al. The spectrum of thyroid disease in a community:the Wickham study. Clin Endocrinol. 1977;7:481-493.

61.Falkenberg M, Kagedal B, Norr A. Screening of an elderly female population for hypothyroidism-and hyperthyroidism by use of a thyroid panel. Acta Med Scand 1983,214:361-365

62.Bahemuba M, Hodkinson HM Screening for hypothyroidism in elderly inpatients Br Med J 1975;2:601-603.

63.Jeffreys PM. The prevalence of thyroid disease in patients admitted to a geriatric department Age & Aging. 1972;1:33-37.

64. Ross DS Subclinical hyperthyroidism. possible dangers of overzealous thyroxine replacement therapy. Mayo Clinic Proc. 1988;63:1223-1229.

65. Platt H. A Progressive Thyroid Profile: The Effects of an Educational Intervention to Encourage Its' Use. (forthcoming publication).

66. United States Preventive Services Task Force

67. Canadian Task Force on the Periodic Health Examination: Periodic health examination, 1990 update. Can Med Assoc J. 1990;142:955-61.