A DESIGN FOR A RANDOMIZED CONTROLLED TRIAL TO ASSESS THE EFFECT OF CLINICAL AUDITS ON AMBULATORY CARE IN HOSPITAL CLINICS

By

CORNELIA JOHANNA BAINES, M.D.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

May 1980

A CLINICAL TRIAL OF AUDITS TO ASSESS IMPACT ON PATIENT CARE

MASTER OF SCIENCE (1980) McMASTER UNIVERSITY
(Medicine)

TITLE: A Design for a Randomized Controlled Trial to Assess the Effect of Clinical Audits on Ambulatory Care in Hospital Clinics.

AUTHOR: Cornelia Johanna Baines, M.D. (University of Toronto)

SUPERVISOR: Dr. Peter Tugwell

NUMBER OF PAGES: ix, 158

ABSTRACT

A design is presented for a randomized controlled trial to answer the question: do audit procedures which hospitals are expected to carry out to meet requirements for hospital accreditation have a beneficial impact on ambulatory patient care in general and special clinics?

Approximately 60 Ontario hospitals reporting general and special ambulatory care clinics to Statistics Canada will be invited to participate in the trial. Outpatient visits to potentially eligible clinics will be documented by hospital medical records staff and submitted to the Hospital Medical Records Institute. Nurse abstractors will identify indicator conditions from this outpatient census and categorically score patient management over three periods of time: a pre-audit period before intervention occurs, a first audit period and a second audit period.

Hospitals stratified according to size and function will have been randomly allocated to three groups. Indicator conditions relevant to caseloads and casemix will be assigned to eligible clinics.

Groups I and II will both have indicator conditions introduced to clinic staff prior to the first audit period. Only Group I will receive feedback about their performance in the first audit period.

After the second audit period it should be possible to separate the effects of audit awareness (Groups I and II), of feedback (Group I) and of extraneous factors (Group III) on patient management. Results will be expressed as mean clinic scores.

If improved performances occur in Groups I and II, the usefulness of current accreditation criteria will have been demonstrated. If no change in performance occurs and the indicator condition criteria approach is accepted as being valid in this setting, then it may be appropriate to consider new approaches to accreditation.

ACKNOWLEDGEMENTS

Writing this thesis has been fun for me and a lot of work for a number of others. The guidance of Dr. Peter Tugwell and Dr. J.C. Sibley is acknowledged with thanks. The patience of Professor C.H. Goldsmith as he shepherded me through the pitfalls of statistical analysis deserves a medal. And the good nature of my husband and son who took over many "household executive" tasks in the last year reminds me again how fortunate I am. It is a privilege to be a student.

TABLE OF CONTENTS

Table of Contents (continued)

Table of Contents (continued)

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

1.1 Introduction

Separating itself from the Joint Commission on Accreditation of Hospitals, the Canadian Council on Hospital Accreditation (CCHA) was incorporated in 1958 under federal law. Participation in its program is entirely voluntary. According to the CCHA annual report (1979), 712 (54%) of 1339 hospitals listed in the Canadian Hospital Directory are accredited and 145,811 (74.5%) of 195,586 beds are accredited. For 334 Ontario hospitals listed by the directory, the corresponding figures are 66.6% and 82% respectively. However, if one looks at only the 234 public and private, active and chronic, hospitals for medical care in Ontario, and excludes the 100 'hospitals' among which are penitentiary hospitals, detoxication centres and children's rehabilitation units, then 81.2% of Ontario hospitals are accredited.

In its attempts to monitor the quality of hospital practice in Canada, the CCHA focussed initially on structure (for example the qualifications of staff or the facilities available), subsequently on process (in particular on medical records) and recently on outcomes.

"Process" refers to the actions of health professionals in the management of patients while "outcome" refers to the end results of health care as it affects the patient's health and satisfaction with care (Tugwell, 1979[1]). This changing of focus parallels the evolution of quality of care appraisal as reflected in the literature over the last two and a half decades. Since 1977, quality of care appraisal has been made mandatory for hospital accreditation in Canada. As is the case in the U.S.A., such requirements are being implemented before the value of such procedures has been unambiguously established (Komaroff, 1978). For this reason, the CCHA approached the department of clinical epidemiology and biostatistics at McMaster University in late 1979 and asked for assistance in determining whether the clinical audits they were requesting hospitals to implement, did result in improved patient care.

Up until now the CCHA has left it to the discretion of individual hospitals to design their own methods for auditing patient care, setting out only general guidelines (CCHA, 1977), and until recently, the emphasis has been on auditing in-patient care. Audits of ambulatory care have been less rigorously performed, if at all. The disadvantages arising when individual institutions generate their own audits include a possible lack of of expertise, a possible lack of enthusiasm for the task and where both expertise

and enthusiasm exist, a possible lack of time due to competing responsibilities. Often, there is a lack of attention to anchoring the standards to patient benefit. Finally, with each institution devising its own audit, comparisons between institutions are not possible.

The majority of patients receiving care from the health care system are ambulatory patients (Christoffel and Loewenthal, 1977). For the year April 1, 1978-March 31, 1979, there were in Ontario, 1,454,552 hospital admissions (excluding psychiatric hospitals) (Ministry of Health, 1979[1]), 7,866,761 outpatient visits to hospitals (Ministry of Health, 1979[2]) and 29.6 million visits (fiscal 1977) to private offices (Lussing, 1980).

Ambulatory care is all care delivered to patients who can arrive at and depart from the medical care services on the same day, and has as its counterpart, non-ambulatory care delivered to patients admitted to hospital for more than one day.

Some ambulatory care is primary in that it is the first patient contact with the health care system, it is the first stage in the treatment of an illness and it is delivered usually by family physicians. Some ambulatory care is secondary or tertiary depending on the degree of specialization associated with it and on how the patient gained access to it.

All types of ambulatory care can be episodic, as in the treatment of an acute infection of short duration, or continuous, as in the management of a diabetic patient.

My particular interest is in the implementation of audits of the quality of care received by outpatients in hospital ambulatory care clinics and in the outcomes of such audits. Examination of the HS-1 (Appendix (App.) A-1) form reveals that hospital ambulatory care includes "General and Special Clinics", "Outpatient Day and Night Programs", "Emergency Units", "Surgical Day Care Programs" and "Home Care Programs". Visits to general and special clinics accounted for 2.54 of the 7.86 million outpatient visits to hospitals in Ontario in 1978 (Lussing, 1980). This thesis will design a trial to determine the effect of audits on general and special clinics. It should be noted that until now, most quality of care research has been directed at hospital inpatient care (Payne, et al. 1976), private and group practices (Spasoff, et al. 1977; Lohr, et al. 1980) and hospital emergency care (Spasoff, et al. 1977, Frazier and Brand, 1979). Thus it seems appropriate to focus on hospital outpatients receiving non-emergency care.

1.2 Research Question

Do audits have a beneficial impact on the quality of ambulatory primary care delivered in general and special clinics of hospitals? A breakdown of this general question into its component parts is to be found in Chapter III, Section (S) 3.1.

1.3 Literature Review

The purpose of an audit imposed in order to establish that a hospital merits accreditation is, first and foremost to make an appraisal of the quality of care (Surridge, 1979) delivered to patients. Secondly it is to ensure not only that the deficient care will be rectified but also that adequate or good care may become better. Such a statement requires that quality of care, quality of care assessment and quality of care assurance be defined.

Quality of care in spite of Donabedian's detailed examination of the subject (American (Am.) Public Health Assoc. 1965[1]) remains difficult to define. In this thesis a concept of quality of care appropriate for hospital ambulatory care clinics in a developed industrial nation will be used. In a given location at a particular point in time, constrained by the number and types of health care professionals, other health resources available and modified by the capacity of the patient to exploit the resources

available to him, the quality of care given to a patient is assumed to be better the more those factors capable of improving his health are implemented.

Quality of care assessment or appraisal occurs when the quality of care is graded categorically or in some way quantified.

Quality of care assurance is the extension of assessment in that it is a mechanism for ensuring that deficits in patient management which are identified will on subsequent assessment be shown to have been improved or corrected.

Both implicit and explicit criteria have been used to assess quality of care, the former depending on detailed objective definitions of criteria, the latter depending on intuitive, unstandardized, individual assessment of quality. While implicit criteria may be useful for application in a restricted setting, for example one practice's assessment of its performance for its own interest, only explicit criteria will permit objective studies involving many physicians and many institutions.

A review of the literature reveals a number of approaches to assessing quality of care, some of which are listed in Table 1.1 below. The nature of "structure" is evident from the table. The "assessment of process is the evaluation of the activities of

physicians and other health professionals in the managment of patients - - - the assessment of outcome is the evaluation of end results in terms of (patient) health and satisfaction." (Am. Public Health Association, 1965[3]).

TABLE 1.1

CATEGORY	METHOD	REFERENCE
STRUCTURE	Degree of skill or level of certification of health professionals.	Am. Pub. Health Assoc. 1965[2].
	Ratio of health workers to patients.	
	Type and number of specific facilities.	
PROCESS	Chart review using implicit criteria.	Brook, 1973
	Direct observation.	Clute, 1963
	Drug utilization reviews	Lohr, et al. 1980 Brook and Williams, 1976 Sibley, et al. 1975[1]
	Indicator conditions	Burdette, et al. 1974
	Peer review of billing claims	Buck and White, 1974
	Review of operative specimens	Fessel and Van Brunt, 1972
	Simulation studies	Barro, 1975[2]
OUTCOME	Functional, emotional, and health status	Kane, et al. 1976
	Mortality and morbidity	Shapiro, et al. 1958 Lipworth, 1963
	Patient satisfaction	Burdette, et al. 1974
	Restoration of normal physiological or biochemical indices	Williamson, et al. 1975 Starfield and Scheff, 1972 Inui, et al. 1976
	Return to work	Brook and Stevenson, 1970
	Symptom level	Ashley, et al. 1971 Burdette, et al. 1974 Brook and Stevenson, 1970

The heterogeneity of methods for assessing quality of care in part reflects the uncertainties associated with a developing field of research and in part may be explained by the multiplicity of purposes for which assessment is undertaken. These include: education - both the person assessed and the assessor may learn from the audit procedure; improvement of patient outcomes; administrative purposes may be fulfilled; improved efficiency or improved medical records may result; the audit may assist in local or regional policy formulation; or the audit may be required for accreditation purposes.

The dominating problem seems to be that good provider-process does not necessarily ensure good patient outcome (Barro, 1975[1]) and good patient outcomes can occur when preceded by deficient or even absent process as Donabedian pointed out (Am. Public Health Assoc. 1965[4]). It would appear that aspects of the quality of care which are assessed by process measures are sometimes different to the aspects assessed by outcome measures (Romm and Hulka 1979).

Fessel and Van Brunt (1972) using both process and outcome measures to assess the management of appendicitis in hospitals, showed that the hospital which scored lowest on process measures scored highest on outcome measures. In contrast, Starfield and

Scheff (1972) focussing on iron deficiency anemia in children, demonstrated a strong relationship between appropriate process and good outcomes.

Process measures generally depend on the medical record for data, a source which is usually accessible, which is less expensive to use compared to sources for outcome measures and which is often incomplete. Interestingly, the degree of documentation in medical records may not correlate with the degree of adherence to essential treatment criteria and the completeness of recording clinical data may be unrelated to immediate outcomes (Sanazaro and Worth, 1978). Thus it is not surprising that explicit process criteria can result in a very negative assessment of quality of care at the same time as implicit outcome criteria produce a positive assessment of the same care (Brook, 1973). More appropriate than long lists of explicit process criteria may be the application of principles of decision-making as described by Greenfield, et al. (1977) with criteria mapping. Criteria are branched conditionally, so that an action indicated by one criterion depends on the results of a previous action. Unfortunately the map given for chest pain lists 218 criteria. Criteria mapping is obviously an unwieldy mechanism, but it results in a more favorable score to the physician.

Outcome measures generate higher costs than process measures and depend largely on the patient who is not always accessible and who may not give reliable information. Further, outcomes are influenced both by the disease severity or disease stage (Gonnella, et al. 1976) on presentation of the patient and by the many factors which influence health outside the health care system. Ultimately outcome assessment rests on two major assumptions (Brook, et al. 1976). The first is that medical care really does have an impact on the outcome one has chosen. The second is that adverse outcomes occur frequently enough that they will be detectable.

Although it is necessary to be aware of the conflicts and ambiguities inherent in process and outcome assessment, the main emphasis in this thesis is on whether audits induce measurable change. There is evidence in the literature that educational activities can change physician performance. However, from the discussion thus far one is aware that physician performance (that is, provider-process) may or may not be shown to influence patient outcome.

Imui, et al. (1976) showed that physician exposure to tutorials on the management of hypertension was associated with a significant improvement in hypertensive patient outcomes when the outcome was adequately controlled blood pressure. On the other

hand, Sanazaro and Worth (1978) in a prospective study involving 50 hospitals concluded that introduction of an audit procedure (in which physicians formally endorsed the criteria applied and in which lists of the relevant criteria were placed on the front of patients' charts) was associated with slightly better adherence to treatment criteria by physicians. (With these two references cited it appears the words audit and education are being used as synonyms, which they are not. Rather they are two intertwined activities. If the purpose of an audit is to assess quality of care and to improve it when it is deficient, the mechanism by which it may be improved is "education" which may induce people to change behaviour. If education is the focus of interest, a measure of its impact is change in behavior. Thus it is reasonable to infer that if education can change physician behavior, audits which incorporate an educative element can also be expected to change performance.)

Although slightly better adherence to treatment criteria was shown, this adherence was unrelated to outcomes when the results for several diagnoses were pooled by Sanazaro and Worth. However, failure to meet treatment criteria for acute bacterial pneumonia and for acute myocardial infarction was associated with unsatisfactory outcomes. The conclusions drawn are that audits can increase adherence to treatment criteria and that provided treatment criteria are valid, adherence to them will be associated with better

outcomes.

It does not seem unreasonable that an audit procedure as outlined in this thesis may be expected to have a beneficial impact on patient outcomes if physicians accept the criteria and if the criteria are valid. However, even if this expectation is not unreasonable, it must be tempered by the observation made by Brook and Appel (1973), namely that having assessed the quality of care for three tracer conditions, urinary tract infection, hypertension and peptic ulcers, one cannot assume one has simultaneously measured the quality of care delivered for other diseases. This inability to generalize is a disadvantage.

Although it is frequently claimed that most attention has been paid to quality assessment and assurance for inpatient hospital care, much has been written about ambulatory care as well (Clute, 1963, Peterson, et al. 1956, Hulka and Cassel, 1973, Payne, et al. 1976, Lindsay, et al. 1976, Pozen and Bonnet, 1976, Romm and Hulka, 1979, Lohr, et al. 1980).

Payne (1979) is a strong advocate for the use of medical records in the evaluation of physician performance, emphasizing the importance of process measures while not ignoring appropriate outcome measures. Based on his analysis of available data in the literature, he concludes good recording is related to good practice

and although the relationship is not perfect, it is statistically significant. He describes the "ultimate" (utopian?) medical record as one which is always legible, quickly retrieving past data and displaying current data, which demands completion, corrects errors in medication or conflicting laboratory requests, is responsive to protocol and is an ideal instrument for quality assurance. Indisputably, such a record would be useful.

In contrast to a utopian medical record, Frazier and Brand (1979) (who say the "medical record remains grossly deficient for audit purposes") describe a computerized clinical algorithm which uses branching criteria to evaluate the quality of care of lacerations in an emergency service. One can evaluate the performance of the provider by measuring compliance with algorithmic criteria. To do this the physician must agree to use a structured checklist rather than write his usual note. The method succeeded in showing that different providers vary significantly in their compliance with algorithmic criteria.

What is needed for the purposes of this thesis however is a method which is ready to be applied to a variety of hospital clinics - not just a single type - and to a variety of disease conditions. The indicator conditions developed by Sibley, et al. (1975[2]) seem to meet those requirements and are discussed in S 2.4.3.3.

At present the essential features of an audit are that there must be identification of instances where agreed upon criteria are not met, a decision made whether failure to meet the criteria was justified or not, release of information such that the group studied is aware that particular problems are or are not occurring and subsequent assessment to determine whether identified problems have decreased in frequency. It may also be appropriate to do long term re-assessment to determine how long any early benefit is maintained.

Kessner (1978) declared "the time for testing pragmatic evaluation schemes in live practice settings is long overdue". Hospital ambulatory care clinics are certainly live practice settings. It is hoped that application of the indicator conditions developed by Sibley, et al. (1975[1]) will provide a pragmatic evaluation.

CHAPTER II

RESEARCH DESIGN

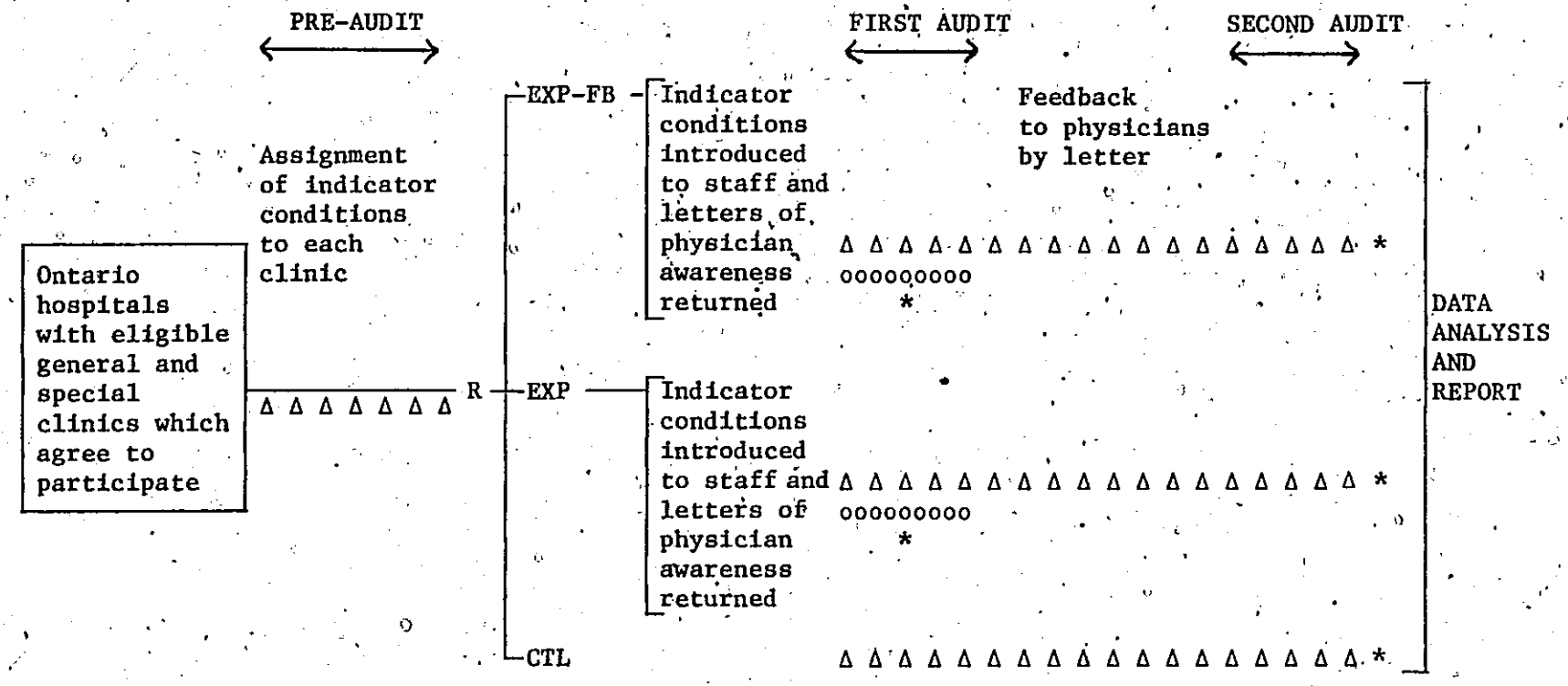
2.1 Introduction

A randomized controlled trial (RCT) (Fig. 2.1) is being used to answer the question posed by the CCHA because an RCT is the best method for identifying an effect attributable to compliance with CCHA regulations. Further, failure to demonstrate a difference between experimental and control groups in an RCT provides the most compelling evidence that a change in CCHA accreditation policy is indicated, provided the study has been properly designed and carried out.

2.2 Research Objectives

The research objective is to determine whether the implementation of audit procedures to meet the requirements of the CCHA for hospital accreditation causes a measurable difference in the quality of ambulatory care received by patients in hospital general and special clinics.

FIGURE 2.1



* = Retrospective scoring by nurse abstractors for any previous study phase not yet scored.
R = Randomization EXP-FB = Experimental with feedback EXP = Experimental and no feedback CTL = Control

2.3 Justification

The implementation of audit procedures can be associated with positive as well as negative consequences.

2.3.1 Possible positive consequences

1. The trial may reveal that audits are associated with improved patient care and thus may facilitate the implementation of audits in the future.

2. The trial may improve the knowledge base of individual physicians and/or their performance.

3. The trial may lead to decreased costs by reducing the number of specific diagnostic procedures performed and possibly by improving patients' health status sufficiently that future need for health services will be diminished.

4. By identifying clinics which do not meet standards, an opportunity for correcting deficiencies or innovating new solutions, will arise. Conversely clinics which have very high performance scores may yield insights into the mechanism of their success.

5. The trial may lead to the development of a national catalogue of audit procedures from which hospitals could select individual ones most suited to their needs. Comparisons between institutions and regions would be possible. Such a system would enhance the prestige of the CCHA and perhaps render more positive attitudes of skepticism towards quality assurance procedures presently held by some physicians.

2.3.2 Possible negative consequences

1. Physician defensiveness, skepticism and hostility all may be evoked in response to an audit or reinforced by the results of an audit.

2. Where deficient performance is identified, the possibility may exist of exploitation of such information either within or without the institution to the detriment of either the institution or the individual.

3. Awareness that an audit is being conducted may lead to the practice of defensive medicine with increased costs generated by an increased number of procedures performed as well as possible associated increased risks to the patient.

The trial is being designed taking into account these possible adverse side effects and it is expected such consequences can be avoided.

2.4 Details of design

2.4.1 Sample

2.4.1.1 Hospitals

2.4.1.1.1 Distribution and types of clinics in hospitals

Data acquired from the Ontario Ministry of Health (Lussing, 1980) lists the types of "General and Special Clinics" reported by Ontario hospitals and the numbers of patient visits to individual clinics annually. This information is submitted to Statistics Canada on HS-1 forms (App. A-1) and since these forms are filled by all Canadian hospitals, similar data are available for all provinces. Twenty-three types of clinics are listed.

In Ontario, of the 234 hospitals categorized by the Ministry of Health as public and private, active and chronic hospitals for medical care, 82 report one or more general or special clinics for a total of 548 such clinics. Twenty of these 82 hospitals meet the criteria of the Ontario Council of Administrators of Teaching Hospitals to be teaching hospitals. (These criteria are that the hospital train both graduates and undergraduates and that it have a university affiliation.) The 20 teaching hospitals operate 343

general and special clinics and report 1,891,670 visits in 1978-79 (App. C-1). The remaining 62 hospitals operate 205 general and special clinics and report 591,744 visits in 1978-79 (App. C-2).

Unfortunately, the type of clinic with the second highest frequency in Ontario is described as "other" and accounts for 325,502 visits (of which 17,760 are to the Toronto Addiction Research Foundation). For Toronto's Scarborough Centenary Hospital, 'other' meant visits for family planning, sutures, local anesthetics and nerve blocks, while for the Brantford General, it included prenatal nutrition counselling and more general nutrition counselling as the sole two services. With such variability in types of services, it seems necessary to exclude this type of clinic from the present trial although at a future time a special investigation of a category which accounts for so many visits to general and special clinics may be warranted.

At issue in the approach to selection of hospitals as just described is generalizability. A nationally drawn sample may not be feasible. A regionally representative study may not be generalizable. A provincial sample may be a satisfactory compromise, especially if, when the study is implemented, it can be demonstrated that the characteristics of Ontario hospitals serving one third of the nation's population are similar to those in the rest of Canada.

2.4.1.1.2 Exclusion criteria for hospitals

1. Beginning with the population of 234 accredited hospitals, the group not reporting general and special clinics on the HS-1 forms will be excluded from the study even though those excluded may report day-care units or emergency units.

2. Of the remaining 82 hospitals, three institutions with unique functions will be excluded, namely the Clarke Institute, the Princess Margaret Hospital and the Queen Elizabeth Hospital. The Clarke Institute serves only one function, psychiatric care, and is not representative of the institutions delivering primary ambulatory care to patients. The Princess Margaret Hospital operates only cancer clinics and again serves a particular purpose for a particular population. The Queen Elizabeth Hospital is a chronic hospital recording one to 11 visits by outpatients yearly to its seven ambulatory care clinics. The patient population served by its clinics is virtually exclusively in-patient in nature.

3. Hospitals will be excluded which report only "other" ambulatory care clinics. "Other" includes a wide diversity of services from nutritional counselling to suture removal to nerve blocks depending on the institution. Such diversity in function does not permit application of the intervention this study will use. Institutions in this category include the Toronto Addiction and

Research Foundation, the Windsor Salvation Army Grace Hospital, Windsor Hotel Dieu, Thunder Bay St. Joseph's and the Brantford General.

4. Another exclusion category will be those hospitals for which the annual number of visits recorded per clinic on the HS-1 form is so small that the required number of patients with indicator conditions could not accumulate within the period of the study. Which institutions will be excluded will not be known until the most recent data prior to implementation of the study have been examined.

5. Hospitals which have consented to participate in the trial but which fail to record acceptably complete data on the daysheets for the two month assessment period will be excluded from the trial. However, if the probe employed to identify patients is an HMRI abstract form, this exclusion criterion will not need to be employed (S 2.4.1.2.2).

6. There are 10 hospitals with "orphan" clinics only, that is, for which none of the currently developed indicator conditions are suitable. These clinics are allergy, cancer, dental, endocrine, ophthalmology, TB (tuberculosis) and VD (venereal disease) in type. The hospitals which report only one or more in this category are: Belleville General, Toronto York West Park, St. Catherines Shaver, Englehart, Geraldton, Guelph General, North Bay Civic, North Bay St.

Joseph's, Sault Ste. Marie Plummer, Toronto City Central and Wingham and District Hospital.

This subset of hospitals will be invited to participate in the pre-audit phase of the pilot study (S 2.4.4) in order to establish whether their caseloads offer opportunities for new indicator conditions to be developed. If the pre-audit reveals the need for new types of indicator conditions, they will be developed and as a result more types of clinics will be able to enter the trial when it commences.

This leaves a maximum of 62 hospitals of which 20 are teaching hospitals. The 42 non-teaching hospitals with a total of 146 potentially eligible clinics have a median number of two clinics and a median number of 310 beds (App. C-3).

2.4.1.1.3 Hospital Compliance

A problem which must be acknowledged is that in the design for this trial it is assumed that most if not all hospitals approached will agree to participate in the trial. This assumption is based on the belief that hospitals which have chosen to be accredited wish to retain this status. Retaining this status requires the implementation of audit procedures. The trial offers hospitals not only the opportunity to apply a ready-made audit

package, but also offers them the possibility of establishing whether or not audits are useful. Nevertheless, it is recognized that a large number of hospitals may decline to participate. Gaining their cooperation will require more than a letter. Personal persuasiveness combined with prominent CCHA endorsement of the study may accomplish what a letter alone would not.

2.4.1.1.4 Exclusion criteria for clinics

1. If during the pre-audit assessment period some of the clinics fail to record data satisfactorily on the daysheet summaries (S 2.4.1.2.2), these clinics will not enter the study. If HMRI forms are the probe used, this exclusion criterion will not be needed.

2. Clinics will be excluded if the probe employed in the pre-audit phase reveals that for available indicator conditions, the frequency of episodes is too low. Should this criterion result in more than 25% of the community hospitals being excluded entirely from the study, consideration will have to be given to prolonging the length of audit in certain institutions so that an adequate sample size can be collected.

2.4.1.1.5 Inclusion criteria for clinics

Clinics in which illness episodes appropriate for the indicator conditions of interest occur frequently enough to produce an adequate sample for analysis within the time constraints of the study will enter the study.

2.4.1.1.6 Numbers of clinics participating per institution

Since in comparison to teaching hospitals, community hospitals have so few clinics, it is planned that every eligible clinic in community hospitals will participate in the trial. In most cases this means one or two clinics will be involved per institution.

In teaching hospitals, the median number of clinics per institution is 18. While not all the clinics are likely to be eligible, it is likely that more than five clinics per institution may prove eligible. If it is accepted that an arbitrary four clinics per institution are sufficient for the purpose of the study, then it will be necessary to randomly choose four of whatever number of eligible clinics exists.

2.4.1.1.7 An alternative approach to sampling

If it is argued that the eligible clinics as described in S 2.4.1 are a non-probability sample and that therefore one cannot apply the usual statistical analyses, a counter-argument is that it is not a sample of clinics which will be participating but rather the population of general and special clinics which generates enough episodes of illness that indicator conditions can be applied and differences detected. The hospital is the allocation unit, and hospitals, not clinics, are accredited.

However, if clinics were to be the allocation unit, a possible method would be:

1. Individual types of clinics could be stratified according to their hospital stratum, that is all general medicine clinics would be stratified into three categories, all surgery clinics would be - and so on.

2. A preliminary census could determine the frequency of each indicator condition in each cell, where a cell would be, for example, all general medicine clinics in teaching hospitals.

3. From the cell containing the entire population of general medicine clinics in teaching hospitals, a random sample could be selected such that the incidence of types of illness episodes in the sample paralleled that in the population.

4. Clinics in the sample thus chosen would be randomly assigned to one of the three experimental groups.

While this approach is tidy, it has two undesirable consequences. It would be expensive and it would facilitate contamination.

First, since Ontario hospitals are not at present submitting HMRI abstracts for outpatient visits, preliminary definition of the sample frame would require a census of the approximately two million visits to general and special clinics which remain after visits to the 'other' category are excluded. Since 55,000 outpatient visits per year to one British Columbia hospital require two full-time medical records staff to complete the HMRI abstracts, to document two million would require 72 full-time staff whose training and salaries would be costly. In addition the HMRI fee of 70 cents/form would be imposed for a total of $2 \times 10^6 \times \$.7 = \$1.4$ million. Further it is estimated that to draw the sample would take a statistician, expert in sampling, one month of full-time commitment.

Secondly, a disadvantage to a method in which clinics rather than hospitals are randomly assigned to the three treatment groups is that within one hospital, two or three experimental groups might be represented. If one clinic were assigned the EXP-FB and another to CTL, contamination of the CTL clinic would be more likely to occur than if all clinics in that hospital were in the CTL group.

I have therefore concluded that the awkwardness inherent in the sample selection previously outlined, must be accepted.

2.4.1.2 Patients

2.4.1.2.1 Criteria for entry

For a period of time sufficient to collect the required number of cases, all patients presenting to the participating clinics with a diagnosis corresponding to an indicator condition assigned to that clinic, will be assessed for eligibility by the nurse abstractors. Eligibility will depend on whether the episode meets the criteria listed by Sibley, et al. (1975[1]) (App. A-2) modified to be appropriate for the proposed trial (S 2.4.3.3.6). The patients may be referred, self-referred or regular clients of the clinics. Collection of episodes will stop when the sample size requirements are met.

2.4.1.2.2 Probes

The term 'probe' denotes the mechanism whereby charts, appropriate for the audit selected, are identified.

Two kinds of probes can be considered. The first, would be daysheets (App. A-3) and duplicate prescription pads as used in the Burlington randomized controlled trial of nurse practitioners (Batchelor, et al. 1975). The second would be HMRI abstract forms (App. A-4).

In the pre-audit phase the probe will reveal the presenting complaints and diagnoses occurring in each clinic and the frequency with which they occur. These data will enable the research team to assign indicator conditions relevant to the caseload and casemix of each clinic.

In the first and second audit phases, the probe would permit the nurse abstractors to identify the charts with the chosen indicator conditions. In the case of first type of probe it would be necessary to have continuous daysheet and prescription monitoring throughout the three phases because a hiatus in the collection process, followed by re-institution would possibly trigger increased awareness of the trial by the physicians, particularly when the second audit began.

However, the major problem associated with this probe will be the degree of compliance of the clinic staff (nurse or secretary) in filling the daysheet forms and gathering reliably the prescription copies. It is easy to imagine that data collection will be more reliable on slack days and less reliable on busy days. The incentive to be conscientious may be very weak, or if initially strong, may attenuate with time.

A second type of probe would be to have HMRI abstract forms filled for outpatient visits just as they are presently filled for in-patient admissions. When one examines the data collected on the Burlington daysheet, it is possible to select out a short list of data which would comprise an adequate probe for the purposes of this study and which could be entered on the HMRI abstract form.

TABLE 2.1

Daysheet Data (Batchelor, et al. 1975)	Short List for HMRI
Patient name	OHIP number
Patient number	Chart number
Doctor's number	Doctor's number
New patient?	New patient?
Fees charged	Reason for visit or complaints
Was doctor seen?	Diagnoses: most responsible; others
Reasons for visit or complaints	Procedures
First visit (Yes/No)	Sex
Diagnoses	Year of birth
Procedures (time, dollars, type)	Referral
Sex	Date
Prescription given?	Clinic ID
Referral (Yes? No? To?)	
Charge to -	
Day of week	
Date	
Type of health professional	
Page of	
Batch	

General information about entering the data on HMRI abstracts is to be found in Appendix A-4.

During the pre-audit phase aggregation of HMRI abstract data will yield a census of diagnoses and complaints for eligible clinics and permit the assignment of appropriate indicator conditions to clinics. Subsequently, during the first and second audit periods, HMRI abstracts will identify charts containing the indicator conditions of interest (S 2.4.3.2 Item #7).

Using HMRI services will entail training and labour costs for hospital medical records staff, the fee for each HMRI abstract submitted to HMRI and the charge for preparing summary data. Nevertheless, the HMRI abstract is the probe preferred for this trial. Preliminary discussions with HMRI staff reveal that a one-half day training program will be needed to prepare hospital medical records staff to complete the HMRI abstracts for outpatient visits.

2.4.1.2.3 Informed consent and confidentiality

Informed consent by the patient is not an issue in this trial since quality of care assurance programs are required by the CCHA for hospital accreditation. However physicians staffing the clinics and the hospital administration personnel will be assured of the confidentiality of all data (S 4.2).

2.4.1.2.4 Case severity and demographic differences in patients

It will be assumed that the ranking, stratification and randomization processes will distribute patient differences (such as case severity, socio-economic status, age, sex and ethnic origin) among treatment groups.

2.4.1.3 Physicians

2.4.1.3.1 General remarks

The number of physicians staffing each clinic will not be known until the pre-audit data are examined. Differences in the qualifications of the physicians should be minimized by randomization. In any case, it is the clinic performance which is being audited and the hospital which is being accredited. A consequence of the study may be a conclusion that physician performances need to be evaluated, but since the research question addresses the effect of audits on the quality of care delivered in clinics, the training of the physicians will be ignored.

It is also recognized that where internes and residents staff clinics, staff changeover may result in individuals present at the time of the first audit not being present at the time of the second. Nevertheless one may argue that this is the normal state of affairs in hospitals and that the research question addresses the value of audits under these circumstances and not under stable and predictable conditions.

2.4.1.3.2 Potentially confounding variables

1. Physician heterogeneity

Intraphysician heterogeneity of performance for different indicator conditions and interphysician heterogeneity in performance within clinics are expected to be distributed evenly across the three experimental groups.

2. Physician characteristics

Physician characteristics such as age and number of years of training and of practice may influence attitudes towards audits, performance and response to feedback. It would be possible but perhaps not desirable to include with the letter indicating audit awareness, a questionnaire which would identify the institution, the clinic description, the age of the physician, years of post-graduate training and years of practice. Given the often defensive posture which audits can elicit from practising physicians, such a request might provoke resistance. If one assumes that randomization will distribute differences in physicians equally in treatment groups, one expects the differences to cancel out.

3. Staffing changes

Changes in physician staff over the time of the study is another confounding factor. As mentioned previously (S 2.4.1.3.1) such changes are inherent in the normal hospital life cycle and are likely to be evenly distributed across experimental groups since it is planned to use simultaneous, not staggered, audit periods for all institutions. Therefore any adverse effect on the study's results due to this factor should be cancelled out.

4. Contamination

Where several hospitals in one community are participating in the trial, contamination may occur if, for example, staff from a CTL institution learn in casual conversation about indicator conditions or are informed about the content of feedback in EXP-FB groups. This will be impossible to control; however physicians in signing the letter which indicates they are informed about the trial will also have agreed not to discuss the trial outside their hospital (App. B-3).

5. Continuing medical education (CME)

Although Palmer and Reilly (1979) report no correlation has been demonstrated between quality of care and participation in CME, CME could theoretically alter physicians' management of indicator

conditions and an improvement in clinic performance might therefore wrongly be attributed to awareness of the audit procedure alone. If CME is a major factor in altering group performance, its effect should be discernible in the control group and any effect over and above that observed in the control group may be assumed to be at least in part due to the experimental intervention. The stratification of the participating institutions may permit identification of discrepancies in the extent of CME influence across strata.

In order to assess the type of CME exposure physicians have had, a letter and questionnaire will be sent (App. B-4) to each physician at the end of the second audit. The questionnaire will be pre-tested in the pilot study (S 2.4.4). From the questionnaire a categorical score for CME impact will be obtained for each physician (high, medium or low) and, in a manner analogous to the development of mean clinic scores (S 2.4.3.5.1 (b)), mean clinic CME impact scores will be derived (S 2.4.4.3 Item #7 (e)).

6. Influence of other health professionals

It is possible that in some clinics, non-physician staff may influence clinic routine in response to the audit. Changes in booking, record-keeping or procedures might occur or nurses may assume the function of aide-memoire to the physicians. Physicians

will be asked to assess the extent of this influence as part of the CME questionnaire.

2.4.2 Allocation to treatment groups

2.4.2.1 Stratification

Hospitals entering the study have been placed in two categories, teaching and non-teaching. The latter will be called community hospitals. The 42 community hospitals when ranked according to their number of beds yield a median number of 310 beds. Community hospitals are stratified into those with more than and less than the median number of 310.

(According to Lussing (1980), there are three ways to determine the number of beds attributed to any hospital. First is "total rated beds". This number represents the capacity of the physical plant but not the number of functioning beds. Second is the 'approved number', the number approved by the Ministry of Health. Third is the number 'staffed and in operation' as of March 31, 1979. It is the opinion of ministry officials that this latter number is the most accurate reflection of hospital operation. Seasonal variation causes the number staffed and in operation in the summer to be less than the number approved. It is the number of beds staffed and in operation which is used in this thesis.)

From the three strata, teaching hospitals, community hospitals with more than 310 beds (Comm >310) and community hospitals with less than 310 beds (Comm <310), there will be random allocation of participating hospitals to the three experimental groups:

EXP-FB	Experimental with feedback to physicians
EXP	Experimental with no feedback
CTL	Control Group

The three strata just described do not take into acount the possibility that indicator conditions may be unevenly distributed across clinics. This problem will be dealt with in S 2.4.3.2, item #3.

2.4.2.2 Randomization

2.4.2.2.1 Method of randomization

Each of the six possible combinations of EXP-FB, EXP and CTL will be given a number from one to six as shown below in Table 2.2.

TABLE 2.2

1. EXP-FB; EXP; CTL
2. EXP-FB; CTL; EXP
3. EXP; EXP-FB; CTL
4. EXP; CTL; EXP-FB
5. CTL; EXP-FB; EXP
6. CTL; EXP; EXP-FB

Using random number tables, the first digit encountered between one and six in the initial digits of each column successively, will indicate which combination of experimental groups will be applied to the first three hospitals in ranked lists (where the ranking is according to the number of beds) for each of the three strata. The second digit between one and six will determine the combination applied to the fourth to sixth hospitals, and so on, until all hospitals have been assigned to treatment groups.

In each stratum there may be a terminal group of hospitals of less than three. In such a case, the hospitals will be assigned to the first two or the first one of the experimental groups in the

trio selected by the randomization procedure. This method ensures that in each stratum of the three treatment groups there will be similar distributions of hospitals with respect to number of beds.

2.4.2.2.2 Time of randomization of hospitals

One could randomize the participating institutions immediately after their initial consent to participate and before the pre-audit period commences or one could randomize after the pre-audit period was completed. The latter point in time will be chosen if daysheets and prescription copies are the mode of census taking. Only institutions which comply with this probe will actually enter the trial, thus ensuring that optimal conditions will prevail in terms of institutional compliance. Then if the audit has an effect, it will be more likely to be demonstrated.

If the HMRI probe is used, randomization could occur either at the beginning or the end of the pre-audit period. If the trial were designed so that the CTL hospitals' medical records staff did not complete HMRI abstracts until after the second audit, then randomization would necessarily occur at the beginning of the pre-audit period. In this way the CTL institutions would escape the census of diagnoses and presenting complaints necessary for EXP-FB and EXP institutions.

However, as is explained in S 2.4.3.2, Item #7 (f) (ii), we have chosen ongoing HMRI abstraction by hospital medical records staff in all three experimental groups throughout the study. Thus the time of randomization is not crucial and could occur at the beginning or the end of the pre-audit phase.

2.4.2.3 Definition of treatment groups

1. EXP-FB

The protocol for this group is easily justified. It represents the experience of physicians submitting to audits and fulfills completely the CCHA criteria previously outlined; the CCHA criteria require feedback. The research question could be applied to this group. However, if the answer to the research question is derived only from comparing the EXP-FB group to the CTL group it will leave uncertain issues such as

- how much of any improvement in performance is due to continuing medical education rather than the audit?
- in the absence of feedback, how important is simple awareness that an audit is occurring in improving performance?

2. CTL

A control group of institutions in which there is no awareness of the audit will permit analysis of the change in performance which occurs due to factors beyond the control of the experimental design, factors which will presumably be acting similarly on all hospitals. Granted that the hospital administration will have received the initial letter requesting participation in the trial it is expected that any awareness is likely to subside with time when no further events occur. Certainly triggers to awareness will be reduced if the probe used is the HMRI abstract. Whatever the awareness of the trial may be, the staff at CTL hospitals will not have been informed which indicator conditions are assigned to their clinics nor will they know the explicit criteria for the indicator conditions.

3. EXP

With the EXP group submitting to the protocol as described, one should be able to answer the question how much is performance influenced by awareness of the study's implementation together with knowledge of the explicit criteria for the indicator condition of interest — but without any reinforcement by feedback.

In order for this approach to truly test for a reactive

effect, it would be preferable to simply inform the physicians that an audit was occurring but not to familiarize them with the indicator conditions. An argument for the proposed design wherein the EXP physicians are introduced to the indicator conditions is that by the time the second audit period is completed, the physicians will in all likelihood have forgotten the indicator conditions, but will have a lingering recollection that an audit has been performed. Comparison between EXP-FB and EXP performances will indicate how important the feedback mechanism actually was in influencing behaviour, and thus how important feedback is in quality assurance programs.

2.4.3 Intervention

2.4.3.1 Introduction

2.4.3.1.1 CCHA Criteria

The CCHA Guide to Hospital Accreditation (1977) sets out the following criteria for hospitals carrying out quality of care evaluation procedures.

1. Criteria of optimal achievable care, set by the hospital's own medical staff, must be measurable, with emphasis on justification for medical intervention and on patient outcomes.

2. Comparison of actual clinical practices against these predetermined criteria must take place.

3. Results must be analyzed by means of peer review. Clinically valid, acceptable variations must be separated from those that cannot be justified.

4. Action must be taken on variations deemed not justified.

5. Followup must occur after an appropriate interval to make sure action has been taken and has resulted in correction of any problems identified.

6. Documented reports of the results of all audit activities must go to the appropriate clinical departments, the medical advisory committee, the chief of the medical staff and to the hospital's governing body.

2.4.3.1.2 Relevance of the proposed trial to CCHA criteria

1. CCHA Criteron #1

"Criteria of optimal achievable care, set by the hospital's own medical staff, must be measurable, with emphasis on justification for medical intervention and on patient outcomes."

Chambers, et al. (1980) comment that "in developing indicator conditions the emphasis is on explicit criteria that define adequate care for the patient as opposed to criteria that define only high standards of care". In selecting indicator conditions (S 2.4.3.4) as a tool for this trial, I have accepted

that "adequate care" is the "optimal achievable care" mentioned in Criterion #1.

Since the trial involves many institutions, the criteria defining adequate care cannot be set by individual physicians within each of the institutions. The research team (having chosen the appropriate indicator condition for an individual clinic after the pilot study and pre-audit data are known) will introduce the criteria developed for the indicator condition(s) to the clinic staff. The clinic staff will be informed that the criteria were evolved by their peers and there will be opportunity for discussion. It must be stressed however that once the trial is in progress, the indicator conditions and their criteria will not be malleable.

Clinics for which several indicator conditions are appropriate will have one assigned randomly to them if the trial proceeds on the basis of only one indication condition per clinic. It is possible that experience with the pilot study (S 2.4.4) will encourage the application of more than one indicator condition to a clinic when the caseload permits.

2. CCHA Criterion #2

"Comparison of actual clinical practices against these predetermined criteria must take place."

The proposed trial will compare the actual records of the management of patients in the clinics with the explicit criteria defined for the specific indicator conditions.

3. CCHA Criterion #3

"Results must be analyzed by means of peer review. Clinically valid, acceptable variations must be separated from those that cannot be justified."

A nurse abstractor (S 2.4.3.2 Item #7, (b)) will abstract data from charts identified by the HMRI probe. The management recorded in each chart will be rated as superior, acceptable or indeterminate. From these ratings will be derived a mean score for for the clinics individually (S 2.4.3.5.1).

Audit committees in EXP-FB hospitals will meet the research staff to determine inter-observer reliability, that is, the agreement between the nurse abstractors' scores and those of the audit committee, the agreement between the nurse abstractors' scores and those of the research staff and finally between the

audit committee's scores and those of the research staff. The scores of the research staff will be regarded as 'truth' (S 2.4.3.2 Item #7).

4. CCHA Criterion #4

"Action must be taken on variations deemed not justified."

Clinic performance ratings after the first audit will be released to the clinic staff in EXP-FB hospitals so that any deficits in group performance will become known. The expectation that their performance should change will be made clear if serious deficiencies in performance are identified. Thus feedback after the first audit will be to individual physicians about group performance. Although one could argue that motivation to improve performance would be greater were physicians to be informed specifically about their individual performances, it is my opinion and that of CCHA staff that such an approach would have an adverse effect on the trial. Foreknowledge of such an assessment might instigate indirect behaviour which would compromise the completion of the trial.

However, if after the second audit, either the group as a whole or individual physicians within the group have not corrected previously identified deficiencies, then it may be appropriate for

the local audit committees to approach individual physicians about their individual performances.

5. CCHA Criterion #5

"Followup must occur after an appropriate interval to make sure action has been taken and has resulted in correction of any problems identified."

Followup will be accomplished by carrying out a second audit three months after the termination of the first audit procedure. This will be done by the nurse abstractors without further notification to the medical staff in any of the three treatment groups.

6. CCHA Criterion #6

"Documented reports of the results of all audit activities must go to the appropriate clinical departments, the medical advisory committee, the chief of the medical staff and to the hospital's governing body."

A summary of the trial's results will be sent to the outpatient department, the medical advisory committee, the chief of the medical staff and to the hospital's governing body for each participating hospital.

2.4.3.2 Sequential steps of the trial

1. Letter of invitation to hospital administrators

The CCHA will send a letter (App. B-1) to administrators of eligible hospitals in Ontario explaining that a randomized controlled trial is to be undertaken to establish whether or not CCHA-required audits have a beneficial effect on outpatient care and requesting that they indicate their willingness to participate.

2. Initial census of outpatient episodes in outpatient clinics

Daysheets and duplicate prescription pads should then be distributed to the appropriate clinics or HMRI abstract forms completed by hospital medical records staff for all outpatient visits to clinics which are presumed to be eligible. This census period, the pre-audit phase of the study, will last approximately three months.

3. Assignment of indicator conditions to clinics

The research staff, using summary data from HMRI or the daysheet and prescription data collected during the pre-audit phase, will identify the indicator conditions appropriate for each eligible clinic. When more than one indicator condition could be

assigned to one clinic, two options are possible. One indicator condition alone could be used having been randomly chosen from those which were appropriate. Alternatively, one could apply all the indicator conditions whose sample sizes could be met by the clinic's caseload. Although this would make analysis more complex and increase costs, it may be that when the study is underway, the assignment of multiple indicator conditions to single clinics is the preferable approach. If less than 10% of clinics could accept multiple indicator conditions, it may be preferable to universally apply only one indicator condition per clinic. If more than 20% of clinics can be assigned multiple indicator conditions, multiple ones should be assigned.

A further issue to consider is matching for indicator conditions. Thus rather than randomly assigning one of several possible indicator conditions to clinics as mentioned above, it may be feasible to choose indicator conditions such that there is matching across experimental groups. Certainly it will be preferable to compare performance scores which are based on common indicator conditions. However the case mixes of clinics may not always permit this.

Finally, since within one institution, it may be feasible for many clinics to submit to audit in this trial, it is necessary

to decide whether there should be an upper limit to the number of participating clinics per institution. On the one hand one can argue that all eligible clinics should participate in order to apply the intervention to as large a number of clinics and as many types as possible thus maximizing the generalizability of the trial. On the other hand, an increased number of clinics will increase costs and if there is any homogeneity of performance at all within institutions, the increased information yielded may not be worth the cost. Again, this is a decision which will have to be deferred until the time of the trial.

4. Allocation of hospitals to treatment groups

Stratification and randomized allocation of the eligible hospitals to the three treatment groups as previously described will be carried out (S 2.4.2.2).

5. Introduction of indicator conditions to clinic staff

Protocols for the assigned indicator conditions will be distributed to staff of the clinics in EXP and EXP-FB hospitals but not to CTL hospitals. Audit committees made up of members of the ambulatory care clinic staff will be asked to introduce the indicator conditions to the clinic staff. EXP-FB and EXP staff must be informed not only of the explicit criteria for the

indicator condition to be used in their clinic, but also that an audit is to be performed. Consultation with research staff will be available as required.

6. Physician awareness

To ensure that all the physicians working in the EXP-FB and EXP groups are familiar with the criteria for management of the selected indicator conditions and are aware that an audit is to be performed, a personalized form letter will be distributed (App. B-2) and each physician will be asked to sign it. The hospital administration will be asked to distribute these letters to staff physicians and to return them promptly to the research staff before the date at which the first audit is to be initiated. A delayed response will require intervention by the research staff to accomplish this task.

It is likely that physician awareness will vary from hospital to hospital and that the indicator conditions will be studied more zealously in some clinics than others. One response to this situation would be to adopt a highly interventionist strategy wherein the research staff energetically thrust explicit criteria "down the throats" of clinicians. Such a strategy would be self defeating for reason of its un-naturalness. Rather it is preferable to allow events to pursue their normal course with

physicians responding in their own fashion to the challenge of participating in an audit procedure rather than being acted upon by an audit procedure.

Ideally the phase encompassing steps one to six will be accomplished simultaneously in all EXP and EXP-FB institutions and steps one to four in CTL institutions.

7. First audit period

If possible, this period will begin simultaneously in all hospitals.

a) The probe

Continuous with the pre-audit and extending until the end of the second audit period, hospital medical records staff will complete HMRI abstract forms for all outpatient visits to participating clinics. As was explained previously, (S 2.4.1.2.2), the continuous collection of data by medical records staff and its transfer to HMRI abstracts is considered to be preferable to depending on clerical or nursing staff in the clinics to fill daysheets. All HMRI abstracts will be collected by the research staff prior to submission to HMRI.

b) Intra- and inter-observer reliability: hospital medical records staff

The performance of the medical records staff will be monitored.

(i) A simple numerical check will compare the number of HMRI abstracts prepared with the number of outpatient visits recorded by the hospital's administrative staff for the participating clinics. No more than a 2% difference is expected. If the discrepancy exceeds 2%, the research staff will determine whether the problem resides in the medical records department or in the administration, and then attempt to correct it.

(ii) A random sample of HMRI abstracts will be drawn by the research staff prior to the abstracts being forwarded to HMRI. An example of sample size estimation is given in (e) below. To test intra-observer accuracy for the recording of complaints and diagnoses the hospital medical records staff will fill duplicate HMRI abstracts for the sample drawn without being informed that they are being tested. A research staff person will also fill out HMRI abstract forms for the sample to test inter-observer agreement.

Kappa statistics will be used to assess intra- and inter-observer agreement at regular intervals throughout the study for quality control (Cohen, 1960). $k = (p_o - p_c)/(1-p_c)$ where p_o is the observed proportion of agreement and p_c is the proportion expected on the basis of chance alone. Thus kappa statistics are a measure of agreement corrected for chance agreement. When $k = 1$, the agreement between observers is perfect; when $k = 0$, there is no agreement and when $k = -1$ there is perfect disagreement. The calculation can be done by running the AGREE.PUB.LIB program (Hewlett Packard 3000, McMaster University).

A kappa statistic of no less than 0.8 is expected for both intra- and inter-observer reliability although the pilot study (S 2.4.4) may modify this expectation. If agreement falls below the criterion set, observer training will have to be improved. If agreement is still unsatisfactory once maximum levels of performance have been attained, the criterion will be modified.

c) Scoring

Using HMRI data abstracted by the hospital medical records staff, nurse abstractors trained by the research staff will assess the management of the assigned indicator conditions for the pre-audit, the first and second audit periods.

For each chart assessed the nurse abstractors will fill out a form similar to that used by Sibley, et al. (App. A-2) in order to arrive at an assessment of the management which will be categorically scored as indeterminate (I), acceptable (A) or superior (S). Then, using an HMRI abstract corrections form (App. A-4) the nurse abstractor will append the following information to the HMRI master file which contains data from the HMRI abstracts previously filled by the hospital medical record staff:

Hospital identity number

Batch and abstract identification

Phase of study (pre-audit, first audit or second audit)

Treatment group (one of three)

Stratum (one of three)

Indicator condition used (one of 20 or more)

Management score (I, A or S)

Since several visits will be associated with one illness episode during any one audit period, any one patient may have had several HMRI forms completed by the hospital medical records staff. (A "visit" represents one encounter of the patient with one clinic on one day, during which he is seen by one or more physicians. An illness episode in general includes all the visits made by the patient for the treatment of one illness. In particular, for each

indicator conditon, "illness episodes" are defined with explicit criteria (Sibley, et al. 1975[1], App. A-2).)

The assessment of the patient's management by the nurse abstractors cannot occur until criteria for an episode of care have been fulfilled. The correction entry appending the scoring information will link to the most recent HMRI abstract form completed by the hospital medical records staff for the illness episode being evaluated.

One patient will not be allowed to generate two separate scores for the management of any one indicator condition within one collection period of the trial. However, it is possible that for some indicator conditions one patient may generate management scores for more than one collection period. For example, two separate illness episodes of otitis media, each in a different audit period, might occur in one patient.

d) Intra- and inter-observer reliability: nurse abstractors

As with the hospital medical records staff, intra- and inter-observer agreement will be assessed using kappa statistics. Intra-observer agreement should be excellent (kappa > 0.8) and will be tested initially during the training program and subsequently

throughout the study to maintain quality control. At regular intervals the research staff will draw random samples of the scoring forms and the nurse abstractors will repeat the scoring procedure for the relevant charts without knowing they are being tested. This intra-observer agreement will be tested. The same random sample will permit testing of inter-observer agreement between research staff and nurse abstractors. Kappa should exceed 0.8. As was the case with the hospital medical records staff, if the criterion is not met, the research staff will seek to improve the training of the abstractors.

Finally it will be necessary to determine the agreement between the score entered on the scoring forms and the score entered on the HMRI abstract corrections form. Here kappa should exceed 0.9.

e) <u>An example of sample size estimation for determination of kappa</u>

To estimate the sample size required to assess agreement on scoring one could use the counts reported by Sibley, et al. for the indicator condition otitis media and calculate the agreement due to chance as shown below.

TABLE 2.3

		Rater A			
Percent Score		I	A	S	
	I	7/84			28
Rater B	A		0.04		2
	S			49	70
		28	2	70	100

The variance of kappa can be calculated with the following formula (Fleiss, 1973):

$$\text{var}(\kappa) = \frac{1}{N(1-p_c)^2}\left[(p_c+p_c)^2 - \sum_{i=1}^{m} p_{i.}p_{.i}(p_{i.}+p_{.i})^2\right]$$

where N is the sample size we wish to estimate, p_c is the proportion of agreement due to chance, $p_{i.}$ is the proportion of all charts assigned to the ith category by rater B and $p_{.i}$ is the proportion assigned to the ith category by rater A so that

$$\text{var}(\kappa) = \frac{1}{N(.43)^2}\left[(.57+.57)^2 - (.28\times.28\times.56 + .02\times.02\times.04 + .7\times.7\times1.4)\right]$$

$$= \frac{.8923}{N}$$

then:

Squaring: $.64 = 3.84\ (.8923/N)$

and $N = 3.84 \times .8923/.64 = 5.3$ or 6.

This number is less than that required to satisfy the assumption that $(k - E(k))/var(k)$ is normally distributed. To satisfy this assumption, the minimum number of observations for $r = 3$ categories is at least $2r^2$, in this case, 18 (Cicchetti and Fleiss, 1977). This will be the sample size used.

f) Some methodological problems

(i) Lost charts

Some charts, after having been identified by the probe may not be retrievable. Since it is possible that charts are more likely to be lost the more clinics a patient visits and this may influence management by an individual physician, it may be necessary to determine whether the management of patients with lost charts differs from that of patients whose charts are not lost. Up to a 5% rate of unfound charts will be considered acceptable. More than that will require investigation.

(ii) The timing and effect of data abstraction

Scoring requires that the patients charts be subjected to two processes of data abstraction. The first is performed by the hospital medical records staff and the second by nurse abstractors.

Examination of Fig. 2.1 will reveal that the design chosen for this thesis requires the hospital medical records staff to function continuously throughout the entire study period in the three experimental groups.

The advantages of doing so are several. The task becomes a routine one in the same way that HMRI abstracts are currently routinely completed for all inpatient admissions. Secondly, there is no need to call a halt, for example at the end of the first audit period, and then to reinstitute the process at the initiation of the second audit. Cessation even for several months may result in loss of skills and resumption in itself would be a signal that "something" is happening. Further, there is no reason to suspect that practising physicians will be any more aware of HMRI abstraction for outpatients than they are now for inpatients.

Since EXP-FB and EXP groups should be the same in all respects with the exception of feedback after the first audit, hospital medical records staff should fill HMRI abstracts for the

pre-audit and first audit periods in both groups. Without the HMRI abstraction in the pre-audit period, indicator conditions could not be assigned to clinics. Without HMRI abstraction in the first audit period, nurse abstractors would not be able to derive scores, necessary for feedback to occur. Since cessation followed by resumption of medical records staff activity may attract attention, continuous functioning until the end of the second audit is proposed.

The same continuous functioning by the hospital medical records staff is considered appropriate for the CTL group for the reasons outlined above, namely it is an extension of a current activity already performed by hospital staff and the activity is unlikely to be noticed by physicians. Another factor supporting this decision is that the alternative — retrospective completion of HMRI abstracts after the second audit period — will give rise to problems. In some hospitals it will be difficult to identify charts which have passed through the clinics of interest at the time periods of interest. Secondly such a process concentrates the work load of the medical records staff very intensively. The consequence may be either speed at the expense of reliability or a protracted period of data gathering which will delay the final analysis of the entire study.

The second process applied to patient charts is the scoring procedure by the nurse abstractors. Scores must be derived for the first audit period in the EXP-FB group, otherwise feedback cannot occur. Since, as previously mentioned, the EXP group should have a parallel experience to the EXP-FB group in all respects except for feedback, it is reasonable to have the nurse abstractors score the first audit period concurrently in the EXP group. Simultaneously, the scores for the pre-audit period should be derived for EXP and EXP-FB groups.

However, the nurse abstractors should not return to these hospitals to score performance while the second audit period is occurring. Unlike the hospital medical records staff who are "part of the scenery", the nurse abstractors are likely to be more visible. Their arrival, especially in smaller hospitals, might heighten physician awareness during the second audit period. In the EXP-FB and EXP hospitals it is proposed that scoring for the second audit period be done after its completion.

In order to reduce the possibility of contamination, no nurse abstractor activity will occur at all in CTL hospitals until after the completion of the second audit. Here the alternative would be to have the CTL hospitals exposed to an experience identical to that of the EXP-FB and EXP hospitals in all respects

save for being introduced to the indicator conditions. However, the advantage of ensuring a homogeneous experience for all three groups is less important than the advantage of avoiding the contamination likely to occur were the nurse abstractors to function in CTL institutions during the first audit period.

(iii) Bias arising from the attitudes or the presence of nurse abstractors

Since the nurse abstractors will be part of the research staff, it is unlikely that they will display prejudice either in favour of or against specific institutions. To avoid triggering attention on the audit, the nurses will be instructed to maintain as low a profile as possible when they visit each hospital and to be uncommunicative about the results of their audit if they converse with hospital staff.

8. Feedback to EXP-FB clinics

Once performance scores have been calculated for the clinics in the EXP-FB group for the first audit period, the feedback process must occur in such a way that a minimum level of feedback uniformity is ensured for each physician. This could be accomplished by a personalized standard letter (App. B-3) sent to each physician by registered mail. Hospitals accustomed to

presenting the results of audits at grand rounds will be encouraged to follow this practice. It will be necessary to record which hospitals augment the impact of the letters by holding rounds as it will also be necessary to record the proportion of clinic physicians which attends rounds (S 2.4.3.5.4).

9. Second audit

Three months after termination of the first audit period and when feedback to the EXP-FB group has been accomplished, a second audit period of equal length to the first will commence.

At the conclusion of the second audit period, questionnaires with introductory letters will be sent to all participating physicians to determine the type and extent of continuing medical education to which they have been exposed (S 2.4.1.3.2 Item #5) and whether or not the CME topics included the conditions assigned to their clinic.

10. Analysis and report

After the second audit period is completed, summary information about the results of the trial will be transmitted to appropriate persons as described in the CCHA Guide (CCHA 1977).

2.4.3.3 Indicator conditions

2.4.3.3.1 Justification for using indicator conditions

By definition, indicator conditions are disorders or diseases which occur relatively frequently and for which there is consensus about diagnosis and management or, as Kessner (1978) says, a clinical entity whose outcome can be affected favourably or adversely by the choice of treatment.

Sibley, et al. (1975[1]) have developed indicator condition protocols which can be used for program evaluation, for continuing education or for accreditation purposes (Chambers, et al. 1980). They suggest that quantitative scores on the adequacy of care can be derived from the untouched medical record retrospectively in primary care private practices. Such a method seems more feasible than evaluation procedures which depend on observation or on detailed questionnaires administered to patients and physicians.

As Tugwell remarked in a discussion of quality of care (1979[2]) "Unfortunately assessment of health outcomes of patients is expensive, time-consuming and the clinical process needs to be measured anyway if differences in patient outcome are to be interpretable." Quality of care assessments which are based only on patient outcomes are handicapped by the rarity of adverse

outcomes and by the profusion of factors outside medical care which influence outcome. If processes are identified which are known to affect patient outcome (either favourably or adversely), one has a mechanism for assessing quality of care which links clinical process and patient outcome. Indicator conditions are a method of implementing such an approach.

Sibley, et al. (1975[1]) developed protocols for a number of indicator conditions (see Table 2.4). After an illness episode has been identified by a probe, it must meet general criteria to establish the eligibility of the episode as a tracer condition (App. A-2). Then for each indicator condition, protocols have been developed which

- define an episode
- list the possible interventions by the physician and
- specify which interventions are necessary for achieving one of three possible scores (App. A-2).

Finally Sibley, et al. have developed an abstract form with which the score may be derived (App. A-2).

Applying indicator conditions to three private family practices it was found that the scores were "in close agreement

with the outcome measures of mortality and physical, social and emotional function done on the same study subjects." (Sibley, et al. 1975[2]). There was internal consistency in the scores derived from three approaches: assessment of indicator conditions, the use of drugs and the opinions of consultants using implicit criteria. Finally there was high inter-observer agreement on scoring. These observations lend credence to the further application of indicator conditions.

2.4.3.3.2 Relevance of indicator conditions to hospital clinics

Table 2.4 lists the types of clinics found on the HS-1 form under the heading "General and Special Clinics" and indicates the at least theoretically appropriate indicator conditions for each clinic. "Orphan" clinics, for which none of the currently developed indicator conditions are appropriate, are marked with an asterisk.

Table 2.4

Types of clinics on MS-I form	Types of Indicator Conditions Theoretically Applicable.
Allergy*	
Arthritis	Arthritis.
Cancer*	
Cardiac	Chest pain, myocardial infarction.
Dental*	
Dermatology	Pityriasis rosea, acne.
Endocrine*	
Family Practice	Hypertension, prenatal care, care of newborn, urinary tract infection, knee injuries, pityriasis rosea, anemia, arthritis, acne, back pain, chest pain, fatigue, headache, myocardial infarction, obesity, The Pill, rectal bleeding.
Gastro-intestinal	Rectal bleeding.
General Medicine	Hypertension, urinary tract infection, pityriasis rosea, acne, anemia, arthritis, back pain, chest pain, fatigue, headache, myocardial infarction, obesity, The Pill, rectal bleeding.
General Surgery	Rectal bleeding.
Gynecology	Vaginal discharge.
Neurology	Headache.
Obstetrics	Prenatal care, The Pill.
Ophthalmology*	

Orthopedics	Back pain, knee injury.
Ear, nose and throat	Otitis media.
Psychiatry	Depression.
Children	Otitis media, care of newborn, immunization.
Tuberculosis*	
Urology	Urinary tract infection.
Venereal disease*	
Other*	

---

*orphan clinics: see text

Table 2.5 lists the indicator conditions which did and did not discriminate among 19 physicians (Goldsmith, 1980). An indicator condition has not discriminated if when applied to physicians' practices, the scores obtained do not permit categorization of individual performances at different levels; that is, it cannot separate favourable from unfavourable performances. For example, if an indicator condition requires a sample of 27 episodes per physician, but the physician managed only eight patients, the indicator condition could not discriminate, but that is due to an excessively small sample, not due to any failing of the indicator condition itself. In contrast, an indicator condition such as hypertension which requires a sample size of 634 episodes per physician to differentiate between two physicians

(alpha = 0.05 and beta = .20) may benefit from having its criteria modified. In the data Goldsmith has analysed, extreme intra-physician variability was observed and this could account for part if not all of the failure of this indicator condition to discriminate.

TABLE 2.5

INDICATOR CONDITIONS WHICH

DID DISCRIMINATE AMONG 19 PHYSICIANS	DID NOT DISCRIMINATE AMONG 19 PHYSICIANS
Otitis media	Hypertension
Prenatal care	Urinary tract infection
Newborn care	Knee injury
Immunization	Pityriasis rosea
Depression	Anemia
Acne	Arthritis
Headache	Back pain
Obesity	Chest pain
	Fatigue
	Myocardial infarction
	The Pill
	Rectal bleeding
	Vaginal discharge

Taking the indicator conditions which did differentiate, one can select those with the lowest sample size requirements (Goldsmith, 1980) as shown in Table 2.6 below..

TABLE 2.6

INDICATOR CONDITIONS WHICH SHOULD DISCRIMINATE AND HAVE FAVOURABLE SAMPLE SIZE

(alpha = 0.05 beta = .20)

INDICATOR CONDITION	NO. OF EPISODES REQUIRED/M.D. FOR A TWO-MAN PRACTICE
Otitis media	45
Prenatal	64
Newborn	8
Immunization	17
Depression	52
Acne	8
Headache	35
Myocardial infarction	6
Obesity	19

Since at this time it is not known what sample sizes are required to detect inter and intra-clinic differences, we have assumed only that sample size requirements will retain the same ranking in the new setting (that is hypertension and urinary tract

infections will still need the largest sample sizes) and that it is likely that sample size requirements will be somewhat smaller for detecting intra- and inter-clinic differences than for detecting inter-physician differences.

Further it is possible that the reduction in sample size requirements may be such that most if not all the indicator conditions developed by Sibley, et al. (1975[2]) may be used in the trial. Thus the research investigator has two options.

a) The first option is to apply all the indicator conditions to the many clinics displayed in Table 2.4. This would be the optimal course in that a maximum number of clinics and the widest range of illness episodes would be involved in the audit.

b) The second option would be to defer to the constraints which at this time are only conjecture, but which may prove to be very real after the pilot study data are analysed. Table 2.7 provides the abbreviated list of clinic types selected on the basis that they can be audited by indicator conditions which are likely to require the smallest sized samples. The six types of clinics thus selected account for only 147 of the 548 clinics functioning in Ontario hospitals in the fiscal year of 1978.

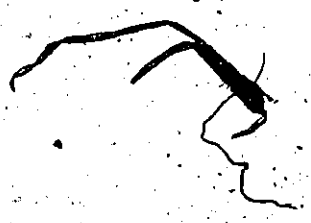


TABLE 2.7

SHORT LIST OF CLINICS

TYPE OF CLINIC	#IN ONTARIO*	PROPOSED INDICATOR CONDITIONS
Obstetrics	16	Prenatal care
Children's	24	Newborn care Immunization
Psychiatry	50	Depression
Dermatology	21	Acne
Family Practice and General Medicine	36	Acne Headache Obesity Prenatal care Newborn care Immunization Depression Otitis Media

*(App. C.2)

c) A third option would involve the application of not only the currently available indicator conditions but also newly developed ones; however the identity of the latter will not be known until pilot study data are available. The approach to developing new indicator conditions is described by Chambers, et

al. (1980).

Only when a census of illness episodes is available for each clinic will it be known whether in fact the indicator conditions proposed for any clinic are at all appropriate. For example children's clinics may not include immunization procedures in their services. It should be noted that about one half of the clinics reported by community hospitals are psychiatric or children's clinics.

2.4.3.3.3. Exclusion criterion for indicator conditions

1. The explicit criteria defined for each of the indicator conditions (App. A-2) by Sibley, et al. for private practices in primary care must be modifiable so that they can be applied to general and special clinics in hospitals. Those which cannot be so modified will not be employed.

2.4.3.3.4 Inclusion criteria for indicator conditions

Depending on the findings of the pilot study, indicator conditions will be preferred which

1. Require smaller rather than larger sample sizes.

2. Differentiate before and after within clinics or between clinics more effectively.

3. Can be applied over a shorter rather than longer period of time.

4. And for which the clinical effectiveness associated with management is highest. A panel of experts will be used to rank all indicator conditions considered for use in the trial (whether current or newly developed) in terms of the evidence for the clinical effectiveness of their management.

2.4.3.3.5 General modifications of indicator conditions

1. The indicator conditions developed for use in primary care practices (Sibley, et al. 1975[2]) will be modified for application to primary ambulatory care delivered by both primary and specialist physicians in general and special clinics in hospitals.

2. It is likely that the time required to collect illness episodes will be longer in the hospital clinic setting than in primary care private practice. Private practice offices may function five to seven days a week while single clinics may function only once or several times a week often for only a few hours per session.

3. While Sibley, et al. used 35 episodes per indicator condition per practice, the present trial will use, where possible, sample sizes estimated from pilot study (S@2.4.4) data for each indicator condition.

4. While Sibley, et al. did not inform the audited physicians of the indicator conditions or the explicit criteria for them until after the audit, in this trial every participating physician (except those in the CTL group) will know both the indicator condition and its explicit criteria before the audit begins. For this reason peer consensus in the development of criteria to justify the choice of criteria by every participating physician is not seen to be necessary.

5. While Sibley, et al. used daysheets, prescription copies, hospital records and direct record searches as probes to identify charts for this trial, we propose using HMRI abstract forms as the probe.

6. While Sibley, et al. used nurse abstractors and an abstract form to arrive at scores for performance, the present trial will use nurse abstractors and a similar abstract form to develop the score but the results will be appended to HMRI abstract forms to centralize data collection.

7. While Sibley, et al. used a categorical score combining data from indicator conditions, referral practices and drug use patterns, the score for this trial will be based only on indicator conditions. Drug use patterns and referral practices would be difficult to monitor in as many as 70 institutions distributed over such a wide geographical area as the province of Ontario. In particular, assuring full compliance of prescription gathering in the clinic milieu would be difficult if not impossible. The multitude of pharmacies involved in filling out clinic prescriptions in urban settings would also add to the complexity of such an undertaking.

Like Sibley, this trial will use a categorical score for the indicator conditions, namely: indeterminate, adequate or superior.

8. While Sibley, et al. correlated the categorical scores with specific measures of physical and social outcomes, this trial will not incorporate any independent outcome measure other than those implied by the indicator condition protocols.

2.4.3.3.6 Specific modification of indicator conditions

a) Criteria for eligibility of an episode

The eligibility of an episode for appraisal as a tracer condition depends on six criteria (App. A-2, Sibley, et al. 1975[1]). These criteria are modified below to meet the needs of the proposed trial.

1. The indicator condition must have been assessed and/or managed as an ambulatory care episode in whole or in part. If management begins in one clinic and the patient is referred to another clinic for continued management within the same institution, the management will be assessed and the score attributed to the referring clinic whether or not the second clinic is participating in the trial.

2. Episodes which are scrutinized must fall within the intervals of interest, namely the pre-audit, first audit and second audit periods of the trial.

3. The identity of the physician managing the patient must be recorded on the chart. Where a clinical clerk has managed the patient, the score for management will be assigned to the supervising staff person.

4. An episode is defined for each indicator condition used.

b) Criteria for specific indicator conditions

Rather than offering detailed protocols for every indicator condition, what follows are two examples of the sorts of modifications which will be necessary.

1. Otitis media (App. A-2)

The definition of an episode will remain unchanged. However, under "categories of intervention", the criterion "tetracycline in a child under the age of 8" would be inappropriate for a clinic the caseload of which was entirely adult. Thus in an adult clinic, option one, item two under "adequate scoring" would read simply "an inappropriate antibiotic would be chloramphenicol".

Furthermore, if this indicator condition were assigned to an ENT clinic (which one would expect to be staffed by or supervised by specialists) option two for "scoring adequate" which is "consultation" is not appropriate and would be deleted. Finally, in ENT clinics, option two, item eight under "scoring superior" would be modified to read "follow-up or audiometric examination."

2. Hypertension

If the initial assessment and management of hypertension must occur within one of the three data collection periods, the number of illness episodes per clinic may be intolerably small and a large number of patients being managed for hypertension would be lost for audit purposes. To overcome this problem it may be useful to develop two sets of criteria: one for the maintenance therapy of hypertension and the other for initial assessment and management.

For the modified indicator conditions to have clinical credibility, it will be necessary to arrive at a consensus on the required changes using a panel of experts.

2.4.3.4. Feedback

Ideally the feedback delivered to the physicians in the EXP-FB clinics would be as intense as possible to achieve a maximum reinforcing effect if such occurs. On the other hand, accurate assessment of the role feedback plays in quality assurance programs might be impeded if differences in the intensity of feedback occurred without the knowledge of the investigator.

The simplest feedback mechanism would be a form letter (App. B-3) sent to all EXP-FB physicians informing them of the results

achieved in their clinics. The letter should also indicate how the clinic's performance compares with its peer clinics. This mechanism will be implemented in all EXP-FB clinics.

The physicians' personal scores might also be included in the letter, only his own, not those of his colleagues. An argument for doing so is that identifying to an individual his own performance measures may have a greater impact on his behaviour than reporting poor group performance to him. In the latter case, the individual may consider himself to be better than the group and therefore not in need of altering his performance. The argument against reporting individual's scores, at least initially, (that is after the first audit phase), is that audits tend to make physicians defensive if not hostile and any negative feedback may provoke resistance to or interference with the trial. Such a possibility is feared by Dr. J. Murray of the CCHA (1980).

In addition to the letter, if a hospital's customary practice is to discuss the results of their audit activities at rounds, it will be their option to use feedback information from the research staff for this purpose. In this case, it will be necessary to categorize EXP-FB clinics into those which simply received the letters and those which received the letters but in addition exposed their staff to discussion of the results at rounds. A further

aspect which would have to be taken into account is that not all clinic physicians may attend rounds and were there a high percentage of absenteeism, one could not atttribute any extra feedback impact to grand rounds, even if grand rounds were held (S 2.4.3.5.4).

2.4.3.5 Outcomes

The outcome of major importance will be a series of mean scores representing:

1. One clinic's performance for a single indicator condition.
2. One clinic's performance for all the indicator conditions assigned to it.
3. One institution's performance based on all the clinics within it.
4. The performance of all institutions within or across strata for each experimental group.

2.4.3.5.1 Mean scores

Two types of mean scores will be calculated for the pre-audit period. The first will be the mean TIC-score (trial-indicator-condition-score) based only on data from the trial.

The second will be a mean SIC-score (Sibley-indicator-condition-score) in which data collected by Sibley, et al., SIB-data, (re-expressed data from Sibley) will be applied to trial data. The reason for so doing is that if analysis (S 3.2) shows that the distributions of the two scores do not differ, it will be an indication that the modified indicator conditions in the general and special clinic settings are producing scores which are similar to the original indicator conditions in the private primary care setting; in other words, a kind of concurrent validity will have been demonstrated.

It will not be appropriate to calculate SIC-scores for the first and second audit phases because SIB-data represent the performance of physicians who did not know what the indicator conditions being applied were, whereas in this trial, physicians in the EXP-FB and EXP group will know beforehand the criteria for the indicator condition(s) to be applied to their clinic.

Since it may be easier to understand the development of mean SIC- and TIC-scores if the method is presented using actual numbers from a previous study, what follows will deal first with mean SIC-scores.

a) Mean SIC-score for one clinic based on one indicator condition

SIB-data are available for each indicator condition and were developed by aggregating the categorical scores from 19 practices and re-expressing them.

An alternative to re-expression (which will be described in detail below) would be to assign scores of I = 0, A = 1 and S = 2. Such scoring implies that the interval between indeterminate and acceptable is the same as between acceptable and superior. It is not known that this is true. Ordinal data are not necessarily interval. In contrast, the method of re-expression used in this thesis takes into account cumulative distribution, can distinguish between U-shaped and bell-shaped distributions (which 0-1-2 cannot) and identifies indeterminate performance with a negative and superior with a positive score.

Using data from Sibley, et al., Goldsmith (1980) applied a method of re-expression described by Mosteller and Tukey (1977[1]) to counts for the indicator condition otitis media as shown in Table 2.8.

TABLE 2.8

Result	Count	p = Fraction<	P = Fraction≤	$\phi(p)$	$\phi(P)$	$\frac{\phi(P)-\phi(p)}{P-p}$*
I	120	.0000	.2777	.000	-.5915	-2.13
A	9	.2777	.2986	-.5915	-.6103	- .90
S	303	.2986	1.000	-.6103	.0000	.87
	432					

* The signs have been reversed to parallel the gradient in categories so that negative values correspond to poor performance and positive to good. The re-expressed values will be referred to as SIB-data.

In the table, p equals the fraction less than the count observed, P equals the fraction equal to or less than the count and formulas for $\phi(p)$ and $\phi(P)$ are:

$$\phi(P) = P \log_e p + (1 - p)\log_e(1 - p)$$

$$\phi(P) = P \log_e P + (1 - P)\log_e(1 - P)$$

A clinic's mean SIC-score for each indicator condition (in this case #1) can be developed for the data collected in the trial, as shown below.

TABLE 2.9

SIB-data	Count in clinic	Product
I = -2.13	n_1	$-2.13n_1$
A = -.90	n_2	$-.9n_2$
S = .87	n_3	$.87n_3$
	n	

The clinic's mean SIC-score will be the sum of the products divided by n. Where one clinic is assigned several indicator conditions, the SIC-scores will be aggregated to obtain an overall mean SIC-score for the clinic.

b) Mean TIC-score for one clinic based on one indicator condition

The reason for calculating a second set of mean scores is that the indicator conditions will have been modified and applied to a new setting; they may yield very different mean scores to

those developed for Sibley, et al.

(i) Re-expressing the trial data

Re-expressing the data is required for the same reason given in a) above. To develop the analogue of SIB-data, a jack-knifing procedure (Mosteller and Tukey, 1977[2]) as described below, will be used. Jack-knifing is a technique which permits one to avoid developing a score from data which will be subsequently analysed; thus jack-knifing validates the score.

Jack-knifing would aggregate all the counts achieved in the pre-audit period by all clinics (say A to Z) for an indicator condition (say #1) as shown in Table 2.10.

TABLE 2.10

	Clinic A	Clinic B	Clinic C	. . . Clinic Z	
I	n_{AI}	n_{BI}	n_{CI}	n_{ZI}	n_I
A	n_{AA}	n_{BA}	n_{CA}	n_{ZA}	n_A
S	n_{AS}	n_{BS}	n_{CS}	n_{ZS}	n_S
					N

In the same way SIB-data were developed (Table 2.8), TRIAL-data (the re-expressed data from the trial) could be developed for Clinic A as shown in Table 2.11.

TABLE 2.11

Result	# = Count	p = Fraction<	P = Fraction≤	$\phi(p)$	$\phi(P)$	$\frac{\phi(P)-\phi(p)}{P-p}$
I	$n_I - n_{AI}$					X_A
A	$n_A - n_{AA}$					Y_A
S	$n_S - n_{AS}$					Z_A

Similarly, for Clinic B, TRIAL-data X_B, Y_B, Z_B could be developed from the difference $n_I - n_{BI}$, $n_A - n_{BA}$, and $n_S - n_{BS}$, and so on for all clinics.

(ii) Developing the TIC-score

For clinic A and indicator condition #1, the method would be:

TABLE 2.12

TRIAL-data	Count in Clinic A	Product
$I = X_A$	n_{AI}	$n_{AI}(X_A)$
$A = Y_A$	n_{AA}	$n_{AA}(Y_A)$
$S = Z_A$	n_{AS}	$n_{AS}(Z_A)$
	n_A	

and clinic A's mean TIC-score for indicator condition #1 would be the sum of the products divided by n_A. Similarly for Clinic B and indicator condition #1 the mean TIC-score would be:

$$\frac{n_{BI}(X_B) + n_{BA}(Y_B) + n_{BS}(Z_B)}{n_B}$$

Having explained the development of mean scores, I will now turn to outcomes in the context of the sequential phases of the trial.

2.4.3.5.2 Pre-audit phase

This phase is expected to reveal a census of the clinic caseload which will allow the research team to choose appropriate indicator condition(s) for each clinic. Subsequently, this phase will be subjected to a retrospective audit.

In this period, for each clinic, the number of sessions per month, the number of patients per month and the frequency of presenting complaints and diagnosis will be tabulated.

2.4.3.5.3 Audit phases

An abstracting from (App. A-2) similar to that used by Chambers, et al. (1980) will be used to record information for each patient chart audited and to derive the score. Depending on the number of explicit criteria which are fulfilled, the score will be

indeterminante, acceptable or superior. The results will be displayed as follows:

TABLE 2.13

HOSPITAL	IND. COND.	PRE-AUDIT MEAN SCORE	FIRST AUDIT MEAN SCORE	SECOND AUDIT MEAN SCORE
CLINIC A				
CLINIC B				
CLINIC C				

Similarly structured tables will display summarized data organized in the following ways:

1. For each experimental group aggregating the mean scores for all institutions within each stratum.
2. For each experimental group collapsing the data across strata.

3. For each type of clinic classified as to experimental group and stratum.

4. For each type of clinic classified as to experimental group but collapsed across strata.

5. By indicator condition classified as to experimental group and stratum.

The question arises, how valid is a comparison of scores aggregated from several physicians in several clinics within one institution, in turn aggregated with scores from other institutions in the same stratum and in same treatment group — with similarly aggregated scores from institutions assigned to the other two treatment groups? Sibley sets a precedent (Noack, 1980) for aggregating scores for several indicator conditions in individual practices. He also reports that within practices there is good score agreement when two different sets of indicator conditions are used to arrive at the aggregate scores.

The outcomes for the CME questionnaire component of the trial are described in S 2.4.4.3.

2.4.3.5.4. Feedback phase

As mentioned previously (S 2.4.3.4), it will be necessary to document the intensity of feedback operating in each clinic. The audit committees in each EXP-FB hospital, or the research staff if necessary, will be responsible for gathering data as organized in Table 2.14 below.

TABLE 2.14

STRATUM	INSTITUTION	CLINIC TYPE	FEEDBACK: LETTER ONLY	FEEDBACK: LETTER WITH ROUNDS	CLINIC PHYSICIANS AT ROUNDS	CLINIC PHYSICIANS NOT AT ROUNDS

2.4.4 Pilot study

2.4.4.1 Purposes

1. To yield an estimate of the time periods necessary to collect specific numbers of episodes for each indicator condition.

2. To reveal which if any new indicator conditions need to be developed.

3. To permit validation of the modified indicator conditions for the general clinics and allow for input from specialists in special clinics as to the appropriateness of the criteria for indicator conditions used in their clinics.

4. To indicate the magnitude in change (if any) of performance associated with an audit.

5. To allow calculations of sample sizes for each indicator condition based on observed changes and their variabilities.

6. To confirm that predicted levels of intra- and inter-observer agreement are attainable.

7. To pretest the CME questionnaire (App. B-4).

2.4.4.2 Convenience sample

A group consisting of one teaching and four community hospitals will be used. Toronto Western Hospital (TWH) could be approached for the pilot study for two reasons. First it has expressed an interest in implementing audits. Secondly, my association with TWH makes it preferable that it participate in the pilot study rather than the actual trial. TWH has a large number of clinics and a moderate caseload and should yield the desired information. The four community hospitals will be chosen randomly from the list of such hospitals reporting general and special clinics. The strata > 310 and < 310 beds will be represented.

Finally the 10 hospitals with "orphan" clinics (S 2.4.2.3.2 Item #5) will be asked to enter the initial assessment phase of the pilot study and if feasible, enter the audit phase.

2.4.4.3 Organization

1. During an initial assessment period of six months, HMRI abstracts will be completed by hospital medical records staff for all outpatient encounters in all general and special clinics (except for those described as "other"). In this way the frequencies of diagnoses and presenting complaints will be determined and it will be possible to assign relevant indicator conditions to clinics.

Within this period, decisions will be made as to whether new indicator conditions can be developed, in particular for the "orphan" clinics but also for the general and special clinics.

Research staff will monitor medical records staff performance in completing HMRI abstracts as previously outlined (S 2.4.3.2 Item #7 (a)).

2. Introduction of indicator conditions to clinic staff will be effected by the research staff. The audit committee for each hospital along with the relevant specialists for each indicator condition, will be invited to comment on the acceptability of the explicit criteria developed for each indicator condition. Where necessary, further modifications will be made to the indicator conditions. However, indicator conditions must be constant across institutions. Acceptance by the clinicians will be considered as evidence of face validity for the indicator conditions.

3. For the audit period which ensues, summary information from HMRI abstracts completed by the hospital medical records staff will be the probe by which nurse abstractors will select charts for auditing. Management of episodes will be scored as indeterminate, acceptable or superior.

4. The management of indicator conditions during the pre-audit phase will be scored by nurse abstractors to permit a before-after comparison.

5. At the end of the audit period, participating physicians will be asked to fill out the CME questionnaire (App. B-4).

6. The outcomes of the pilot study will be as follows:

a) The frequency of diagnoses and presenting complaints occurring in each clinic during the initial assessment period will be tabulated. If the pre-selected indicator conditions are irrelevant to the caseload in that the most prevalent illness episodes are other than those represented by the indicator conditions, then consideration will be given to developing new indicator conditions.

b) Mean clinic TIC-scores will be developed for the before and after phases of the pilot study.

c) The identity of the physicians and/or the interns staffing each clinic will be tabulated.

d) The mean clinic CME-scores will be developed using the method described in S 2.4.3.5.1 (b) (i) and (ii) with CME-data and mean CME-scores being developed analogously to TRIAL-data and TIC-scores.

7. The analysis of the results will be as follows.

a) The monthly frequencies of diagnoses per clinic will be assessed in light of the estimated sample sizes required for each indicator condition, that is the sample sizes estimated from the Sibley data. Appropriate indicator conditions will thus be identified for each clinic involved in the pilot study. The minimum expectation is that at least one clinic in any one hospital will have a caseload permitting the assignment of at least one indicator condition. When more can be applied they will be.

b) The two types of mean clinic management score having been calculated, a Chi-square analysis ($\alpha = 0.05$) will be used to establish whether the distributions of the two scores are similar. In this way the concurrent validity of the developed scores will be assessed.

c) Adjusted t-tests for unpaired samples will be used to determine the statistical significance of audit effect for each indicator condition. The difference in mean sources (before-after) will be used to make more accurate estimates of sample sizes for the main trial.

If the pilot study reveals no change in mean scores as a consequence of introducing the indicator conditions to the physicians, modification of the overall design of the trial will have to be contemplated. New strategies such as appending the explicit criteria for the assigned indicator condition to every outpatient chart may heighten awareness of the audit advantageously. On the other hand one might conclude that the explicit criteria in the modified indicator conditions were not appropriate for the outpatient setting at all. Such matters will have to be resolved when more information is available.

d) Sample sizes will have to be calculated for each indicator condition and will be based on the type of difference which is to be detected. The estimated sample sizes (Goldsmith 1980) for individual indicator conditions (Sibley, et al. 1975[1]) are based on detecting interphysician differences at one point in time. The number calculated is the number of illness episodes which each physician must see. In this trial, the differences of interest are group differences rather than individual. Thus the number of episodes of illness which must occur in each clinic rather than the number which must be seen by each physician will be calculated.

The way in which the sample size could be estimated from pilot study data for one indicator condition, otitis media, in one clinic, follows (see Tables 2.8 and 2.9).

TABLE 2.15

Score	SIB-data	Pre-audit count	Product	First audit count	Product
I	-2.13	15	15 x -2.13	5	5 x -2.13
A	-.90	5	5 x -.90	5	5 x -.90
S	+.87	10	10 x .87	20	20 x .87

From this means and variances can be calculated:

TABLE 2.16

	Pre-audit	First audit
$\bar{X}$	- 0.925	- 0.07
ΣX^2	79.6725	41.87
$(\Sigma X)^2/n$	25.66	0.168

for which the pooled variance

$$s_p^2 = \frac{(79.67-25.66) + (41.87-0.168)}{30 + 30 - 2} = 1.65$$

Since, for the comparison of two independent means

$$n = 2\left[\frac{(z_\alpha - z_\beta)\sigma}{\delta}\right]^2$$

and the difference one wishes to detect $\delta = 0.8$, $\alpha = 0.05$, $\beta = 0.20$ and $\sigma = \sqrt{1.65}$, then

$$n = 2\left[\frac{(1.96-(-.84))\sqrt{1.65}}{.8}\right]^2$$

$$= 40.14$$

Therefore 40 episodes of otitis media would have to be scored for management in the pre-audit period and another 40 in the first audit period in each clinic.

e) CME scores will be dealt with as follows:

(i) The overall mean CME-score and standard deviation will be displayed for each of the nine treatment cells.

(ii) Correlation coefficients will be calculated (S 3.3.1) first for the pre-audit clinic TIC-scores and secondly for the audit TIC-score and the CME-score.

In general one will be able to determine whether there is a positive, negative or absent linear relationship between CME and TIC scores. In particular one is interested in knowing how much any observed response to audit is due to the audit and how much is due to CME. If the correlation coefficient is less than 0.5, it will be concluded that the CME reported did not contribute to the response attributed to audit. However if the correlation coefficient equals or exceeds 0.5, it will not be possible to separate the effect of CME from the effect of audit.

(iii) Until the data are available, it is not possible to predict how the CME data should be analysed. If one can show that there is no interaction between experimental group and CME, an analysis of covariance could be used with CME the covariable. But since the design outlined in this thesis has not taken CME directly into account, this analysis may not be very useful. Similarly unhelpful would be a situation where CTL clinics had reported a lot of CME while EXP-FB and EXP reported very little.

Thus, until the data are gathered it appears acceptable to analyse CME effect by groups individually as suggested in e) (i) previously.

CHAPTER III

ANALYSIS

3.1 Questions to be answered

1. Do audits have a beneficial impact on the quality of ambulatory primary care delivered in general and special clinics of hospitals?

2. Did awareness that an audit was being performed improve the scores achieved by clinics in the EXP-FB and EXP groups compared to their pre-audit performance?

3. Did awareness that an audit was being performed improve the scores achieved by clinics in the EXP-FB and EXP groups compared to CTL performance in the first audit period?

4. Did feedback render the EXP-FB scores different to EXP scores for the second audit period?

5. In the second audit period did EXP-FB or EXP scores differ from CTL scores?

6. Did CTL performance change over the three audit periods?

7. Within each stratum and across experimental groups were mean scores similar in the pre-audit period?

8. What differences between strata existed in the pre-audit scores?

9. If a response to audit does occur, is the response more marked in one stratum than others? In one type of clinic than others? With one type of indicator condition than others?

10. Is the impact of CME as assessed by individual physicians related to clinic performance?

3.2 Comparing SIC-scores and TIC-scores

In order to determine whether the SIC- and TIC-scores are similar, categorical performance scores from the pre-audit period will be used. It is only in this collection period that the two scores should be compared since only in this period are the physicians in the trial unaware of the indicator conditions in a manner paralleling the physicians in private practice.

For N clinics using indicator condition #1, the following table could be set up.

TABLE 3.1

		Counts from Sibley study	Counts from proposed trial
Score	I	n_{SI}	n_{TI}
	A	n_{SA}	n_{TA}
	S	n_{SS}	n_{TS}

A Chi-square analysis ($\alpha = 0.05$) will be used to test the null hypothesis that there is no difference in the distributions of the two scores. If this hypothesis is not rejected it would suggest concurrent validity for indicator conditions in the two types of practice settings.

3.3 General analysis

3.3.1 Correlation coefficients

As a descriptive statistic, interpretable as an index of association between pre-audit and first audit scores and between pre-audit and second audit scores, Pearson product-moment correlations of clinic mean scores will be obtained using the Statistical Package for the Social Sciences (SPSS) (Nie, et al.

1975) sub-programme SCATTERGRAM. First audit scores will be compared with pre-audit scores for each of the three treatment groups stratified as well as with strata collapsed. Second audit scores will be compared with pre-audit scores for each of the three treatment groups stratified and with the strata collapsed.

Using Fisher's Z-transformation (Kleinbaum and Kupper, 1978) it will be possible to test the hypothesis that two correlation coefficients are equal. To deal with the problem of multiple comparisons, the Bonferroni inequality method (Miller, 1966) would be applied wherein the proposed significance level is divided by the number of comparisons proposed.

Looking at the twelve correlation coefficients which could be developed for the correlation between pre-audit and first audit scores, (namely four coefficients developed for each treatment group with one being for the data collapsed across strata and the other three being for data stratified), the multitude of possible comparisons becomes evident. And a further twelve could be compared for the second category of correlation, that between pre-audit and second audit scores.

3.3.2 Analysis of covariance

The SCATTERGRAM programme will yield slopes as well as correlation coefficients where linear relationships exist. If the slope $\beta>0$, the Biomedical Package (BMDP) sub-program P2V (Dixon, et al. 1979) will be used to perform an analysis of covariance ($\alpha = .05$) to show whether there are significant differences among the true adjusted mean scores in the first audit period and in the second audit period. What follows is a description of ANCOVA for first audit scores adjusting for pre-audit scores. In the same way an ANCOVA will be done for second audit scores adjusting for pre-audit scores.

The complete model for this analysis is:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z_1 + \ldots + \beta_9 Z_8$$
$$+ \beta_{10} Z_1 X + \beta_{11} Z_2 X + \ldots + \beta_{17} Z_8 + E$$

where Y is the dependent variable, the first audit score; X is the covariate, the pre-test score; and Z_1 and Z_8 are dummy variables representing the nine levels.

Thus:

$Z_1 = 1$ if group = EXP-FB/Teaching and $Z_1 = 0$ if otherwise

$Z_2 = 1$ if group = EXP-FB/Comm > 310 and $Z_2 = 0$ if otherwise

$Z_3 = 1$ if group = EXP-FB/Comm < 310 and $Z_2 = 0$ if otherwise

$Z_4 = 1$ if group = EXP/Teaching and $Z_4 = 0$ if otherwise

- -

- -

$Z_8 = 1$ if group = CTL/Comm > 310 and $Z_8 = 0$ if otherwise.

Further:

Ho: $\beta_{10} = \beta_{11} = \cdots \beta_{17} = 0$

must be satisfied. Only if it can be reasonably assumed that there is a common slope is it useful to apply an analysis of covariance.

The adjusted mean scores which are developed will permit comparison of first audit scores as if they all had the same pre-audit mean score.

TABLE 3.2

Group	Adjusted first audit mean score where $\bar{X}$ is the overall mean pre-audit score
EXP-FB	
Teaching	$\beta_0 + \beta_2 + \beta_1\bar{X}$
Comm > 310	$\beta_0 + \beta_3 + \beta_1\bar{X}$
Comm < 310	$\beta_0 + \beta_4 + \beta_1\bar{X}$
EXP	
Teaching	$\beta_0 + \beta_5 + \beta_1\bar{X}$
Comm > 310	$\beta_0 + \beta_6 + \beta_1\bar{X}$
Comm < 310	$\beta_0 + \beta_7 + \beta_1\bar{X}$
CTL	
Teaching	$\beta_0 + \beta_8 + \beta_1\bar{X}$
Comm > 310	$\beta_0 + \beta_9 + \beta_1\bar{X}$
Comm < 310	$\beta_0 + \beta_1\bar{X}$

To test the equality of the nine true adjusted means, a multiple-partial F test is applied to Ho: $\beta_2 = \beta_3 = \ldots = \beta_9 = 0$. F will have k - 1 and n-p-k df where k = 9 (the number of groups) and p = 1 (the number of covariables).

If the null hypothesis is rejected, one will conclude that there are significant differences among the true adjusted mean scores attributable to the groups.

3.3.3 Multiple Comparisons

Multiple comparisons are necessary since we are interested in knowing how the true adjusted mean scores developed in the

previous section compare with each other. In particular the comparisons needed are

- EXP-FB/Teaching with EXP/Teaching
- EXP-FB/Teaching with CTL/Teaching
- EXP/Teaching with CTL
- EXP-FB/Comm > 310 with EXP/Comm > 310
- EXP-FB/Comm > 310 with CTL/Comm > 310

and so on to

- EXP/Comm < 310 with CTL / Comm < 310

According to Winer (1962) a test on the difference between two adjusted means has the form

$$F = \frac{(\bar{T}'_j - \bar{T}'_m)^2}{MS'_e\left[\frac{2}{n} + \frac{(\bar{X}_j - \bar{X}_m)^2}{\Sigma_{XX}}\right]}$$

with 1 and k(n-1)-1 df,

where $\bar{T}_j$ and $\bar{T}_m$ are the two adjusted mean scores being compared, $\bar{X}_j$ and $\bar{X}_m$ are the corresponding mean pre-audit

scores, MS'_e is the error term from the model given in the previous section, n is the cell frequency and

$$\Sigma_{XX} = \sum_{j=1}^{9} \sum_{i=1} (X_{ij} - \bar{X}_j)^2$$

Since the cell frequencies are unlikely to be equal n may have to replaced with the harmonic mean n' where

$$n' = \frac{2}{(\frac{1}{n_1} + \frac{1}{n_2})}$$

3.3.4 Analysis of Variance (ANOVA)

a) The BMDP P2V sub-programme will also do an analysis of variance for the following model:

$$Y_{ijk\ell} = \mu + \alpha_i + \beta_j + \pi_{(ij)k} + \gamma_\ell + \alpha\gamma_{i\ell} + \beta\gamma_{j\ell} + \alpha\beta\gamma_{ij\ell} + \beta\pi_{(ij)k\ell} + \varepsilon_{(ijk\ell)}$$

where

α_i denotes the group with α_1 = EXP-FB,

α_2 = EXP and α_3 = CTL

β_j denotes the stratum with β_1 = Teaching,

β_2 = Comm>310 and

β_3 = Comm<310

γ_k denotes three times with γ_1 = the pre-audit collection period, γ_2 = the first audit and γ_3 = the second audit period.

$\pi_{(ij)k}$ = kth hospital within the ith group and jth stratum and k = 1,2, ...n_{ij}

$\varepsilon_{(ijkl)}$ denotes error and the products of the coefficients indicate interactions among factors.

The ANOVA table which could then be developed includes nine possible sources of variation for mean clinic scores (Table 3.3). Corresponding to these nine sources are seven null hypotheses.

1. Ho: $\mu_{EXP-FB} = \mu_{EXP} = \mu_{CTL}$ (no main effect of treatment)

2. Ho: $\mu_{TEACHING} = \mu_{COMM>310} = \mu_{COMM<310}$

(no main effect of stratum)

3. Ho: no interaction between treatment and stratum

4. Ho: $\mu\gamma_1 = \mu\gamma_2 = \mu\gamma_3$ (no time effect)

5. Ho: no interaction between time and treatment group

6. Ho: no interaction between time and stratum

7. Ho: no interaction between time, stratum and treatment group.

The analysis of variance displays the sum of squares (SS) and the mean sum of squares (MS) for each source and the MS is an estimate of the true variance.

If the null hypothesis is true, the MS say for treatment group, will be the same as that of the general population. The ratio MS treatment group: MS error (which is called an F statistic with two and 51 degrees of freedom (df)), has an expected value of one when the null hypothesis is true and will be greater than one when the null hypothesis can be rejected. In this way all seven null hypotheses can be tested.

For among hospital differences the denominator df will be 51 and for within hospital differences the denominator df will be 102.

TABLE 3.3

SOURCE	df	SS	MS	F
AMONG HOSPITALS	59			
Exp. Group	2	SS_{EG}	$SS_{EG/2}$	$F = \frac{SS_{EG/2}}{SS_{E/51}}$
Strata	2	SS_{S}	$SS_{S/2}$	
Exp. Group × Strata	4	SS_{EGS}	$SS_{EGS/4}$	
Error	51	SS_{E}	$SS_{E/51}$	
WITHIN HOSPITALS	120			$F = \frac{SS_{T/2}}{SS_{E/102}}$
Time	2			
Time × Exp. Group	4			
Time × Strata	4			
Time × Exp. Group × Strata	8			
Error	120			
TOTAL	179			

b) To answer the question whether or not different degrees of response are associated with individual indicator conditions another ANOVA based on a similar model to that in a) will be done.

TABLE 3.4

SOURCE	df	SS	MS	F
AMONG INDICATOR CONDITIONS	n-1			
Exp. Group	2			
Strata	2			
Exp. Group × Strata	4			
Error	n-9			
WITHIN INDICATOR CONDITIONS	2n			
Time	2			
Time × Exp. Group	4			
Time × Strata	4			
Time × Strata × Exp. Group	8			
Error	2n-18			
TOTAL	3n-1			

This ANOVA would be based on the mean scores for single indicator conditions per institution.

3.3.5 Analysis of CME-impact

The method outlined for the pilot study (S 2.4.43) will be applied in the major trial.

CHAPTER 4

CONCLUSION

4.1 Criteria for success

The principal hyptheses in this trial are:

1. That in general, performance will be better in teaching than in community hospital clinics before the indicator conditions are introduced to the physicians.

2. That performance scores will be higher in the first audit period than in the pre-audit period in all EXP-FB and EXP clinics.

3. That clinics with the lowest scores in the pre-audit period will change more than clinics with higher scores.

4. That improvement in performance will be better maintained in clinics which receive feedback than in clinics which do not.

If the trial produces data sufficient to enable the investigators to accept or reject these hypotheses ($\alpha = 0.05$ and $\beta = 0.20$) the trial will have been successful.

Three major impediments to success are recognized.

1. A large number of hospitals may refuse to participate so that random allocation of hospitals to three treatment groups may not be possible. At worst one would be left then with using the available institutions and using before-after comparisons. This would still permit an assessment of the impact of audits on performance.

One cogent reason for not participating would be the costs engendered by using HMRI services. Therefore it is planned to apply for grant money to cover HMRI costs with hospitals beng asked to provide only necessary medical records staff and space.

2. Notes on charts may be so cryptic that the hospital medical records staff may not be able to identify the "most responsible diagnosis". This may be a problem in all or only in some hospitals, and the extent of the problem may be revealed during the pilot study. If the HMRI probe cannot be used, the investigators will consider using daysheets (S 2.4.1.2.2) and will also explore the possibility of using billing data.

3. Again, the notes on the charts may be so cryptic that applying the explicit criteria for the indicator conditions will yield only the lowest score for all charts assessed. Should this

happen two options will be considered. One will be modification of the indicator conditions. The other will be to consider replacing the indicator condition with another mechanism for audit.

4.2 Ethics

"Ethical principles are properly invoked to defend the rights, dignity and health of individuals and groups." (Ingelfinger, 1975).

A brief review of the literature on ethics as it relates to clinical research reveals several foci of concern. Concern that the research —

1. Will not respect the "sanctity of life".
2. Will expose the patient to an unfavourable risk benefit ratio as when patients receiving placebo are subjected to invasive procedures.
3. Will benefit the researcher more than the patient.
4. Will use coercion to assure participation in the trial.
5. Will breach "confidentiality".

The trial described in this thesis poses no threat to the sanctity of any person's life. Nor can it be argued convincingly

that it poses a physical risk to the patient. One could imagine that after introduction to the criteria associated with an indicator condition, say hypertension, a physician might prescribe a drug he would otherwise not have done, and the patient might subsequently have an adverse drug reaction. Nonetheless it is generally accepted that the benefits of treating hypertension exceed the risks of conventional therapy (Ontario Council of Health, 1977). It is much easier to argue that if the introduction of indicator conditions has any effect, it will likely be to improve, not diminish the standards of practice. And it is unlikely that patients would object to such a manoeuvre.

The third concern, that of the trial benefiting the researcher more than the patient is harder to address. Certainly being involved in a trial results in the researcher earning a living doing what interests him. But that is not the purpose of the study. Rather the purpose is to reduce current uncertainty about the value of an activity (audits) which is being required of many institutions.

In this trial the issue of coercion is irrelevant with respect to the patient in that the patient will be unaware that the study is occurring. Although it has been suggested that peer review should only occur with fully informed patient consent (Sullivan,

1977) such an opinion seems tenable to me only in the narrow context of psychiatry, if at all; and even there it can be argued against when peer reviewers assure privacy safeguards.

However, if the agency which is responsible for giving or with-holding hospital accreditation is associated with the trial, coercion is implicitly present although it is not the patient who is being coerced but rather the hospital and the physicians. There is a fine line to be drawn between the trial being supported because CCHA endorsement lends it credibility and a trial being supported because what is perceived is the CCHA wielding an accreditation stick.

It has been generally recognized that physicians do not like the concept of peer review. But this cannot be a reason to dispense with peer review studies when government, the public and professional organizations are demanding quality assurance programmes. If one accepts the inevitability of peer review in the current state of society, one must then accept that involved physicians may to a greater or lesser extent feel themselves, quite correctly, to be coerced. The need for peer review mechanisms will not go away.

With regard to confidentiality, both patient and physician must be protected. In this trial the patients' names will not be recorded, only their chart number and possibly their OHIP number.

In any case, the invasion of their privacy will be no more than that currently endured by all hospital inpatients in Ontario. It is the physicians for whom the issue of confidentiality is most important because the way they manage their patients is going to be graded. Those whose performance is graded "indeterminate" must by CCHA guidelines, sooner or later be confronted with an invitation to change their performance. This must be the responsibility of the audit committees of individual hospitals and not of the research staff. It will be the obligation of the research staff to release information about physician scores only to the local audit committee. Further the research staff will maintain the anonymity of clinic and hospital scores. Under no circumstances will scores for individual physicians, clinics or hospitals be revealed to outside officials or agencies.

When analysis is complete, all identity codes on the data will be erased. This is particularly important since physicians have a very justified fear that audit data could be exploited by lawyers in medico-legal disputes between patients and attending physicians (Murray, 1980).

"The social contract that facilitates the existence of individuals within social groups requires that each individual occasionally yield some of his rights, including privacy and freedom

of action, for the benefit of society as a whole, . . . with proper safeguards, individually identifiable data from medical records (must) continue to be made accessible for medical and epidemiologic research." (Gordis and Gold 1980). This goes for doctors as well as patients.

REFERENCES

American Public Health Association (1965[1]): A guide to medical care administration. Vol. II: Medical care appraisal (prepared by A. Donabedian). Part I: Frame of reference: 8. U.S. Public Health Service.

-------- (1965[2]): Part I: Frame of reference: 2.

-------- (1965[3]): Part I: Frame of reference: 3.

-------- (1965[4]): Part II: Selected methods of appraisal: 32.

Ashley, J.S.A., Howlett, A., and Morris, J.N. (1971): Case fatility of hyperplasia of the prostate in two teaching and three regional-board hospitals. Lancet 2: 1308-1311.

Barro, A.R. (1975[1]): Survey and evaluation of approaches to physician performance measurement. (Suppl.), Introduction: 1053. Washington, D.C. Association of American Medical Colleges.

-------- (1975[2]): Chapter III: Simulated process approaches: 1069.

Batchelor, G.M., Spitzer, W.O., Hay, W.I., Sibley, J.C., Sweeny, G.P. and Gent, M. (1975): Burlington randomized controlled trial of nurse practitioners; methodological manual and final report. McMaster University, Faculty of Health Sciences.

Brook, R.H. and Stevenson, R.L. Jr. (1970): Effectiveness of patient care in an emergency room. NEJM 283: 904-907.

Brook, R.H. (1973): Quality of care assessment: a comparison of five methods of peer review. (DHEW Publication No. HRA-74-3100) Rockville, Maryland. Department of Health, Education and Welfare, U.S.A.

Brook, R.H. and Appel, F.A. (1973): Quality of care assessment: choosing a method for peer review. NEJM 288: 1323-1329.

Brook, R.H., Avery, A.D., Greenfield, S., Harris, L.J., Lelah, T., Solomon, N.E. and Ware, Jr., J.E. (1976): Quality of medical care assessment using outcome measure: an overview of the method. R-2021/1-HEW Santa Monica, CA. Rand Corporation.

Brook, R.H. and Williams, K.N. (1976): Effect of medical care review on the use of injections. Ann. Int. Med. 85: 509-515.

Buck, C.R. and White, K.R. (1974): Peer review: impact of a system based on billing claims. NEJM 291: 877-883.

Burdette, J.A., Babineau, R.A., Mayo, F., Hulka, B.S. and Cassel, J.C. (1974): Primary medical care evaluation. The AAFP-UNC Collaborative Study. JAMA 230: 1668-1673.

Canadian Council on Hospital Accreditation (1977): Guide to Hospital Accreditation, Ottawa.

-------- (1979): Annual report, Appendix VII: The comparative chart of total accredited and non-accredited hospitals and hospital beds by provinces.

Chambers, L.W., Sibley, J.C. and Tugwell, P. (1980): Quality care assessment: how to set up and use an indicator condition. Submitted for publication.

Christoffel, T. and Loewenthal, M. (1977): Evaluating the quality of ambulatory health care: a review of emerging methods. Medical Care XV: 877-897.

Cicchetti, D.V. and Fleiss, J.L. (1977): Comparison of null distribution of weighted kappa. Appl. Psychological Measurement 1: 195-201.

Clute, K.F. (1963): The general practitioner. Toronto. University of Toronto Press.

Cohen, J. (1960): A coefficient of agreement for nominal scales. Education and Psychological Measurement 20: 37-46.

Dixon, W.J. and Brown, M.B. (1979): BMDP-79: Biomedical Computer Programs P-series. Berkeley. University of California Press.

Fessel, W.J. and Van Brunt, E.E. (1972): Assessing quality of care from the medical record, NEJM 286: 134-138.

Fleiss, J.L. (1973): Statistical methods for rates and proportions: Chapter 12: The measurement and control of misclassification error: 140. New York, John Wiley and Sons.

Frazier, W.H. and Brand, D.A. (1979): Quality assessment and the art of medicine. The anatomy of laceration care. Med. Care XVII: 480-490.

Goldsmith, C.H., Professor, Dept. of Clinical Epidemiology and Biostatistics, McMaster University (1980): personal communication.

Gonnella, J.S., Louis, D.Z. and McCord, J.J. (1976): The staging concept: an approach to the assessment of outcome of ambulatory care. Med. Care XIV: 13-21.

Gordis, L. and Gold, E. (1980): Privacy, confidentiality, and the use of medical records in research. Science 207: 153-156.

Greenfield, S., Nadler, M.A., Morgan, M.T., and Shine, K.I. (1977): The clinical investigation and management of chest pain in an emergency department. Med. Care: XV: 898-905.

Hulka, B.S. and Cassel, J.C. (1973): The AAFP-UNC study of the organization, utilization and assessment of primary medical care. Am. J. Public Health 63: 494-501.

Ingelfinger, F.J. (1975): The unethical in medical ethics. Am. Int. Med. 83: 264-269.

Imai, T.S., Yourtee, E.L. and Williamson, J.W. (1976): Improved outcomes in hypertension after physician tutorials: a controlled trial. Ann. Int. Med. 84: 646-651.

Kane, R.L., Woolley, F., Gardner, H.J., Snell, G.F., Leight, E.H. and Castle, C.H. (1976): Measuring outcomes of care in an ambulatory primary care population. A pilot study. J. Comm. Health 1: 233-240.

Kessner, D.M. (1978): Quality assessment and assurance: early signs of cognitive dissonance. NEJM 298: 381-386.

Kleinbaum, D.G. and Kupper, L.L. (1978): Applied regression analysis and other multivariable methods: Chapter 6: The correlation coefficient and its relationship to straight-line regression analysis: 71, North Scituate, Mass. Duxbury Press.

Komaroff, A.L. (1978): The PRSO, quality assurance blues. NEJM 298: 1194-1196.

Lindsay, M.I., Hermans, P.E., Nobrega, F.T. and Ilstrup, D.M. (1976): Quality of care assessment.
1. Outpatient management of acute bacterial cystitis as a model. Mayo Clin. Proc. 51: 307-312.

Lipworth, L., Lee, J.A.H. and Morris, J.N. (1963): Case fatality in teaching and non-teaching hospitals, 1956-1959. Med. Care 1: 71-76.

Lohr, K.N., Brook, R.H. and Kaufman, M.A. (1980): Quality of care in the New Mexico Medicaid Program (1971-1975). Med. Care Suppl. 18: 1-129.

Lussing, F. Administrative consultant, Institution Division, Ontario Ministry of Health (1980): personal communication.

Miller, R.G. (1966): Simultaneous statistical inference: Chapter 1: Introduction: 8, New York, McGraw-Hill Book Company.

Ministry of Health (1979[1]): Hospital statistics 1978/79. Public and Private Hospitals. Mental Health Services, Table A: Volume of hospital care given in Ontario April 1, 1979 - March 31, 1979: xiv.

-------- (1979[2]): Table 2: Outpatient activity, April 1, 1978 - March 31, 1979: 19.

Mosteller, F. and Tukey, J.W. (1977[1]): Data analysis and regression: A second course in statistics, Chapter 5: The practice of re-expression: 106. Reading, Mass., Addison-Wesley.

-------- (1977[2]): Chapter 8: A method of direct assessment: 133.

Murray, J.H., Associate executive director, Canadian Council on Hospital Accreditation (1980): Personal communication.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. and Bent, D.H. (1975): SPSS: Statistical Package for the Social Sciences. New York, McGraw Hill.

Noack, H. (ed.) (1980): Medical education in primary care, Chapter 14: The assessment of physician performance in primary health care: 202, London, Croom Helm.

Ontario Council of Health (1977): Hypertension: A report of the Ontario Council of Health. Toronto, The Ontario Council of Health.

Palmer, R.H. and Reilly, M.C. (1979): Individual and institutional variables which may serve as indicators of quality of medical care. Med. Care XVII: 693-717.

Payne, B.C., Lyons, T.F., Dwarshius, L., Kolton, M. and Morris, W. (1976): The quality of medical care: evaluation and improvement. Chicago, Hospital Research and Educational Trust.

Payne, B.C. (1979): The medical record as a basis for assessing physician competence. Ann. Int. Med. 91: 623-629.

Peterson, O.L., Andrews, L.P., Spain, R.S. and Greenberg, B.G. (1956): Analytic study of North Carolina general practice, 1953-1954. J. Med. Educ. 31 (12) (part 2): 1-165.

Pozen, M.W. and Bonnet, P.D. (1976): Effectiveness of educational and administrative interventions in medical outpatient clinics. Am. J. Pub. Health. 66: 151-155.

Romm, F.J. and Hulka, B.S. (1979): Care process and patient outcome in diabetes mellitus. Med. Care XVII: 748-757.

Sanazaro, P.J. and Worth, R.M. (1978): Concurrent quality assurance in hospital care. Report of a study by private initiative in PSRO. NEJM 298: 1171-1177.

Shapiro, S., Andrews, L.P., Spain, R.S. (1958): Comparison of prematurity and perinatal mortality in a general population and in a population of a prepaid group practice medical care plan. Am. J. Pub. Health 48: 170-185.

Sibley, J.C., Bell, J.D., Bethune, R.D., Hay, W.I., Rudnick, K.V., Sackett, D.L., Spitzer, W.O. and Sweeny, G.P. (1975[1]): Indicator conditions, drug utilization and consultant questionnaire for development of a quality of care appraisal method for primary health care projects. McMaster University, Faculty of Health Sciences (mimeo).

Sibley, J.C., Spitzer, W.O., Rudnick, K.V., Bell, J.D., Bethune, R.D., Sackett, D.L. and Wright, K. (1975[2]): Quality of care appraisal in primary care: a quantitative method. Ann. Int. Med. 83: 46-52.

Spasoff, R.A., Lane, P. and Steele, R.(1977): Quality of care in hospital emergency departments and family physicians' offices. CMAJ 117: 229-232.

Starfield, B. and Scheff, D. (1972): Effectiveness of pediatric care: the relationship between process and outcome. Pediatrics 49: 547-552.

Statistics Canada (1979): Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures. Ottawa, Statistics Canada.

Sullivan, F.W. (1977): Peer review and professional ethics. Am. J. Psychiatry 134: 186-188.

Surridge, D. (1979): Working papers in medical audit. CMAJ 120: 1323-1326.

Tugwell, P. (1979[1]): A methodological perspective on process measures of the quality of medical care. J. Clin. & Investigative Med. 2: 113-121.

--------(1979[2]): How to assess quality of care studies that employ "process measures". McMaster University, Medical Centre (mimeo).

Williamson, J.W., Aronovitch, S., Simonson, L., Ramirez, C. and Kelly, D. (1975): Health accounting: an outcome-based system of quality assurance — illustrative application to hypertension. Bull. N.Y. Academy of Medicine (2nd series) 51: 727-738.

Winer, B.J. (1962): Statistical Principles in Experimental Design: Chapter 11: Analysis of Covariance: 578, New York. McGraw-Hill Book Co.

World Health Organization (1975): International Classification of Diseases, 1975 Revision. Geneva, World Health Organization.

APPENDIX A

APPENDIX A-1: HS-1 FORM FROM STATISTICS CANADA

STATISTICS CANADA — HEALTH AND WELFARE CANADA

1979 - 80 ANNUAL RETURN OF HEALTH CARE FACILITIES – HOSPITALS – PART ONE

Name of hospital — City, town, etc. — (08)

AMBULATORY CARE UNITS

General and Special Clinics

CLINIC	Number of visits: Outpatients 1	Number of visits: Inpatients 2
01. Allergy	021	022
02. Arthritis and rheumatism	031	032
03. Cancer	041	042
04. Cardiac diseases	051	052
05. Dermatology	061	062
06. Dermatology	071	072
07. Endocrine and metabolic diseases (excl. Diabetic Day Care)	081	082
08. Family practice	091	092
09. Gastroenterology	101	102
10. General medicine (incl. diagnostic and screening)	111	112
11. General surgery	121	122
12. Gynecology	131	132
13. Neurology	141	142
14. Obstetrics (incl. Well-baby clinics)	151	152
15. Ophthalmology	161	162
16. Orthopaedics	171	172
17. Otorhinolaryngology	181	182
Psychiatry and/or Mental Health (excl. day and night care)		
18. Adults	191	192
19. Children (specify upper age limit ____ yrs.)	201	202
20. Tuberculosis	211	212
21. Urology	221	222
22. Venereal diseases	231	232
23. All others (not specified above)	241	242
24. TOTAL VISITS		

Outpatient Day and/or Night Care Programs

PROGRAM (Excl. Surgical Day Care – See line 38 below)	Number of persons: Outpatients 3	Number of persons: Inpatients 4	Number of visits 5
25. Diabetic Day Care	253	254	255
26. Geriatric Day Care	263	264	265
27. Psychiatric Day Care and/or Night Care	273	274	275
28. Alcoholism Day and/or Night Care	283	284	285
29. Drug Addiction Day and/or Night Care	293	294	295
30. Dialysis Day and/or Night Care	303	304	305
31. Other (specify)	313	314	315
32.	323	324	325
33. TOTAL	333	334	335

EMERGENCY UNIT

Number of visits to the Emergency Unit (include ALL visits for appointments, treatments or examinations done by the Emergency Unit)	345
Outpatient visits to the Emergency Unit	
34. – Surgical day care visits	355
35. – All other outpatient visits	365
36. Inpatient visits to the Emergency Unit	375
37. TOTAL VISITS TO EMERGENCY UNIT	

SURGICAL DAY CARE PROGRAM*

38. Total number of visits to all units	385

*Note that the procedures performed on surgical day care patients will also be recorded as outpatient visits in the appropriate counts of activities of the Surgical Suite, Emergency Unit, Surgical day care and other units (refer to the Instructions & Definitions).

HOME CARE PROGRAMS

Hospital-based	
39. Number of patients admitted to all hospital-based home care programs during the year	395
40. Number of patients discharged from all hospital-based home care programs during the year	405
41. Number of patients from your hospital admitted to other home care programs during the year	415

PART ONE – P. 8 of 18.

APPENDIX A-2

ELIGIBILITY OF AN EPISODE FOR APPRAISAL AS TRACER CONDITION (Sibley, et al. 1975[1])

1) The Indicator Condition must have been assessed and/or managed as an ambulatory care episode, in whole or in part.

   In the case of the hospitalized patient, it must be evident that the primary care physician rendered all or most care prior to referral to hospital. This may only involve a telephone intervention, which would not necessarily be recorded on the primary care records, but which may be indicated on the correspondence from the hospital.

2) Episodes scrutinized must fall within pre-defined dates that identify the interval of interest.

3) There must be evidence that the Primary Health Care Professional under assessment has intervened in the management of the episode being studied.

4) Target of evaluation AND/OR identification of evaluee -- For the purpose of assessment, the Primary Health Care Professional, who is the target of health evaluation process, must have made a decision in the episode under study. This could be -

   (a) the Primary Health Care Professional with whom the patient made initial contact,

   (b) the Primary Health Care Professional who made an assessment of the patient apart from the decision to hospitalize,

   (c) the first Primary Health Care Professional making a decision in an eligible episode.

5) In evaluating Drug Utilization, conditions other than the Indicator Conditions may be implicated.

6) An episode is defined separately for each indicator condition under study.

INDICATOR CONDITION #1 - OTITIS MEDIA

(Sibley, et al. 1975[1])

DEFINITION OF AN EPISODE - OTITIS MEDIA (2 Yrs. - 12 Yrs.)

Otitis Media episode begins when patient first consults physician or nurse about complaints related to the ear, or when diagnosis is recorded in the chart, or a note is made that the eardrum is red or inflamed, within the study period. This must be a new condition, or there must be reasonable evidence that a prior episode had been resolved satisfactorily. The episode ends on the last recorded visit concerning this problem or at the end of the study period, whichever occurs first.

CATEGORIES OF INTERVENTION

1. Follow-up visit within one month of initial episode.
2. An appropriate antibiotic (Erythromycin, Penicillin, short acting Sulpha, Amoxicillin, Hetacillin, Cephalexin, Ampicillin, Dicloxacillin, and Cloxacillin). An inappropriate antibiotic would be Tetracycline, Chloramphenicol, Clindamycin, or Lincomycin.
3. Antibiotics administered for at least ten days.
4. Continue antibiotics plus a further repeat visit.
5. Consultation.
6. Statement that patient is cured or a clear statement of patient's status.
7. Myringotomy plus further repeat visit.
8. Third or subsequent visits with evidence or a statement that hearing has been checked.
9. Late follow-up or late consultation or audiometric examination.

SCORING

ADEQUATE

OPTION 1

1. Follow-up visit requested or occurred within one month of initial episode.
2. An appropriate antibiotic (Erythromycin, Penicillin, Sulpha, Ampicillin, Amoxicillin, Hetacillin, Cephalexin, Dicloxacillin, and Cloxacillin.

An inappropriate antibiotic would be Tetracycline, Chloramphenicol, Clindamycin, or Lincomycin.

3. Antibiotics administered for at least 10 days.
6. Statement that patient is cured or a clear statement of patient's status.

OPTION 2

5. Consultation.

INDETERMINATE

Less than adequate.

SUPERIOR

Adequate - plus - any of the following:

4. Continue antibiotics plus a further repeat visit.
7. Myringotomy plus further repeat visit.
8. Third or subsequent visits with evidence or a statement that hearing has been checked.
9. Late follow-up or late consultation or audiometric examination.

OTITIS MEDIA (2 Yrs. - 12 Yrs.)

Episode begins when the patient first consults physician or nurse about complaints related to the ear, or when the diagnosis is recorded withhin the study period. This must be a new condition, with reasonable evidence that a prior episode has been resolved satisfactorily. The episode ends on the last recorded visit concerning the problem or at the end of the study period, whichever occurs later.

R.T. ☐☐ (1 2) DR. ☐☐ (4 5) I.D. ☐☐☐☐ (6 7 8 9) 1st V. d d m m y y ☐☐☐☐☐☐ (14 15 16 17 18 19)

CAT. ☐ (3) AGE ☐☐ (10 11) Last V. d d m m y y ☐☐☐☐☐☐ (20 21 22 23 24 25)

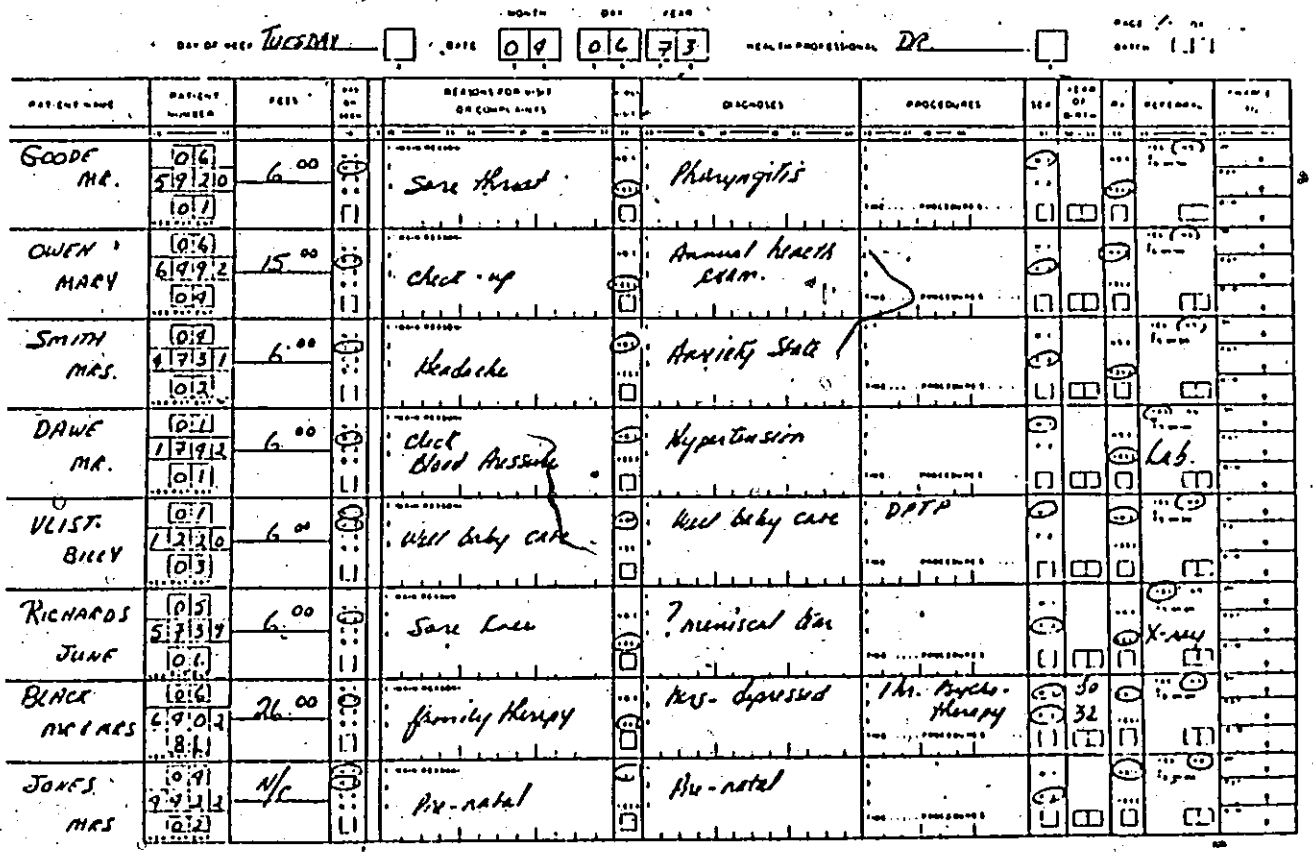
V. ☐☐ (12 13) N.A. ☐ (26)

PROBE ☐ (27)

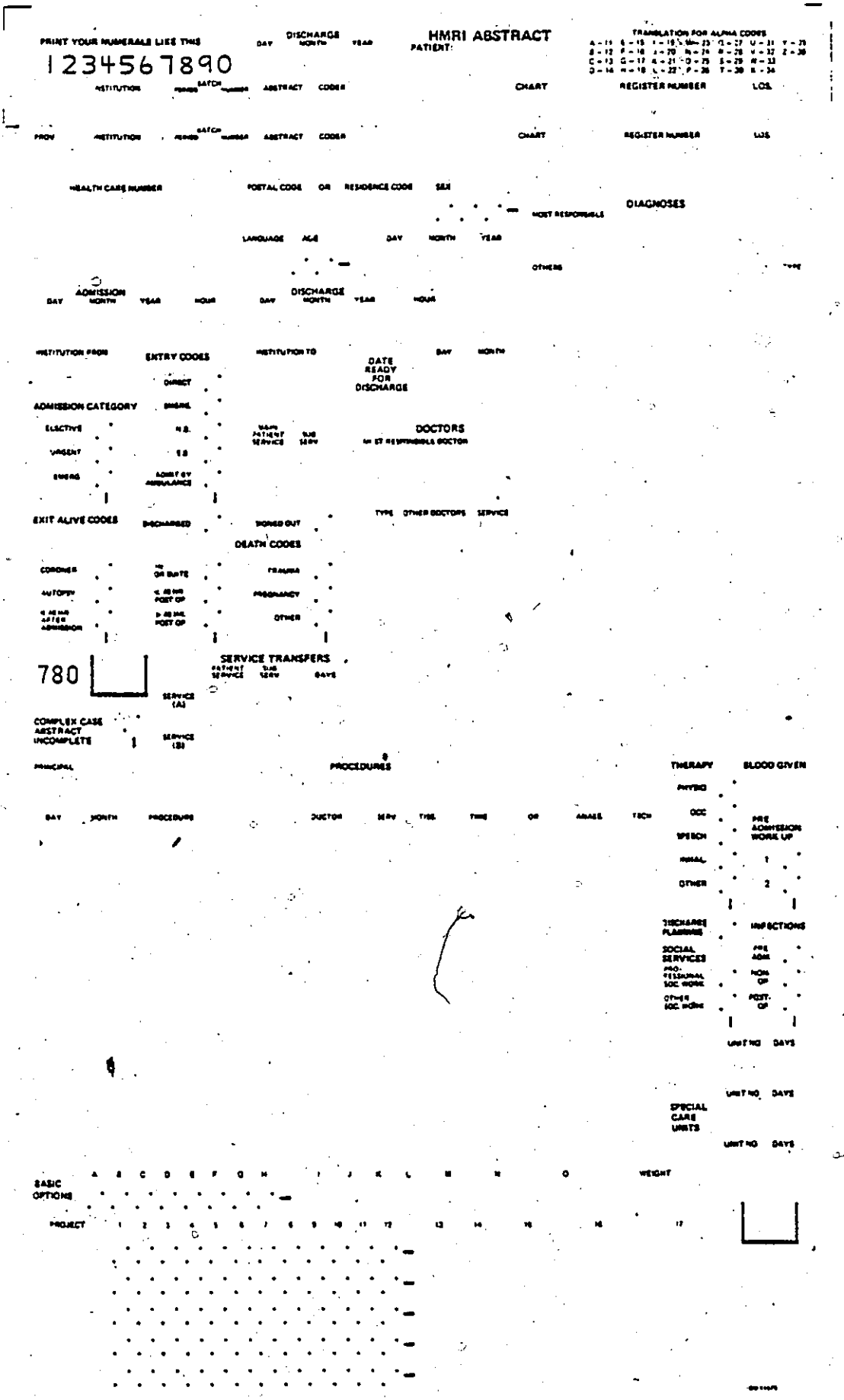
CATEGORIES OF INTERVENTION

- Follow-up visit (within one month). ☐ (28)
- Antibiotics:
  - Appropriate - Penicillin, Erythromycin short acting Sulphas, Ampicillin, Amoxicillin, Hetacillin, Cephalexin, Dicloxacillin, and Cloxacillin. ☐ (29)
  - Inappropriate - Tetracycline, Chloramphenicol, Clindamycin, and lincomycin. ☐ (30)
  - Antibiotic administered for at least ten days. ☐ (31)

- Continue antibiotics plus a further visit. ☐ 32
- Myringotomy plus further visit.
- Three or more visits and hearing checked. ☐ 33
- Late follow-up or late consultation or Audiometric examination. ☐ 34
- Consultation. ☐ 35
- Statement that patient is cured or a clear statement of patient's status. ☐ 36



APPENDIX A-3: DAY SHEET FROM THE BURLINGTON RANDOMIZED CONTROLLED TRIAL OF NURSE PRACTITIONERS (BATCHELOR, ET AL. 1975)


HMRI ABSTRACT
PATIENT:
PRINT YOUR NUMERALS LIKE THIS
1234567890
DISCHARGE
DAY
MONTH
YEAR
TRANSLATION FOR ALPHA CODES
INSTITUTION
BATCH
NUMBER
ABSTRACT
CODER
CHART
REGISTER NUMBER
LOS
PROV
HEALTH CARE NUMBER
POSTAL CODE OR RESIDENCE CODE
SEX
DIAGNOSES
MOST RESPONSIBLE
LANGUAGE
AGE
OTHERS
TYPE
ADMISSION
HOUR
INSTITUTION FROM
ENTRY CODES
INSTITUTION TO
DATE READY FOR DISCHARGE
ADMISSION CATEGORY
ELECTIVE
URGENT
EMERG
DIRECT
EMERG
N.B.
T.D.
ADMIT BY AMBULANCE
MAIN PATIENT SERVICE
SUB SERV
DOCTORS
MOST RESPONSIBLE DOCTOR
TYPE
OTHER DOCTORS
SERVICE
EXIT ALIVE CODES
DISCHARGED
SIGNED OUT
DEATH CODES
CORONER
AUTOPSY
TRAUMA
PREGNANCY
OTHER
780
SERVICE TRANSFERS
PATIENT SERVICE
SUB SERV
DAYS
SERVICE (A)
SERVICE (B)
COMPLEX CASE ABSTRACT INCOMPLETE
PROCEDURES
PRINCIPAL
DAY
MONTH
PROCEDURE
DOCTOR
SERV
TYPE
TIME
OR
ANAES
TECH
THERAPY
PHYSIO
OCC
SPEECH
NURSE
OTHER
BLOOD GIVEN
PRE ADMISSION WORK UP
1
2
DISCHARGE PLANNING
SOCIAL SERVICES
PRO-FESSIONAL SOC WORK
OTHER SOC WORK
INFECTIONS
PRE ADM
NOW OP
POST-OP
UNIT NO
DAYS
SPECIAL CARE UNITS
BASIC OPTIONS
A B C D E F G H I J K L M N O
WEIGHT
PROJECT
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

HMRI

ABSTRACT CORRECTIONS

PROV (1) INSTIT (2–3) BATCH YEAR (6–7) BATCH PERIOD (8–9)

	ABSTRACT IDENTIFICATION (10–15)						FIELD IDENTIFICATION (16–21)						DATA (22–31)									
	BATCH NUMBER		ABSTRACT			CHECK DIG	GROUP		FIELD		OCCURRENCE		1	2	3	4	5	6	7	8	9	10
1																						
2																						
3																						
4																						
5																						
6																						
7																						
8																						
9																						
10																						
11																						
12																						
13																						
14																						
15																						
16																						
17																						
18																						
19																						
20																						
21																						
22																						
23																						
24																						
25																						
													1	2	3	4	5	6	7	8	9	10

HMRI USE ONLY	
DATE REC'D:	AC #:

HMRI FORM 78 10

GENERAL INFORMATION ABOUT THE HMRI ABSTRACT

Fields which are not useful for the purposes of this study (but which individual hospitals could use to meet their own requirements) have a line drawn through them.

Entry codes will indicate the type of clinic the patient visited.

Service transfers will indicate referrals made for the patients.

Diagnoses can incorporate three categories of diagnoses. The first is "most responsible" and is flagged by "M". The second type is coexisting primary diagnoses and these are flagged by the number "1" under Type.

Finally, presenting complaints can be entered flagged by the number "5" under Type.

Coding for these three types of entries is possible (WHO, 1975).

Procedures will permit specification of procedures ordered (Statistics Canada, 1979).

APPENDIX B

APPENDIX B-1

LETTER TO HOSPITAL ADMINISTRATORS/MEDICAL DIRECTORS

INVITING PARTICIPATION IN TRIAL

Dear -

Since 1977, quality of care appraisal has been made mandatory for hospital accreditation in Canada. Such requirements are being implemented before the value of such procedures has been unambiguously established. For this reason, the Canadian Council of Hospital Accreditation approached the Department of Clinical Epidemiology and Biostatistics at McMaster University and asked for assistance in determing whether the clinical audits imposed on hospitals did result in improved patient care.

Until recently, the emphasis has been on auditing in-patient care. Audits of ambulatory care have been less rigorously performed, if at all. Therefore a randomized controlled trial has been designed to determine whether audit procedures have a beneficial impact on out-patient care in general and special clinics in Ontario hospitals. Attached is a summary of the proposed design. Its success depends on the cooperation of all eligible institutions.

In order for you to become more informed about the trial, we propose a meeting so that discussion is possible. We will call you in the near future so that a meeting can be arranged at your convenience.

Yours sincerely

APPENDIX B-2

FORM LETTER INDICATING PHYSICIAN

AWARENESS OF AUDIT (FOR EXP-FB and EXP only)

TO: - The principal investigator

The audit committee in my hospital has informed me of the purposes of the CCHA study. I know that the specific indicator condition(s) chosen for my clinic(s) are

_____

_____

_____

and I am aware of the explit criteria for each. I understand that patients with these indicator conditions will have their charts examined and a rating will be assigned to each. The ratings will be: indeterminate, acceptable or superior care.

I understand that all information will remain confidential, that it is the performance of groups, not individuals, which is being assessed and that in order to preserve the randomized controlled nature of this trial, I should not discuss it with colleagues from other institutions.

Yours sincerely

_____(name)_____

_____(clinic)_____

_____(hospital)_____

APPENDIX B-3

FORM LETTER FOR FEEDBACK TO EXP-FB PHYSICIANS

Dear Doctor -

We have now calculated the management scores for the indicator conditions which were assigned to your clinic, namely _____ .

The mean score for your clinic was ___ and for your hospital _____ . The average score for all similar clinics in other hospitals was _____ and for all other similar hospitals _____ . Thus your clinic's performance does/does not compare favourably with others. If you would like to know more about how these scores were derived, we will be happy to send you more information.

When the trial is completed, your audit committee will receive a summary report. In the meantime we would ask you not to discuss these findings with medical colleagues not associated with your hospital.

Yours sincerely,

APPENDIX B-4

LETTER ACCOMPANYING CME-IMPACT QUESTIONNAIRE

Dear Dr.

As you will remember, a clinical trial on the effect of audits on patient care has recently been completed in your hospital along with many others in Ontario. Analysis of the results is not yet complete, but the conclusions of the study will soon be forwarded to your hospital.

In the meantime we would like to ask your cooperation to help us answer one more question: to what kinds of continuing medical education have doctors in the trial been exposed?

We would be very grateful if you would complete the enclosed questionnaire and return it to us in the self-addressed envelope.

Thank you very much for your help.

Yours sincerely,

CME QUESTIONNAIRE

Which of the following types of continuing medical education have you used in the last year?

1. Grand Rounds.

   No. If Yes: Once a week ☐ 1-3 times/month ☐ Less frequently ☐

2. Sub specialty Rounds.

   No. If Yes: ☐ ☐ ☐

3. Journals.

		Scan only	Scan & read some articles	Read most articles
No. If Yes: List	_____	☐	☐	☐
	_____	☐	☐	☐
	_____	☐	☐	☐
	_____	☐	☐	☐
	_____	☐	☐	☐

4. Local meetings other than rounds.

   No. If yes, estimate hours attended per month: _____ hrs.

5. Out of town meetings.

   No. If yes, estimate number of days in last year: _____ days.

6. Cassettes with CME programmes.

   No. If yes, estimate number listened to: _____ cassettes/month.

7. Personal communication with colleagues or consultants.

   No. If yes, estimate number of verbal consultations/month ____ .
   And estimate number of letters from consultants
   received per month _____ .

8. What type of clinic are you working in?

   (e.g.: general medicine, obstetrics, dermatology) _____ .

9. In which hospital is it located? _____ .

APPENDIX C

APPENDIX C-1

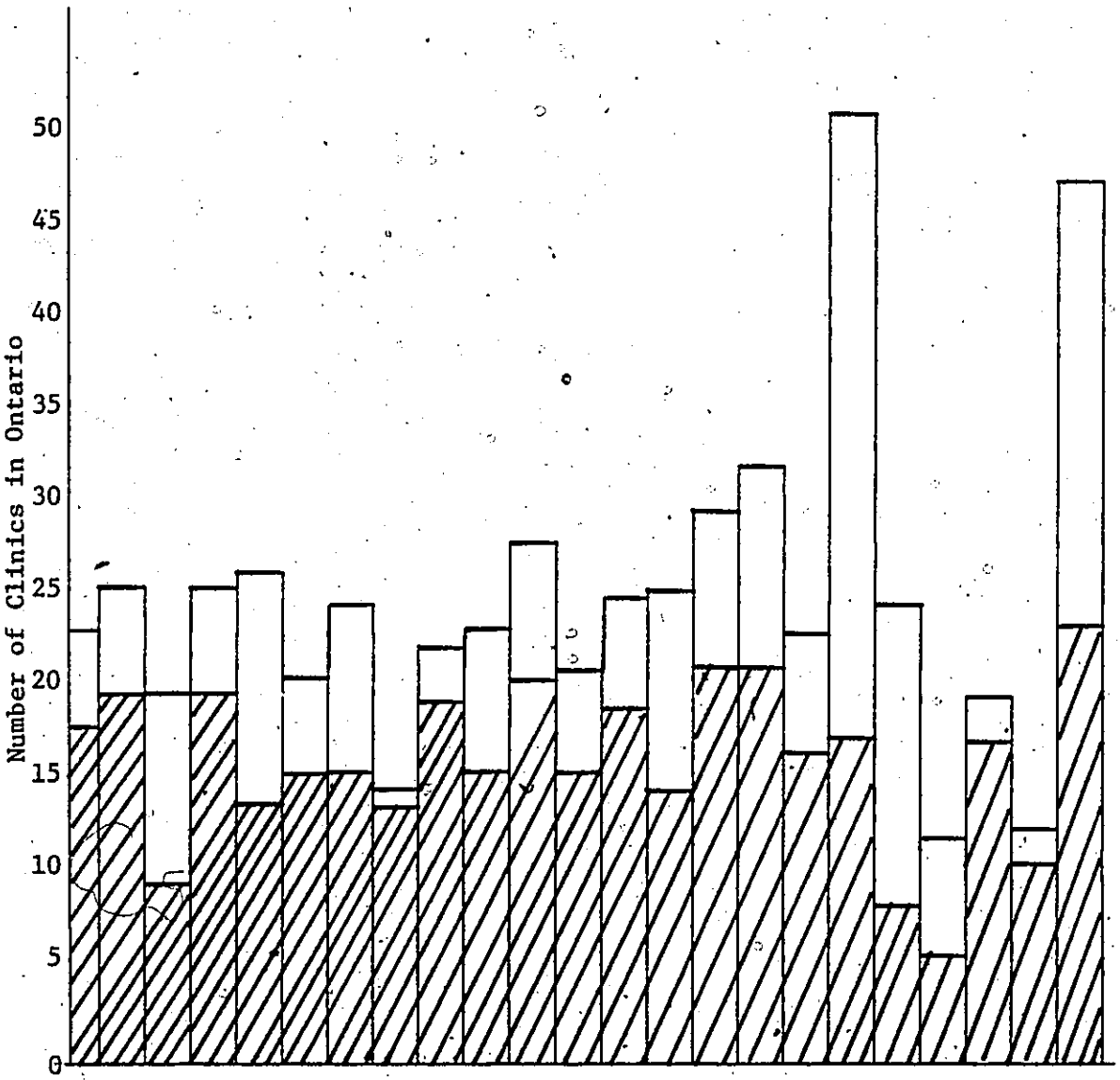
Summary Information About Ontario Teaching Hospitals ('78-'79)*

Name	# beds	# clinics	# visits/yr.
London Victoria	1282	20	131,265
Toronto General Hospital	1089	20	112,042
Hamilton Civic	1031	12	16,618
Toronto Sunnybrook	935	19	176,128
Ottawa Civic	930	17	106,115
Toronto Western Hospital	732	22	88,692
Hospital for Sick Children	705	18	160,115
Toronto St. Michael's	701	21	122,711
Hamilton St. Joseph's	617	11	22,410
Toronto Wellesley	584	20	77,232
London St. Joseph's	534	18	62,584
Kingston General	518	17	97,386
Ottawa General	514	18	102,626
Toronto Mt. Sinai	510	20	159,387
London University	391	18	76,460
Women's College	391	20	64,342
Hamilton Chedoke	356	3	3,533
Hamilton MUMC	342	13	138,912
Ottawa Children's	272	18	103,490
Kingston Hotel Dieu	219	18	69,645
		343	1,891,670

(*Lussing, 1980)

APPENDIX C-2

Histogram Showing Total Number of Each Type of General and Special Clinic in Ontario
(Shaded Area Represents Clinics in Teaching Hospitals)
(MOH Data 1978-79)



	Allergy	Arthritis	Cancer	Cardiac	Dental	Dermatology	Endocrinology	Family Practice	Gastro-Enterology	General Medicine	General Surgery	Gynecology	Neurology	Obstetrics	Ophthalmology	Orthopedics	E.N.T.	Psychiatry	Children	Tuberculosis	Urology	V.D.	Other	
	22	25	19	25	25	21	23	14	21	22	27	20	24	16	28	31	22	50	24	10	22	11	47	= 548
	17	19	8	19	13	16	16	13	18	15	19	15	17	13	19	19	15	17	7	3	17	8	20	= 343

APPENDIX C-3

Community hospitals identified as potential participants

Almonte
Barrie: Royal Victoria
Toronto: Etobicoke General
Chatham Public General
Toronto: North York General
Mississauga
Cornwall General
Thunder Bay: McKellar
Cambridge: South Waterloo
Hearst: Notre Dame
Kingston: St. Mary's
Kitchener-Waterloo
Richmond Hill York
Little Current: St. Joseph's
Burlington: Joseph Brant
Oakville-Trafalgar
Oshawa General
Windsor Western IODE
Ottawa: The Salvation Army Grace
Toronto Scarborough Centenary
Pembroke General
Pembroke Civic
Peterboro Civic
North York: Ontario Crippled Children
Ottawa: Queensway
Picton: Prince Edward
St. Catharines General
Sault Ste. Marie General
Toronto Scarborough General
Southamptom Saugeen
Sudbury General
Timmins: St. Mary's
Toronto: North York Baycrest
Toronto York: North Western
Toronto City: Orthopedic and Arthritis
Toronto: St. Joseph's
Toronto: East General
Welland County General
Toronto York: Humber Memorial
Toronto: Branson
Woodstock General
Toronto Etobicoke: Queensway